MIT-WPU

SINCE 1983

Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

FoET

Springer

## SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY

## 3rd SPRINGER INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND APPLICATIONS-ICCSA21

# Certificate

This is to certify that

**Sachin Bhoite, Anuradha Kanade, Punam Nikam and Deepali Sonawane**

has presented a paper titled

**Predictive Analytics Model of an Engineering and Technology Campus Placement**

at 3rd Springer International Conference on Computational Science and Applications-ICCSA21 held during 10th-11th December 2021 at School of Computer Engineering and Technology, FoET at Dr. Vishwanath Karad MIT World Peace University, Pune, India

| Prof. Sumedha Sirsikar | Prof. Rashmi Phalnikar | Prof. Vrushali Kulkarni | Prof. Prasad Khandekar |
|---|---|---|---|
| TPC Co-chair | TPC Chair | Organizing Co-chair | Organizing Chair |
| SCET, MIT-WPU | SCET, MIT-WPU | HoS, SCET, MIT-WPU | Dean, FoET, MIT-WPU |

# Chapter 21
# Predictive Analytics Model of an Engineering and Technology Campus Placement

**Sachin Bhoite, Anuradha Kanade, Punam Nikam, and Deepali Sonawane**

## 1 Introduction

According to the requirement of industry, colleges must update their curriculum and provide necessary technical and practical knowledge to the students. It will help in fulfilling the requirement of skilled and qualified students of the industries. DM and machine learning (ML) scholars have studied classification problems most recurrently [1]. In which the value of a dependent variable can be predicted based on the values of other independent variables [2]. This paper aims to determine the features impacting on prediction of placement and also students will get to know the placement status and get help in improving their weaker areas in advance.

Basically, this model will help to make training and placement officers (TPO) work easy and increment the total number of placements. Hence, it will directly lead to an increment in the rank of engineering and technology institutions. As our objective is to predict the placement of a student, in such a way that either he will get placement or not. It is a binary classification problem. To get good accuracy with minimum error, we have experimented with various classification ML algorithms with K-fold cross-validation techniques and trained and tested the data splitting techniques. The Value of K is tested for better results though most of the time it has

S. Bhoite (✉) · A. Kanade · P. Nikam · D. Sonawane
School of Computer Science, MIT-WPU, Pune, Maharashtra, India
e-mail: sachin.bhoite@mitwpu.edu.in

A. Kanade
e-mail: anuradha.kanade@mitwpu.edu.in

P. Nikam
e-mail: punam.nikam@mitwpu.edu.in

D. Sonawane
e-mail: deepali.sonawane@mitwpu.edu.in

S. Bhalla et al. (eds.), *Proceeding of International Conference on Computational Science and Applications*, Algorithms for Intelligent Systems,
https://doi.org/10.1007/978-981-19-0863-7_21

considered as 10. Also, we used EL techniques, which are comparatively faster and give better accuracy for classification projects.

## 2 Related Work

The researchers have studied several connected national and international research papers, thesis to understand datasets, data pre-processing methods, features selection methods, type of algorithms used in the existing studies.

Authors in [3] performed a step-wise analysis based on specific statistical frameworks for the placement. The analysis concluded with student datasets including academic and selection subtleties is important for forecasting future selection possibilities. Authors in [4] proposed the campus placement prediction work using the classification algorithms Decision Tree and Random Forest. The accuracy obtained after analysis for Random Forest is greater than the Decision tree. Authors in [5] used different ML algorithms to analyze students' admission preferences. They found Random Forest classifier is a good classifier as its accuracy is very high. Authors in [6] used different ML models to analyze students' placement, they found AdaBoost classifier along with the Bagging and Decision Tree as Base Classifier gives high accuracy. The student placement analyzer recommendation system, built using classification rules-Naïve Bayes, Fuzzy C Means techniques, to predict the placement status of the student to one of the five categories, viz., Dream Company, Core Company, Mass Recruiters, Not Eligible, and Not Interested in Placement. This model helps weaker students and provides extra care toward improving their performance henceforth [7]. Authors in [8] presented student career prediction using advanced ML techniques. In this paper, Advanced ML algorithms like SVM, Random Forest decision tree, One Hot Encoding, XG boost are used. Out of all, SVM gave more accuracy with 90.3%, and then the XG Boost with 88.33% accuracy.

Authors in [9] presented student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics. They used Support Vector Machine with RBF Kernel (SVM), Support Vector Machine with Linear Kernel (SVML), Logistic regression (LR), Decision tree (DT), Random Forest (RF) techniques. ML is used to predict placement results and the programming skill level. The researcher created a classification model with precision, recall, and F-measure.

Authors in [10] presented the study on educational data mining for student placement prediction using ML algorithms. ML algorithms are applied in the weka tool and R studio which are J48, Naïve Bayes, Random Forest, Random Tree, Multiple Linear Regression, binomial logistic regression, Recursive Partitioning, Regression Tree, conditional inference tree, Neural Network. In the weka tool, Random Forest and Random Tree algorithms are giving 100% accuracy on the student placement dataset. Authors in [11] presented a survey on placement prediction systems using

ML. The author has suggested ensemble methods, which is a Machine Learning technique that combines several base models in order to produce one optimal predictive model.

## 3 Research Methodology

The proposed work was carried out by performing experiments on the pass-out student's dataset with various ML algorithms.

### 3.1 Algorithms Used

The objective of research needs to use classification methods. Hence, researchers have used the following ML classification algorithms.

1. Logistic Regression
2. K-Nearest Neighbors
3. Decision Tree
4. Random Forest
5. Support Vector Machine
6. Naive Bays

   Also, used following advanced EL algorithms.

7. Adaptive Boosting,
8. Extreme Gradient Boosting (XGBoost) and
9. Grid Search CV

## 4 Steps in Building Predictive Models Using ML

We followed the Cross-Industry Standard Process (CRISP) methodology.

**Understanding of problem and objectives of the research:** Understanding dataset of already placed students and selection of the appropriate features for placement prediction.

**Data Understanding:** Data of already placed students were collected. All the attributes of the dataset were analyzed based on their importance and relevance based on the placement prediction. Point 5, About the dataset of this topic explains details about the dataset.

**Feature Engineering:** In this phase, the data from multiple data sources were integrated into one dataset. The next step is that the data were cleaned by removing unwanted columns, handling missing values, creating unique classes, performing transformation for numerical data, and all the cleaning activities on the data. Point 6, Feature engineering of this topic explains details about the same.

**Table 1** Univariate feature selection for placement prediction

| Feature name | Feature score | Feature name | Feature score |
|---|---|---|---|
| Sem_IV_Aggregate Marks | 311.517737 | Sem_VI_Pending Back Papers | 22.920021 |
| Aggregate Present Marks | 223.533006 | Sem_V_Pending Back Papers | 20.066178 |
| Sem_III_Aggregate Marks | 198.255768 | Sem_III_Back Papers | 16.854853 |
| Sem_VI_Aggregate Marks | 151.002450 | Back Papers | 12.203902 |
| Sem_V_Aggregate Marks | 147.823502 | Pending Back Papers | 7.822886 |
| Sem_II_Aggregate Marks | 142.692722 | Sem_I_Back Papers | 5.458014 |
| Sem_I_ Aggregate Marks | 138.860021 | Sem_II_Back Papers | 2.265504 |
| College Name | 55.624319 | Sem_II_Pending Back Papers | 0.701740 |
| Sem_VI_Back Papers | 53.893101 | Sem_IV_Pending Back Papers | 0.558951 |
| SSC Aggregate Marks | 42.917186 | Sem_III_Pending Back Papers | 0.200665 |
| Defense Type | 27.490582 | Sem_I_Pending Back Papers | 0.124713 |
| Category | 27.385665 | Gender | 12.518480 |
| 12th/Diploma marks | 26.351629 | Branch | 0.103342 |

**Experimenting:** A number of ML algorithms were tested and experimented with parameter tuning mentioned in Table 2 and 3 to predict the college, and its results are discussed in point 9, result and discussion.

**Evaluation:** Models developed were evaluated based on their performance for accuracy metric [13]. More information is presented in point 8.

**Result and Discussion:** Result and discussion are discussed in point 9.

**Implementation:** Once the model is evaluated, it is used to evaluate unseen data, which is discussed in point 10.

## 5 About the Dataset

Researchers have collected 16 engineering colleges' 9766 data records. A number of columns in the dataset per college was varied from 20 to 46. We merged the dataset by considering common and important columns from our objective point of view in the excel file format, and then converted it into a CSV file. Which is essential to read by the python code to implement ML algorithms.

## 6 Feature Engineering

In general, every ML algorithm takes some input data to generate desired outputs. These input data are called features, which are usually presented in structured columns. As per goal or objective algorithms require input features with some specific

**Table 2** List of experiments with model combinations

| Sr No | Nameof Algorithm | Data splitting method used | Datasplitting folds/ratio | | | Parameter tuned | No. of parame ter Tested |
|---|---|---|---|---|---|---|---|
| 1 | Logistic Regression | K-FCV | 3 | 5 | 10 | label encoding | 6–10 |
| | | | | | | onehot encoding | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | label encoding | 6–10 |
| | | | | | | onehot encoding | 6 to 10 |
| 2 | Support Vector Machine (SVC) | K-FCV | 3 | 5 | 10 | estimator | 6–10 |
| | | | | | | param_grid | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | estimator | 6–10 |
| | | | | | | param_grid | 6–10 |
| 3 | Decision Tree | K-FCV | 3 | 5 | 10 | max_depth | 6–10 |
| | | | | | | min_impurit y_decrease | 6–10 |
| | | | | | | max_leaf_no des | 6–10 |
| | | | | | | min_leaf_no des | 6–10 |
| | | | | | | max_feature s | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | max_depth | 6–10 |
| | | | | | | min_impurit y_decrease | 6–10 |
| | | | | | | max_leaf_no des | 6–10 |
| | | | | | | min_leaf_no des | 6–10 |
| | | | | | | max_feature s | 6–10 |
| 4 | Random Forest | K-FCV | 3 | 5 | 10 | max_depth | 6–10 |
| | | | | | | min_impurit y_decrease | 6–10 |
| | | | | | | max_leaf_no des | 6–10 |
| | | | | | | min_leaf_no des | 6–10 |
| | | | | | | max_feature s | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | max_depth | 6–10 |
| | | | | | | min_impurit y_decrease | 6–10 |

(continued)

**Table 2** (continued)

| Sr No | Nameof Algorithm | Data splitting method used | Datasplitting folds/ratio | | | Parameter tuned | No. of parameter Tested |
|---|---|---|---|---|---|---|---|
| | | | | | | max_leaf_nodes | 6–10 |
| | | | | | | min_leaf_nodes | 6–10 |
| | | | | | | max_features | 6–10 |
| 5 | Gaussian NB | K-FCV | 3 | 5 | 10 | | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | | 6–10 |
| 6 | K Neighbors Classifier | K-FCV | 3 | 5 | 10 | leaf_size | 6–10 |
| | | | | | | n_neighbors | 6–10 |
| | | T-TS | 70:30 | 80:20 | 90:10 | leaf_size | 6–10 |
| | | | | | | n_neighbors | 6–10 |

**Table 3** List of experiments with advanced algorithms

| Sr. No | Name of the Algorithm | Data splitting method used |
|---|---|---|
| 1 | Ada Boost Classifier (DT) | T-TS |
| 2 | Extreme Gradient Boosting (XGBoost) Classifier | T-TS |
| 3 | Grid Search CV | T-TS |

characteristic to get the desired output. Hence, there is a need of feature engineering. Feature engineering efforts mainly have two goals:

1. Generating the proper input dataset, as per the requirement of the ML algorithm.
2. Improving the performance of ML models.

As per the experience of the researcher, we need to spend more than 70% of the time on data preparation. The following steps are carried out to achieve the same.

1. Missing Values
2. Handling categorical data (Label Encoder)
3. Change the data type
4. Drop columns

## 7 Feature Selection

Every time domain experts may not be available to decide independent features to predict the category of the target feature. Hence, before fitting model, we must make sure that all the features that we have selected are contributing to the model properly and weights assigned to it are good enough so that our model gives satisfactory accuracy. For that, we have used 3 feature selection techniques: Univariate Selection,

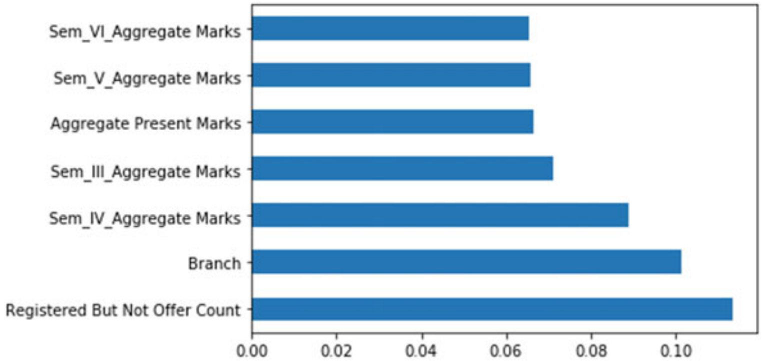


**Fig. 1** Feature selection using feature importance for placement prediction

Recursive Features Importance, and Feature importance. We used the python scikit-learn library to implement it.

The Univariate Selection method shows the highest score for the following features (Table 1).

While using the Recursive Feature Importance method, the following features are selected, and the remaining are rejected.

Selected Features: [‘Pending Back Papers’, ‘Sem_III_Pending Back Papers’, ‘Sem_IV_Aggregate Marks’, ‘Sem_IV_Pending Back Papers’, ‘Sem_V_Pending Back Papers’, ‘Sem_VI_Back Papers’].

Inbuilt class Feature importance comes with Tree based Classifiers; we used Extra Tree Classifier from python scikit-learn library for extracting the top 7 features of the dataset (Fig. 1).

Hence, as per all the above methods and also as per domain our knowledge, we have chosen 25 important features which are as follows to predict target feature ‘Job Offer’.

[‘Branch’, ‘Aggregate Present Marks’, ‘Back Papers’, ‘Pending Back Papers’, ‘Sem_I_ Aggregate Marks’, ‘Sem_I_Back Papers’, ‘Sem_I_Pending Back Papers’, ‘Sem_II_Aggregate Marks’, ‘Sem_II_Back Papers’, ‘Sem_II_Pending Back Papers’, ‘Sem_III_Aggregate Marks’, ‘Sem_III_Back Papers’, ‘Sem_III_Pending Back Papers’, ‘Sem_IV_Aggregate Marks’, ‘Sem_IV_Back Papers’, ‘Sem_IV_Pending Back Papers’, ‘Sem_V_Aggregate Marks’, ‘Sem_V_Back Papers’, ‘Sem_V_Pending Back Papers’, ‘Sem_VI_Aggregate Marks’, ‘Sem_VI_Back Papers’, ‘Sem_VI_Pending Back Papers’, ‘12th/Diploma_Aggre_marks’, ‘SSC Aggregate Marks’].

## 8 Experimentation

There are adequate models that are studied and tested for the objective with optimal values for K-fold cross-validation (K-FCV), Train-Test split (T-TS), parameters tuning, and testing. In this process, the python sklearn library has played a very important role. So detail is mentioned in the table below.

Apart from the above methods while doing parameter tuning, we have used the following ensemble algorithms.

After the discussion of the accuracy results researcher has suggested a web module named 'Free guide to notify the campus placement status (FGNCPS)' through which students will get to know their placement status in advance and also come to know to work more on weaker areas.

## 9 Result and Discussion

After implementing data cleaning process, removing all the noise, selecting relevant features and encoded it into ML form, the next step is building a predictive model by applying various ML techniques to find out the best model which gives us more accuracy for train data and test data.

**Model selection for placement prediction**: After implementing all the above methods mentioned in Table 4 and 5, we found XGBoost classifier is the best classifier to predict campus placement.

**Table 4** Results of placement prediction using ML techniques with K-fold cross validation

| Sr. No | Name of Algorithm | Train Accuracy | Test Accuracy |
|---|---|---|---|
| 1 | Logistic Regression | 0.7251336898395722 | 0.7371794871794872 |
| 2 | Support Vector Machine | 0.7235294117647059 | 0.7393162393162394 |
| 3 | Decision Tree Classifier | 0.8165775401069518 | 0.782051282051282 |
| 4 | Random Forest Classifier | 0.823663101604278 | 0.7162393162393162 |
| 5 | Gaussian NB | 0.5294117647058824 | 0.49145299145299143 |
| 6 | K Neighbors Classifier | 0.8235294117647058 | 0.7606837606837606 |

**Table 5** Results of placement prediction using Ensemble Learning

| Sr. No | Algorithm | Train Accuracy | Test Accuracy |
|---|---|---|---|
| 1 | AdaBoostClassifier(DT) | 0.85 | 0.82 |
| 2 | XGBoost | 0.88 | 0.84 |
| 3 | GridSearchCV | 0.851336898395 | 0.82478632478632 |

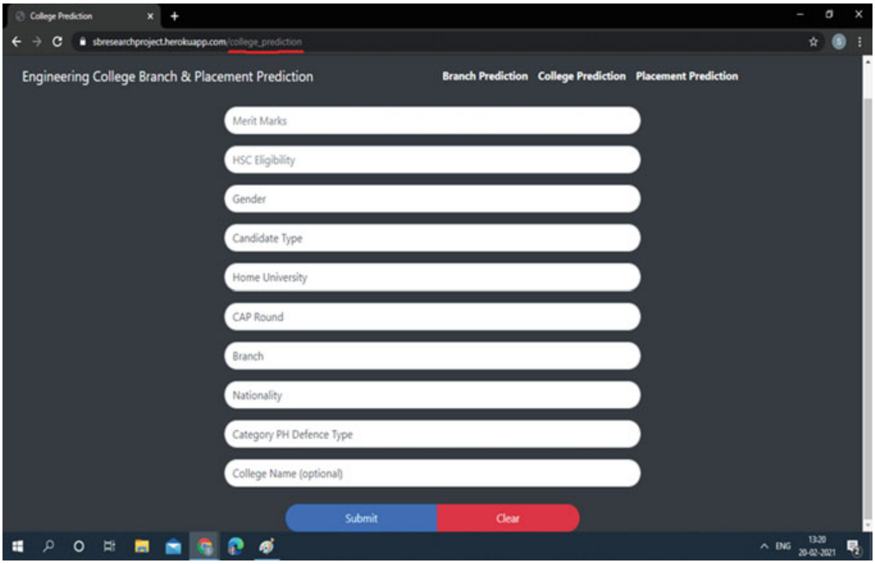


**Fig. 2** Placement prediction web module

We can see the result of placement prediction using ensemble classifier XGBoost with 0.88 training accuracy and with 0.84 testing accuracy, which is comparatively very high. Hence, we have chosen the XGBoost classifier to implement the model.

## 10 Implementation

### *10.1 A Free Guide to Notify the Campus Placement Status (FGNCPS)*

While predicting the campus placement of Engineering and Technology students, we have proposed the following FGNCPS web module. The aspirant student has to submit some basic information which is nothing but selected input features to predict their placement status in the early stage of academics (Fig. 2).

## 11 Conclusion

In this research, to predict the campus placement of Engineering and Technology students, all the ML model building steps are rigorously implemented on the dataset. Python, various libraries played a vital role during whole this process. In this study, 25 input features are selected out of the existing 46 features of the dataset. These features are very important, according to Univariate Selection, Recursive Features Importance, Lasso feature selection methods, and researchers' domain knowledge. To predict the campus placement, suit of ML and EL methods are experimented and compared. This suit contains Logistic Regression, K-Nearest Neighbors', Decision Tree Classifier, Random Forest Classifier, Naive Bayes, and Support Vector Machine classifiers. Under EL, we have experimented with Adaptive Boosting, Gradient Boosting, and GridSearchCV methods. After a comparison of all algorithms' accuracy, we found that the XGBoost classifier has greater accuracy for this project. Also, it has been observed that feature engineering is a very important step in a model building because, after it, results have been more improved. At the end researchers have suggested, 'A free guide to notify the campus placement status (FGNCPS)' web module for placement aspirant students.

## References

1. Kabakchieva D, Stefanova K, Kisimov V (2011) Analyzing university data for determining student profiles and predicting performance. In: 4th International conference on educational data mining (EDM 2011). The Netherlands, pp 347–348
2. Nie M, Yang L, Sun J, Su H, Xia H, Lian D, Yan K (2017) Advanced forecasting of career choices for college students based on campus big data. Higher Education Press and Springer-Verlag Berlin Heidelberg
3. Kumar N, Singh AS, Thirunavukkarasu K, Rajesh E (2020) Campus placement predictive analysis using machine learning. In: 2nd International conference on advances in computing, communication control and networking (ICACCCN), ISBN: 978-1-7281-8337-4/20/$31.00 ©2020. IEEE
4. Pothuganti M, Swaroopa N (2019) Campus placement prediction using supervised machine learning techniques. Int J Appl Eng Res 14(9):2188–2191, ISSN 0973-4562
5. Kalathiya D, Padalkar R, Shah R, Bhoite S (2019) Engineering college admission preferences based on student performance. Int J Comput Appl Technol Res 8(09):379–384, ISSN:-2319–8656
6. Khandale S, Bhoite S (2019) Campus placement analyzer: using supervised machine learning algorithms. Int J Comput Appl Technol Res 8(09):379–384, ISSN:- 2319–8656, 358–362
7. Apoorva Rao R, Deeksha KC, Vishal Prajwal R, Vrushak K, Nandini (2018) Student placement analyzer: a recommendation system using machine learning. IJARIIE 4(3), ISSN(O)-2395-4396
8. Roy KS, Roopkanth K, Teja VU, Bhavana V, Priyanka J (2018) Student career prediction using advanced machine learning techniques. Int J Eng Technol 7:26–29
9. Ishizue R, Sakamoto K, Washizaki H, Fukazawa Y (2018) Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics. Res Pract Technol Enhanced Learn 13. https://doi.org/10.1186/s41039-018-0075-y
10. Sreenivasa Rao K, Swapna N, Praveen Kumar P (2017) Educational data mining for student placement prediction using machine learning algorithms. Int J Eng Technol, [S.l.] 7(1.2):43–46, ISSN 2227-524X
11. Bangale M, Bavane S, Gunjal A, Dandhare R, Salunkhe SD (2019) A survey on placement prediction system using machine learning. IJSART 5(2), ISSN [ONLINE]: 2395-1052

# VIRTUAL INTERNATIONAL CONFERENCE

Jointly Organized by

PIRENS INSTITUTE OF BUSINESS MANAGEMENT AND ADMINISTRATION, LONI, INDIA and
PROWESS UNIVERSITY, USA

PARTICIPATED AND PAPER PRESENTATION
IBMA

## CERTIFICATE OF PAPER PRESENTATION

This is to certify that

*Dr. Rijwan M. Shaikh*

has Participated and Presented a Research Paper Entitled

Emerging Economic Problems Before India

in International Conference on 'Paradigm Shift in Economy: Its altercation on Trade, Commerce, Management, Engineering and Social Science' held on 25th & 26th February 2022.

| Dr. Rajesh R | Dr.M.A.Tamboli | Dr.N.U.Bankar | Dr.V.N.Sayankar |
|---|---|---|---|
| India Coordinator-PROWESS UNIVERSITY,USA | Convener | Dy.Director-IBMA | Director-IBMA |

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

सत्यं शिवं सुन्दरम्

Estd. 1949

# Journal of
# The Maharaja Sayajirao University of Baroda

## Certificate of Publication

Certificate of publication for the article titled:

**AWARENESS LEVEL ABOUT EMOTIONAL COMPETENCE AMONG MBA STUDENTS**

Authored by

**Dr. Rijwan M. Shaikh**

**Associate Professor, Sinhgad Institute of Management (SIOM), Vadgaon, Pune.**

Volume No .56 No. 2(V) ·2022

in

Journal of The Maharaja Sayajirao University of Baroda

ISSN : 0025-0422

(UGC CARE Group I Journal)

MSU of Baroda

Editor

Journal MSU of Baroda

# EMERGING ECONOMIC PROBLEMS BEFORE INDIA

**Dr. Rijwan M. Shaikh**
Associate Professor, Sinhgad Institute of Management, Vadgaon Bk., Pune.

**Abstract-** Indian economy has evolved very fast in last few decades; improving infrastructure has been the policy of the government to accelerate the economic growth. Eradication of poverty and further raising the standards of living of each of the individual of different social class could be made only possible with the offering of the employment to them. It's a reason why the focus is mainly on developing of infrastructure by the government. Indian economy has shown time and again its potential by contributing in world's economic growth. The per capita GDP based on purchasing power parity has raised more than 3 fold when compared with that of year 2000. Moreover, there is also increase of almost 10 percent of economic activities against that of year 2005. There is further divide in incomes of rich and poor and the citizens of urban and rural India. This divide is also evident among number of states within the country. Almost 8 crores of Indians seems falling in extremely poor category. Major chunk of this population resides in five states of India. Also, because of lots of economic issues arisen like Covid 19 pandemic, demonetization and change of tax regime to GST, economic growth was slowed down recently. Using India's huge population to accelerate economic growth can make India great again. Still there are villages in the country which are looking for constant electric supply to households, public works and even for the sake of farming. There are also newly emerged factors like internet connectivity and data consumption defining the economic growth of the country. Therefore, the present study focuses on how the social progress and building of infrastructural facilities could help India achieve economic growth against developed nations. The shortcomings in the sectors like health, skill development and providing of education platform are the main reasons of India's pulling back in economic development. The present research work also looks for identification of the opportunities and challenges for better prospects of India.

**Keywords:** Economic development, Social Development, Human Capital Development, Indian Economic State, Economic Problems, India.

## I. INTRODUCTION

The economic progress of any country is defined with its social progress and building of infrastructural facilities. Physical infrastructural facilities will further define where GDP will head whereas social infrastructural facilities will define where Human Development Index would head. Social infrastructure would ensure the citizens are healthy, educated and skill enough. The economic development of many western nations is evident because of their self-sufficiency in the health and education sector. Human development is equally important of infrastructural development to achieve stable economic growth. A poor person can be made self-reliant by educating him and by no means of social upliftment program. Educated population means less or no poverty and economic growth is in reach. A country with good physical infrastructure can fructify the hidden talent and potential of its skilled and educated population.

1.1 Indian Economy Types

- Market economy
- Traditional economy
- Mixed economy

Part of Indian economy is regulated and part is deregulated in few areas. Government is on selling spree of its stake in some of PSU’s and PSB’s. It helps them reduce fiscal deficit significantly and unearth the true potential of these business entities. The Indian agricultural sector still offers 50% of the employment to countrymen whereas service sector employs 1/3rd of its population. But service sector contributes to 2/3rd of India's output. The FDI laws have been reframed in sick sectors like telecom for their better competitiveness.

### 1.2 Challenges to Indian Economic Development and Opportunities

1. The ever disturbing geopolitical situation in the world is the most significant factor that may harm the Indian economy. The tension at the border between India-China, India-Pakistan, Russia-Ukraine and between US and other gulf nations is keeping economy a tensed one.
2. A challenge to deal with Covid -19 pandemic and manage its further occurrences so that Indian economy remains least affected in short as well as long run.
3. To keep rate of inflation within comfortable level is another challenge as WPI stood at 13.56% in December 2021 whereas CPI was at 7 month high of 6.01% in January 2022.
4. The continuous reduction in demand by consumers and business customers is evident with sharp decline in seeking of new loans and thus restricting further new investments.
5. Vaccinations of children below 18 years of age of Covid 19 and further restrict the spread of corona virus.
6. To provide combined effect of fiscal policy and monetary policy through increase of spending’s and controlling of inflation rates.
7. The increasing NPA’s of banks making it difficult for them to survive in the times of crisis like Covid and others.
8. The control of rate of unemployment and eradicate poverty while achieving economic growth.
9. Gaining self-sufficiency in providing of healthcare services and increasing reach of education hampered again by covid 19 pandemic.
10. Indian unorganized sector has suffered worst besides the shaking of micro to medium scale of enterprises with back to back setbacks like demonetization, GST and covid 19.
11. To reduce the gap widened between incomes and wealth of rich and poor because post covid 19 pandemic.
12. To maintain the social stability and harmony across different states of the country to invite more FDI.

### 1.3 India's Strengths

1. Strength of the Indian economy lies in its ability to adapt on its backing by reserves of foreign currencies and fiscal and monetary policies.
2. Indian economy still remains the fastest growing one.
3. Indian economy has been always groomed by best of policymakers and administrators to reach to its state of desire.
4. Robust Indian IT infrastructure making its contribution to the economy.
5. India recently achieved a landmark of receiving 3rd country in the world after US and china to have most nos. of tech start-ups. Many of them have turned unicorns.
6. More than 60% of Indian population is falling into youth category is another demographic dividend.

7. India is known economical labour supply and equipped with a variety of skilled and semiskilled human capital.
8. The population of India is strength in itself if directed with set of objectives it could turn the things around in favour of the nation.
9. Indian stock market has witnessed IPO of 65 companies in calendar year 2021. The trend is expected to continue even in year 2022. Altogether these companies raised capital of almost 1.25 lakh crores.

## II. LITERATURE REVIEW

**Dr. Shubhra Aanand et al.** Indian economy has been unified with the global that has been experienced in past many times. The accounts like capital one and current are the reasons for their unification. The flexible nature of Indian economy because of its adaptability has gained over the best of its administrative and policymaking initiatives. It has been able to sustain its tag of one of the fastest growing economy in the world. The author tries to figure out impact of many such financial crises over Indian economy and the future prospects could be gained out of it.

**Aamir Firoz Shamsi et al.** The countries like Japan owning a successful economy besides Germany as well as china have reached this feat only because of their collaboration with the neighbouring economies. It is presented as one of the successful economic model in the present research paper. The author also examines factors to make India a global economic power. The research paper throws light upon means to achieve sustainable economic growth. It also highlights the fact that economic potential may diminish with troubled relations with the neighbours and it may become come across as biggest hurdle in India's dream of becoming an economic power house of the world.

**Mohammad Reza NORUZI et al.** The opportunity an economy may receive with the ever increasing globalization has been studied in the present research paper. It will lead to better employment rate and positive economy. Customers will have access to variety of the products at competitive prices whereas organizations can rest assured of uninterrupted supply and demand. On the other hand globalization may cause emergences of some threats. It calls for skilled, literate and qualified human capital. It further increases competition and calls for innovation.

**Rajiv Kumar Bhatt et al.** The 1st part of paper deals with economic recession. Next part talks about adaptability of Indian economy in such times of crises whereas last part of the research paper focuses upon suggestions to deal with such economic situations. It is also seen that India confronted economic recession with by pouring of liquidity instead of just relying on financial policy. The author in the end suggest that government spending's should be more on agriculture and infrastructure sectors.

**Suraj Walia et al.** The author identifies different sectors which were affected because of worldwide recession. Globalization has made whole world economy fragile and suscptible to such economic crises. The author tries to figure out how Indian economy is suffering because of global economic tremors. At least they lead to slowing down of Indian economy if not completely melting it down.

**Dr. Amit Kumar Khare et al.** The paper highlights how Indian economy even during global turmoil has been able to be on its track. It also point out the need to open FDI path for service infra and aviation industry. The step is necessary to accelerate the growth in service as well as manufacturing sector and they are backed by 1991 economic reforms and its benefits India had by author.

**Kalim Siddiqui et al.** The author has studied transition the economies are going through its dynamics mainly in reference with the developing economies. Author stresses upon need of job creations that an economy shall achieve with the act of FDI and not just mere encouraging of investments. Author also tries to examine related practical and conceptual studies. Author also makes economies aware of ill effects of capital liberalization and neoliberalism which is making economies of developing nations more vulnerable.

**K G. Viswanathan et al.** Author tells that financial troubles are unavoidable for economies. Author also states that after economic recession economies are bound to bounce back. Author compares all the past economic recessions with each other and identifies that the Great Depression only was the toughest one which has left every country affected of several economic problems.

## III. THE STATE OF INDIAN ECONOMY

India has managed to maintain forex reserve to increase despite economy being under trouble because of back to back major economic issues like nationwide lockdown caused of covid 19 pandemic. By December' 2021, Indian forex stood at US $ 633.6 Billion.

**Table 1: Foreign Exchange Reserves** (in US $ Billion)

| 2018-19 | 2019-20 | 2020-21 | 2021-22* |
|---|---|---|---|
| 412.9 | 477.8 | 577 | 633.6 |

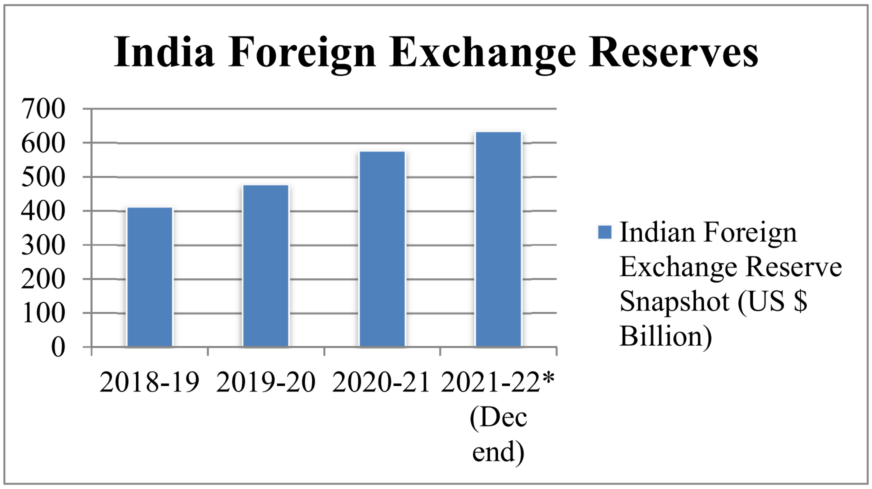


**Fig 1: Indian Economy at a Glance**

Indian economy is set to experience V shaped recovery after FY 20 and 21 were the toughest one because of covid 19 pandemic. Estimated growth of 9.2% in GDP in year FY 2022 is estimated to further reach 8.5% which is itself is promising enough.

**Table 2: GDP Growth** (in %)

| 2019-20 | 2020-21 | 2021-22* | 2022-23# |
|---|---|---|---|
| 4 | -7.3 | 9.2 | 8.5 |

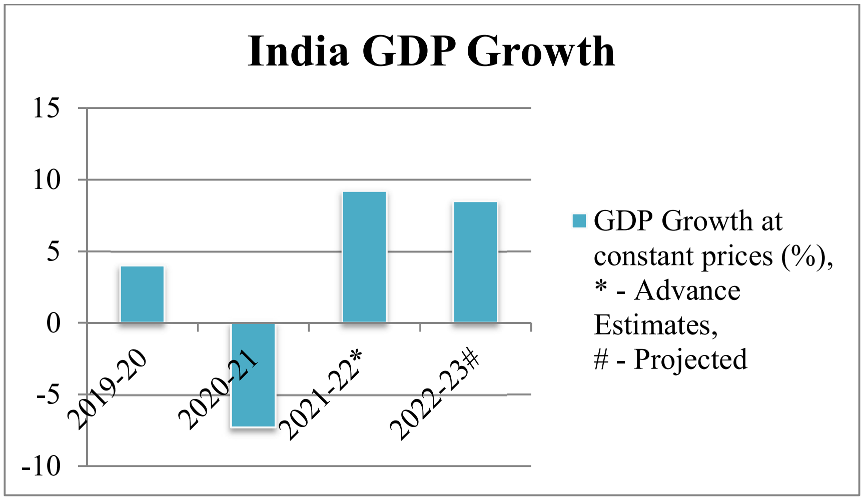


**Fig 2: GDP Growth**

The privatization policy of Indian government will help diminish the rate of growth of fiscal deficit to controllable one. It could be witnessed by selling of government stake I Air India, disinvestment of LIC in March 2022, proposed privatization of HPCL and PSB's like IDBI.

**Table 3: Fiscal Deficit** (% of GDP)

| 2018-19 | 2019-20 | 2020-21 | 2021-22* |
|---|---|---|---|
| 3.4 | 4.6 | 9.2 | 6.8 |

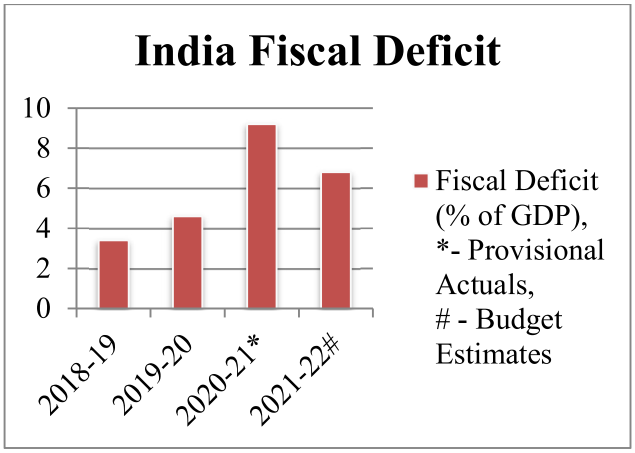


**Fig 3: Fiscal Deficit** (% of GDP)

There is surge in demand to pre covid level at the industry and may even further see the boost in terms of industrial growth. Because of disturbed supply chain in automobile sector and electric 2 wheeler segment further growth in industrial sector may also continue in coming years.

**Table 4: Industrial Growth** (%)

| 2018-19 | 2019-20 | 2020-21 | 2021-22* |
|---|---|---|---|
| 5.3 | -1.2 | -7 | 11.8 |

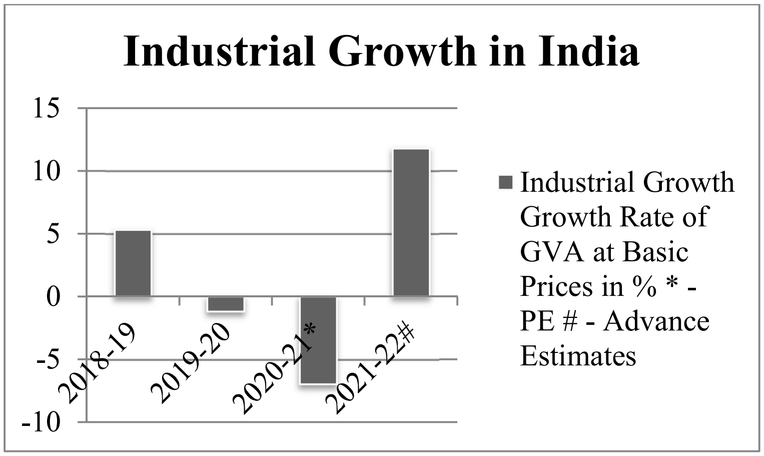


**Fig 4: Industrial Growth** (%)

The increasing role of Indian service sector has also shown the good recovery post covid nationwide lockdowns. In FY22 it is estimated to grow at a rate of 8.2% from -8.4% from its preceding year.

**Table 5: Services** (%)

| 2018-19 | 2019-20 | 2020-21 | 2021-22* |
|---|---|---|---|
| 7.2 | 7.2 | -8.4 | 8.2 |

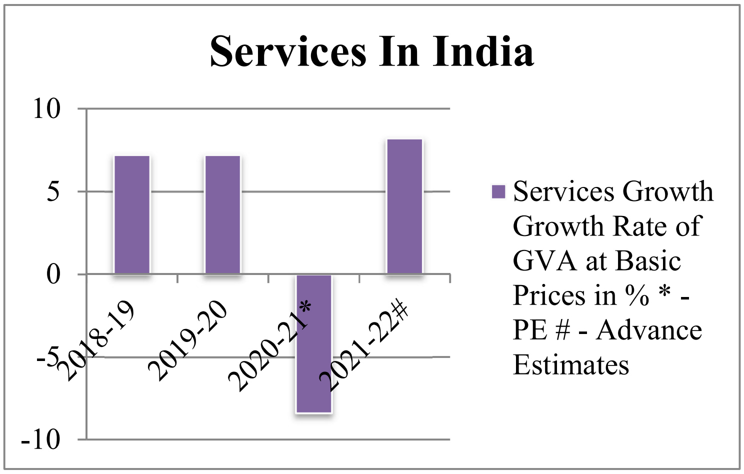


**Fig 5: Services** (%)

In balance to achieve economic growth through monetary and economic policy, inflation has been one thing which is on the rise. The WPI index has reached to uncomfortable position of double digit. The quarter results of many listed companies have shown sharp decline in profit despite growth in revenue. It may further lead to rise in inflation.

**Table 6: Inflation in India**

| Parameter | 2018-19 | 2019-20 | 2020-21 | 2021-22* |
|---|---|---|---|---|
| CPI Combined | 3.4 | 4.8 | 6.6 | 5.2 |

| | | | | |
|---|---|---|---|---|
| WPI | 4.3 | 1.7 | 1.95 | 12.5 |

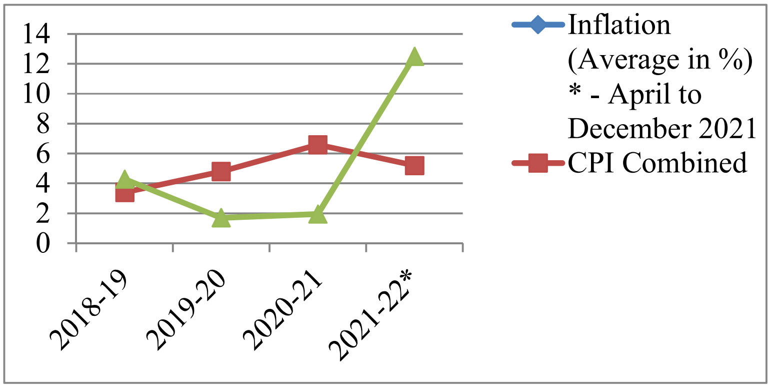


**Fig 6: Inflation in India**

Moreover, evolution of India in global competitiveness rank could serve as a booster dose to the Indian economy.

**Table 7: Evolution of India's Ranking In Global Competitiveness**

| Year | India Rank |
|---|---|
| 2007-08 | 48 |
| 2008-09 | 50 |
| 2009-10 | 49 |
| 2010-11 | 51 |
| 2011-12 | 56 |
| 2012-13 | 59 |
| 2013-14 | 60 |
| 2014-15 | 71 |
| 2015-16 | 55 |
| 2016-17 | 63 |
| 2017-18 | 58 |
| 2018-19 | 68 |
| 2019-20 | 43 |
| 2020-21 | 43 |

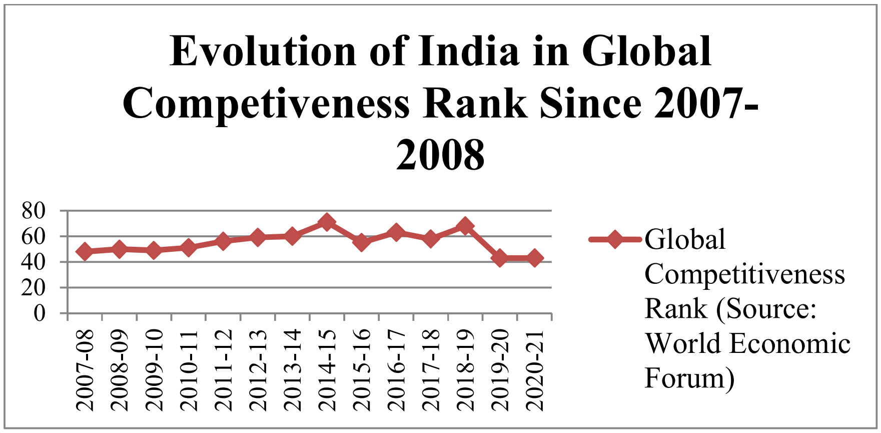


**Fig 7: Evolution in Global Competiveness Rank since 2007-2008**

India's global competitiveness rank as shown in data released by world economic forum in 2019 shows that India's rank is sustained by its ever increasing market size, innovation capability and robust financial system mainly. The macroeconomic factor of one which is political has also help in taking India great leap in global competitiveness rank.

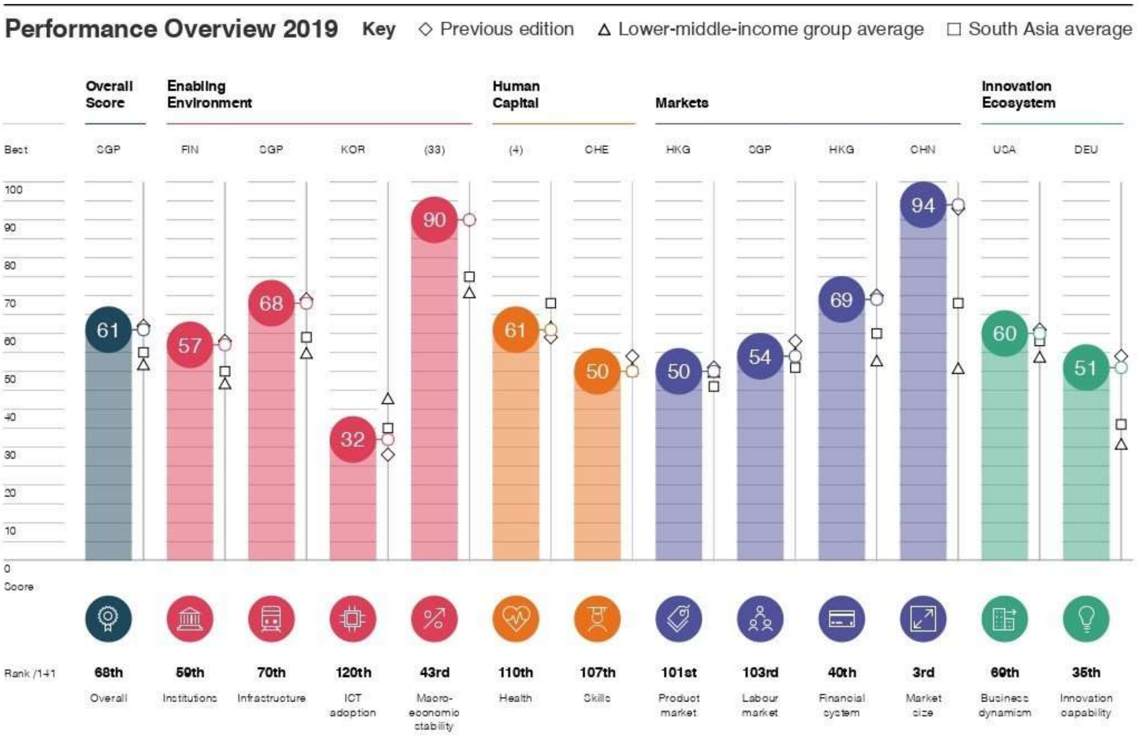


**Fig 8: India's Performance across the 12 Pillars of Global Competitiveness**

## IV. DEVELOPMENT ISSUES OF INDIAN ECONOMY

1. The per capita income (PPP) of India stood at Rs. 1,35,000 in 2020 whereas China had 5 times higher.
2. Agriculture sector employs more than 50% of its population but its contribution to GDP has restricted to 20%. The over dependency of India in generating jobs over Agricultural sector shows that Indian economy is still of traditional nature.
3. Indian population size has reached to 135 cr. It is both boon and issue to an economy as the use of such a huge human capital will decide the type of outcome it may give.
4. To maintain the present level of standard of living can be only achieved with the boost in capital formation.
5. In world inequality report of FY22, it is mentioned the 10% affluent elite class in the country own more than 57% of total wealth whereas bottom 50% of the population holds just 13% of total wealth.

## CONCLUSION

The research paper takes into consideration the emerging economic problems before India. To discuss the same it becomes important to study what type of economy India has and what type of economic stance India has taken in recent past. The challenges, opportunities and the strengths of Indian economy are studied at length.

The current state of Indian economy reflects that it is on path of economic success. The factors like inflation rate and fiscal deficit are the immediate issues Indian economy faces. The geopolitical instability remains perennial economic factor of concern. Better state of foreign exchange reserve, recovery in GDP, Industrial growth rate and service growth rate makes one even more hopeful and positive about Indian economy. The better performance of Indian economy is evident with sharp increase in rank in global competitiveness of Indian economy.

The robust economic development, a nation can achieve and sustain with its social development only. The human capital development can be acquired through providing of better educational and healthcare facilities. Increasing per capita income and bridging the increasing gap between rich and poor divide remains challenging though.

## REFERENCES

[1] Dr. Shubhra Aanand "Global Economic Crisis: Impacts, Challenges and Opportunities for India" Pacific Business Review International Volume 6, Issue 5, November 2013

[2] Rajiv Kumar Bhatt "Recent Global Recession and Indian Economy: An Analysis" International Journal of Trade, Economics and Finance, Vol. 2, No. 3, June 2011

[3] Suraj Walia "Impact of Global Economic Crisis on Indian Economy: An Analysis" International Journal of Latest Trends in Engineering and Technology (IJLTET)

[4] Aamir Firoz Shamsi "India as an Emerging Economy" Transnational Corporations Review Volume 6 Number 1 March 2014 www.tnc-online.net info@tnc-online.net

[5] Mohammad Reza NORUZI "A literature Review of Global Economy and Globalization Era" Research gate https://www.researchgate.net/publication/47348022 25 October 2016.

[6] Dr. Amit Kumar Khare "Global Economic Turmoil A Road Map of Indian Economic Growth" International Research Journal of Marketing and Economics Vol. 4, Issue 5, May 2017 Impact Factor-

5.671 ISSN: (2349-0314) © Associated Asia Research Foundation (AARF) Website: www.aarf.asiaEmail : editor@aarf.asia, editoraarf@gmail.com

[7] Kalim Siddiqui “Financialization and Economic Policy:” World Review of Political Economy 8 (4): 564-590, winter, Pluto Journals. DOI: 10.13169/worlrevipoliecon.8.4.0564.

[8] K G. Viswanathan “The Global Financial Crisis and its Impact on India” Journal of International Business and Law, Vol. 9, Iss. 1 [2010], Art. 2

[9] Dr.M.Sivakumar “2008 Global Economic Crisis and Its Impact on India’s Exports and Imports” Munich Personal RePEc Archive t https://mpra.ub.uni-muenchen.de/40950/ MPRA Paper No. 40950, posted 30 Aug 2012 09:19 UTC

[10] Nanto, Dick (2009), "The Global Financial Crisis: Analysis and Policy Implications", Congressional Research Service 7-5700.

# VIRTUAL INTERNATIONAL CONFERENCE

Jointly Organized by

PIRENS INSTITUTE OF BUSINESS MANAGEMENT AND ADMINISTRATION,LONI,INDIA and
PROWESS UNIVERSITY,USA

PARTICIPATED AND PAPER PRESENTATION

IBMA

## CERTIFICATE OF PAPER PRESENTATION

This is to certify that

*Dr. Rijwan M. Shaikh*

has Participated and Presented a Research Paper Entitled

Awareness level about Emotional Competence among MBA students

in International Conference on 'Paradigm Shift in Economy: Its altercation on Trade, Commerce, Management, Engineering and Social Science' held on 25th & 26th February 2022.

| | | | |
|---|---|---|---|
| Dr. Rajesh R | Dr.M.A.Tamboli | Dr.N.U.Bankar | Dr.V.N.Sayankar |
| India Coordinator-PROWESS UNIVERSITY,USA | Convener | Dy.Director-IBMA | Director-IBMA |

# AWARENESS LEVEL ABOUT EMOTIONAL COMPETENCE AMONG MBA STUDENTS

**Dr. Rijwan M. Shaikh**
Associate Professor, Sinhgad Institute of Management (SIOM), Vadgaon, Pune.

**Abstract**

This is a descriptive study to assess awareness level about Emotional Competence. A survey questionnaire was used to collect primary data from 400 MBA students from Pune. Responses were measured on a 5-point Likert scale for the two sections. The sample mean was compared against the hypothesized population mean of "2" and was tested for statistical significance at 95% confidence level. The mean was found to be well below the hypothesized population mean (Mean = 1.30 and 1.27; SD = 1.10 and 1.05).The results indicate that the awareness level about Emotional Competence among MBA students is significantly lower.

**Keywords:** Emotional Competence, Personal Competence, Social Competence.

## 1.0 Introduction

Emotional competence depicts the capacity of an individual to communicate their own emotions with complete freedom, and it is gotten from emotional intelligence, which is the capacity to distinguish emotions. Competence is the degree of expertise with which somebody interacts constructively with others. This individual emotional competence depends on an individual's acknowledgment of individual emotions and how emotions influence others, and it is additionally based on the capacity to keep up emotional control and adapt. The individual must be fit for understanding their own emotions before they assess the emotions of others.

Another individual aspect of emotional competence is social competence, which alludes to empathy towards others. It includes the abilities we should be effective in a work atmosphere and relationships. It is critical to utilize viable correspondence and to realize how to oversee conflicts. Through emotional competence, individuals can respond to their own emotions and those accomplished by others. An individual can react accurately when somebody encounters emotions like anger, fear, and pain. Perceiving one's own emotions, opens up the chance of reacting appropriately to the emotions that others experience. Without knowing one's own emotions, it is hard to help or feel empathy for another.

Given the importance of emotional competence, we decided to survey MBA students to assess their awareness level.

The research questions to be addressed are as follows:
RQ1: Is the awareness level of MBA students regarding Personal Competence significant?
RQ2: Is the awareness level of MBA students regarding Social Competence significant?

The term "significant" was operationalized at the mid-point of a 5-point Likert scale at the value of "2".

## 2.0 LITERATURE REVIEW

The examination by Ikavalko et al. (2020), investigated emotional competence at work and explained emotional competence comparable to sociocultural aspects of emotions at work. Emotional competence at work was investigated through interviews, surveys, and observations. The investigation was directed for more than one year, during which an emotion training intervention was led within a medium-sized organization, operating in the medical care area. The investigation shed light on emotional competence at work, identifying three domains: individual emotional competence, emotional competence within interactions, and emotional competence embedded in workplace practices.

Social-emotional competence is a basic factor to focus with universal preventive interventions that are led in schools because the construct (a) partners with social, behavioral, and academic outcomes that are significant for healthy development; (b) predicts significant life outcomes in adulthood; (c) can be improved with possible and cost-effective interventions; and (d) assumes a basic job in the behavior change process. This article (Domitrovich et al., 2017) audits this examination and what is thought about effective intervention approaches. Given that, an intervention model is proposed for how schools should improve the social and emotional learning of students to advance versatility. Proposals are likewise offered for how to help the usage of this intervention model at scale.

The advancement of social-emotional competence and execution of social-emotional learning programs have expanded considerably in schools; however, little is thought about instructors' impression of such programs. This subjective examination (Humphries et al., 2018) investigated youth (3 to 8 years of age) instructors' impression of classroom-based social-emotional learning programs for young, urban-dwelling children. A focal point of the examination included learning what educators accept were the basic segments and difficulties of such programs. Five topics came out of the content analysis: responsibility, curricula design, contextual relevance, support, and barriers.

According to Cornell et al. (2017), kids are becoming familiar with social-emotional competencies, for example, self-awareness, self-regulation, and social awareness. Sustaining these abilities is significant for positive developmental results. In this study, the researchers' layout tools which distinguished age proper utilization of kids' adapting language in the early learning setting, the development and validation of tools to quantify the coping construct, and its relationship with youngsters' anxiety, strengths, and difficulties. This exploratory examination found that support in the program helped youngsters' social-emotional competencies. On the whole, the part features how social-emotional aptitudes can be evaluated and educated in an early learning setting.

A contextual research gap clearly exists and hence this study was undertaken by surveying MBA students in Pune.

## 3.0 METHODOLOGY

The research methodology adopted is outlined below:

1. A survey questionnaire was administered to 400 (Bullen, 2016) MBA students.
2. The selection of the 400 students was based on the judgment of the writer of getting an adequate response in a reasonable time. Convenient sampling was used.
3. The survey questionnaire had two sections – first on Personal competence and second on Social competence.
4. Responses were sought on 5-point Likert-scale.
5. For assessment of awareness regarding Personal Competence, the scale used was: 0-Not at all aware, 1-Little bit aware, 2-Somewhat aware, 3-Well aware, 4-Highly aware. The 10 sub-questions were: 1 –emotional awareness, 2 –accurate self-assessment, 3 – self-confidence, 4 – self-control, 5 – trustworthiness, 6 – adaptability, 7 – innovativeness, 8 – achievement drive, 9- commitment, and 10 – initiative.
6. For assessment of awareness regarding Social Competence, the scale used was: 0-Not at all aware, 1-Little bit aware, 2-Somewhat aware, 3-Well aware, 4-Highly aware. The 10 sub-questions were: 1 – empathy, 2 – service orientation, 3 – developing others, 4 – leveraging diversity, 5 – political awareness, 6 – influence, 7 – communication, 8 – leadership, 9- conflict management, and 10 – building bonds.

### 3.1 Statement of Hypothesis

Ho1: MBA student's awareness of Personal Competence is not significant
Ha1: MBA student's awareness of Personal Competence is significant

Ho2: MBA student's awareness of Social Competence is not significant
Ha2: MBA student's awareness of Social Competence is significant

Data analysis included descriptive analysis specifying features of the sample and the inferential analysis to test the hypothesis. A t-test was used given the fact that the SD of the population is not known, in which case, a Z-test could have been applied. The use of a t-test in practice is widely done as a substitute for the Z-test wherein the SD of the sample is taken as the SD of the population (given unknown population SD).

The survey instrument returned a Cronbach's alpha of 0.93 that is better than 0.70 (the standard) and hence was considered as reliable.

## 4.0 Data Analysis

### *4.1 Descriptive analysis*

244 out of the 400 respondents were male and 156 were female. In terms of specialization, 150 students belonged to Marketing, 129 were Finance students, while 64 were from Operations and 57 were HR students.

### *4.2 Inferential analysis*

The null hypothesis was set as the sample mean (x̄) equals the hypothesized population mean (μ). The alternate hypotheses were: (Ha1 and Ha2) x̄<2.

Summary of the ratings for awareness regarding Personal Competence are given in Table 1 below:

**Table 1: Summary of responses for awareness regarding Personal Competence**

| Sub-questions | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average Awareness | 1.34 | 1.33 | 1.32 | 1.30 | 1.34 | 1.31 | 1.29 | 1.26 | 1.24 | 1.25 | 1.30 |

Summary of the ratings for awareness regarding Social Competence are given in Table 2 below:

**Table 2: Summary of responses for awareness regarding Social Competence**

| Sub-questions | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average Awareness | 1.27 | 1.34 | 1.28 | 1.31 | 1.24 | 1.25 | 1.25 | 1.28 | 1.26 | 1.21 | 1.27 |

Table 3 shows the testing of the hypothesis at 95% confidence level.

**Table 3: Testing of the hypothesis**

| Parameter | H1 value | H2 value |
|---|---|---|
| Sample Mean ($\bar{x}$) | 1.30 | 1.27 |
| Hypothesized population mean (μ) | 2 | 2 |
| SD of sample | 1.10 | 1.05 |
| n (sample size) | 400 | 400 |
| t-value=abs(($\bar{x}$ - μ) / (s/√n)) | 12.69 | 13.90 |
| p-value =tdist(t,(n-1),1) | 0.000 | 0.000 |
| Decision | Reject Null | Reject Null |

The null hypothesis was rejected in favor of the alternate that the sample mean is significantly different from the hypothesized population mean.

## 5.0 Discussion and conclusion

### *5.1 Discussion on results*

The sample mean for awareness about Personal Competence was 1.30 (SD = 1.10) and was significantly lower than the midpoint of the awareness scale. Moreover, the sample mean for awareness about Social Competence was 1.27 (SD = 1.05) and was significantly lower than the midpoint of the awareness scale.

### *5.2 Conclusion*

The research shows that awareness regarding Emotional Competence is significantly lower. As MBA students step out of their schools and enter the corporate world emotional competence will be a key attribute for their success or failure. It is evident that special efforts must be made to familiarize students with Emotional Competence.

As the study was based on sampling, limitations of sampling, in general, apply to it as well. Researchers are encouraged to study factors that drive such lower awareness which will add to the earlier research.

## References

1. Bullen P B (2016). How to choose a sample size (for the statistically challenged). Accessed from http://www.tools4dev.org/resources/how-to-choose-a-sample-size/
2. Cornell, C., Kiernan, N., Kaufman, D., Dobee, P., Frydenberg, E., & Deans, J. (2017). Developing social emotional competence in the early years. In *Social and emotional learning in Australia and the Asia-Pacific* (pp. 391-411). Springer, Singapore.
3. Domitrovich, C. E., Durlak, J. A., Staley, K. C., & Weissberg, R. P. (2017). Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. *Child development*, *88*(2), 408-416.
4. Humphries, M. L., Williams, B. V., & May, T. (2018). Early childhood teachers' perspectives on social-emotional competence and learning in urban classrooms. *Journal of applied school psychology*, *34*(2), 157-179.
5. Ikävalko, H., Hökkä, P., Paloniemi, S., &Vähäsantanen, K. (2020). Emotional competence at work. *Journal of Organizational Change Management*.

BENGAL, PAST & PRESENT

Journal of the Calcutta Historical Society.

# CERTIFICATE OF PUBLICATION

This is to certified that the article entitled

EMERGING ECONOMIC PROBLEMS BEFORE INDIA

Authored By

Dr. Rijwan M. Shaikh
Associate Professor, Sinhgad Institute of Management, Vadgaon Bk., Pune.

Published in Vol. 140, Issue: (V) January-June 2022
Bengal, Past and Present with ISSN : 0005-8807
UGC-CARE List Group I

Impact Factor: 5.70

3rd National Level Students' Research Conference on "Innovative Ideas and Inventions in Computer Science & IT with Its Sustainability"
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

# Farming as a Service (FaaS) Through IoT Based Indo Green Agri Drone

Mr. Shubham Kaundinya[1], Miss Vaishnavi Pande[1], Dr. Ankush Kudale[2] [1]Student, MCA-III, Sinhgad Institute of Management, Pune, Maharashtra, India [2]Assistant Professor, MCA, Sinhgad Institute of Management, Pune, Maharashtra, India

ABSTRACT

This paper describes the problems may be caused by poor management and organization within the scheme and poor technology usage in the agriculture. Latest technology developments have turned present-day unmanned systems into realistic alternatives to traditional water supply survey methods. Technological Solution: Flying robot suitable for monitoring water leakages to canal system of irrigation. We describe the technical requirements for each of these monitoring types and discuss the operational aspects. The selection of a specific sensor/platform combination depends critically on the target species and its behavior. The technical specifications of unmanned platforms and sensors also need to be selected based on the surrounding conditions of a particular project, such as the area of interest, the survey requirements and operational constraints.

Keywords: Sensors, Remote control, Motor, Electronic speed control, battery, Radio transmitter and receiver.

## I. INTRODUCTION

This document is completing as from my winter 2021 distributed research experience as a postgraduate student. We will promote the product as per the identified customers and will make sure that sufficient research has been done in identifying the needs of the customers. Since, this product will be a unique product in the market, we will make sure that customers are satisfied by the facts we present before them, and so we will do the best for identifying and presenting the best part of the product. It can applicable to drip irrigation agriculture farming also.

## II. BACKGROUND AND LITERATURE REVIEW

Complexities in water distribution for the use of Agriculture through irrigation canal. effecting in water wastage and farmers crises against water distribution authority. The objectives of this project is as under -

- Identify leakages to canal of water supply.
- Measure the quantity of water supplied to agriculture farm and actual water received in farm.
- Billing of water supply at actual water received in farm.
- Quantify water consumption pattern by farm and by crop.

## III. METHOD AND MATERIALS

A. About material:

1) Actuators and motors
2) Sensors
3) Software –Python 4) Remote control 5) Frame.
6) Motor
7) Electronic Speed Control (ESC) 8) Flight Control Board.
9) Radio transmitter and receiver.
10) Propeller (2 clockwise and 2 counter-clockwise)
11) Battery & Charge

Fig. 1 Agri Drones

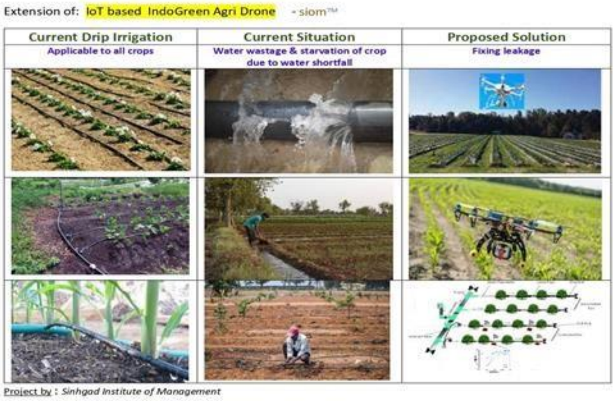

Extension of: IoT based IndoGreen Agri Drone - siom™
Current Drip Irrigation
Applicable to all crops
Current Situation
Water wastage & starvation of crop due to water shortfall
Proposed Solution
Fixing leakage
Project by : Sinhgad Institute of Management

Fig. 2 Extension of Agri Drones

B) About method (Agriculture Drone system using GPS): In this device there are eleven content. We intend to protect your idea we will apply for provisional patent with prior art and claims to patent and trademarks office Govt. of India.

There is no any competitors for this in market. Technical specifications of this drone is as under a drone is bit more complex than the accepted definition of a thing in the IoT. The flying Drones can be considered as IoT. Drones are currently used in two standard agricultural applications tracking and distribution.Tracking (and subsequent analysis) is used in both plant and livestock agriculture and helps farmers understand the status, resources, and productivity of their farms. Distribution using drones involves physically moving resources across a farm, including spreading agricultural chemicals such as leakage of cannal water. The Agriculture Wonder Drone System is designed by making use of GPS where the automatically controlled drone based on aerial leakage of water cannel. Where the drone was behaved at required altitude, and then it is switch to altitude hold mode, which maintains the same altitude until it is switched back.

## IV. MARKET AND SALES ANALYSIS

We have already developed device which has been published in IPR gazette of Govt. of India and team of researchers and students whom have experience of research projects and execution, implementation experience.

Our team capacities:

Sales: For sales we will contact irrigation department of government also for farmer’s community.

Marketing: we will go for digital marketing as well as agricultural exhibition, news and media.

Operations: For development and implementation we have microcontroller development expert in industry and required peripherals we will import and assemble and also for software development we have MCA students and Alumni who will work on this project.

Technical Knowledge: In case of technical knowledge we will ask to incubation support and also industry experts for more technical details.

Finance: we are in process of searching the funding agency or interested business or startup who can help is for raising funds for developing this device.

## V. GENERATE PYTHON-PACKAGE AND PYTHON-CODE

Setting Up the Path for Windows:

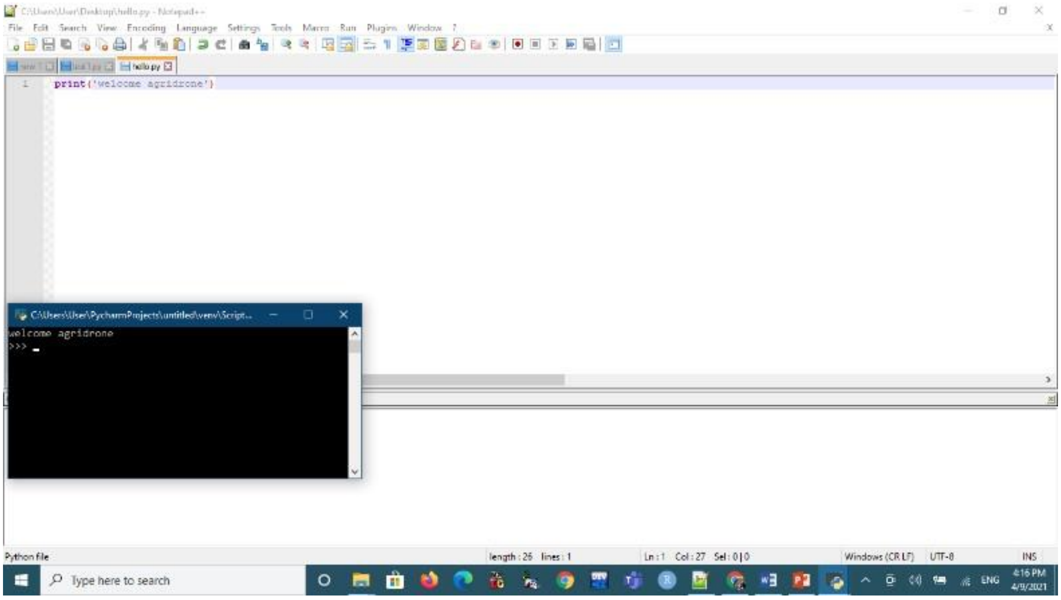
Assuming you have installed python in c:\Program Files\python\python32-37 Right-click on 'My Computer' and select 'Properties'.

Click the 'Environment variables' button under the 'Advanced' tab.

Now, alter the 'Path' variable so that it also contains the path to the python executable.

Example, if the path is currently set to 'C:\WINDOWS\SYSTEM32', then change your path to read 'C:\WINDOWS\SYSTEM32; c:\Program Files\python\python32-37

If you use bash as your shell, then you would add the following line to the end of your '.bashrc: export PATH=/path/to/python:$PATH' Program:



C:\Users\User\Desktop\hello.py - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
hello.py

```
print('welcome agridrone')
```

C:\Users\User\PycharmProjects\untitled\venv\Script...

```
welcome agridrone
>>> _
```

Python file
length : 26 lines : 1
Ln : 1 Col : 27 Sel : 0 | 0
Windows (CR LF) UTF-8 INS
Type here to search
ENG
4:16 PM
4/9/2021

Fig. 3 Output: welcome Agridrone

## VI. RESULT

The idea of execution is simple. A risk involved is mainly with the trust of customers. Flying Drone will designed and operationalized. Break-even point –No Loss no Profit –complete for social cause for initial 2 years.

After 2 years based on utility it has been estimated on 10% on manufacturing cost

For 1 flying Drone cost approx. 557000/- and initial we will develop one drone as a pilot project. Rs.500000/- (Rs. Five Lakh only) required from Funding agency/Incubation center as a support seed money. Balance fund will raise aprox.Rs.57000/- we will ask to other service providers and industry/business partners/vendors who are interested to contribute social as well funding agencies.

Below support required from incubator center apart from funds, Mentoring, Technical support for development, Government permissions, Peripherals space, IP protection.

| Sr. | Particulars | Ist Year(Rs.) | 2nd Year(Rs.) | Total(Rs.) |
|---|---|---|---|---|
| 1 | Drone peripherals | 100000 | 100000 | 200000 |
| 2 | Survey | 5000 | 5000 | 10000 |
| 3 | Assembly | 25000 | 25000 | 50000 |
| 4 | Legal permission | 0 | 5000 | 5000 |
| 5 | Software | 25000 | 0 | 25000 |
| 6 | Salaries: | 70000 | 70000 | 140000 |
| 7 | Supporting Technical Staff | 60000 | 1000 | 61000 |
| 8 | Expert | 25000 | 25000 | 50000 |
| 9 | Books | 2000 | 2000 | 4000 |
| 10 | Travel | 5000 | 5000 | 10000 |

| | | | | |
|---|---|---|---|---|
| 11 | Other staff, if any | 1000 | 1000 | 2000 |
| | Total-> | 318000 | 239000 | 557000 |

Table 1 Estimation

## VII. CONCLUSIONS

Our device is very important to our country and Government also because it has various techniques to handle their work. It has complex structure and Lightweight size. The device has been successfully Carry required work of area of fix customer problems. With the help of IoT they can access all information. In this manuscript different types of system useful for Agriculture wonder drone system using electronic speed-controller and Agriculture drone system using GPS were discussed. Mainly the paper focused on selection of best compatible design for Drone system for Agriculture purpose. Some of the exiting implementation was discussed with their advantages and disadvantages. In line to this the experimentation and expected result also discussed for further implementation.

## VIII. ACKNOWLEDGMENT

I would like thank Dr. Ankush Kudale sir who guided us for completion of this idea for how to build and how to use of this device in future. I would also like thanks for my all faculty member and respective departments throughout the completion of this idea and how to develop more and more information will update for sensors and robotic device.

## IX. REFERENCES

[1]. Prof. K. B. Korlahalli, Mazhar Ahmed Hangal, Nitin Jituri, Prakash Francis Rego, Sachin M. Raykar, "An Automatically Controlled Drone Based Aerial Pesticide Sprayer", Project Reference No.39S_BE_0564. [1][2] [2]. S. R. Kurkute, C. Medhe, A. Revgade, A. Kshirsagar, "Automatic Ration Distribution System A Review". Intl. Conf on Computing for Sustainable Global Development, 2016. [2][3]

[3]. Vardhan, P. H., Dheepak, S., Aditya, P. T., & Arul, S. (2014) "Development of Automated Aerial Pesticide Sprayer." International Journal of Engineering Science and Research Technology, vol 3, issue 4.[1][2]

[4]. Aditya S. Natu., Kulkarni, S., C. (2016) "Adoption and Utilization of Drones for Advanced Precision Farming: A Review." published in International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169, Volume: 4 Issue: 5 PP.563 – 565[1[2]

[5]. https://www.geeksforgeeks.org/python-programming-language/[2[4]

[6]. https://en.wikipedia.org/wiki/Agricultural_drone [1]

[7]. https://www.engpaper.com/agriculture-drone.html[1]

[8]. Swapnil R. Kurkute, Dipak Patil, Priyanka V. Ahire, Pratikha D. Nandanvar, "NFC Based Vehicular Involuntary Communication System", International Journal of Advanced Research in Computer Science, ISSN No. 0976-5697 Volume 8, No. 5, May-June 2017.[1][2]

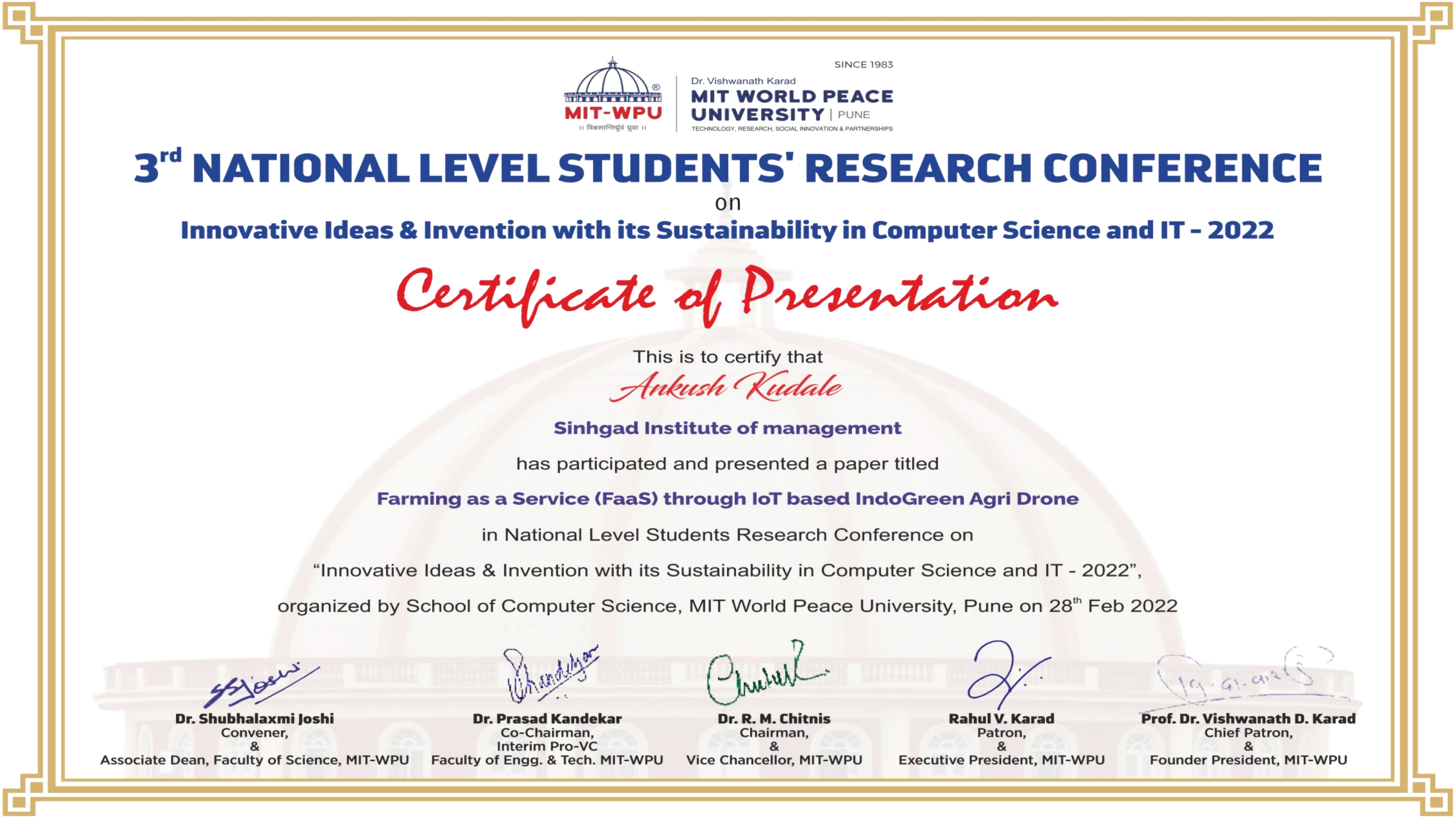
[9]. Abdullah Tanveer, Abhishek Choudhary, Divya Pal, Rajani Gupta, Farooq Husain, "Automated Farming using Microcontroller and Sensors". International Journal of Scientific Research and Management Studies

(IJSRMS), ISSN: 2349371, Volume 2, Issue 1, Page No.-21-30[1][2]

Sinhgad Institute of Management
PUNE
411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM



MIT-WPU

॥ विश्वशान्तिर्ध्रुवं ध्रुवा ॥

SINCE 1983

Dr. Vishwanath Karad

MIT WORLD PEACE UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

# 3rd NATIONAL LEVEL STUDENTS' RESEARCH CONFERENCE

on

## Innovative Ideas & Invention with its Sustainability in Computer Science and IT - 2022

## Certificate of Presentation

This is to certify that

*Ankush Kudale*

**Sinhgad Institute of management**

has participated and presented a paper titled

**Farming as a Service (FaaS) through IoT based IndoGreen Agri Drone**

in National Level Students Research Conference on

"Innovative Ideas & Invention with its Sustainability in Computer Science and IT - 2022",

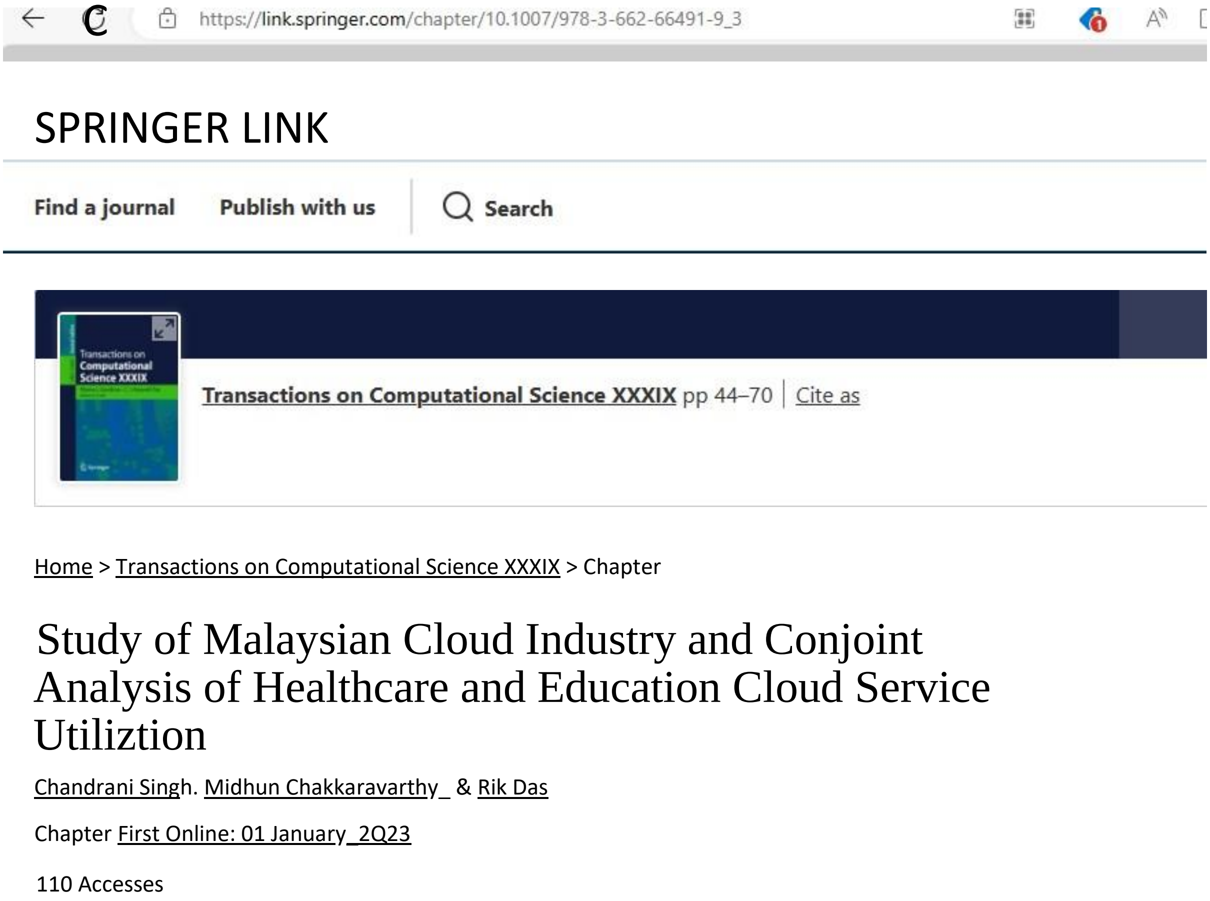
organized by School of Computer Science, MIT World Peace University, Pune on 28th Feb 2022

| | | | | |
|---|---|---|---|---|
| SSJoshi | | | | |
| **Dr. Shubhalaxmi Joshi**<br>Convener,<br>&<br>Associate Dean, Faculty of Science, MIT-WPU | **Dr. Prasad Kandekar**<br>Co-Chairman,<br>Interim Pro-VC<br>Faculty of Engg. & Tech. MIT-WPU | **Dr. R. M. Chitnis**<br>Chairman,<br>&<br>Vice Chancellor, MIT-WPU | **Rahul V. Karad**<br>Patron,<br>&<br>Executive President, MIT-WPU | **Prof. Dr. Vishwanath D. Karad**<br>Chief Patron,<br>&<br>Founder President, MIT-WPU |

Sinhgad Institute of Management
PUNE
411 041.

Singh

Dr.Chandrani Singh ,Director –MCA,SIOM



https://link.springer.com/chapter/10.1007/978-3-662-66491-9_3

SPRINGER LINK

Find a journal Publish with us Search

Transactions on Computational Science XXXIX pp 44–70 | Cite as

Home > Transactions on Computational Science XXXIX > Chapter

# Study of Malaysian Cloud Industry and Conjoint Analysis of Healthcare and Education Cloud Service Utiliztion

Chandrani Singh. Midhun Chakkaravarthy_ & Rik Das

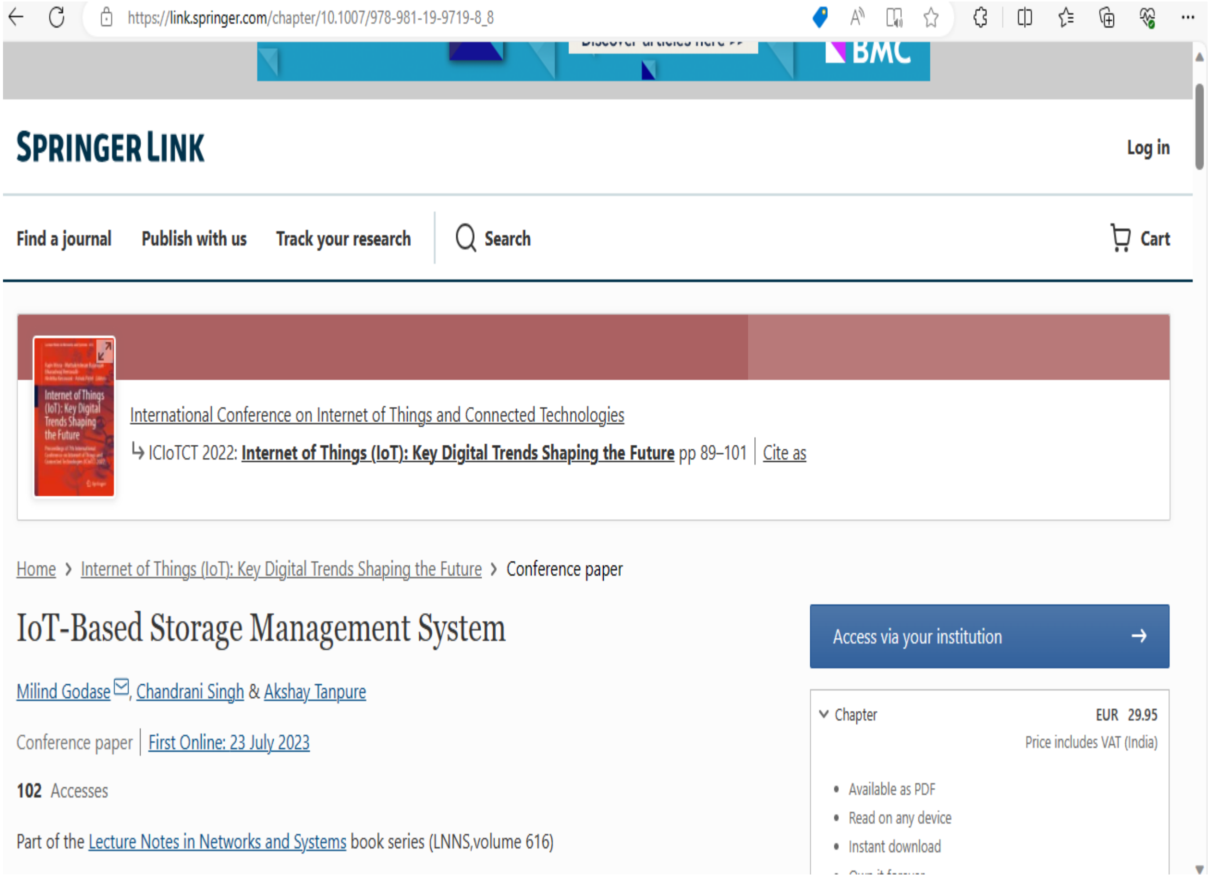
Chapter First Online: 01 January_2Q23

110 Accesses

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM



https://link.springer.com/chapter/10.1007/978-981-19-9719-8_8
Discover articles here >>
BMC
SPRINGER LINK
Log in
Find a journal
Publish with us
Track your research
Search
Cart

International Conference on Internet of Things and Connected Technologies

↳ ICIoTCT 2022: **Internet of Things (IoT): Key Digital Trends Shaping the Future** pp 89–101 | Cite as

Home › Internet of Things (IoT): Key Digital Trends Shaping the Future › Conference paper

# IoT-Based Storage Management System

Milind Godase, Chandrani Singh & Akshay Tanpure

Conference paper | First Online: 23 July 2023

**102** Accesses

Part of the Lecture Notes in Networks and Systems book series (LNNS,volume 616)

Access via your institution →

Chapter EUR 29.95

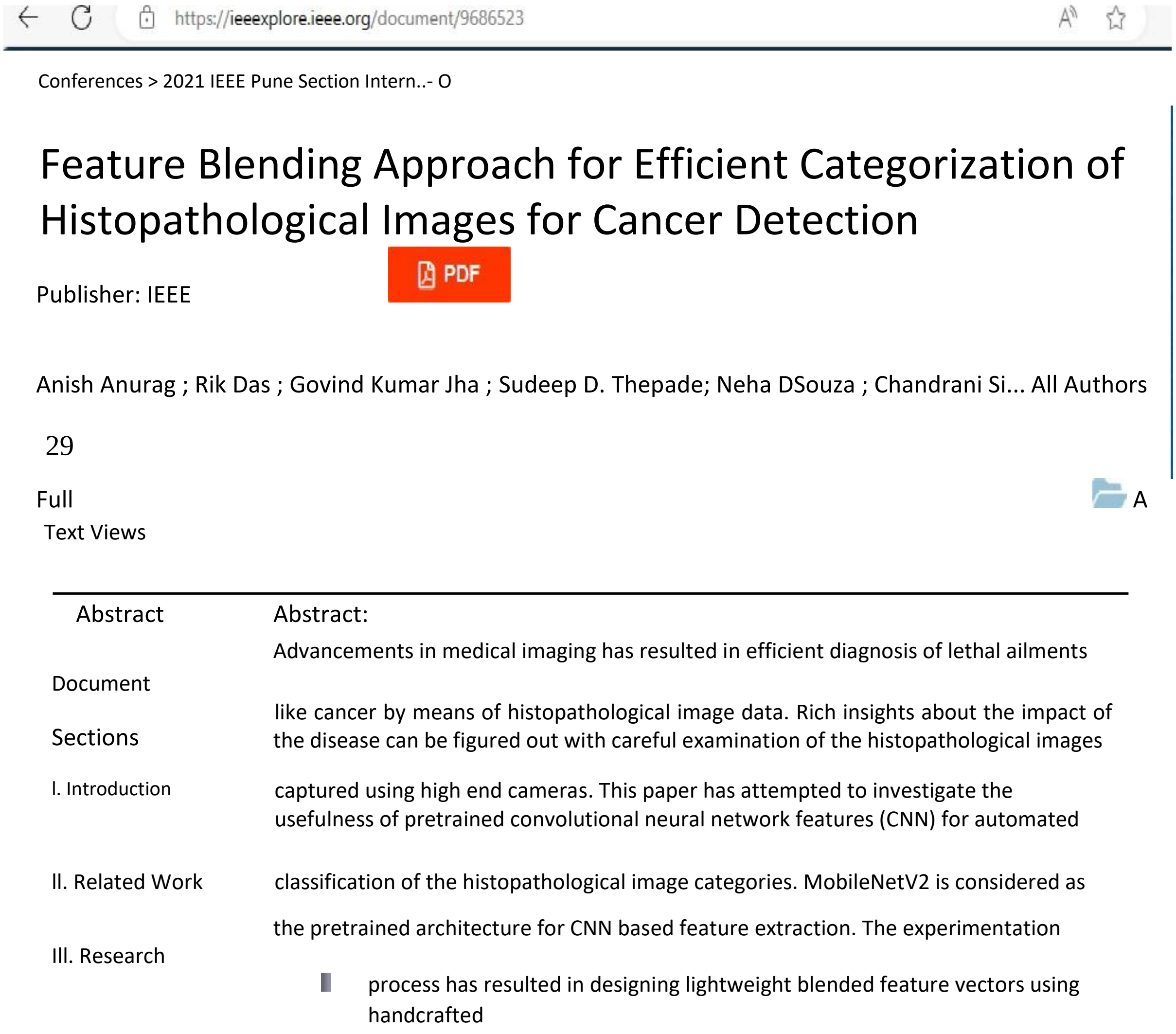
Price includes VAT (India)

- Available as PDF
- Read on any device
- Instant download

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM



https://ieeexplore.ieee.org/document/9686523

Conferences > 2021 IEEE Pune Section Intern..- O

# Feature Blending Approach for Efficient Categorization of Histopathological Images for Cancer Detection

Publisher: IEEE

Anish Anurag ; Rik Das ; Govind Kumar Jha ; Sudeep D. Thepade; Neha DSouza ; Chandrani Si... All Authors

29

Full
Text Views

Abstract

Document

Sections

I. Introduction

II. Related Work

III. Research

Abstract:

Advancements in medical imaging has resulted in efficient diagnosis of lethal ailments like cancer by means of histopathological image data. Rich insights about the impact of the disease can be figured out with careful examination of the histopathological images captured using high end cameras. This paper has attempted to investigate the usefulness of pretrained convolutional neural network features (CNN) for automated classification of the histopathological image categories. MobileNetV2 is considered as the pretrained architecture for CNN based feature extraction. The experimentation

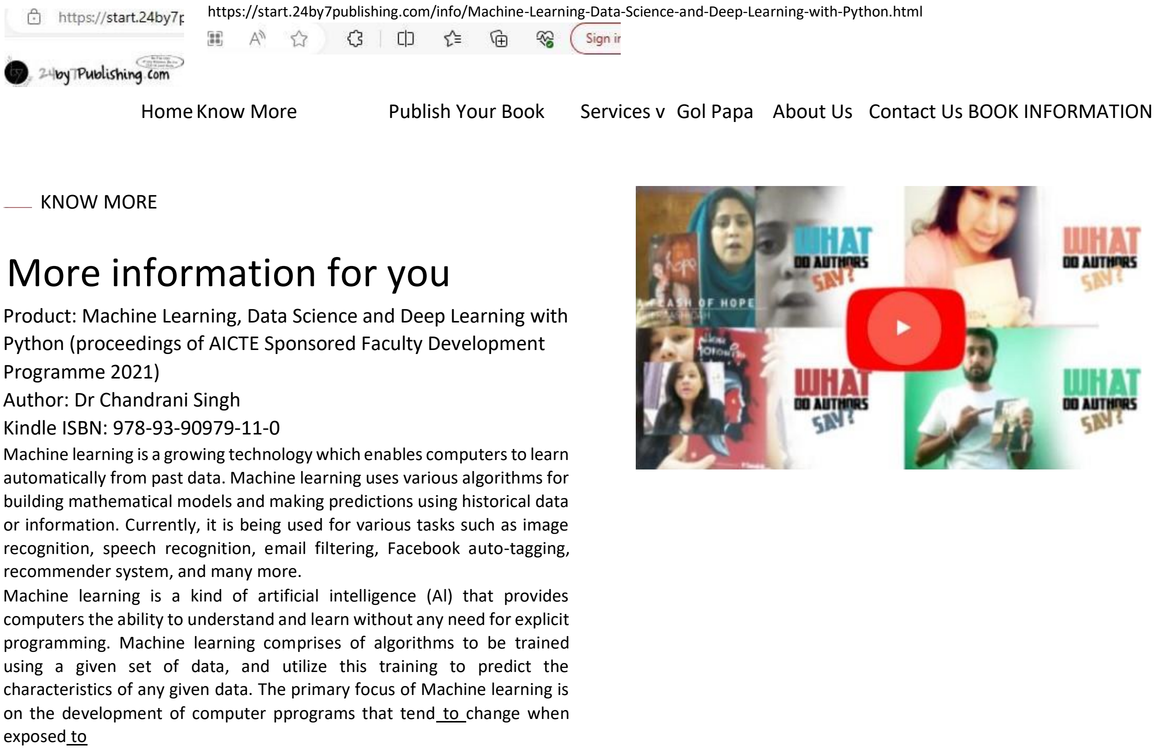
process has resulted in designing lightweight blended feature vectors using handcrafted

Sinhgad Institute of Management
PUNE
411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

https://start.24by7publishing.com/info/Machine-Learning-Data-Science-and-Deep-Learning-with-Python.html



KNOW MORE

## More information for you

Product: Machine Learning, Data Science and Deep Learning with Python (proceedings of AICTE Sponsored Faculty Development Programme 2021)

Author: Dr Chandrani Singh

Kindle ISBN: 978-93-90979-11-0

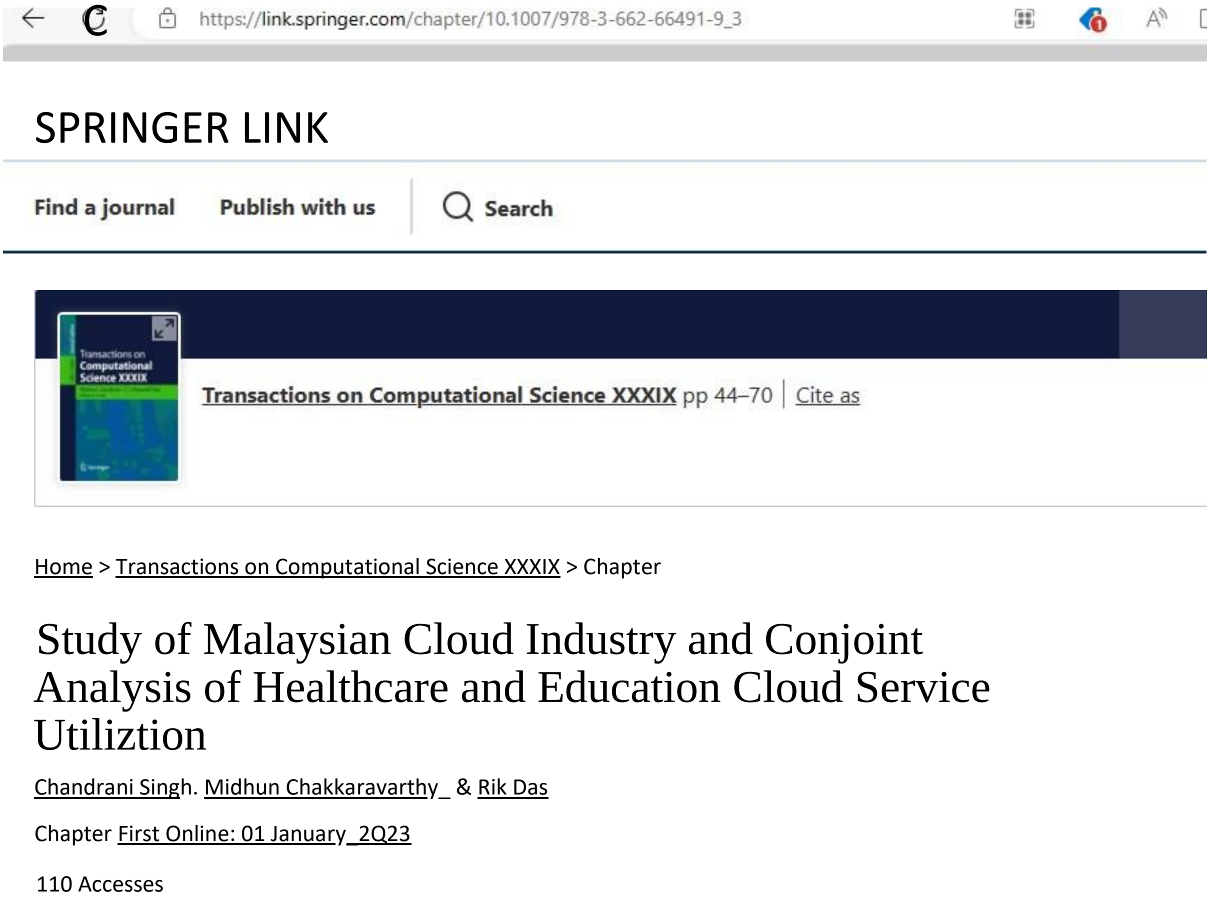
Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more.

Machine learning is a kind of artificial intelligence (AI) that provides computers the ability to understand and learn without any need for explicit programming. Machine learning comprises of algorithms to be trained using a given set of data, and utilize this training to predict the characteristics of any given data. The primary focus of Machine learning is on the development of computer pprograms that tend to change when exposed to

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM



https://link.springer.com/chapter/10.1007/978-3-662-66491-9_3

SPRINGER LINK

**Transactions on Computational Science XXXIX** pp 44–70 | Cite as

Home > Transactions on Computational Science XXXIX > Chapter

# Study of Malaysian Cloud Industry and Conjoint Analysis of Healthcare and Education Cloud Service Utiliztion

Chandrani Singh. Midhun Chakkaravarthy_ & Rik Das

Chapter First Online: 01 January _2Q23

110 Accesses

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

# HOLISTIC ANALYSIS OF PROBLEM SPACES, ALGORITHMIC SOLUTIONS AND ITS CORRESPONDING ICT IMPLEMENTATION IN COMPUTATIONAL CHEMISTRY

**Dr. Sunil Khilari**[*]

[*]**Sinhgad Institute of Management, Pune**

## ABSTRACT

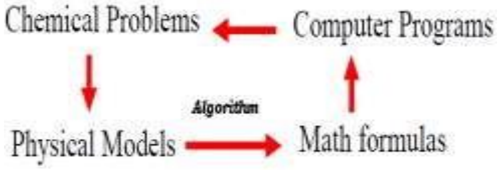
Digital Transformation and scientific computing will help to address some of the challenges in computational chemistry and process systems particularly computational tasks that scale exponentially with various computational problems from chemistry domain. To design Knowledge Management ( KM) base therefore has a lot to offer to a chemistry lab. Our main purpose is to identify computational problems of high priority to progress in chemical knowledge management initiatives that should be undertaken with support provided in the development of open source ICT tools for the computational chemistry domain. This paper presents algorithms, software development and computational complexity analysis for problems arising in the Computational Chemistry domain.

This assembled information on problem spaces, algorithmic solution and corresponding ICT implementation will be valuable resource to those charged for sharing, reusing, creating and researching new theories, principles in computational chemistry. This paper also insight the mapping of solution, problem and implementation spaces with interrelated attributes which will help to the chemical and related scientific research and software development community.

**Keywords:** Information Communication Technology (ICT), Knowledge Management(KM),Computational Chemistry(CC), problem space.

## I. COMPUTATIONS IN COMPUTATIONAL CHEMISTRY

Use of methodical approximation and computer programs to obtain results relative to chemical problems. Computational chemistry is simply the application of chemical, mathematical and computing skills to the solution of interesting chemical problems. It uses computers to generate information such as properties of molecules and/or simulated experimental results. Some common computer software used for computational chemistry includes MATLAB, Nlopt , TINKER , Gaussian etc.(8). Also computational chemistry is based on an approximations and assumptions. Computational Chemistry Calculate Energy, Structure, and Properties. Computations of this type are derived directly from theoretical principles, with no inclusion of experimental data. Mathematical approximations are usually a simple functional form for an approximate solution to a differential equation. A mathematical method that is sufficiently well developed that it can be automated for implementation on a computer.



The input data for these computational problems are laboratory experiments, where few lead compounds were recognized. The problem is to produce new laboratory experiments that will rush the likelihood of discovering new, more influential compounds/substances. In order to do so we have to solve inverse problems based on specific indices. One wants several solutions for the inverse problem that are as diverse (i.e. different chemical structure) as possible. Based on them, a new combinatorial library is created, and new lead compounds are discovered. (1).

## WHAT CAN COMPUTE IN CHEMISTRY?

1. Electronic structure determinations
2. Geometry optimizations
3. Frequency calculations
4. Electrostatic potential
5. Enthalpies of formation
6. Orbital energy levels
7. Ionization energy
8. Reaction path
9. Reaction rate
10. Thermodynamic calculations- heat of reactions, energy of activation
11. calculation of electron and charge distributions
12. Molecular geometry

## II. PROBLEM SPACES IN COMPUTATIONAL CHEMISTRY

The problem space, which corresponds to well-defined computational problems, is the center of the computational chemistry problems in research domain.

Storing and searching for data on chemical entities. Identifying correlations between chemical structures and properties. Electronic structure determinations and geometry optimizations. Frequency calculations, transition structures and protein calculations. Computing electron and charge distributions. The prediction of the molecular structure of molecules by the use of the simulation of forces to find stationary points on the energy surface as the position of the nuclei is varied.

Thermodynamic calculations involving heat of reactions and energy of activation

| Sr.No. | Types of calculations | Problems Spaces |
|---|---|---|
| 1 | Molecular Geometry | 1. What is the energy for a given geometry?<br>2. How does energy vary when geometry changes?<br>3. Which geometries are stable?<br>4. How does energy change w/r external perturbation? |
| 2 | Reaction Mechanism | 1. How to represent chemical reactions using balanced chemical equations?<br>2. How to calculate the quantities of material involved in a chemical reaction?<br>3. The original reactants must contain atoms of which |
| | | element? |

| | | |
|---|---|---|
| 3 | Determination of Bond Energies | 1. Fin energy needed to break one mole of the bond to give separated atoms.<br>2. Finding enthalpy changes of reaction from bond enthalpies<br>3. Detect state of Bond Breakage and Formation |
| 4 | Quantum Mechanics | 1. What is mathematical description of the behavior of electrons<br>2. Computing electrostatic properties<br>3. What is attraction of electrons to nuclei |

Table No.1 –Identified Problem Spaces and corresponding computations

## III. CHALLENGES IN COMPUTATIONAL CHEMISTRY DOMAIN

- Invent new algorithms to globally optimize at the worldwide level the use of raw materials, energy, and environmental impact of chemical processes.
- Develop computer methods that will accurately predict the properties of unknown compounds.
- Develop reliable computer methods to calculate the detailed pathways by which reactions occur in both ground states and excited states, taking full account of molecular dynamics as well as quantum and statistical mechanics(8)
- Devise experimental tests to establish the reliability of new theoretical treatments.
- Invent new computer tools and logistics methods to reduce significantly the time needed for commercializing new drugs.
- Develop new and powerful computational methods, applicable from the atomic and molecular level to the chemical process and enterprise level that will enable multiscale optimization.

## IV. ALGORITHMIC SOLUTIONS FOR COMPUTATIONAL PROBLEMS IN CHEMISTRY

Important features of algorithms are fitness, definiteness, input, output and effectiveness

**Finiteness** -it must terminate after finite number of steps.

 **Definiteness** -It must have each and every step of procedure to be precisely defined.

 **Input/output**- Algorithm must communicate to the environment in which it operate

 **Effectiveness** - Algorithms must be practical .i.e. must be capable of implementation

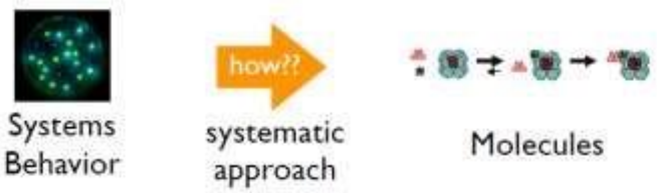
 **Degree of goodness** -algorithms is of its speed of execution/generates correct result

Faster and cheaper computers will extend the range of high-level methods .To begins; we must understand that programs and algorithms are not the same thing. We prove theorems and analyze algorithms, not their implementations.

Algorithmic understanding is defined as the ability to match up or recall an appropriate mathematical formula and a Strategy to compute a numerical answer (5) .We begin with an overview of the algorithm.

The entire computational chemistry problems are solved with the help of algorithmic solutions. Some of these solutions are very crude and others are expected to be more accurate than any experiment that has yet been conducted. There are several implications of this situation.

Computational chemistry end-users require knowledge of each algorithm being used and how accurate the results are expected to be.



Following are the list of computational chemistry algorithms identified and used for solving various computational problems. Some of algorithms are implemented with the help of ICT tools.

| Sr No | Computational Chemistry Algorithm | SrNo | Computational Chemistry Algorithm | SrNo | Computational Chemistry Algorithm | Sr No | Computational Chemistry Algorithm |
|---|---|---|---|---|---|---|---|
| 1 | ab initio Algorithm | 35 | EMBED Algorithm | 69 | Molecular Recognition and Docking Algorithms | 103 | Simple Left to Right (SLR) backtracking algorithm |
| 2 | Accelerated Random Search (ARS) Algorithm | 36 | Euclidean Algorithm | 70 | Monte Carlo Algorithm | 104 | SKETCH Algorithm |
| 3 | ACRB algorithm | 37 | Fletcher±Powell ( FP) Algorithm | 71 | Morgan Algorith m | 105 | SMILES2 Algorithm |
| 4 | ACRN Algorithm | 38 | Floyd’s Algorithm | 72 | Morgan's Algorithm | 106 | Smolyak's Algorithm |
| 5 | Adaptive Substituent Reordering Algorithm (ASRA) | 39 | Floyd-Warshall Algorithm | 73 | MTL Algorithm | 107 | Solovay-Kitaev Algorithm |
| 6 | AHM Algorithm | 40 | Fruchterman and Reingold Algorithm | 74 | Nelder-Mead (NM) Algorithm | 108 | Spearman or Kendall Algorithm |
| 7 | Arvis-Patrick Algorithm | 41 | Gear Algorithm | 75 | NEO Algorithm | 109 | Steepest Decent Algorithms |
| 8 | ASD Algorithm | 42 | Genetic Algorithms | 76 | Nesbet's Algorithm | 110 | Stochastic Simulation Algorithm (SSA) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9 | Beeman's Algorithm | 43 | Gillespie Algorithm | 77 | Netwton-Raphson Algorithm | 111 | Strassen Algorithm |
| 10 | Berny Algorithm | 44 | Google's PageRank Algorithm | 78 | NiceGraph Algorithm | 112 | Support Vetor Machines(SVM) Algorithm |
| 11 | Black Box Algorithms | 45 | GOTS Algorithm | 79 | NLP Algorithm | 113 | SVM Algorithm |
| 12 | BLTF Algorithm | 46 | GPLHR Algorithm | 80 | NOMAGIC Algorithm | 114 | SYMMLQ Algorithm |
| 13 | Branch-and-Bound Algorithm | 47 | Graph-theoretic Algorithm | 81 | NOVA Algorithm | 115 | Tarjan Algorithm |
| 14 | Bron-Kerbosh Algorithm | 48 | Grover’s Algorithm | 82 | Numerical Advection Algorithm | 116 | Tau-leap Algorithm |
| 15 | Bucket Evaluations (BE) Algorithm | 49 | HPVK Algorithms | 83 | Opentox Algorithm | 117 | Teepest Decent Algorithm |
| 16 | CABASS Algorithm | 50 | Hybrid genetic Algorithm (HGA) | 84 | Parareal Algorithm | 118 | Tensor Algebra Algorithm |
| 17 | CANON Algorithm | 51 | IRC Algorithm | 85 | Optimal Damping Algorithm(ODA) | 119 | TG Algorithm |
| 18 | Cauchy's Steepst Descent Algorithm | 52 | Jaguar Algorithm | 86 | PCT Algorithm | 120 | Thomas Algorithm |
| 19 | CHAIN Algorithm | 53 | Jarvis-Patrick Algorithm | 87 | Polak±Ribiere Algorithm | 121 | TNKMS Algorithm |
| 20 | Chameleon Algorithm | 54 | KP Algorithm | 88 | PSO Algorithm | 122 | Tree Algorithm |
| 21 | CLIQUE Algorithm | 55 | KvasnickaPospichal's Algorithm | 89 | Python Algorithm | 123 | Trotter-Suzuki Algorithms |
| 22 | Common Subexpression Elimination (CSE) Algorithm | 56 | Lancsoz Algorithm | 90 | QSSA Algorithm | 124 | TruncatedNewton Algorithm |
| 23 | Conjugate gradient Algorithm | 57 | Lanczos Algorithm | 91 | Quantum Algorithm | 125 | Tversky search Algorithm |
| 24 | Coulomb hole Algorithm | 58 | Lazar Algorithm | 92 | Quantum phase estimation (QPE) Algorith | 126 | Tweaking Algorithm |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 25 | Crank-Nicolson Algorithm | 59 | Leader Algorithm | 93 | quasi-decision Algorithm | 127 | UCK Algorithm |
| 26 | Davidson Algorithm | 60 | Levenberg-Marquardt | 94 | quasi-Newton Algorithms | 128 | Ullman's Algorithm |
| | | | Algorithm | | | | |
| 27 | De Novo Algorithms | 61 | LevenbergMarquardt minimization Algorithm | 95 | Quasi-optimal Algorith | 129 | Universal Chemical Key (UCK) Algorithm |
| 28 | Dijkstra Algorithm | 62 | Mass correction Algorithm | 96 | Random forest Algorithm | 130 | verlet Algorithm |
| 29 | Distancegeometry Algorithms | 63 | McLean Algorithm | 97 | RASP Algorithm | 131 | Vertical Mixing Algorithms |
| 30 | DIIS Algorithm | 64 | MD5 message digest Algorithm | 98 | Reciprocal nearest-neighbor (RNN) Algorithm | 132 | Winograd and Strassen's Algorithm |
| 31 | DNG Algorithm | 65 | Metaheuristic Algorithm | 99 | Sequential Minimal Optimization (SMO) mining Algorithm data | 133 | Winograd's Algorithm |
| 32 | Eades Algorithm | 66 | Metropolis Algorithm | 100 | Roothaan Algorithm | | |
| 33 | Elbert's Algorithm | 67 | MH sampling Algorithm. | 101 | SHAKE and RATTLE Algorithm | | |
| 34 | Ellipsoid Algorithm | 68 | Modified GramSchmidt Algorithm (MGS) | 102 | SHAVITT Algorithm CR | | |

Table No.2 List of Algorithms used in Computational Chemistry domain

Following are some properties of computational chemistry algorithms -

| Sr.No. | Computational Problem Category | Algorithmic Solution | Properties of Algorithm |
|---|---|---|---|
| 1 | Chemical Reactivity | ab initio Algorithm | 1. Chemical Reactivity<br>2. Valance Calculation<br>3. Total Energy Calculations |
| 2 | Reaction Mechanism | ACRB algorithm | 1. chemical Reaction Balancing<br>2. To Know interlocked relations among the coefficients of elements in the reaction functions |
| 3 | Geometry optimization | Berny Algorithm | 1. Rational function optimization |

| | | | |
|---|---|---|---|
| | | | 2. Finding transition structure using the single-ended methods |
| 4 | Chemical Modeling | NOVA Algorithm | 1. Molecular Fragmentation in Quantum Chemical Calculations<br>2. Chemical reaction simulation<br>3. Chemical compound modelling |

Table No.3 Computational Problem space and corresponding algorithmic solutions

## V. ICT IMPLEMENTATIONS OF ALGORITHMIC SOLUTION OF COMPUTATIONAL PROBLEMS OF CHEMISTRY

There are new potential applications from software and Internet-based computing. Many computational chemistry techniques are extremely computer-intensive. Depending on the type of calculation desired, it could take anywhere from seconds to weeks to do a single calculation. There are many calculations, such as ab initio analysis of biomolecules that cannot be done on the largest computers in existence. Likewise, calculations can take very large amounts of computer memory and hard disk space. In order to complete work in a reasonable amount of time, it is necessary to understand what factors contribute to the computer resource requirements. Ideally, the user should be able to predict in advance how much computing power will be needed.(8).

The demand for software development may increase as chemistry and chemical engineering move to new areas in which there are no standard software packages. For Internet-based computing an exciting possibility will be to share more readily new software developments directly from the developers, bypassing the commercial software vendors. Another area of sharing leading to powerful new computational opportunities in the chemical sciences is the use of peer-to-peer computing in the form of sharing unused cycles on small computers. The description of the advantages and limitations of each software package is again a generalization for which there are bound to be exceptions. Some software packages can be run on a networked cluster of workstations as though they were a multiple-processor machine. However, the speed of data transfer across a network is not as fast as the speed of data transfer between them.

The researcher is advised to carefully consider the research task at hand and what program will work best in addressing it. Both software vendors and colleagues doing similar work can provide useful suggestions. Today, advances in software have produced programs that are easily used by any chemist.

| Sr.No. | Algorithm | ICT Tool | ICT Tool Properties |
|---|---|---|---|
| 1 | ab initio Algorithm | Nlopt 2.4.2 | Unix<br>Open Source<br>Fortran |
| 2 | Beeman's Algorithm | TINKER 2.0 | DOS<br>Open Source<br>Fortran77 |
| 3 | Berny Algorithm | Gaussian 3.0 | Windows<br>Open Source<br>C |

| | | | |
|---|---|---|---|
| 4 | NOVA Algorithm | YASARA | Linux<br>Open Source<br>Android |

Table No.4 Computational Chemistry algorithms-ICT implementation software's and software properties

There is a big difference in the software packages available for performing these computations. The most complex software packages require an input specifying many details of the computation and may require the use of multiple Input and executable programs.

The most user-friendly software packages require little more work than a molecular mechanics calculation. The price for this ease of use is that the program uses many defaults, which may not be the most appropriate for the needs of a given research project.

## VI. HOLISTIC ANALYSIS

| **SrNo** | **Computational Problem Category** | **Problem Spaces** | **Algorithmic Solution** | **ICT Implementation** | **ICT Tool Properties** | | |
|---|---|---|---|---|---|---|---|
| 1 | Reaction Mechanism | chemical Reaction Balancing | ACRB algorithm | MATLAB | Empirical | trial-anderror | Unix |
| 2 | Chemical Reactivity | Valance Calculation | ab initio Algorithm | Nlopt 2.4.2 | opensource | Fortran | Unix |
| 3 | Molecular Simulation | designed to allow high numbers of particles in simulations of molecular dynamics | Beeman's Algorithm | TINKER 2.0 | opensource | Fortran77 | DOS |
| 4 | Geometry optimization | identify linear connection between gradient and coordinate changes | Berny Algorithm | Gaussian 3.0 | opensource | | Windows |
| 5 | Chemical Modeling | Studying the macroscopic and experimental influences on microscopic structure chemical kinetics and thermal decomposition | NOVA Algorithm | YASARA | opensource | Android | Linux |

Advances in scientific computing will help to address some of the challenges in computational chemistry and process systems engineering, particularly computational tasks that scale exponentially with size. While single-threaded execution speed is important and needed, coordination of multiple instruction multiple data (MIMD) computer systems is rapidly becoming the major challenge in scientific computing. The optimal parallel organization is application dependent: synchronous systems execute multiple elementary tasks per clock cycle while asynchronous models use clusters, vector units, or hyper cubes. On the positive side, parallel computing is becoming almost routine as individual researchers, groups, universities, and companies embrace low-cost cluster parallel computers made from commodity off-the-shelf processors and network interconnects. From a computational science point of view, however, this multiplies the complexity of delivering higher performance computational tools to practicing researchers. Even when the number of parallel supercomputer vendors peaked in the mid-1990s, the number of manufacturers, processors, and network architectures was limited to a handful of such systems; by contrast, the number of possible cluster configurations is enormous.

Holistic Analysis of Problem spaces corresponding algorithmic solution and ICT Implementation.

Table No.5 –Holistic analysis of CC Problems, Algorithms and ICT software with their properties

This study may helpful for changing the nature of computational chemistry software development.

We have to stop learning how to make molecules or materials. There’s plenty more to do in the area, even without considering flow chemistry, cascade reactions, and other such dynamic synthetic strategies (10).

## VII. BENEFITS OF STUDY

1. Computational chemistry has become a useful way to investigate materials that are too difficult to find or too expensive to purchase.
2. It also helps chemistry end-users for make predictions before running the actual experiments so that they can be better prepared for making observations.

## VIII. CONCLUSION

Computational chemistry and process systems engineering play a major role in providing new understanding and development of computational procedures for the simulation, design, and operation of systems ranging from atoms and molecules to industrial-scale processes. Knowledge management (KM) aims to maximize efficiency, nurture creativity, and even enhance coincidence in computational chemistry.

The applications of standard software engineering methods like automatic code generation, simplified the code, reduce the development, testing and debugging time. These are may be the challenge’s to specific development of computational chemistry software. It is now so easy to do computational chemistry that calculations can be performed with the knowledge of the underlying principles. As a result, many people do not understand even the most basic concepts involved in a calculation. To date, the field has neither sufficient tools nor enough trained people to pursue computational chemistry and chemical engineering across all these scales. The field will qualitatively change—in new insights, in what experiments are done and how chemical products and processes are designed—when this is achievable.

## REFERENCES

1. Algorithmic Strategies in Combinatorial Chemistry, Deborah Goldman,Sorin Istrail
2. R. Leach, Molecular Modelling, 2nd Ed.,Prentice Hall, 2001

3. J. Cramer, Essentials of Computational Chemistry, 2nd Ed., Wiley, 2004
4. Bemis, Guy W. and Kuntz, Irwin D., A fast and efficient method for 2D and 3D molecular shape description, J. Computer-Aided Molecular Design, Vol.6 (1992) 607-628
5. BAYRAM COSTU,Algorithmic, Conceptual and Graphical Chemistry Problems: A Revisited Study, Asian Journal of Chemistry, Vol. 22, No. 8 (2010), 6013-6025
6. F. Jensen, Introduction to Computational Chemistry John Wiley & Sons, New York (1999).
7. David C. Young,Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems.Copyright 2001 John Wiley & Sons, Inc.ISBNs: 0-471-33368-9 (Hardback); 0-471-22065-5 (Electronic)
8. H.S. Chan, and K.A. Dill, Physics Today, 46 (2): 24-32 Feb. 1993
9. Robinson, Daniel D.; Lyne, Paul D.; Richards, W. Graham, J. Chem. Inf. Comput. Sci., 40(2), 503-512, 2000.
10. Bruce C. Gibb is in the Department of Chemistry at Tulane University, New Orleans, Louisiana 70118, USA. Nature Chemistry | VOL 4 | October-2012 | www.nature.com/naturechemistry
11. https://www.shodor.org/chemviz/overview/ccbasics.html
12. http://people.chem.ucsb.edu/kahn/kalju/chem226/public/comput_programs.html
13. http://www.chemaxon.com/products/jchem-for-office/
14. http://www.ncbi.nlm.nih.gov/books/NBK207665/

Sinhgad Institute of Management
PUNE
411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

# DEVELOPING AND CULTIVATING AN INNOVATIVE AGRICULTURE 4.0 FARMING SYSTEM

**Dr. Sunil Khilari**[*] [*]Sinhgad Institute of Management, Pune

**ABSTRACT:**

Digital Transformation in Indian agriculture and agricultural applications should follow the roots of the plants. As technology developers look forward to 2022 and beyond, we get through a few ag-tech companies to gain an understanding of the trends that shape mobile app development to transform agriculture. We need to address key structural challenges, such as lack of infrastructure, technology and funding, and the acceptance of digital technologies as these services will be available on mobile phones, applications and the web. There is a basic need for a successful re-establishment of existing agricultural farming systems, combined with new technological advances. Developing new technologies to strengthen Indian agricultural research and production is greatest significant requirements for agricultural growth .In order to recover from economic crisis, natural disasters, Indian farmers are increasingly adopting smart farming technologies such as Farming-as-a-Service (FaaS), Food-as-Service (FaaS), Agriculture Drone-as-a-Service (DaaS), Euipment-asService (EaaS) and Software-as-a-Service (SaaS) models of the Sustainable Agricultural Center to address emerging issues. There is poor access to the agricultural software system available to all agricultural stakeholders and no separate software system has all the FaaS available in one place. In this paper the researcher focuses on the importance of developing new farming practices - such as Services. As a technical solution for all agricultural stakeholders such as farmers, beginners, Farmepreneurs, governments, agribusinesses, machinery suppliers, agronomists and IT dealers etc.

**Keywords:** Farming-as-a-Service (FaaS) , Food-as-a-Service (FaaS), Agriculture Drone-asa-Service (DaaS), Equipment-as-a-Service (EaaS), Software-as-a-Service (SaaS)

## 1. INTRODUCTION

India is primarily regarded as a nation where agriculture and related programs are considered to be the core source of living for more than 80 percent of the population. The portion of agriculture in GDP (GDP) has reached about 20 percent by 2021 due to the stronghold of farming communities among the current diversity. The agricultural domain has been the only one contributing to the positive growth in recent times and the constant provision of basic services has helped to provide food security to Indians and citizens of the world. Agriculture is a livelihood, there is a basic need for the re-establishment of the best farming practices, combined with new technological advances in this sector to ensure sustainability and eliminate scarcity and hunger. Promoting new technologies to strengthen Indian agricultural research and production is greatest important requirements for a sustainable agricultural system. To ensure efficiency, productivity, quality, capacity and continuous supply of basic inputs, Indian farmers are increasingly adopting smart farming technologies using drones and robots. Subsequently with the launch of Farming-as-a-Service (FaaS), various models were developed to develop a sustainable eco-system to address emerging issues in the sector. Awareness, accessibility and access to infrastructure, strong and soft resources for most Indian farmers to date are rare and very serious and there is a great need to present and disseminate ideas, ideas, lessons and research on agricultural use. 4.0 focus on knowledge and technologies that enable farming methods. In this chapter the authors focus on introducing appropriate technological resources that will ensure economic sustainability, enhance food security through data-driven decision making by various stakeholders such as farmers, agribusinesses and agricultural startups, farm entrepreneurs, government and nongovernment. agencies, equipment suppliers, agronomists, forestry IT professionals and retailers. The analyzed information will be used by farmers as a good opportunity to choose the right farming methods to help produce, empower workers to provide timely assistance, and industries to use real-time monitoring using sensors and equipment. The chapter will help to build concepts, methods, processes, benefits and introduce a few scenarios for successfully deploying a farming service approach that will incorporate a payment model as you travel ensures cost-effectiveness and ease of operation.

## 2. AGRICULTURE ECOSYSTEM

The natural ecosystem typically makes up thousands of species of living things and therefore a particular combination in their functioning. In contrast, the agricultural ecosystem is relatively competitive and human-controlled. The agricultural ecosystem is an enduring managed ecosystem, often producing crops and animal feed. Agricultural ecosystems are man-made, and are based on experimentation and performance. A. ecosystem practices within agricultural schemes can deliver services that support the facility of services, including fertilization, pest control, genetic diversity for upcoming agricultural use, soil conservation, soil fertility control and cycling nutrients. So agriculture produces more than just crops and food. Agricultural processes have an environmental impact that affects a variety of ecosystem services, comprising water quality, carbon dioxide, pollination, nutrient circulation, soil conservation, and biodiversity conservation etc.

## 3. AGRICULTURE ECOSYSTEM SERVICES

Agriculture shows an significant role in all human life. Agriculture is the pillar of India's economic system. In addition to providing food and agricultural products, agriculture also provides employment opportunities for the massive majority of the population. Agriculture is important to people because it accomplishes the basic dietary need. It helps people to grow the most suitable food plants and to reproduce suitable animals according to natural features. It also helps people to know how to use the land effectively to prevent disasters. A diversity of agriculture provides people with food and textiles, firewood for living and fuel, plants and tree roots, as well as natural oil and livelihood products. With a disciplined and regulated ecosystem, agriculture shows a significant role in providing and challenging other ecosystem resources. Agriculture provides all five main groups of ecosystem services - Agricultural Services, Provisioning Services, Cultural Services, Management Services, Support Services while also emphasizing support services that allow it to establish. Ecosystem processes of ecosystem services that directly and/or indirectly advantage people and social well-being. The services of the agricultural ecosystem are to benefit people to have access to adequate food and water, human health and well-being depending on these services and the environment from the environment. The ecosystem has many species that cooperate with each other. They are all important parts of the ecosystem.

## 4. CLASSIFICATION OF AGRICULTURE ECOSYSTEM SERVICES

Ecosystem resources from agriculture include water management, climate systems and cultural services, and advanced support services. Agriculture is an significant engine in the Indian economy. It supports the livelihoods of large numbers of people across the country and is essential for rural development and poverty alleviation, as well as food and other agricultural products. A major contest for the agricultural sector is access to adequate agricultural, agricultural and productive food to meet human needs; conserving biodiversity and utilizing natural resources and refining human health and well-being. The improved demand for food and food crops requires careful management of biodiversity and agricultural systems to confirm ecological health and ecosystem services that produce more productivity and less land. The ecosystem will establish wildlife habitats and unique land cover on farms. Ecosystem experience has shown that agricultural management and environmental management systems are an investment for farmers. Benefits and rewards of promoting agricultural development and the implementation of sustainable food production systems. **Table No.1: Integration of Agriculture Ecosystem and Services provided**

| Types of ecosystem à<br>Types of Services ¯ | Agro-ecosystem | Agro-forestry | Canal/Tank ecosystem | climate regulation |
|---|---|---|---|---|
| Farming Services | Farming-as-a-Service(FaaS) | Machinery-as-aService(MaaS) | Food-as-a-Service(FaaS) | Robot-as-a-Service(RaaS) |
| Provisioning Services | Food, medical plants,fiber, bioenergy, | Food, timber,Medical plants,Fiber | silt collection,Food, fiber and | timber, medicinal plants,Fish, firewood,fodder |
| Cultural services | Agro-tourism,aesthetic,landsca pes | Cultural and amenity | Festivals and other recreational | Ecotourism |
| Regulating services | Soil conservation,Air quality and climate regulation | Carbon sequestration,biodrainage,natural hazard regulation,air quality | Ground water recharge,Soil and water conservation, flood control, surface | Carbon sequestration, waste assimilation, nutrient recycling,protection , shore-line protection |
| Supporting services | Biodiversity conservation,soil enrichment,wildlife habitat,soil fertility | Biodiversity conservation, nutrient cycling | Cropping diversity | Fish breeding nursery (ground) |

*Source: Compiled from seminar proceedings of ICAR-National Institute of Agricultural Economics and Policy Research*

India has the second largest population in the world and is considered the world's largest diversity of climates. In the past, agriculture was widely regarded as the leading basis of revenue for the unemployed in rural areas, as a result of which donations from the agricultural sector were overlooked and regarded as a small technology industry that could not afford to contribute significantly to the economy and development. . So many people in agriculture are forced to move to nearby cities in pursuit of work. Farmers who have the heart and the heart of the economy are now struggling throughout India to sell their product at affordable prices. They work day and night to cultivate fine plants, but they often go to bed empty-handed.

The fact is that the Indian Agricultural Industry contributes 13.7% to GDP. It provides food to 1.30 Billion people and is the seventh major exporter of agricultural food in the world. Currently more than 52% of Indians are involved in agriculture.

There are many reasons for Farmer’s financial distress. Some of them are as under-

1. Repeated crop failures
2. Lack of access to insurance credit system
3. Poor government support
4. Poor food security
5. Poor technology available for agriculture sector
6. Lack of availability of irrigation water
7. Unfavorable climatic conditions
8. Lack of transparency in supply chain
9. Inadequate and poorly distributed rainfall
10. Frequent crop failures

Due to the above problems there are growing pressures from climate change, soil erosion and loss of biodiversity, pests and diseases as well as consumer fluctuations in food and concerns about how they are produced. Although modern agriculture offers many technological solutions, the result is not always the same because each farm is different: different location, soil, existing technology and potential production.

Responses to major global changes such as population growth, changes in eating habits, and climate change. It is important for decision-makers to understand the possible trade-offs between these goals in order to achieve equity and environmental impact. On the other hand in India the ancient farm practices are based on the values of sustainability, innovative farmers, hard workers and entrepreneurs. We as Indians need a change in our way of thinking and thinking about agricultural transformation and consider the great challenges of the 21st century by: Innovation & Technology in agriculture

1. Employment Generation
2. Sustaining food
3. Nutrition security
4. Mitigation of climate change
5. Sustainable use of critical resources - water, energy and land.

All of these challenges underscore the need for a new agricultural vision as we move forward in the 21st century.

The paradigm shift should start by altering the mindset of policymakers from shifting productivity to focusing on sustainability. Dissimilar the green revolution, the technology bundle under agro-ecology couldn’t be the same. Each farm is different. Therefore, farmers should be uncovered to certain basic agro-ecology principles and simple strategies, and then be left free to explore, develop and apply what is relevant to local conditions.

Changing the view that this sign is simple, but basic. If we are interested in development, and if we acknowledge that development is about change, let us not worry too much about the provision of new information and technologies from research and instead focus on the conditions necessary to seek and use information to bring about that change. There are now a number of programs that are part of agricultural development, many of which highlight their use of innovation, so it is not possible to list them all here.

## 6. AGRICULTURE 4.0

Agriculture 4.0 is a term for the succeeding major trends facing the agricultural industry, which include a strong focus on accurate agriculture, internet of things (IoT) and big data use to drive greater business success in the aspect of population growth and climate change.

Agriculture 4.0 is the term for the succeeding key trends facing the industry, which include a strong focus on accurate agriculture, Internet of Things (IoT) and big data use to sustain greater business success in the face of population growth and climate change.

In 2018, the World Government Summit published their testimony entitled Agriculture 4.0 - The Future of Agricultural Technology, in collaboration with Oliver Wyman. The report discourses four key developments that put pressure on agriculture in the upcoming: Population census, Lack of natural resources, Climate change, and food insecurity. The term "Agriculture 4.0" has entered public awareness.

**Table No.2 : Agriculture 4.0 service's**

| Technology Type | Category of Service | Device/System | Use Case |
|---|---|---|---|
| of<br>Internet Things (IoT) | • FaaS<br>• ADaaS<br>• EaaS<br>• RaaS<br>• MaaS | • Drone<br>• Robot<br>• Autonomous tractor<br>• Sensors<br>• pH probe<br>• Capacitance hygrometer | • greenhouse monitoring<br>• Drip Irrigation leakage monitoring<br>• Canal Water Supply<br>• Plant & Soil Management |
| Blockchain | • FaaS<br>• ADaaS<br>• EaaS<br>• RaaS<br>• MaaS | • Farm Management Software (FMS)<br>• Immutable ledger system<br>• | • Farm Inventory Management<br>• Agricultural Supply Chain<br>• Microloans<br>• Agricultural Subsidies<br>• Payment from Consumer to Farmers |

| | | | |
|---|---|---|---|
| Artificial Intelligence(AI) | • FaaS<br>• ADaaS<br>• EaaS<br>• RaaS<br>• MaaS | • AI Sensors<br>• AI Chabot's | • detecting diseases in plants<br>• pests, and poor plant nutrition on farms<br>weather forecasting<br>• improve crop yields<br>•<br>• reduce food production costs |
| Data Science | • FaaS<br>• ADaaS<br>• EaaS<br>• RaaS<br>• MaaS | • MyCrop<br>• real-time system<br>• RFID chip | • Optimize their production cycles<br>• Yield Predictions<br>• Digital Soil and Crop Mapping<br>• Fertilizers Recommendation |

In order to meet the above challenges and seize the opportunity we will need the concerted effort of governments, investors, and new agricultural technologies. Agriculture 4.0 will no lengthier rest on on the use of water, fertilizer, and pesticides alike in all grounds. Instead, farmers will use the smallest required amounts and direct the most specific areas. Farms and agricultural activities will have to be conducted in a very different way, mainly due to advances in technologies such as sensors, metals, machinery and information technology. Future agriculture will use sophisticated technologies like robots, temperature and humidity sensors, aerial photography, and GPS technology. These advanced equipment and accurate agricultural systems and robots will permit farms to be additional profitable, efficient, safe, and environmentally friendly.

You are friendly. By using smart data, farmers can better understand their output practices and understand what changes can produce the greatest value. Agriculture 4.0 is more than just a movement. The term has come to be used as a term to capture everything for the next step forward in agriculture: an intelligent, highly efficient industry that utilizes fully data-intensive and new technologies to benefit the entire supply chain.

## CONCLUSION

New technologies have already disrupted traditional farming practices, in a way that was previously unprofitable and equipment that is now accessible and used on farms throughout India. IoT-based drones provide a third eye in the sky, an insect investigation in the field that requires extra attention. Modern advances in sensory technology mean that robots, drones, and Chabot are now able to use the higher wavelengths to detect plants, insects, insects, weeds and sick plants from the air.

Blockchain-like technology is also growing, creating a new way of collaborating on a supply chain. This technology eliminates the need for a consultant. And that's not all. Blockchain can reduce unemployment and significantly improve food safety and security. Tracking is also being improved, with regulators being able to quickly track food sources and determine the extent of any pollution and market problems. These and other technologies for AI, Data Science, IIoT, Quantum Computing, RFID etc. The innovations serve as an important distraction, driving force and the great efficiency of thousands of areas from agriculture such as fishing, poultry farm, dairy farming etc. Awareness, accessibility and access to infrastructure, strong and efficient resources for most Indian farmers to date are rare and extremely critical and there is a great need to present and disseminate ideas, ideas, lessons and research on the use of Agricultural 4.0 with a focus on agricultural technology artificial intelligence (AI), IoT, Data Science, Blockchain, Quantum Computing provides farmers with a way to broader automation of manual agri work. successful IT companies also play a key role in redefining the agricultural sector through innovative solutions such as AI, Data Science, IIoT, Quantum Computing, RFID and - FaaS, DaaS, EaaS, RaaS, MaaS and performing well with advanced access to technology, finance and business skills.

A new wave of new technologies is moving towards the agricultural sector and it looks very similar to the disruption of technology that brings everywhere tools such as communication platforms to advanced technologies such as self-driving tractors, digital farming, and participants will have more. of new solutions to choose from in the coming years.

## REFERENCES

1. Kapil Shah, (January, 2021),” The future of Indian agriculture lies with ‘atmanirbhar' farmers” , https://www.downtoearth.org.in/blog/agriculture/the-future-of-indian-agriculture-lies-with-atmanirbhar-farmers75228
2. Ramesh Chand,( December,2019,”102 Annual Conference Indian Economic Association (IEA) Transforming Agriculture for Challenges of 21st Century”, December,2019
3. R. Hinz1 , T. B. Sulser,( December 2019) “Agricultural Development and Land Use Change in India: A Scenario Analysis of Trade-Offs Between UN Sustainable Development Goals (SDGs)”,https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019EF001287
4. Andrew John Hall, (February, 2007),” Challenges to Strengthening Agricultural Innovation Systems: Where Do We Go From Here? “ ,The Commonwealth Scientific and Industrial Research Organization
5. Mr. Sandeepa, Dr. K. S. Sarala, “Opportunities and Challenges of Agripreuership in India-Some Reviwes, Palarch’s Journal Of Archaeology Of Egypt/Egyptology 17(7), ISSN 1567-214x
6. Kate,( January,2020), ”How Agriculture Affects Our Daily Life”, Get Social with Us, Uncategorised,http://makebakegrow.com/2020/01/08/how-agriculture-affects-our-daily-lives/
7. Saraswathi P and Kaushik K (January, 2018),”Use Cases and Challenges in Smart Farming System on the Verge of Cloud and Internet of Things (IoT)”,” American Journal of Computer Science and Information Technology”,2018,Vol.6. No.3:S1
8. S. J. Yelapure et al, “Literature Review on Expert System in Agriculture “, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (5) , 2012,5086-5089
9. Manish Mahant, Abhishek Shukla, Sunil Dixit, Dileshwer Patel, (2012) , "Uses of ICT in Agriculture",International Journal of Advanced Computer Research (IJACR) ISSN (Print):2249-7277 ISSN (Online):2277-7970 Volume-2 Issue-3 March-2012
10. Sanjeev S Sannakki, Vijay S Rajpurohit, V B Nargund, Arun Kumar R, Prema S Yallur, (2011), "Leaf Disease Grading by Machine Vision and Fuzzy Logic",Arun Kumar R et al, Int. J. Comp. Tech. Appl., Vol 2 (5), 1709-1716,ISSN:2229-6093

11. Marina García-Llorente,Radha Rubio-Olivar,Inés Gutierrez-Briceño (2018),"Farming for Life Quality and Sustainability: A Literature Review of Green Care Research Trends in Europe", ", Int J Environ Res Public Health v.15(6); 2018 Jun ,PMC6025610, doi: 10.3390/ijerph15061282

12. Marion Olubunmi Adebiyi, (March,2020),"Machine Learning–Based Predictive Farmland Optimization and Crop Monitoring System", Hindawi,Volume 2020, https://www.hindawi.com/journals/scientifica/2020/9428281/

Sinhgad Institute of Management PUNE 411 041.

Singh

Dr.Chandrani Singh ,Director –MCA,SIOM

# Non-Fungible Tokens (NFT's): The Future of Digital Collectibles

Mr.Vimu Ram Kale
vimurk111@gmail.com
Dr. Chandrani Singh
singh.chandrani@gmail.com
Dr. Sunil Khilari sunilkhilari@sinhgad.edu
Sinhgad Institute of Management, Pune, India

**Abstract:** Non-Fungible Tokens (NFT's) indicate the creation of a blockchain-based digital certificate of authenticity that is comparable to other virtual crypto assets and currencies. The use of blockchain technology and the exchange of digital currency have become increasingly widespread in recent years. Having said that, as has been shown in recent years, the NFT market is also booming. The very idea of NFT is derived from an Ethereum token standard that aims to separate and recognise each token with its distinct signature being tied with digital attributes. India has also seen increased interest in this digital sector, particularly from the future new-age investors and digital innovators, as a result of the spectacular return on its quickly expanding global market. However, due to the early stage of the NFT ecosystem's growth, India lacks a regulatory legislative framework to oversee such immature digital crypto assets. There are several legal complexities surrounding them, which has made it difficult to determine their legal legitimacy and sanctity. New artists could have a tendency to become lost in this chaotic growth in the absence of comprehensive descriptions. This paper aims to examine the idea of NFT in comparison to bitcoin and copyright, as well as its operational and technological elements. It attempts to examine the legal hazards that affect its operation as well as the potential and difficulties the Indian legal system has with regard to crypto-assets.

**Keywords:** NFT, Digital Collectibles, Legal Validity.

## 1. Introduction

The digital world has seen a rapid transformation due to the development of technology and the pandemic that has affected the entire planet. Digital investments and digital currencies are two examples of this revolution brought about by the new digital environment. The cryptocurrency market has grown significantly, increasing the number of investors around the nation who are eager to participate in some of these crypto assets and cryptocurrencies.

Everything is now publicly displayed online as a result of the digital revolution. As a result of individuals being utterly ignorant of the ownership and originality of the works exhibited online, the digital platform has had a negative impact on the invention and development of original artists. Lack of legitimacy and security leaves artists and creators with less money for their work. The NFTs symbolize ownership of a special object, which is subsequently connected to a token via the blockchain, in order to address this issue.

As 2022 approaches the mainstream art world, the market for NFTs has experienced a boom, increasing from 41 million in 2018 to approximately 338 million today. A new generation of traders, or digital natives with wealth and reputation who are prepared to invest in asset classes outside of the traditional asset markets, is being seen in the NFT market.

Millions of artists and investors worldwide are now involved in these digital assets because of the rising popularity of non-fungible tokens. It is reasonable to assume that NFTs are poised to forge their way into the modern digital age. Selling for $69.3 million, Beeple's "Everydays: the First 5000 Days" is the most expensive NFT ever sold. After posting a new piece of art every day for 5,000 days — from May 2007 to February 2021 — Beeple put them all into one picture and sold it at the auction house Christie's.Vignesh Sundaresan eventually won the auction on March 11 for 42,329 ether (**Robin Barber, 2022)** to the well-known artist Nuclaya choosing to post his album on the internet platform.

The legitimacy of these digital assets, however, is still up for debate. Due to its rising popularity, there have been concerns about its survival and legal viability in India. India has provided evidence of the validity of these tokens. The viability of the tokens in India has drawn severe criticism from the Indian populace, calling them into question.

## 2. An Overview on Non-Fungible Tokens (NFT's)

NFT's can be thought of as a certification of ownership of digital or physical assets on a blockchain-powered digital marketplace. A blockchain is a digital ledger that keeps track of transactions in the form of a decentralized database.

Technology has made digital information more widely used than ever before. The original content loses value as a result of the rising replication of digital content because it is so difficult to trace down and further identify the creators of works that are displayed on online platforms. Once made available online, they can be cheaply copied, duplicated, and distributed.

Even if the owner has been found, it is a very difficult task to demonstrate such ownership using the documents kept by institutions. NFT intends to address the problems associated with decentralization, ownership tracking, monitoring, and value storage by making the declaration of the legal owner over the original work plain and transparent in the event of any duplication and further ensuring their grant of legitimate royalties.

Some of the unique features of NFT's constitute of the following

- Indivisible i.e., incapable of being further splitted or separated into smaller denominations
- Cannot be demolish, reproduced or removed from the blockchain
- Traceability: easy tracking of original owner, eliminates the need for third party verification
    - Scarce: purposely limited to add additional value

### Semantics of Non-Fungible Tokens

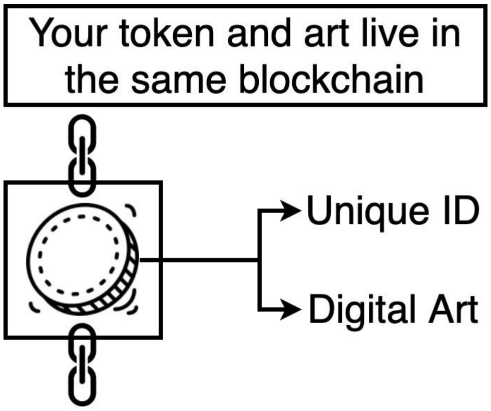


**Figure 1: Representing how NFT's Works**

**(Source: https://vonnie610.medium.com/everything-you-need-to-know-about-nfts-be2601d09cf5)** A Non-Fungible Token is operated as well as stored on a blockchain ledger. The token and its associated ownership is locked in a block as shown in Figure 1, with the token data being recorded through a blockchain method. Therefore, whenever a token or an item is stored on the blockchain, its adjoining data pertaining to its actual ownership is recorded. Subsequently whenever there is a transaction of any sale or purchase, the blockchain records the same on that very block. This system provides the original artist / creator with a certain percentage of Royalty with every sale or resale of that item. The main purpose of the system of blockchain is to provide a portal for execution of systematic transactions and thereby prevent the reproduction and piracy of original creations. A unique token is assigned to every artist/creator who decides to portray her work on a digital platform, this unique token along with ownership gets stored in the blockchain. The NFT's are sold on a digital platform through the method of auction, wherein the bidder with the highest amount becomes the owner of the token while the trading takes place on a public blockchain through a cryptocurrency wallet wherein NFT’s are purchased and sold mainly using crypto assets.

### 3. Token Standard Used By NFT

The Ethereum token protocol is now the most widely utilized token standard among NFT’s. The ERC-721 and ERC-1155 token protocols are typically used to generate various NFT’s.

Unfortunately, some initial NFT’s create NFT’s using a hybridized ERC-20 token standard. To sell them, they must be enclosed or packaged.Along with Ether, other cryptocurrencies include EOS, Flow, Tezos, and so on. Also provide a special token standard for NFT creation.

### NFT Marketplace

Users must register for an account and link their wallets to their account in order to buy and sell NFT across NFT marketplaces. Different cryptocurrencies are supported by different marketplaces. Ethereum is currently the most widely used currency. Users can purchase NFT’s with ether or any other cryptocurrency. Users must get a crypto wallet to buy NFT and fund it using a flat-to-cryptocurrency exchange in order to buy any kind of cryptocurrency.

**WazirX NFT Platform**

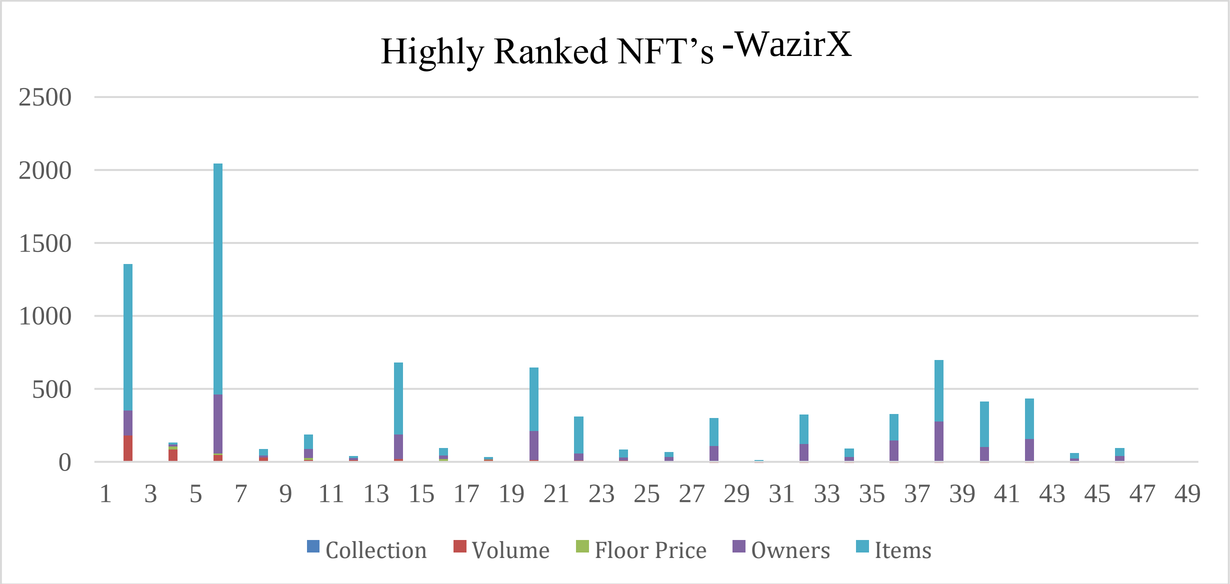


**Figure 2: NFT Ranking**

WazirX, one of India's most reliable exchanges for Bitcoin and other cryptocurrencies, enables trading in Bitcoin, Ethereum, Ripple, Litecoin, and many more cryptocurrencies. The country's first marketplace for NFTs was introduced by WazirX and highly ranked NFT's are shown in Figure 2, India's largest cryptocurrency exchange forum, in response to the recent boom. It creates a system for the seamless exchange of digital assets and intellectual property like works of art, videos, tweets, audio files, tweets, and programmes, in addition to other digital goods and services. The Indian inventors can use this to auction off their digital goods on the blockchain-based NFT marketplace and then get their royalties. By offering their digital assets for auction over the NFT marketplaces and assisting them in earning royalties, the platform seeks to meet the demands of all craftsmen and creators while without charging any fees from users for either producing or listing the non-fungible tokens. However, since NFTs are based on the blockchain, which supports smart contracts, a gas fee must be paid to the miners in the appropriate currency. According to Nischay Shetty, the founder of the platform WazirX, such a market would soon revolutionize the industry in the fast digitizing world due to the increasing interest and acceptance of NFT among individuals all over the world.

## 4. NFT Risk Factors Theft Risk

The NFT risk and issues with intellectual property rights imply that buyers are the only owners of NFTs and only have the right to show them. The restrictions are very clear when it comes to the standards of conduct that customers should adhere to when utilizing NFT markets.

### CyberSecurity Risk

There are significant cyber-security and fraud threats as a result of the development of the digital world and the astounding increase in popularity of NFTs. NFT replica shops that resemble the actual NFT shop and use the same logo and materials as real shops Fake NFT stores are another significant issue connected to the dangers and difficulties of using NFTs in cyber security. Copyright theft, imitation of well-known NFTs or false airdrops, and NFT giveaways are some of the other significant non-fungible token dangers and issues connected to cybersecurity and fraud.

### Money Laundering Risk

From the perspective of money laundering and financial crime, this technology may cause concern.

This NFT, titled "The Pixel," was created by an artist going by the name of Pax and sold at a Sotheby's auction in April 2021 for about $1.3 million. Theoretically, a person wishing to "clean" filthy money may create an anonymous NFT, offer it for sale on the blockchain, buy it from oneself from an unregulated, anonymous digital wallet, and then recognize the money as genuine cash from the sale of the artwork. Trend of illicit use of NFT in the form of stolen money, cryptojacking NFT's is shown below from 2017 onwards

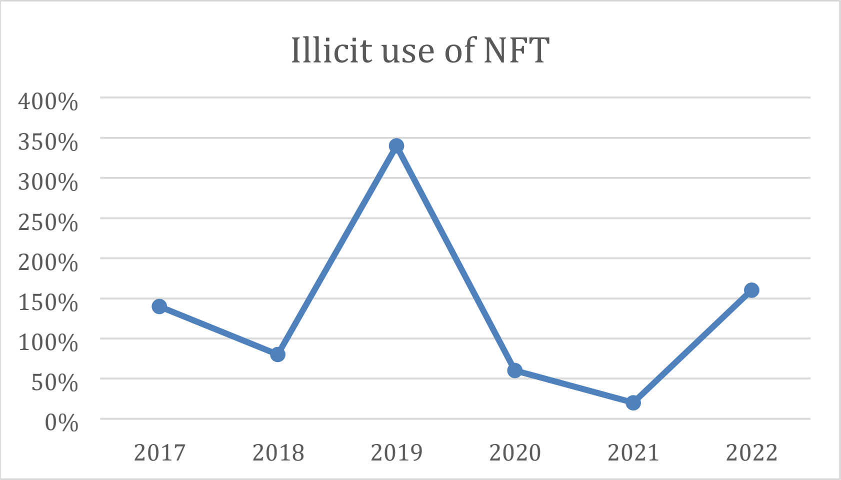


**Figure 3: Illicit value received by NFT platforms**

According to Chainalysis, and as it appears in Figure 3, money laundering poses a significant risk to developing trust in NFTs and should be constantly monitored by markets, regulators, and law enforcement. This is especially true with transfers from authorized cryptocurrency enterprises (**Chainalysis, 2022**).

**Legal Risk**

NFT has no recognized legal definition anywhere in the world. Different nations, including the UK, Japan, and the EU, are going forward with various classification schemes for NFT. As a result, it becomes vital to establish a global organization of non-fungible tokens for establishing laws and legislation everywhere. Tokenization of NFT's can implicate several U.S. laws, and below segments are as follows:

- Licensing
- Securities
- Anti-money laundering
- Sanctions

- Intellectual property
- Gambling, and others.

NFTs may implicate securities laws where:

- NFTs represent presales of digital assets and the proceeds of the sale are used to build the platform
- Pooling or Fractionalization of digital assets
- NFTs represent a license to a digital asset and a share of the revenue from the asset [7] The NFT market has seen a significant increase, which is why it is crucial to establish a regulatory agency. The use cases for NFT's have significantly increased. As a result, a regulatory body must adjust to the laws and policies of NFT's.

Finding the right terminology for NFT is still a problem for the laws that are now in place. It is getting more and more challenging to establish a firm foundation for compliance in NFTs as the market and diversity of NFT's are expanding so quickly.

All platforms in India that have opened trading in NFT's to date are cryptocurrency exchanges, and NFT's can only be exchanged in cryptocurrencies. Trading in NFT's is hazardous because there is, regrettably, still uncertainty over the country's legal stance on cryptocurrencies. In India, there is also no distinct legal structure for NFT's. By extrapolating from current FEMA laws, crypto-assets and NFT's may be regarded as intangible assets under the law, along with software and intellectual property. The precise position of an NFT, however, remains unknown. The Supreme Court has acknowledged that crypto-assets "cannot be stored anywhere," and blockchains are global ledgers.

Since there is currently no specific legal structure in place for NFT's in India, just the fundamentals of contract law apply. Though the majority of stakeholders are of the opinion that the Cryptocurrency and Regulation of Official Digital Currency Bill, 2021.We would make an exception for NFT's given their enormous popularity, one can only wait and speculate about the potential effects NFT's may have if the Indian Government issues a definite ban on cryptocurrencies.

## 5. Evaluation Challenge

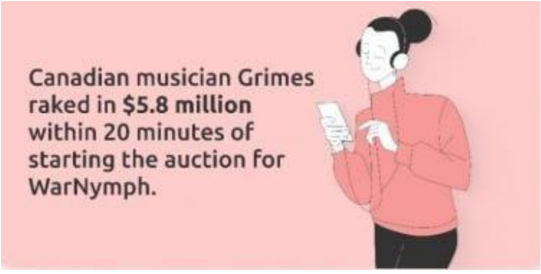
Risks and obstacles associated with non-fungible tokens include the difficulty in estimating their worth. The scarcity, perception of owners and buyers, and accessibility of distribution channels all play a significant role in NFT valuation. It is quite impossible to predict who will acquire an NFT next or the potential motivators for their purchase. As a result, the value of NFTs would essentially rely on how the buyer views their pricing, which would cause volatility.

**Advantages of NFT's**

- Decentralized marketplace
- Unique Collectibles
- Transferability
- Immutable
- Authenticity
- Copyright
- Security
- Identity Management and many more things.

### 6. Industries that Leverage NFT's

**Art & Music NFT** –Music NFT Market Trends

| | |
|---|---|
| <br>In February 2021, 3LAU generated almost **$11.7 million** from 33 Music NFTs. | <br>Canadian musician Grimes raked in **$5.8 million** within 20 minutes of starting the auction for WarNymph. |
| Media NFT | Buying a music NFT |
| <br>Kings of Leon released their album When You See Yourself as NFTs, and the revenue from the NFT sales went over **$2 million.** | <br> <br>In June 2021, the music NFT market turned bleak, dropping by more than 90% in value from a high **$26.9 million** in March to just a little over **$1.7 million** in July. |

| | |
|---|---|
| Crypto-native music rollouts | Philanthropic vessels |
| Art NFT Market Trends | |
| Trend-1: New patronage artists | Trend-2: Evolution of new utilities |
| Trend-3: Growth of communities | Trend-4: Path to cryptocurrency |
| Trend-5: Crypto Wallet | Trend-6: Innovative art projects |

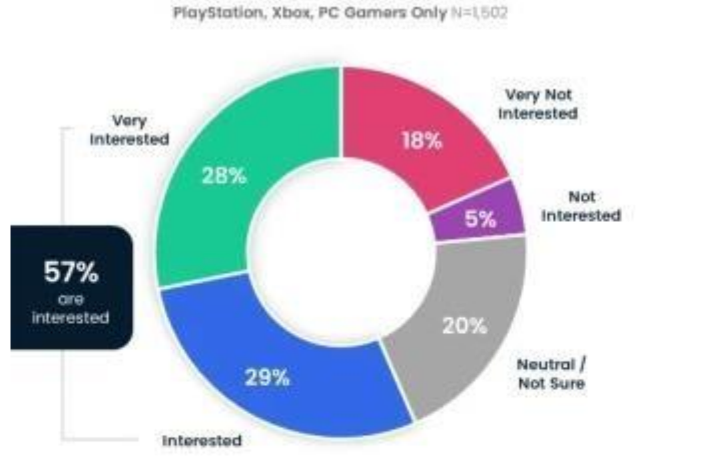
**Figure 4 Art and Music NFT trends**

**(Source: https://influencermarketinghub.com/nft-music/)**

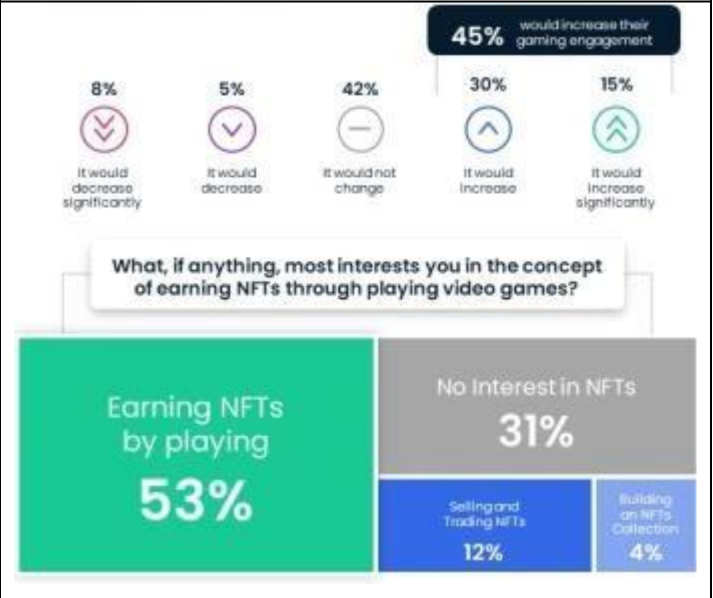
Non-fungible tokens (NFTs) are being embraced more frequently as a fresh and independent revenue stream for musicians and artists as in Figure 4 in a sector where organizations like record labels, galleries and streaming platforms take a significant part of the revenues from performing artists. With no need for a middleman, music-specific NFTs are giving artists a new way to become independent (indie) and monetise their work on the blockchain.Numerous well-known artists have already started using NFTs to monetise their songs. As an illustration, the electronic musician 3LAU tokenized the record "Ultraviolet," which was later sold as NFTs and brought in a total of more than $11.6 million (**Cryptomedia, 2022**).

## 7. NFT In Gaming Industry

Video game skins have evolved into digital works of art. NFT’s might make it possible for digital artists to market their works. They provide the crucial element of getting the necessary exposure, which aids in helping artists determine the market value of their creations. One of the factors driving high prices for cards and limited-edition releases among sneaker heads and sneaker collectors is scarcity and earning of NFT’s by playing video games is presented in Figure 5.It encourages game players to get involved and create unique in-game goods that might possibly inspire the next wave of digital artists. With in-game merchandise that features meta-jokes, memes, and pop culture, it would help boost community participation.



**Statistics on earning on NFT's through playing video games**



**Statistics on NFT's impact on gaming habit**

**Figure 5: Poll for play & earn NFT**

**(Source: https://venturebeat.com/2022/01/19/interpret-studys-says-56-of-gamers-are-interested-in-earning-nfts-in-games/)**

A study of 1,500 console and PC gamers found that 56% of them are interested in earning nonfungible tokens (NFTs) through gaming (A**ndrew Wilson, Interpret, 2022**), according to market research firm Interpret as seen in Figure 4.

**NFT in Metaverse**

The metaverse is a virtual, three-dimensional cosmos that offers consumers and companies countless chances to import goods and services from the real world. A blockchain-supported open and equitable economy is offered by metaverses. NFTs will specifically be used to engage and empower players of blockchain games in the play-to-earn gaming economy.

NFTs serve as the gateway to the metaverse and support social, communal, and identity experiences there. Users can access gaming metaverses through collecting in-game NFTs through collections.

### 8. NFT In Internet Of Things (IOT)

Each IoT device has a unique BCA (Blockchain Account), which allows each one to sign for individual transactions. The researchers employ an ERC-721 NFT, which can distinguish between managers, users, and manufacturers (owners and approvers). Then, each device within the blockchain may validate both its own identification and the data it generates.

An initial seed is generated and not saved on the device in order to establish the NFT. Instead, the researchers generate the private and public keys related to the BCA using a PUF (Physical Unclonable Function) in the device together with other parameters from memory. This is different from many other approaches that keep the private key on the gadget (and where the private key could be leaked).

The device then generates a public key and a BCA_SD and actively joins the blockchain. The NFT (together with the owner) may then be created using these and a smart contract. This will return a Token ID, which is then saved in the firmware of the device (Prof **Bill Buchanan OBE, 2022**).

### 9. AI and ML in NFT

Fetch.ai, an artificial intelligence lab  non-fungible token (NFT) platform, Colearn Paint, allows creators to automatically produce NFTs using a machine learning algorithm and below in Figure 5 is the winner statistics of generate shared NFT artworks minted by the colearn pAInt address: 0xC73Bb3F183fbeAdC23bCB70460006c30Fc4c689A[8]

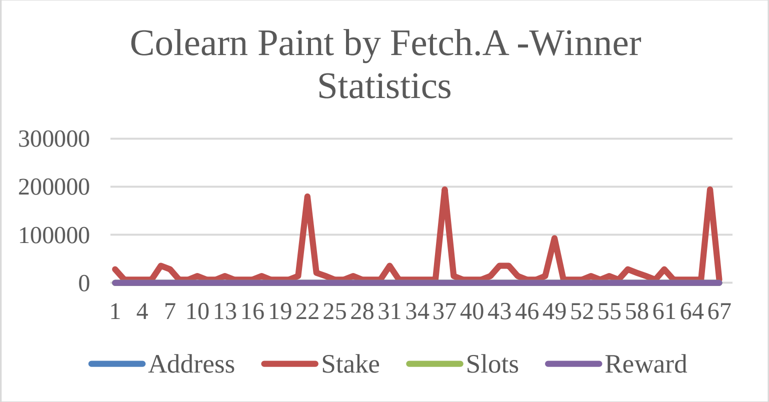


**Figure 6 NFT artworks minted by Colearn Paint 10. Conclusion**

NFT's, which are rapidly expanding, have the potential to bring about a paradigm shift in the digital sphere by giving investors and creators of digital content a secure online marketplace with a variety of options for intellectual property protection. This is extremely beneficial to digital developers and should not be disregarded in favor of a general ban. However, once they are exchanged for fiat money or when virtual currencies are recognized by the legal system, the likelihood of success of digital assets like NFTs will be determined. NFTs, which have the potential to be a tool in the fight against identity theft, are still in their infancy, with the bitcoin community being their main application area. While they are anticipated to grow and expand significantly in 2023, experts predict that their future could either hold a highly regulated, disruptive digital market that skyrockets and seeks to protect the virtual copyrights, or it could hold an asset that ends up experiencing the same fate as various cryptocurrencies have, and explode. NFT traders in India tread carefully because of the country's tumultuous history with the financial instrument. A balanced legal and regulatory framework is required, which is crucial for its efficient trade. In addition, buyers should exercise caution when purchasing these digital tokens because of the caveat emptor, or "buyer beware," principle. The lack of permanence that NFT portrays poses a problem for the environment. It cannot be denied that artists are eager to join the NFT bandwagon in order to profit from their works of art. The sale and purchase of such tokenized artwork raises questions about the ambiguity surrounding its legal validity, as well as a number of other legal issues, including the enforcement of copyright holders' and NFT holders' rights, creator and holder liability, the applicability of a number of other intersecting laws, and the exhaustion of copyright holders' rights following the first sale.

People who engage in digital transactions with NFTs should be aware of the risks and hazards involved. The numerous advantages offered for the protection of digital works of art, however, cannot be blatantly disregarded. Not banning it but regulating it is what is required right now. The Indian legislature should make an effort to put into place a regulatory legal structure designed to solve and accommodate the difficulties encountered in this area of law. The general public still hasn't fully embraced the use of non-fungible tokens. The ability to commercialize works of art in the digital world may be made possible by NFTs, but much will depend on how well these digital assets are received by the traditional art market. NFT traders in India tread carefully because of the country's tumultuous history with the financial instrument. A balanced legal and regulatory framework is required, which is crucial for its efficient trade.

### *11. References:*

1. Cryptomedia,Copyright-2022,Gemini Trust Company, LLC, https://www.gemini.com/cryptopedia/nft-crypto-blockchain-music-industry
2. Vinay Butani and Faiz Thakur,"India: Decoding NFTs And Their Regulation In India","Economics Lasw Practise",January 2022, https://www.mondaq.com/india/fintech/1149624/decoding-nfts-and-their-regulation-in-india
3. Omang Baheti,"Applications Of NFTs In The Gaming Industry","Digit.in",Feb 2022 , https://www.digit.in/features/gaming/applications-of-nfts-in-the-gaming-industry63043.html
4. Wikipedia article,"Non-fungible token", April,2022,https://en.wikipedia.org/wiki/Nonfungible_token
5. Veronica Coutts,"A blockchain believer & Ethereum developer. Trying to spread knowledge, peace and critical thinking",March,2021,https://vonnie610.medium.com/everything-you-need-to-knowabout-nfts-be2601d09cf5
6. Mitchell Clark ,"NFTs, explained",The Verge,Jun,2022,https://www.theverge.com/22310188/nft-explainer-what-is-blockchaincrypto-art-faq
7. Games G.Gatto,Yasamin Parsafar,"Tokenization and the Law: Legal Issues with NFTs",The National Law Review,July-2022 Volume XII, Number 200, https://www.natlawreview.com/article/tokenization-and-law-legal-issues-nfts
8. Walkthrough-CoLearn pAInt,August,2021,https://colearn-paint.fetch.ai/
9. Prof Bill Buchanan OBE,"Trusted IoT: Linking IoT Devices to NFTs",ASecuritySite: When Bob Met Alice, March-2022,https://medium.com/asecuritysite-when-bob-metalice/trusted-iot-linking-iot-devices-to-ntfs-a03dbc7de7b8
10. Andrew Wilson,"Gamers are embracing NFTs,Interpret data finds",Interpret,Jan-2022,,https://interpret.la/gamers-are-embracing-nfts-interpret-data-finds/

11. Robin Barber,"NFT Statistics, Facts & Trends in 2022: All You Need to Know About NonFungible Tokens",Cloudwards.net ,July,2022,]https://www.cloudwards.net/nftstatistics/#Sources

***12.*** Chainalysis,"NFT Transaction Activity Stabilizing in 2022 After Explosive Growth in 2021",May,2022,https://blog.chainalysis.com/reports/chainalysis-web3-report-previewnfts/

***13.*** Monty Preston,"Five trends shaping the future of art NFTs",Fastcompany,April,2022 https://www.fastcompany.com/90742358/five-trends-shaping-the-future-of-art-nfts

***14.*** Genç, Ekin (October 5, 2021). "Investors Spent Millions on 'Evolved Apes' NFTs. Then They Got Scammed". Vice Media. Retrieved November 9, 2021.

***15.*** Hawkins, John (January 13, 2022). "NFTs, an overblown speculative bubble inflated by pop culture and crypto mania". The Conversation. Retrieved May 7, 2022.

***16.*** Bosselman, Haley (March 31, 2021). "'Godzilla vs. Kong' to Have First Major Motion Picture NFT Art Release". Variety. Retrieved April 14, 2021.

***17.*** Vigna, Paul (May 3, 2022). "NFT Sales Are Flatlining". The Wall Street Journal. Retrieved May 5, 2022.

***18.*** Wilson, Kathleen Bridget; Karg, Adam; Ghaderi, Hadi (October 2021). "Prospecting nonfungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity". Business Horizons: S0007681321002019. doi:10.1016/j.bushor.2021.10.007. S2CID 240241342.

***19.*** D'Alessandro, Anthony (April 13, 2021). "Kevin Smith To Sell Horror Movie 'Killroy Was Here' As NFT, Launches Jay And Silent Bob's Crypto Studio". Deadline. Retrieved April 14, 2021.

***20.*** Goldsmith, Jill (September 8, 2021). "Enderby Entertainment's Pandemic Film 'Zero Contact' To Premiere On New NFT Platform Vuele; Watch The Trailer – Update". Deadline Hollywood. Retrieved September 15, 2021.

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

# Prediction of IPL match  performance based on batsman category using Machine Learning Algorithm

Dr. Chandrani Singh,  directormca_siom@sinhgad.edu
Dr. Ramesh D Jadhav,   rameshdjadhav@gmail.com,
Dr. Bharti P  Jagdale, bj70@rediffmail.com,
Dr. Sunil Khilari,  sunilkhilari@sinhgad.edu
Sinhgad Institute of Management, Vadgon(Bk), Pune-411041, India

## Abstract

In the world of business, firms have started adopting technologies to develop  and grow their businesses. To sustain in this competitive market there is a dire need for  better performance and precision in prediction .To achieve this goal various businesses, require to accurately analyze and predict the patterns for data driven decisions. .This research paper themes around sports as a business and takes into consideration cricket which is a very popular and well-known game  that is played and watched in 104 countries.Predicting winner teams and performance of each individual team depends on many factors for example batsman's approach ,bowler's skill and tactics ,the pitch statistics,the weather conditions. and many more.The previous research studies show player's probability of winning against the opponent. Study will find out various attributes to predict  a model for IPL match game changer.Score prediction has been in existence from many years  based on certain parameters taken into consideration ,but this research study will specifically focus on machine learning algorithms to predict a team's performance and score  based on  batsman category, The study will make use of Kaggle data to build a predictive model. The results from conducted the research shows that random forest exuded the best accuracy in terms of score prediction taking into consideration the parameters. **Keyword: Multi-collinearity,accuracy,cross validation ,regression,random forest**

## Introduction

In India cricket is a prevalent sport among all, mainly for  T20 leagues. When millions of  people are watching IPL ("the Indian Premier League") then forecasting of the results become very difficult altogether as the game may change its course altogether.In the game of cricket many aspects and features  control the final result of  the game, and each feature has its own weightage from one  needs to find out impact of the factor on  the result of the T20 cricket match. Many ML (machine Learning) models are used to predict the performance of the team during a specified time interval I.e. from the time  of toss  to the  beginning of the match to forecast victory.Prediction models are useful for regulatory bodies and cricket boards of the states and provinces to carry out strategic decision on the team composition,on the bowler and batsman categories and preferences, on scheduling matches for multiple teams etc wherein score forecast  continuously helps in the management of games and accordingly

Plan the strategy of the game. Player's characteristics depend on various factors like present form, previous record,nature of the pitch,game format and venue etc. The models are helpful in classifying the players priority and plan towards win.In forecasting/prediction linear regression plays a major role with intervention of AI (Artificial intelligence) to gives more accurate result. The famous football bludgeons in Portugal, Game name 'Lisboa e Benfica' [1-4] used ML (machine learning) for information handling and identifying patterns by taking into account practise duration of each player, the resting time, the work out sequence ,their diet strategies etc.. Cricket in India is popular game and many researchers are using functional statistical methods for predicating the game outcome.

Winning & Score Prediction archetypal was established by "Mr. Samuel " back in 2011 this model was recycled by ICC in 2012, and the match against India. New-Zealand was envisaging 282 agreeing to the prediction model and in case if degree of run would have been hand-me-down the possibility of mark would have been 226.[5]

But there are many loopholes in the WASP (Winning and Score Predictor) method which needs to be fixed for better outcome. As the game changes, many external factors have an impact on game performance. Prediction is based on the the pitch ,weather and the boundary size. The team who is batting first sets the forecast trend while the team who is batting second,the performance of the said team can be extracted based on a relative analysis with the first and further predictions are based on relative average of the previous perforances..[6]

The Generic function for classifier model used to measure the points earned by teams based on their past performances included team1, team2, venue, toss winner etc. The random forest classifier and decision rree provided 89.151% accuracy for such categories of predictions[7].

## Research Objectives

- To find the team with the greatest number of wins per period
- To find venue to host he maximum number of IPL Cricket matches
- To find out the team with maximum wins for each season
- To find the most valuable player in the IPL
- To predict win/loss using Machine Learning algorithm

## Literature Review

G. Sudhamathy & G. Raja Meenakshi' (2020) in their "Prediction on IPL data using machine learning techniques in R package" performed the analysis on the 10 year old IPL data-set. In this research study they had undertaken the implementation of four machine learning algorithms and used the training ,testing data-set to help create the model which then classified the data & compared the results with respect to correctness, error-rate, precision, recall, understanding and specificity. [1]

"Ch Sai Abhishek, P Yuktha, Ketaki V Patil, Meghana K S, & MV Sudhamani" [2019] in their research had conducted a study on the "Predictive Analysis of IPL Match Winner using Machine Learning Techniques" and have presented the prediction of the winner results by incorporating machine learning, where different catalogue based machine learning algorithms as logistic-regression, ‘decision-trees, random-forest and K-nearest neighbors were implemented of which the random forest provided an highest accuracy of 89.15%. [2]

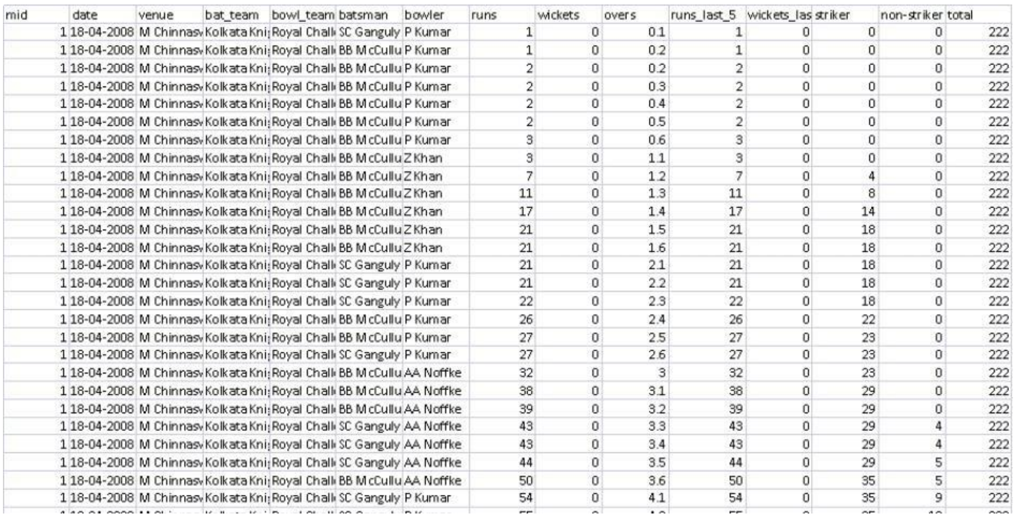
Sonu Kumar, Sneha Roy [2018] in the study title “Score Prediction and Player Classification Model in the Game of Cricket Using Machine Learning” have done the predictive modeling and regression analysis and concludes the significant aspects of Forecast and Organisational Modelling. Using such modelling techniques it can be demonstrated that forecasting the inning mark and players cataloguing is based on the numerical data [3].

Akhil Nimmagadda, Manigandla Venkatesh, Nidamanuri Venkata Kalyan, Nuthi Naga Sai Teja, & Chavali Gopi Raju” [2018] presented the study on the “Cricket score and winning prediction using data mining” and developed a model to forecast the outcome of an ODI cricket game.The model had been developed using multiple variable linear regression. Efficiency and error checking was done. Using multiple linear regressions, each innings score was predicted at regular intervals and also the final the winner of the match. [4]

Vrushali Y Kulkarni, Pradeep K Sinha [2014] in their research work on “Effective Learning and Classification using Random Forest Algorithm” combined the different attributes as measures of evaluation..The accuracy of the model is improved by a hybrid decision tree model. A new similar approach in which, specific tree as well as whole forest is made in related research study. [5]

**Data Collection and Analysis**

In this research study collected data is from Indian Premier League (IPL) Kaggle database.The dataset covers of data of IPL cricket games played from the year 2008 to 2019. The league has eight teams on behalf of eight different Indian cities or states. A snapshot of the relevant data and stats of IPL is presented below.



| mid | date | venue | bat_team | bowl_team | batsman | bowler | runs | wickets | overs | runs_last_5 | wickets_las | striker | non-striker | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 1 | 0 | 0.1 | 1 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 1 | 0 | 0.2 | 1 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 2 | 0 | 0.2 | 2 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 2 | 0 | 0.3 | 2 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 2 | 0 | 0.4 | 2 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 2 | 0 | 0.5 | 2 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 3 | 0 | 0.6 | 3 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 3 | 0 | 1.1 | 3 | 0 | 0 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 7 | 0 | 1.2 | 7 | 0 | 4 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 11 | 0 | 1.3 | 11 | 0 | 8 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 17 | 0 | 1.4 | 17 | 0 | 14 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 21 | 0 | 1.5 | 21 | 0 | 18 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | Z Khan | 21 | 0 | 1.6 | 21 | 0 | 18 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 21 | 0 | 2.1 | 21 | 0 | 18 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 21 | 0 | 2.2 | 21 | 0 | 18 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 22 | 0 | 2.3 | 22 | 0 | 18 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 26 | 0 | 2.4 | 26 | 0 | 22 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | P Kumar | 27 | 0 | 2.5 | 27 | 0 | 23 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 27 | 0 | 2.6 | 27 | 0 | 23 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | AA Noffke | 32 | 0 | 3 | 32 | 0 | 23 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | AA Noffke | 38 | 0 | 3.1 | 38 | 0 | 29 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | AA Noffke | 39 | 0 | 3.2 | 39 | 0 | 29 | 0 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | AA Noffke | 43 | 0 | 3.3 | 43 | 0 | 29 | 4 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | AA Noffke | 43 | 0 | 3.4 | 43 | 0 | 29 | 4 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | AA Noffke | 44 | 0 | 3.5 | 44 | 0 | 29 | 5 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | BB McCullu | AA Noffke | 50 | 0 | 3.6 | 50 | 0 | 35 | 5 | 222 |
| 1 | 18-04-2008 | M Chinnasv | Kolkata Kni | Royal Chall | SC Ganguly | P Kumar | 54 | 0 | 4.1 | 54 | 0 | 35 | 9 | 222 |

**Table 1: IPL Matches played from the years 2008 to 2019 - Snapshot**

**Research Methodology**

The researchers have taken into consideration linear regression,ridge regression and random forest based on the aspects of diminishing the prediction discrepancy,making analysis easy on the aspects of multi-collinearity and considering algorithm accuracy in case of all sorts of data.

Following are the algorithms which were executed on the Kaggle data sets.

### Linear regression

Linear regression analysis is used to estimate the value of a dependant variable based on one or more independent variable.The variable that the researchers want to predict/forecast is called the dependent variable whereas the variables that are on the right side of the equation that help predict the variable on the LHS are called the independent variable.Linear regression fits a straight line that decreases considerably the deviations between the predicted and the actual values.

### Ridge regression

Ridge regression is the technique used for assessment of high correlation among the independent variables and for this research study it was found that there exists high correlation between several of the predictor variables. Hence using ridge regression analysis such situations can be analyzed and in this case the predictor variables suffer from multi -collinearity.

### Random Forest

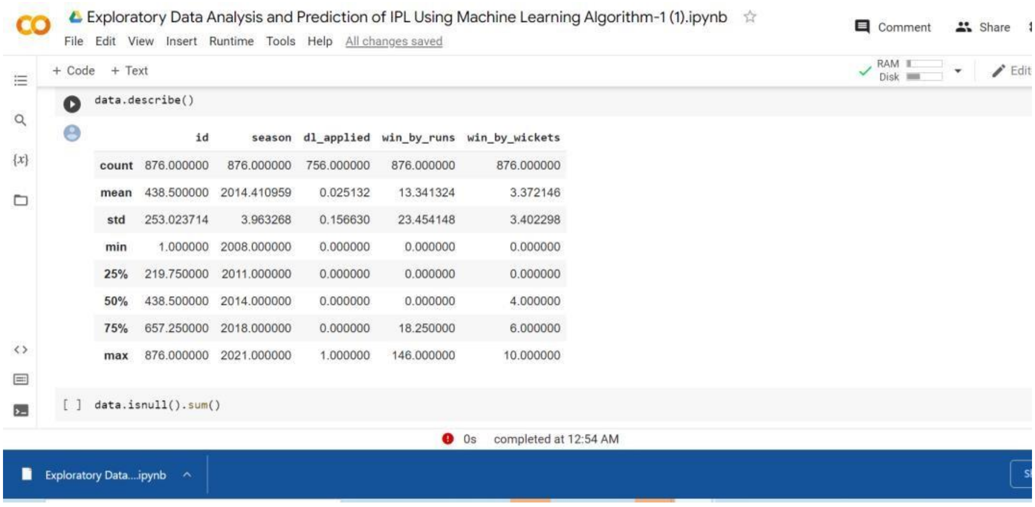
A random forest comprises of multiple decision trees, and this algorithm can handle almost all forms of datasets and gives optimum results in terms of prediction.

**Cross-Val-Score :** Cross value scores are obtained by training the model I using K-1 of the folds as training data and then the resulting model is validated on the remaining data and accuracy computed.

The researchers have used the Google Colab environment to execute the above algorithms on the data set, used the cross-val scoring to ensure accuracy in prediction using k fold approach.

### Observations

The following inferences can be made from the describe () method as shown in Fig 1 and the tables 2,3 and 4.


Exploratory Data Analysis and Prediction of IPL Using Machine Learning Algorithm-1 (1).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Edit
Exploratory Data....ipynb

```
data.describe()
```

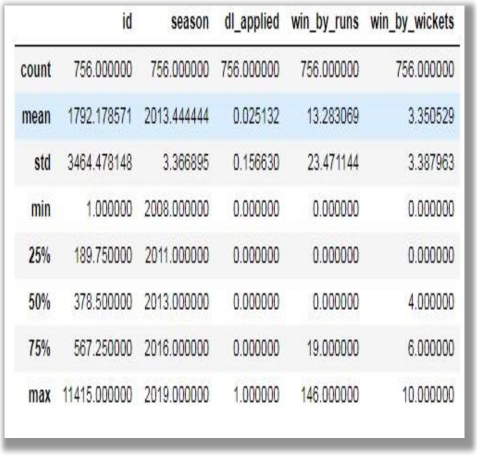
| | id | season | dl_applied | win_by_runs | win_by_wickets |
|---|---|---|---|---|---|
| count | 876.000000 | 876.000000 | 756.000000 | 876.000000 | 876.000000 |
| mean | 438.500000 | 2014.410959 | 0.025132 | 13.341324 | 3.372146 |
| std | 253.023714 | 3.963268 | 0.156630 | 23.454148 | 3.402298 |
| min | 1.000000 | 2008.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 219.750000 | 2011.000000 | 0.000000 | 0.000000 | 0.000000 |
| 50% | 438.500000 | 2014.000000 | 0.000000 | 0.000000 | 4.000000 |
| 75% | 657.250000 | 2018.000000 | 0.000000 | 18.250000 | 6.000000 |
| max | 876.000000 | 2021.000000 | 1.000000 | 146.000000 | 10.000000 |

```
[ ] data.isnull().sum()
```

0s completed at 12:54 AM

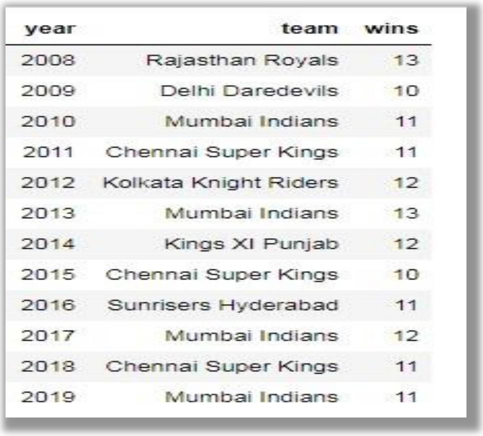
**Fig 1 The colab environment and the describe function on the data set**

The .csv file has data of IPL matches starting from the season 2008 to 2019.The biggest margin of victory for the team batting first (win_by_runs) is 146 runs. The biggest victory of the team batting second (win_by_wickets) is by 10 wickets. 75% of winning teams who bat in the beginning won by a margin of 19 runs.75% of champion players who played in later sequencer won by a boundary of six wickets.There were 756 IPL matches hosted from 2008 to 2019. Table 2 to 6 provide other relevant details and statistics with respect to the matches.



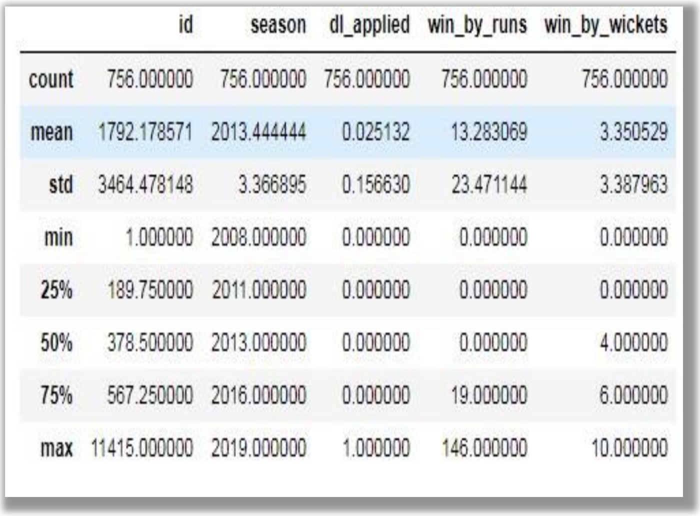
| | id | season | dl_applied | win_by_runs | win_by_wickets |
|---|---|---|---|---|---|
| count | 756.000000 | 756.000000 | 756.000000 | 756.000000 | 756.000000 |
| mean | 1792.178571 | 2013.444444 | 0.025132 | 13.283069 | 3.350529 |
| std | 3464.478148 | 3.366895 | 0.156630 | 23.471144 | 3.387963 |
| min | 1.000000 | 2008.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 189.750000 | 2011.000000 | 0.000000 | 0.000000 | 0.000000 |
| 50% | 378.500000 | 2013.000000 | 0.000000 | 0.000000 | 4.000000 |
| 75% | 567.250000 | 2016.000000 | 0.000000 | 19.000000 | 6.000000 |
| max | 11415.000000 | 2019.000000 | 1.000000 | 146.000000 | 10.000000 |

**Table 2. Statistical Analysis of the IPL data**



| year | team | wins |
|---|---|---|
| 2008 | Rajasthan Royals | 13 |
| 2009 | Delhi Daredevils | 10 |
| 2010 | Mumbai Indians | 11 |
| 2011 | Chennai Super Kings | 11 |
| 2012 | Kolkata Knight Riders | 12 |
| 2013 | Mumbai Indians | 13 |
| 2014 | Kings XI Punjab | 12 |
| 2015 | Chennai Super Kings | 10 |
| 2016 | Sunrisers Hyderabad | 11 |
| 2017 | Mumbai Indians | 12 |
| 2018 | Chennai Super Kings | 11 |
| 2019 | Mumbai Indians | 11 |

**Table 3: IPL Matches -Team with greatest number of wins per season.**



| | id | season | dl_applied | win_by_runs | win_by_wickets |
|---|---|---|---|---|---|
| count | 756.000000 | 756.000000 | 756.000000 | 756.000000 | 756.000000 |
| mean | 1792.178571 | 2013.444444 | 0.025132 | 13.283069 | 3.350529 |
| std | 3464.478148 | 3.366895 | 0.156630 | 23.471144 | 3.387963 |
| min | 1.000000 | 2008.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 189.750000 | 2011.000000 | 0.000000 | 0.000000 | 0.000000 |
| 50% | 378.500000 | 2013.000000 | 0.000000 | 0.000000 | 4.000000 |
| 75% | 567.250000 | 2016.000000 | 0.000000 | 19.000000 | 6.000000 |
| max | 11415.000000 | 2019.000000 | 1.000000 | 146.000000 | 10.000000 |

**Table 4 Teams by runs/wickets**

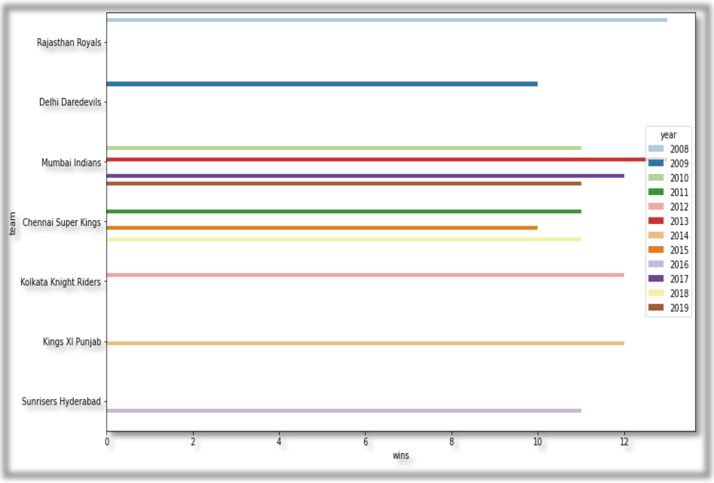


**Fig 2: Bar Plot IPL Matches year 2008 to 2019**

From the bar plot in Fig 2,3,4 and 5 inferences can be drawn regarding the years under consideration , the maximum number of matches hosted,, the top players etc.the total victories in which a particular team has scored, the maximum wins (and also the number of wins). The study reveals that Mumbai Indians had had the most wins in 4 seasons (2010, 2013, 2017, & 2019).

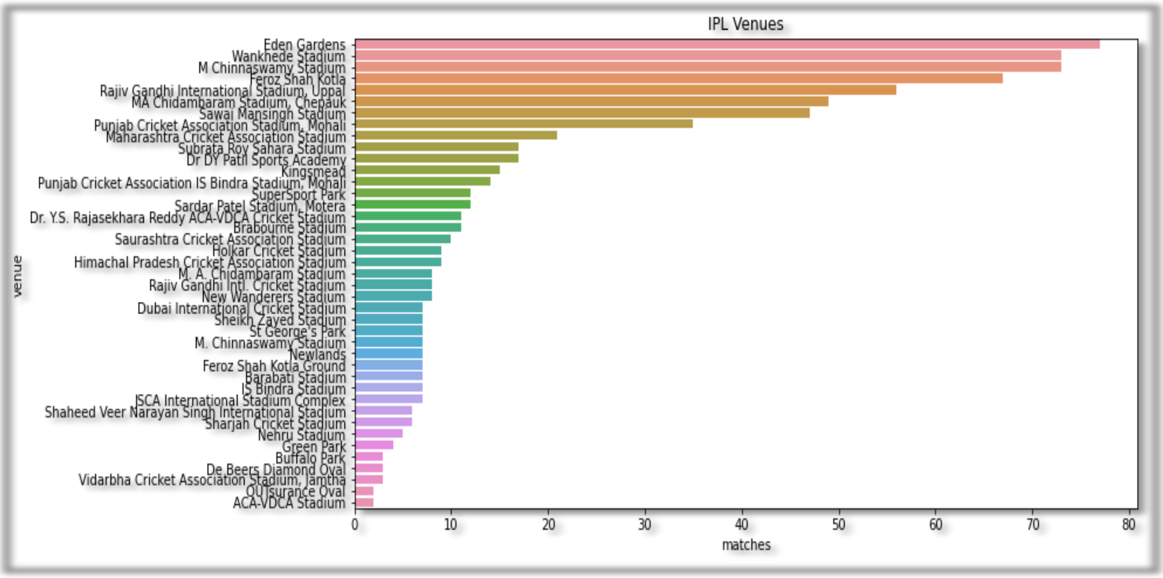
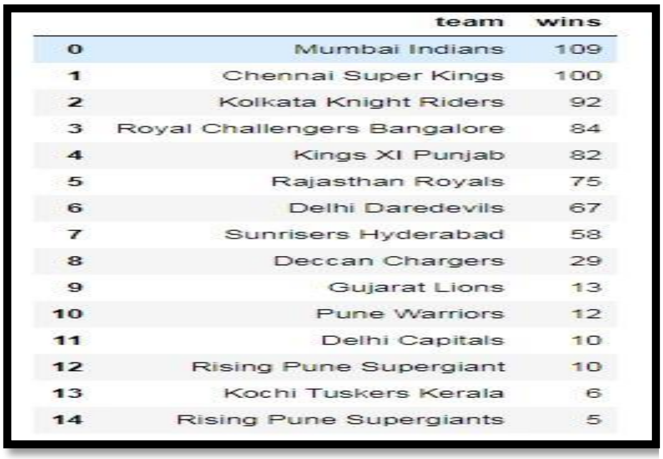
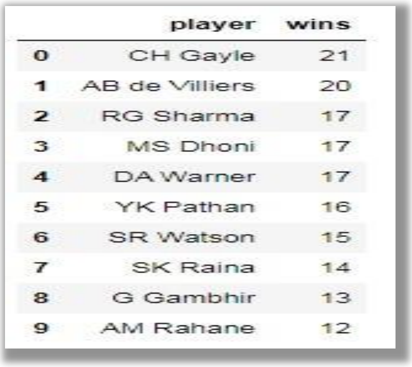


**Fig 3: Bar Plot IPL Matches venue that hosted the maximum number of matches**

Eden-Gardens had hosted the highest number of 'IPL-matches' followed by Wankhede and MChinnaswamy pitch.Till 2019, IPL-matches were hosted with 40 venues. **The most successful IPL team is determined as follows:**In a match of sports, each team competes for win. Hence, the team that has registered the greatest number of victories is the most successful.



| | team | wins |
|---|---|---|
| 0 | Mumbai Indians | 109 |
| 1 | Chennai Super Kings | 100 |
| 2 | Kolkata Knight Riders | 92 |
| 3 | Royal Challengers Bangalore | 84 |
| 4 | Kings XI Punjab | 82 |
| 5 | Rajasthan Royals | 75 |
| 6 | Delhi Daredevils | 67 |
| 7 | Sunrisers Hyderabad | 58 |
| 8 | Deccan Chargers | 29 |
| 9 | Gujarat Lions | 13 |
| 10 | Pune Warriors | 12 |
| 11 | Delhi Capitals | 10 |
| 12 | Rising Pune Supergiant | 10 |
| 13 | Kochi Tuskers Kerala | 6 |
| 14 | Rising Pune Supergiants | 5 |



| | player | wins |
|---|---|---|
| 0 | CH Gayle | 21 |
| 1 | AB de Villiers | 20 |
| 2 | RG Sharma | 17 |
| 3 | MS Dhoni | 17 |
| 4 | DA Warner | 17 |
| 5 | YK Pathan | 16 |
| 6 | SR Watson | 15 |
| 7 | SK Raina | 14 |
| 8 | G Gambhir | 13 |
| 9 | AM Rahane | 12 |

**Table 5 : Most successful IPL team wins**

**Table 6 Most Valuable Player**

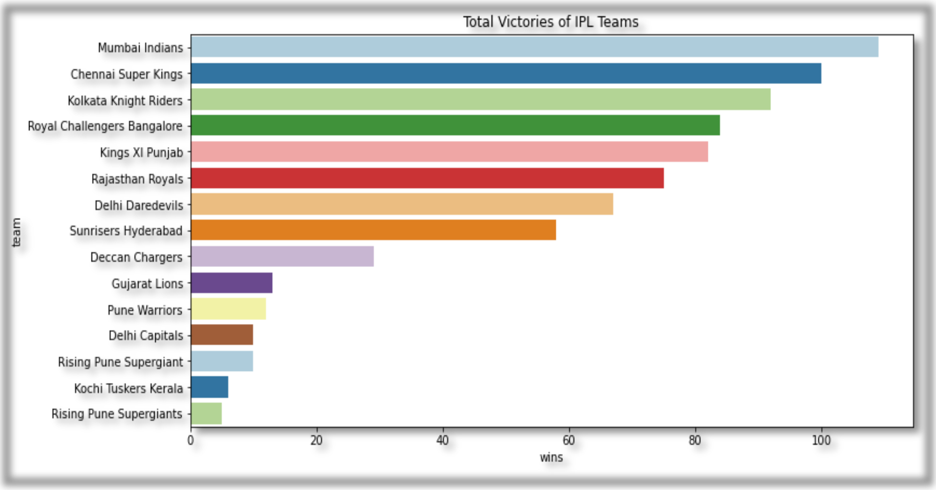


**Fig 4: Total Victories of IPL Teams**

Mumbai-Indians had won the maximum I.e. 109 ,followed by Chennai Super-Kings & Kolkata Knight Riders. The famous cricketer Chris Gayle won the valuable player of the match awards. Six Indian players have figured in the top ten IPL players list.

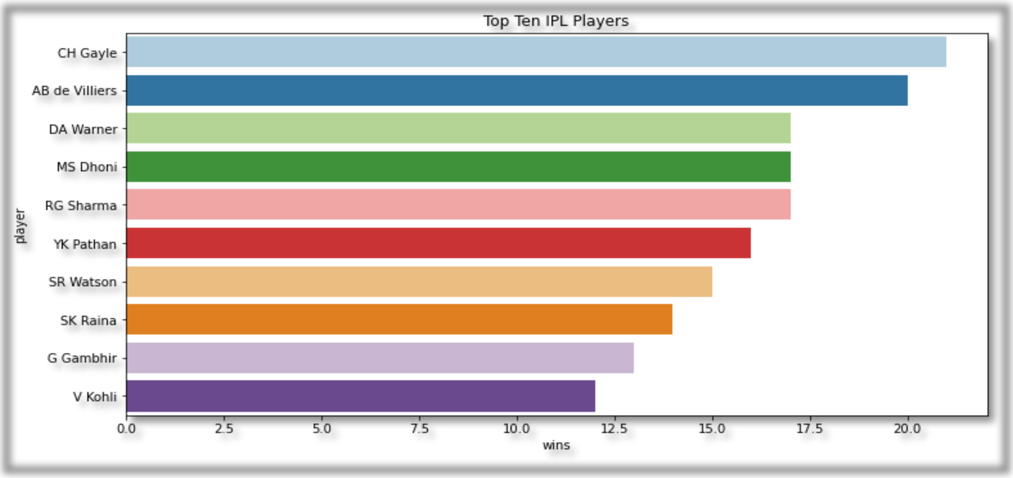


**Fig 5: Bar Plot of Top Ten IPL Players Predictive Modelling**

In this research Linear Regression, Random Forest and Ridge Regression algorithms have been used to predict the total score by taking independent variables like -'runs', 'wickets', 'overs',

“from

sklearn.linear_model the

'striker', 'non-striker'.

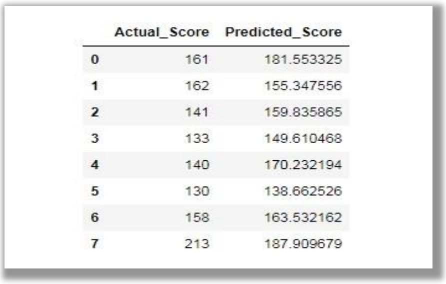
**Linear Regression**

Linear Regression Model is imported

LinearRegression()”

“lin_model.fit(x_train, y_train) ”

**Table 6 Linear-**



| | Actual_Score | Predicted_Score |
|---|---|---|
| 0 | 161 | 181.553325 |
| 1 | 162 | 155.347556 |
| 2 | 141 | 159.835865 |
| 3 | 133 | 149.610468 |
| 4 | 140 | 170.232194 |
| 5 | 130 | 138.662526 |
| 6 | 158 | 163.532162 |
| 7 | 213 | 187.909679 |

**Table 7 Comparing Actual Score and Predicted Score**

“from sklearn.model_selection import cross_val_score’

print(cross_val_score(lin_model, x_train,y_train).mean()*100)

**Output**

**50.58%**

**Table 8 To calculate cross-valscore**

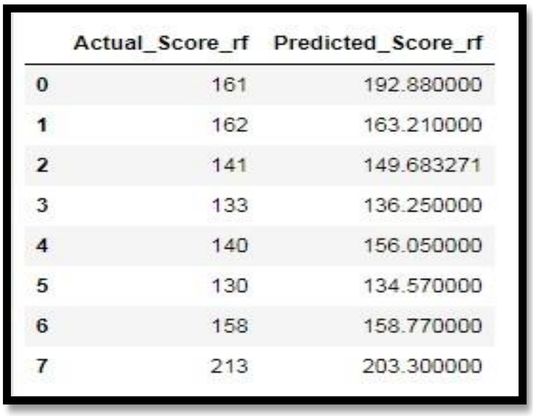
**Regression**

Table 6 to Table 8 showcases the action of linear regression with and without cross validation and prediction accuracy. The researchers then considered Random Forest Algorithm with and without cross validation evaluation function as show cased below

**Random Forest Algorithm**

"from sklearn.ensemble import RandomForestRegressor"
'rfr_model =

"RandomForestRegressor( n_estimators=100, max_features=none)" 'rfr_model.fit(x_train, y_train)



| | Actual_Score_rf | Predicted_Score_rf |
|---|---|---|
| **0** | 161 | 192.880000 |
| **1** | 162 | 163.210000 |
| **2** | 141 | 149.683271 |
| **3** | 133 | 136.250000 |
| **4** | 140 | 156.050000 |
| **5** | 130 | 134.570000 |
| **6** | 158 | 158.770000 |
| **7** | 213 | 203.300000 |

"from sklearn.model_selection import cross_val_score" "Print

(cross_val_scroe(rfr_model,x_train, y_train).mean()*100)
**Output**
**64.86%**

**Table 9: Random Forest Algorithm**

**Table : 10 Random Forest Algorithm: Comparing Actual Score and Predicted Score**

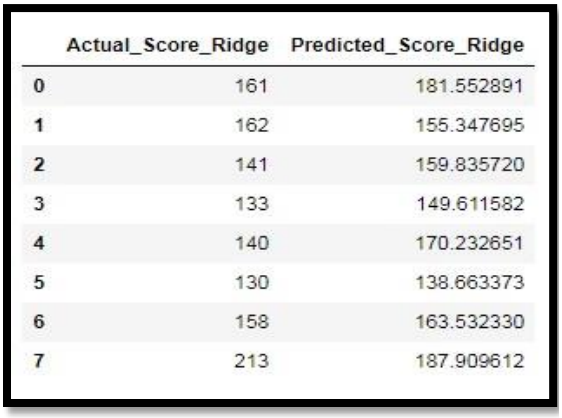
**Table 11: Random Forest Algorithm: Calculation of Cross Val Score**

It is envisaged that calculation of cross value score for the Random Forest Algorithm with respect to the score prediction shows considerable accuracy as can be inferred from Table 11.

The ridge regression algorithm was taken into consideration since there exist significant correlation between runs_made and strikers in the team which was ascertained to be 86% by using CORREL function .Similarly runs and strikers suffered a low correlation whereas pitch and team had high correlation in terms of wins. Incidentally Ridge Regression also could not predict accurately the score.

### Ridge-Regression Algorithm

from sklearn.linear_model sklearn.model_selection ‘ridge_regressor = ‘Ridge(alpha is equal to .01) ‘Print(cross_val_scroe(rid ‘ridge_regressor.fit(x_train, y_train’) y’_train.mean()*100)



| | Actual_Score_Ridge | Predicted_Score_Ridge |
|---|---|---|
| 0 | 161 | 181.552891 |
| 1 | 162 | 155.347695 |
| 2 | 141 | 159.835720 |
| 3 | 133 | 149.611582 |
| 4 | 140 | 170.232651 |
| 5 | 130 | 138.663373 |
| 6 | 158 | 163.532330 |
| 7 | 213 | 187.909612 |

“from import Ridge

import cross_val_score”

ge_regressor, x_train,

Output
**50.58%**

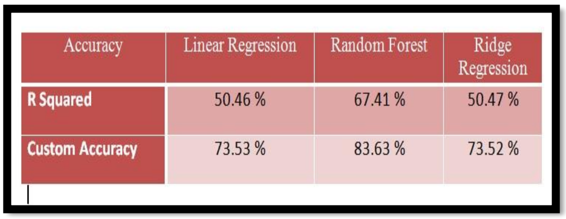
**Table 11: ‘Ridge-Regression**

**Table 12: Ridge Regression: Comparing Actual Score and Predicted Score**

**Table 13: Ridge Regression: Calculate Cross Val Score**

### Result Interpretation

By comparing three algorithms, random forest algorithm is predicting with considerable accuracy as compared to linear regression and ridge regression.



| Accuracy | Linear Regression | Random Forest | Ridge Regression |
|---|---|---|---|
| R Squared | 50.46 % | 67.41 % | 50.47 % |
| Custom Accuracy | 73.53 % | 83.63 % | 73.52 % |

**Table 14: Comparing three algorithms: linear Regression, Random Forest, and Ridge Regression**

### Findings and Observations

In this research study the following observations were made :Mumbai-Indians’ had secured the largely wins in 4-seasons (‘2010, 2013,2017,& 2019’).Eden-Gardens’ had hosted the greatest number of IPL-matches. Till-2019, IPL matches’ be hosted with 40 venues’.Mumbai-Indians is the largely winning team (as they have won the highest number of IPL-matches’ -109) followed with Chennai Super - Kings & Kolkata Knight Riders.Chris Gayle has won the maximum number of players of the match title.In this research it was proven that Random Forest Algorithm provided good accuracy as compared to Linear Regression and Ridge Regression.

### Conclusion

Predicting’ the winning team in ‘sports-cricket is very ‘challenge & complicated. Technology like machine learning and other latest tools made it much simple and easy. The factor like toss, venue and player habits has influence on result. Past performance play vital role in predictive model. Future scope of study can be expanding more attributes of an opponent team, change skill set of batsmen along with opponent team players and weather condition. Predicative model can be used for other indoor game and personal game.

## References

Wired' (2017), "The unlikely secret behind benfica's fourth consecutive primeira liga title".

T. A. Severini' (2014), "Analytic methods in sports: Using mathematics & statistics to understanddata from baseball, football, basketball, & other sports. Chapman & Hall/CRC".

H. Ghasemzadeh' & R. Jafari (2010), "Coordination analysis of human movements with body sensor networks: A signal processing model to evaluate baseball swings," "IEEE Sensors Journal, vol. 11, no. 3, pp. 603–610,".

'R. Rein & D. Memmert (2016), "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science," "SpringerPlus, vol. 5, no. 1, p. 1410".

'Sonu Kumar & Sneha Roy (2018)." "Score Prediction & Player Classification Model in the Game of Cricket Using ML", international journal of scientific and engineering research, volume" 9, issue 8, august-2018 ISSN 2229-5518 pg 237 , at https://www.ijser.org/researchpaper/"Score-Prediction-and-Player-Classification-Model
-in - the-Game-of-Cricket-Using-Machine-Learning.pdf"

Anik Shah, Dhaval Jha, Jaladhi Vyas (2016), "winning and score predictor (wasp) tool",
International conference onrecent innovation in scince, technology, amangemnt and Environmnet Indain federation of United Nation Association New Delhi ( India ) (ICRISTME-16) PG 460-464

'Ch Sai Abhishek, P Yuktha, Meghana K S, & MV Sudhamani (2019) , " Predictive Analysis of IPL Match Winner using Machine Learning Techniques", "International Journal of Innovative Technology & Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume 9 Issue 2S, pg 430-434".

'G. Sudhamathy' & G. Raja eenakshi (2020)," Prediction on IPL-data using machine learning(ML) techniques in R-package", "ICTACT Journal on soft computing, October 2020, Volume-11, Issue: 01".

Akhil Nimmagadda, Manigandla Venkatesh, Nidamanuri Venkata Kalyan, Chavali Gopi Raju, Nuthi Naga Sai Teja [2018],"Cricket score & winning prediction using data mining(DM)", "International Journal of Advanced Research and development ,Volume- 3, Issue-3, Available online at: www.ijarnd.com".

'Vrushali Y Kulkarni, & Pradeep K Sinha[2014]'. "Effective Learning & Classification using Random Forest Algorithm (RFA)", "International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May-2014".

Sinhgad Institute of Management
PUNE
411 041.

Singh

Dr.Chandrani Singh ,Director –MCA,SIOM

# A Case Study on Cryptocurrencies a Good Investment:The Predictive and Perspective Analysis by Using R-Language

Dr. Chandrani Singh[1] , Dr. Ramesh D Jadhav[2,] , Dr.Sunil Khilari [3] , Mr.Ravi Mourya[4]

Sinhgad Institute of Management, Pune (India)

## Abstract

Cryptocurrency which evolved a few years back, is built around facilitating electronic exchange of digital encrypted currency using a peer-to-peer network. Bitcoin being the most popular cryptocurrency, is paving the way and creating disruptions to the existing financial payment systems that have been there for over several decades.While such contemporary types would never replace the traditional, they could pave the way for the global markets to operate seamlessly by clearing away the global barriers specifically in the segment of the exchange rates.There has been considerable advancement in technology but the success of incorporating a given technology is solely governed by the market demands. Cryptocurrencies may in due course revolutionize trading practices with practically minimal to no trading fees in particular.This specific working paper will help investigate the metrics as growth in price and volume of trade for three popular coins namely Bitcoin (BTC), Ethereum (ETH), and Dogecoin (DOGE) and help establish insights for researchers and users to further their study.

**Keyword**: Cryptocurrency, Bitcoin (BTC), Ethereum (ETH), Dogecoin (DOGE).

## 1. Introduction of Cryptocurrency

Cryptocurrency is a technology entailing revolution around the digital payment system that helps verify transactions independently and not with the support of banks.Cryptocurrency being a peer-to-peer system ensures that one can assist an individual to send and receive payments from any time anywhere.Instead of physical money being exchanged, cryptocurrencies exist purely in digital format in wallets and can be transacted and recorded in the public ledger(Kaspersky, 2021).

The name Cryptocurrency came from the concept of making use of encryption ensuring safety and security to confirm the transactions.This makes use of advanced coding concepts and practices involved in storing and sending cryptocurrency data between wallets and public ledgers.The first cryptocurrency that received popularity was Bitcoin that was used to trade for profits, with speculators at times driving prices skyward (Sheikh Owais ,2022)

### Working Methodology of Cryptocurrency

Cryptocurrencies that run on distributed public ledger, units of such cryptocurrencies are mined,that involves using processor power to solve complex mathematical problems that generate coins (Kaspersky,2021).Users can buy such currencies from the brokers, and can spend those currencies by making use of wallets that are encrypted.Generally when one owns a cryptocurrency, a key enables the transfer of that currency from a person to the other without a third party's intervention.Although Bitcoin has been around since 2009 (Jake Frankenfield, 2021), the applications of block chain technology and evolution of cryptocurrency are still emerging more usage in terms of currencies are expected in the future.Transactions that can happen through cryptocurrencies include bonds, stocks, and other financial assets that can eventually be traded ( Chizoba Morah, 2020).

## 2. Research Methodology

### 2.1 Data Collection

The datasets of BTC, ETH, DOGE cryptocurrencies were collected from Kaggle that spanned years from 2013 to 2015.The data was validated and up-to date and the codes were written and executed in RStudio.

### 2.2 Data Exploration

All the three datasets were merged to form one, and it consisted of 7911 observations across 10 attributes as listed below:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| Sr.No | Name | Symbol | Date | High | Low | Open | Close | Volume | Market cap |

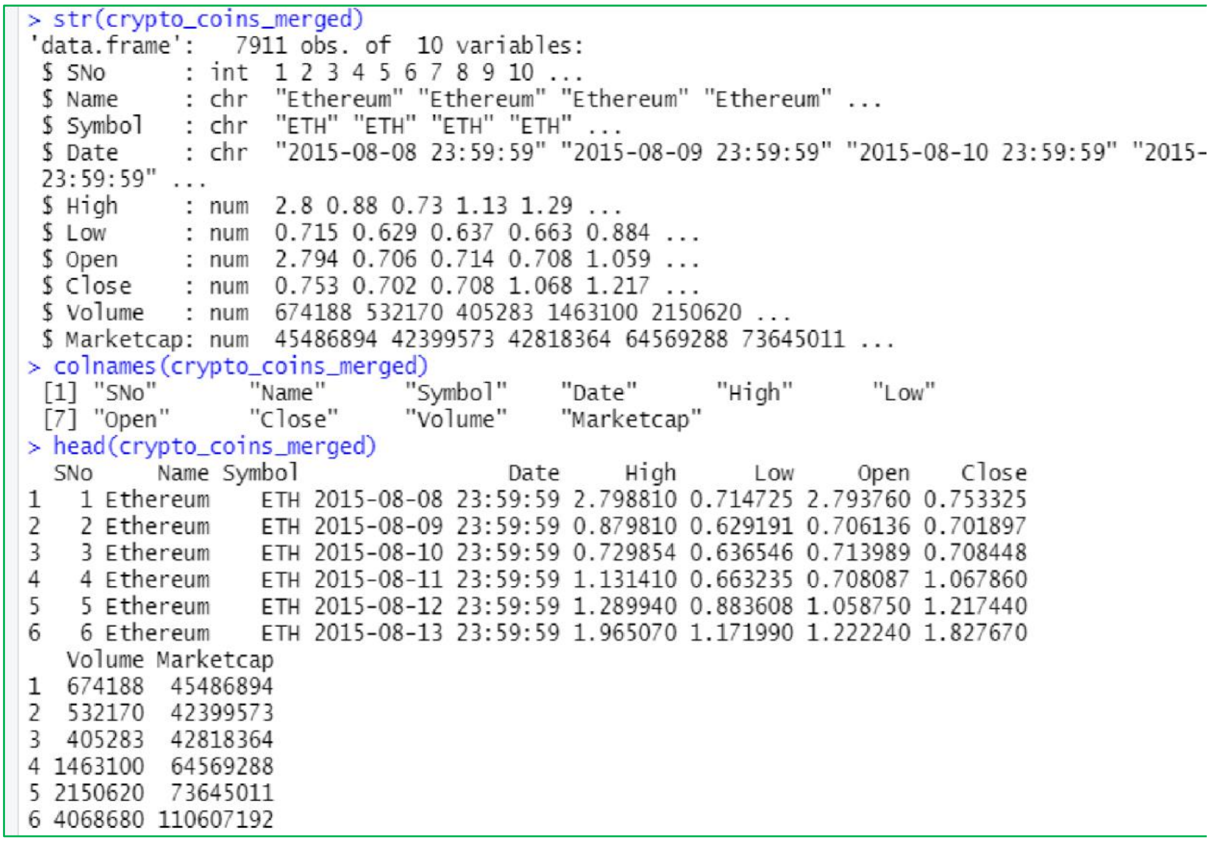
### 2.3 Data Analysis

In order to have a robust data analysis and visualization in representing the results,Tidyverse, was installed using the following command and then running library(Tidy verse) as shown below :

```
install.packages('tidyverse') library(tidyverse)
```

Next, it was required to mount three .csv files that were considered for the study and form a cohesive data frame from the same.The requisite lines of code then helped create a data frame by reading and binding the csv files and creating a single merged file which was then inspected with the str(), head(), and colnames() functions.

```
coin_Bitcoin <- read.csv("E:/R_Scripts/Case study/coin_Bitcoin.csv")

coin_Dogecoin <- read.csv('E:/R_Scripts/Case study/coin_Dogecoin.csv')

coin_Ethereum <- read.csv('E:/R_Scripts/Case study/coin_Ethereum.csv')

crypto_coins_initial <- rbind(coin_Ethereum, coin_Dogecoin)

crypto_coins_merged <- rbind(crypto_coins_initial, coin_Bitcoin)

str(crypto_coins_merged) colnames(crypto_coins_merged)


head(crypto_coins_merged)
```



```
> str(crypto_coins_merged)
'data.frame':   7911 obs. of  10 variables:
 $ SNo      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Name     : chr  "Ethereum" "Ethereum" "Ethereum" "Ethereum" ...
 $ Symbol   : chr  "ETH" "ETH" "ETH" "ETH" ...
 $ Date     : chr  "2015-08-08 23:59:59" "2015-08-09 23:59:59" "2015-08-10 23:59:59" "2015-
 23:59:59" ...
 $ High     : num  2.8 0.88 0.73 1.13 1.29 ...
 $ Low      : num  0.715 0.629 0.637 0.663 0.884 ...
 $ Open     : num  2.794 0.706 0.714 0.708 1.059 ...
 $ Close    : num  0.753 0.702 0.708 1.068 1.217 ...
 $ Volume   : num  674188 532170 405283 1463100 2150620 ...
 $ Marketcap: num  45486894 42399573 42818364 64569288 73645011 ...
> colnames(crypto_coins_merged)
 [1] "SNo"       "Name"      "Symbol"    "Date"      "High"      "Low"
 [7] "Open"      "Close"     "Volume"    "Marketcap"
> head(crypto_coins_merged)
  SNo     Name Symbol                Date     High      Low     Open    Close
1   1 Ethereum    ETH 2015-08-08 23:59:59 2.798810 0.714725 2.793760 0.753325
2   2 Ethereum    ETH 2015-08-09 23:59:59 0.879810 0.629191 0.706136 0.701897
3   3 Ethereum    ETH 2015-08-10 23:59:59 0.729854 0.636546 0.713989 0.708448
4   4 Ethereum    ETH 2015-08-11 23:59:59 1.131410 0.663235 0.708087 1.067860
5   5 Ethereum    ETH 2015-08-12 23:59:59 1.289940 0.883608 1.058750 1.217440
6   6 Ethereum    ETH 2015-08-13 23:59:59 1.965070 1.171990 1.222240 1.827670
   Volume Marketcap
1  674188  45486894
2  532170  42399573
3  405283  42818364
4 1463100  64569288
5 2150620  73645011
6 4068680 110607192
```

**Table 1: R language code and Output**

The merged file and the output data set is shown in Table 1. On this data set the analysis and visualizations are performed of which the first is to examine closing prices over time using ggplot2.

ggplot(data=crypto_coins_merged)+geom_point(aes(x=Date,y=Close,color=Name))+labs(title="Popular Crypto Prices Over Time", y="Price in USD", x="Time") +facet_wrap(~ Name)

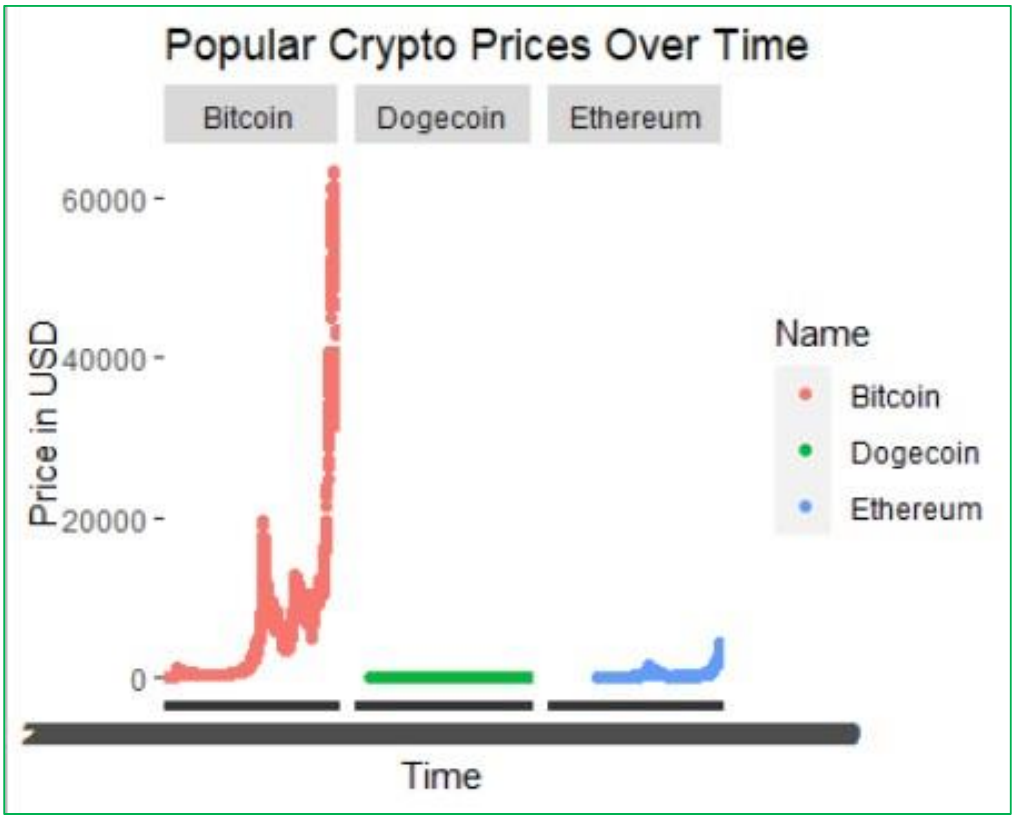


**Figure 1: Popular Cryptocurrency Prices over Time**

**Analysis:** On this graph of Figure 1, the price of Bitcoin is so high that it totally outperforms the most useful information on the other two coins. To view the trend of  the two cheaper coins (DOGE + ETH), let's assume that most casual investors probably did not have 30000 to 60000 dollars just to put in cryptocurrency,so individual cheaper coins trend had to be extracted and analyzed as has been done in the succeeding sections of the paper.

```
ggplot(data = coin_Dogecoin, aes(x=Date,y=Close)) + geom_point() + labs(title="Dogecoin", y="Price in USD", x="Time")
```

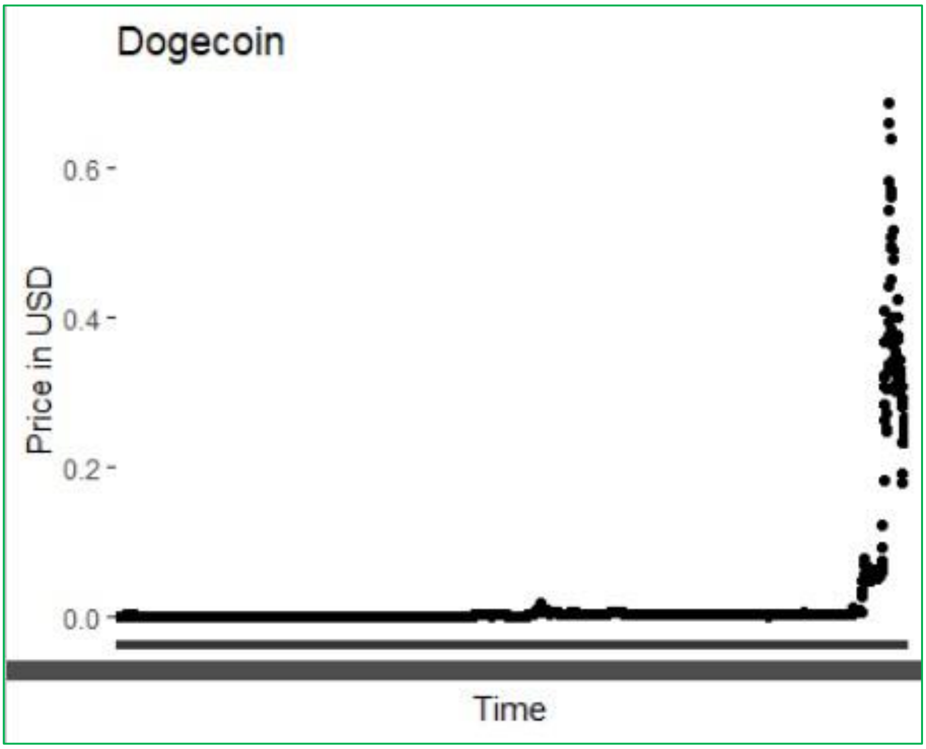


**Figure 2: Dogecoin and change in price by time**

**Analysis:** Additionally, let's add another layer to the ggplot as viewing lots of scattered points can obscure trends, so geom_smooth() is used as shown in Figure 2. Additionally the smoothing method "gam" is used. Loess smoothing would be a more accurate choice, but it takes significantly more time to compute with high observation counts.

Since N > 1000, so Loess smoothing will not be considered. There is another problem with smooth (), and that it is expecting continuous variables along the x axis. So the most ideal is to just connect points between days, and get a line that is technically not mathematically smooth.Below is the code for the same.

ggplot(data=coin_Ethereum,aes(x=Date,y=Close))+geom_point()+geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) + labs(title="Ethereum", y="Price in USD", x="Time")

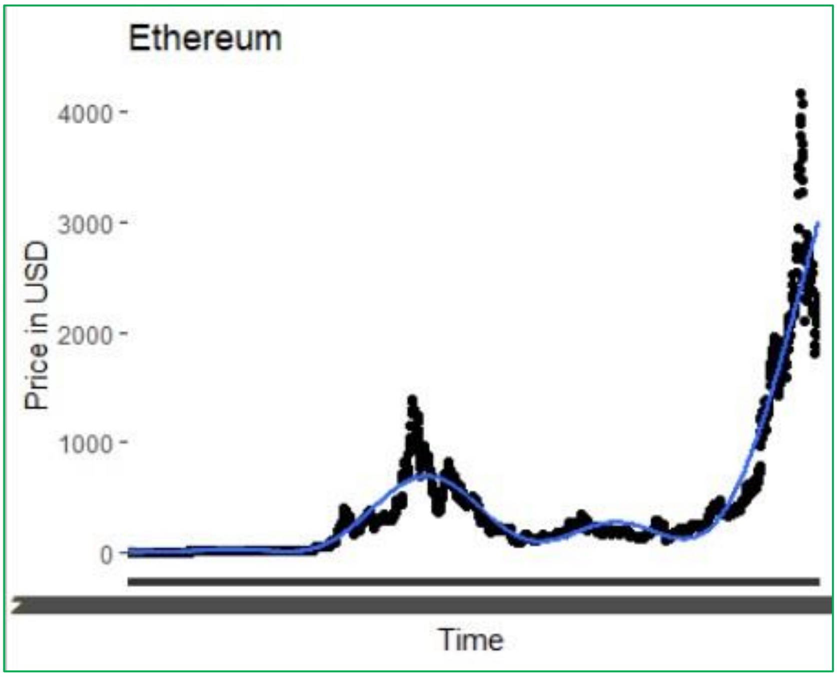


**Figure:3 Ethereum and change in price over time**

**Analysis:** When  the three cryptocurrencies, from visual inspection they tend to have explosive growth and relatively high volatility as shown in Figure 3. In all following graphs splitting BTC results from the other two coins, to overcome the domain issue has been considered. Now evaluating trading volume against closing price the graph is depicted as follows:

```
ggplot(data = crypto_coins_initial,aes(x=Volume,y=Close, color=Name)) +
        geom_point() +geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) + labs(title="Price vs
Volume", y="Price in USD", x="Volume traded per Day") + facet_wrap(~ Name)
```

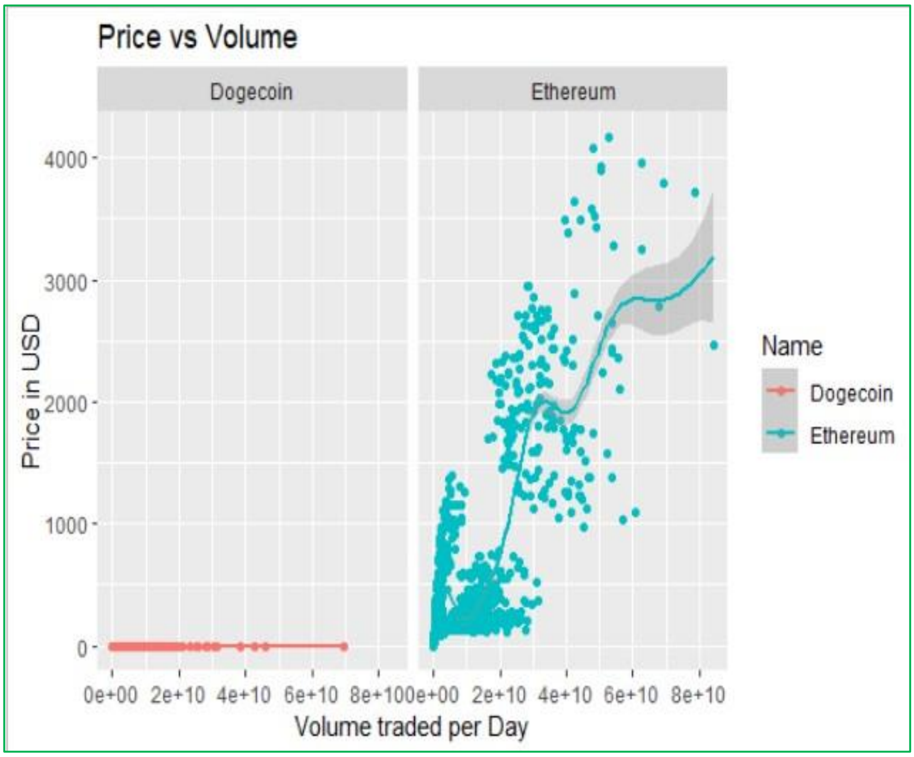


**Figure 4: Price vs Volume (Dogecoin, Ethereum)- volume traded per day**

**Analysis:** Ethereum has a relatively nice volume to price correlation, but Dogecoin doesn't, even though it seems to have a comparable volume of daily trades to Ethereum as shown in Figure 4.There is an interesting curved band where there seems to be neither price nor volume.This graph contains no data on  volume over time,and Bitcoin's behavior with regard to the model is investigated and followed.It is seen that the regression begins to fail at the very high end and linear fitment is  better where data points are clustered tightly.

```
ggplot(data = coin_Bitcoin,aes(x=Volume,y=Close)) +geom_point() +
geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) +labs(title="Price vs Volume", y="Price in USD", x="Volume traded per Day") +facet_wrap(~ Name)
```

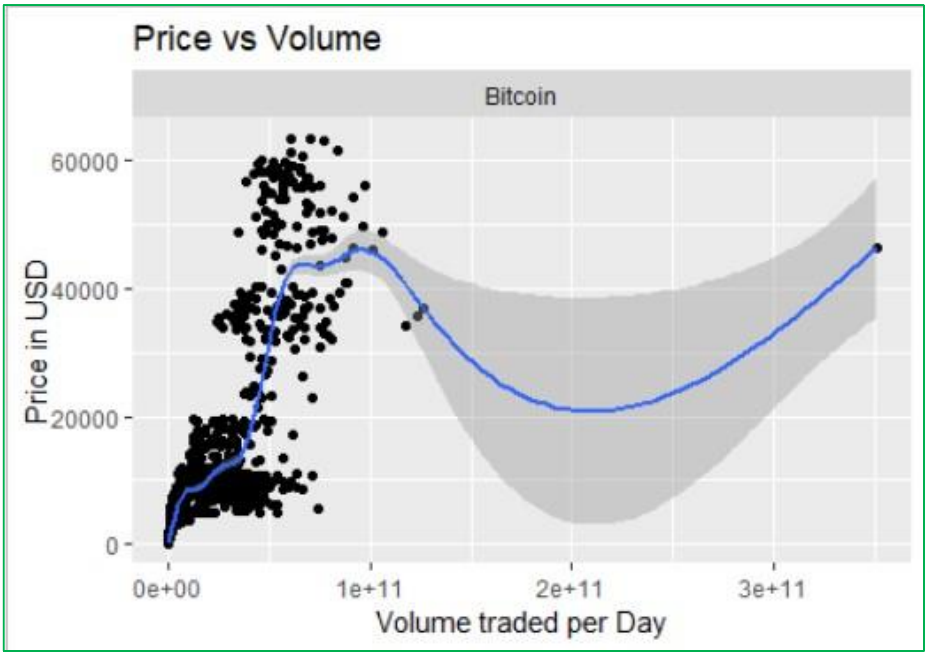


**Figure 5: Price vs Volume (Bitcoin)**

Analysis : Here one failure of smooth('gam') is noticed in Figure 5.A spike up in trade volume, broke the regression and this formed an absolute maximum for the BTC data.The method 'gam' did not handle the outlier well. Instead of 'gam', 'lm' is used as a linear model.

```
ggplot(data=coin_Bitcoin,aes(x=Volume,y=Close))+geom_point()+geom_smooth(aes(group= -1), method ='lm') + labs(title="Price vs Volume", y="Price in USD", x="Volume traded per Day") +facet_wrap(~ Name)
```

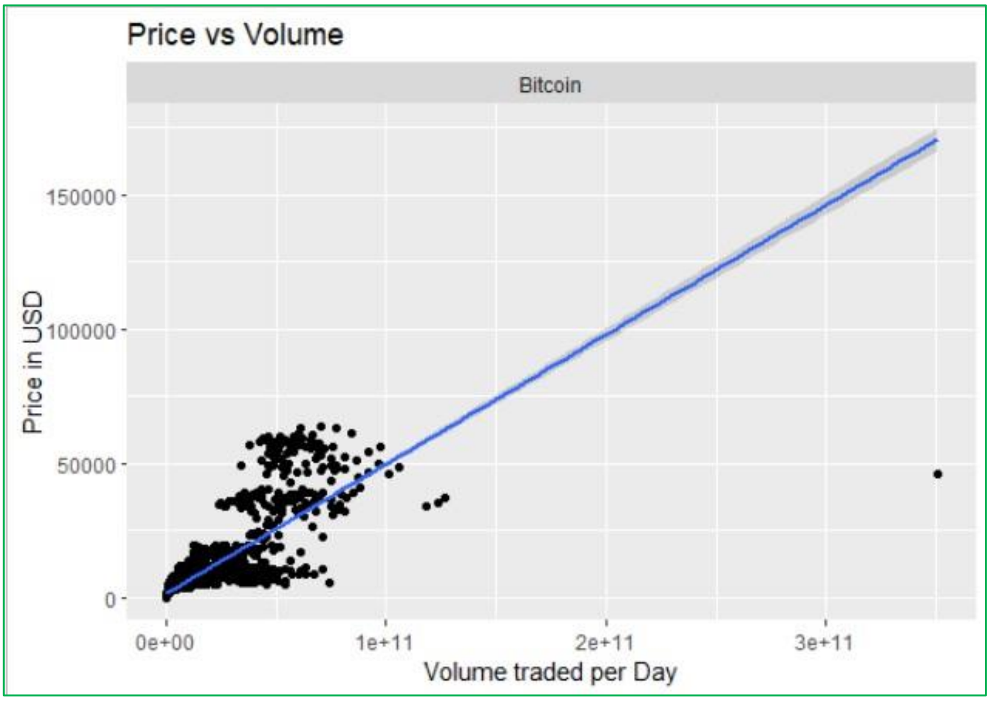


**Figure 6: Price vs Volume-Bitcoin (Volume traded per Day)**

**Analysis :** Here a line of best fit for BTC price compared to volume is envisaged as in Figure 6. In general, as trade demand increases, so does BTC price.The volume is trending over time for the three coins is as follows as :

ggplot(data=coin_Ethereum,aes(x=Date,y=Volume))+geom_point()+geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) +labs(title="Trade Volume over Time", y="Volume Traded per Day", x="Time") + facet_wrap(~ Name). This below figure 7 shows that ETH is being traded fairly frequently, and is trending upward in general.

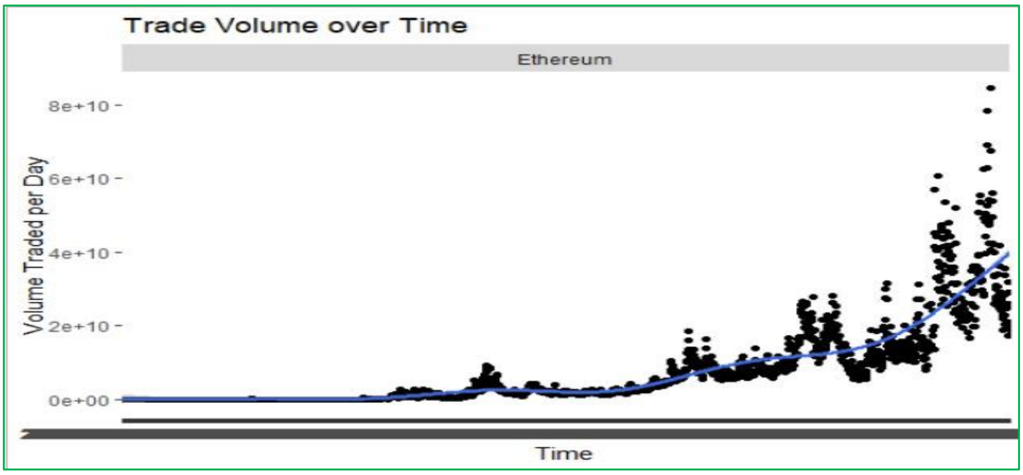
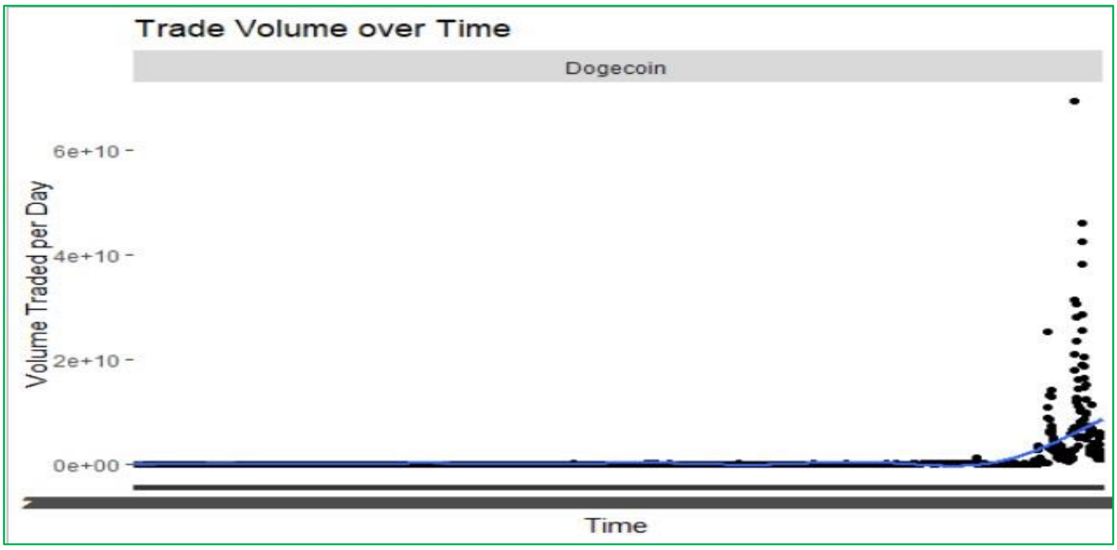


**Figure 7:Ethereum (Trade Volume over time)**

ggplot(data=coin_Dogecoin,aes(x=Date,y=Volume))+geom_point()+geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) +labs(title="Trade Volume over Time", y="Volume Traded per Day", x="Time") + facet_wrap(~Name)

**Figure 8: Dogecoin (Trade Volume over time)**



ggplot(data = coin_Bitcoin,aes(x=Date,y=Volume)) +
geom_point() +

```
  geom_smooth(aes(group= -1), method ='gam', formula =y ~ s(x) ) +
labs(title="Trade Volume over Time", y="Volume Traded per Day", x="Time") +
facet_wrap(~ Name)
```

**Figure 9: Bitcoin (Trade Volume over time) 3. Conclusion**

This predictive and perspective analysis shows that trade volume is trending up for all of the coins. Additionally it was found that the prices of ETH and DOGE were trending up. Through this analysis, it was established that many cryptocurrencies were prone to explosive growth in price and high volatility.Many coins had also seen strong growth in market trade volume It can be advised that a safe investment strategy might focus on taking a few hundred or thousand dollars and investing in a diverse range of currently low-cost cryptocurrencies. Widely investing in different coins will help mitigate the risk of the highly volatile prices. Over a period five years many coins will make the transition from low cost to high. More investigation is needed into factors and signs of explosive coin price growth to choose the best low-cost coins to purchase as investments.

## 4. References

1. Mi Yeon Hong, Ji Won Yoon(2022)The impact of COVID-19 on cryptocurrency markets: A network analysis based on mutual information
2. Peter D. DeVries (2016) An Analysis of Cryptocurrency, Bitcoin, and the Future
3. Giancarlo Giudici1, Alistair Milne (2020) Cryptocurrencies: market analysis and perspectives
4. Jonathan Chiu, Thorsten Koeppl (2017),"The Economics of Cryptocurrencies – Bitcoin and Beyond"
5. Kaggle Data set: https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory
6. Shailak Jani(2018) The Growth of Cryptocurrency in India: Its Challenges & Potential Impacts on Legislation.
7. Sheikh Owais,Feb,2022,"Introduction to Cryptocurrency",https://www.risingkashmir.com/-Introductionto-Cryptocurrency-100414
8. The Goldman Sachs Group, Global Macro Research (2021), CRYPTO: A NEW ASSET CLASS?
9. Chizoba Morah, August,2020,"Why Are Most Bonds Traded on the Secondary Market "Over the Counter"?,https://www.investopedia.com/ask/answers/09/bond-over-the-counter.asp
10. Congressional Research Service(2020) Cryptocurrency: The Economics of Money and Selected Policy .
11. Jake Frankenfield,November,2021,"What Is Bitcoin?", https://www.investopedia.com/terms/b/bitcoin.asp
12. R for Data Science: https://r4ds.had.co.nz/index.html
13. https://www.kaspersky.com/resource-center/definitions/what-is-cryptocurrency
14. Data Visualization in R: https://r4ds.had.co.nz/data-visualisation.html
15. https://en.wikipedia.org/wiki/Cryptocurrency

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

# Business Intelligence Tool-Power BI for Performance Management

Mr.Bharat Mane[1], Dr.Chandrani Singh[2] , Dr.Sunil Khilari[3]

Sinhagad Institute of Management, Pune (India)

------------------------------------------------------------------------------------------------------------------

**Abstract:** An experiment of the available methods of data gathering, storing, processing, Analysis and visualization for real-time information must be carried out in order to achieve the most understandable, efficient, effective and accurate visualization of information. Customizing platforms and designing unique console are two crucial tasks to do in order to accurately and effectively visualize the data. In this paper, dashboard platforms and methods are explored. Researcher created a console and dynamic dashboard based on real-time data. On-time, on-budget and market driven data at the conclusion are used to evaluate the influence of the built dashboards as novel Data Visualization methods. Henceforth dashboard users will thus be suited for interact with the data, which is supported by a unique collection of tables, maps and reports created by the dashboard itself. This will enable everyone to try a number of data visualization techniques while demonstrating how dashboards may be a novel and significant method to deliver accurate and effective information to decision makers for enhancing business intelligence with the help of Power-BI tool.

**Keywords:** Power-BI, Data Mining, Data Visualization, Tableau, Dashboard,

## 1. Introductions

This essay has discussed how safe hybrid configurations and simple IT system integration are advantageous to IT experts and innovators. Judges in the business world may use key analysis skills to make data search, dissection, and report generation simple. Business addicts no longer have to base their calculations exclusively on BI obtained from third parties because live dashboards and reports allow them to access and analyses all of their data in one location. By incorporating Power BI into your association's toolkit, you can make BI accessible to those who require it at the time they require it.

## 2. Problem Statement:

A Power BI study on business intelligence tools and characteristics of business intelligence software. In PowerBi there are Lots of Dashboard Options so new users are not able to understood The Proper Data

## 3. Objectives of the Study

1. To do the real-time stream analyses with the help of Power BI.
2. To access real-time analytics and data from a variety of sensors and social media sources,

3. Analyse and benchmark Power-BI with other data visualization tools.

**4. Motivation for the study**

The conventional method would be to read through and evaluate the dense facts of both scenarios. Obviously, doing this will pass a lot of time.

Information is easier for people to understand when it is visualized. Helps decision-makers understand a narrative, saving time and ensuring that they receive reliable information.

Motivates me to select this topic.

**5. Scope of the Study**

The purpose of this article is to learn how to use tools like Power BI, Qlik Sense and Tableau for data visualization. Tools can assist to the business in the few of below ways as-

- Understanding business requirements
- Immediate action
- Significant analysis
- Identifying patterns
- Identifying defects
- Understanding current trends

Technology's growing presence in the agricultural industry is obvious. There is a change in many robotically unfolding the content of pictures using natural language is core and challenging task. With the progression in computing power along with the accessibility of massive datasets, construction of models that can create legends (**Gounder, M.S., Iyer, V.V., Al Mazyad, A.,2016**) for an image has become promising. On the other side humans are able to easily describe the environments they are in. Given a picture, it's natural for a person to explain an immense amount of details about this image with a fast glance. Though big development has been made in computer vision task like recognizing of an object, classification, image sensing / classification, attribute classification and scene recognition are conceivable but it is a relatively creative task to let a computer define a picture that is forwarded to it in the form of a human-like sentence. Techniques adopted in the agriculture application. Through IoT and automation; we demonstrate the present state of the art and the potential of agricultural technologies in this study. A large-scale agricultural system needs a lot of upkeep, expertise, and oversight. We want to automate the following garden procedures using the provided model.

**6. Systematic Literature Survey**

a. Microsoft Power-BI (**Amrapali Bansal, A. K. Upadhyay**), this paper discusses how safe hybrid configurations and simple IT system integration are advantageous to IT experts and innovators. Judges in the business world may use key analysis skills to make data search, dissection, and report generation simple. Business addicts no longer have to base their calculations exclusively on BI obtained from third parties because live dashboards and reports allow them to access and analyses all of their data in one location. By incorporating Power BI into your association's toolkit, you can make BI accessible to those who require it at the time they require it.

b. Research Data Analysis with Power BI (**Vijay Krishnan, S Bharanidharan, G Krishnamoorthy**), This paper has enlighten proficiency with Power BI and demonstrated that it's a radical path to simplifying the business intelligence and data analytics space, whereby individuals and associations can easily give data, make reports or command them to be automatically created, aggregate them in dashboards, and participate in it with the least amount of time and trouble spent on it. It is clear that Power BI is a special opportunity for research institutes and professionals to meet their data analysis demands when this service is provided by a company of the calibre of Microsoft and with independent verification by Gartner that has compared the competition.

c. Operational Management in a Digital Environment (**Korotkova Kseniia, Kyiv, Ukraine**), this study has described tool to optimize pupil conditions reporting, the university will be suitable to contemporize and speed up the information analysis process. This will help bring the university's conditioning to a substitute creative position, as easily as deliver resources

d. Integrate Power-BI with WPF Desktop Applications (**Maria Dobreva, Nikolay Pavlov, Asen Rahnev**); This paper explores the functional conditions and perpetration of embedding Power BI resources in FDBA operation as a standard functionality to the Framework for Distributed Business Applications.

   - Sourcing of Data
   - Transforming Information
   - Creating Dashboards
   - Reports and Publish

The approach we've chosen provides users with a fluently accessible way to perform and view complex data analysis within the native operation. In future, we will probe styles to allow druggies to execute reports under selection of data.

**7.Power-BI Online Services and Technology Architecture**

The key function of power-BI is to create and share interactive reports and dashboards within and outside organizations. Further this tool enables searching / transforming / visualizing data from numerous types of the data sources especially cloud data or API's. Below figure shows some online services of Power-BI tool as-

Figure: 1- Power-bi online services

Total four phases are in Power-BI architecture, that delivers comprehensive information about each of them.

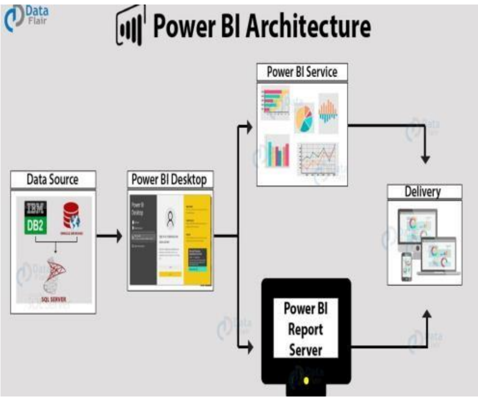


Figure: 2-Architecture of Power-BI

**8. Transforming information through Power-BI**

Power-BI deliver a screening window by choosing columns and entities subsequently the information is imported in the Power-BI environment. Further query can be edited if required and there are many transformation selections available to do such work. All these scenarios are elaborated with the help of below stated some Power-BI components and services. **a) Data Sourcing**

Power-BI could provide data from a large range of internet resources, formats and types of documents. To obtain the information, the files or documents should be imported into PowerBI and also this can be achieved by installing live service connection. If you import a PowerBI document then data sets that are compressed are limited to large up to one gigabyte. This can be shown in below figure.

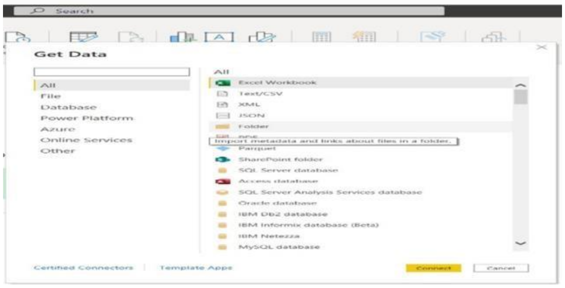
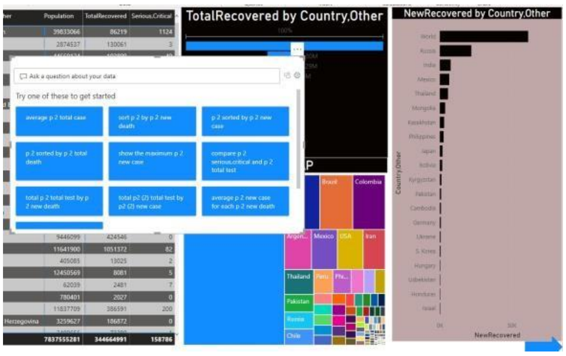


Figure: 3-Data sourcing

There is an important and efficient arrangement of custom visualization available by generating reports; we have to publish this with the Power-BI facilities. Further you can publish this on the cloud server also. **b) Dashboard Creation:**

With the help of Power-BI tool build dashboards by the single elements and/or by holding the page of the live report. If the report is saved and holding of single components, the visual retains the filter setting selected which has been shown in below figure.

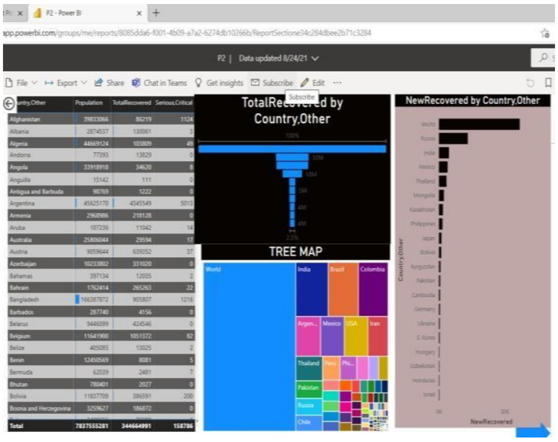

TotalRecovered by Country,Other
NewRecovered by Country,Other
Ask a question about your data
Try one of these to get started
average p 2 total case
sort p 2 by p 2 new death
p 2 sorted by p 2 new case
p 2 sorted by p 2 total death
show the maximum p 2 new case
compare p 2 serious,critical and p 2 total test
total p 2 total test by p 2 new death
total p2 (2) total test by p2 (2) new case
average p 2 new case for each p 2 new death
Brazil
Colombia
Argen...
Mexico
USA
Iran
Thailand
Peru
Phi...
Pakistan
Russia
Chile
World
Russia
India
Mexico
Thailand
Mongolia
Kazakhstan
Philippines
Japan
Bolivia
Kyrgyzstan
Pakistan
Cambodia
Germany
Ukraine
S. Korea
Hungary
Uzbekistan
Honduras
Israel
Country,Other
NewRecovered

| Population | TotalRecovered | Serious,Critical |
|---|---|---|
| 39833066 | 86219 | 1124 |
| 2874537 | 130061 | 3 |
| 9446099 | 424546 | 0 |
| 11641900 | 1051372 | 82 |
| 405085 | 13025 | 2 |
| 12450569 | 8081 | 5 |
| 62039 | 2481 | 7 |
| 780401 | 2027 | 0 |
| 11837709 | 386591 | 200 |
| 3259627 | 186872 | 0 |
| 7837555281 | 344664991 | 158786 |

Figure: 4 –Dashboard creation **c.Report and Publish:**

Once sourcing, retrieving and editing of the files or data, tool can create MIS reports.

Reports are the analysed data visualizations shown below.


P2 - Power BI
app.powerbi.com/groups/me/reports/8085dda6-f001-4b09-a7a2-6274db10266b/ReportSection34c284dbee2671c3284
P2 | Data updated 8/24/21
File
Export
Share
Chat in Teams
Get insights
Subscribe
Edit
TotalRecovered by Country,Other
TREE MAP
World
India
Brazil
Colombia
Argen...
Mexico
USA
Iran
Thailand
Peru
Phi...
Pakistan
Russia
Chile
NewRecovered by Country,Other
Country,Other
NewRecovered

| ...ntry,Other | Population | TotalRecovered | Serious,Critical |
|---|---|---|---|
| Afghanistan | 39833066 | 86219 | 1124 |
| Albania | 2874537 | 130061 | 3 |
| Algeria | 44669124 | 103809 | 49 |
| Andorra | 77393 | 13829 | 0 |
| Angola | 33918918 | 34620 | 8 |
| Anguilla | 15142 | 111 | 0 |
| Antigua and Barbuda | 98769 | 1222 | 0 |
| Argentina | 45625170 | 4345549 | 5013 |
| Armenia | 2968986 | 218128 | 0 |
| Aruba | 107236 | 11042 | 14 |
| Australia | 25806044 | 29594 | 17 |
| Austria | 9059644 | 639352 | 37 |
| Azerbaijan | 10233802 | 331020 | 0 |
| Bahamas | 397134 | 12035 | 2 |
| Bahrain | 1762414 | 265263 | 22 |
| Bangladesh | 166387872 | 905807 | 1216 |
| Barbados | 287740 | 4156 | 0 |
| Belarus | 9446099 | 424546 | 0 |
| Belgium | 11641900 | 1051372 | 82 |
| Belize | 405085 | 13025 | 2 |
| Benin | 12450569 | 8081 | 5 |
| Bermuda | 62039 | 2481 | 7 |
| Bhutan | 780401 | 2027 | 0 |
| Bolivia | 11837709 | 386591 | 200 |
| Bosnia and Herzegovina | 3259627 | 186872 | 0 |
| Total | 7837555281 | 344664991 | 158786 |

Figure: 5 –Reports Publishing

### d. Dashboards

Power-BI dashboard is location for data visualizations, graphs and charts from multiple essential reports, produced by fetching way that makes it easy to effective insights. The benefits of Power-BI are that dashboards are live and real-time.

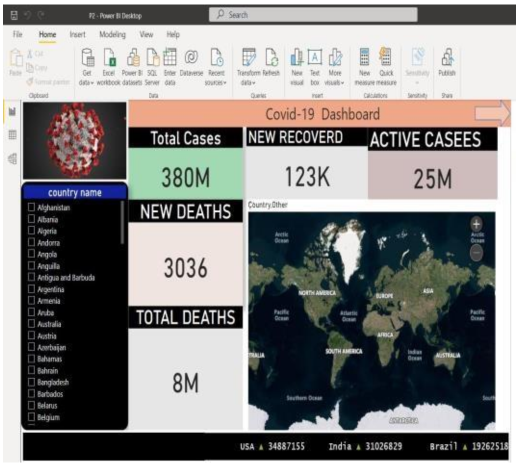


Figure: 6 - Dashboard

### e.Natural language query

Natural language query is a distinctive characteristic of Power-BI that let you ask queries to your dataset expressed in basic English which derives answers in the form of fresh visualizations, graph and charts.

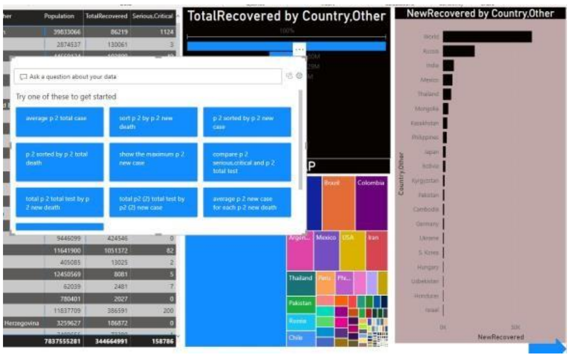

Population
TotalRecovered
Serious,Critical
TotalRecovered by Country,Other
NewRecovered by Country,Other
Ask a question about your data
Try one of these to get started
average p 2 total case
sort p 2 by p 2 new death
p 2 sorted by p 2 new case
p 2 sorted by p 2 total death
show the maximum p 2 new case
compare p 2 serious,critical and p 2 total test
total p 2 total test by p 2 new death
total p2 (2) total test by p2 (2) new case
average p 2 new case for each p 2 new death
Brazil
Colombia
Mexico
USA
Iran
Thailand
Pakistan
Russia
Chile
NewRecovered
Country,Other

Figure: 7- Natural language query **f.Subscribe, Comments, Share:**

It’s never been easier to stay up-to-date on your most important dashboards and reports.

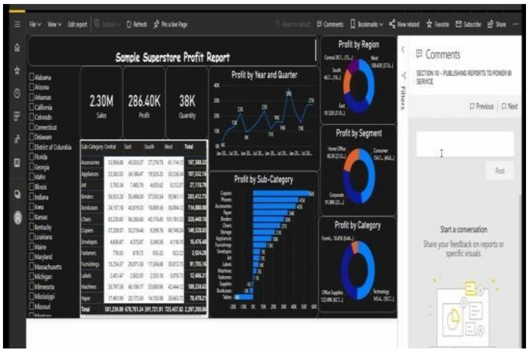


Figure: 8-Comments Dashboard

To share the data and report online administrator generate and manages site collections which can further facilitates the functions that available across site collections such as  InfoPath Forms Services. **g.Power Query**

Power Query is data processing software. Power Query could use to link to numerous types of data sources like web pages, databases, social media, API's ,internet log files, cloud storage etc. which can gather ,store, process ,analyse and combine data (append, merge, join etc.) from variety of locations.

Results of uploaded data:-

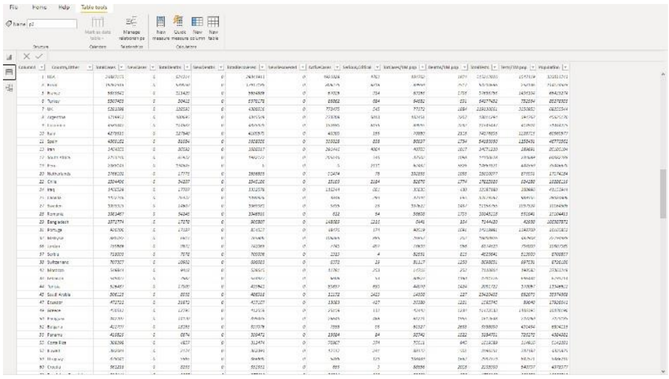
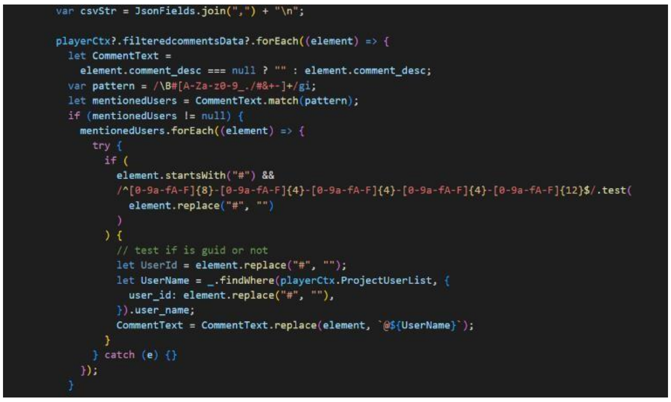
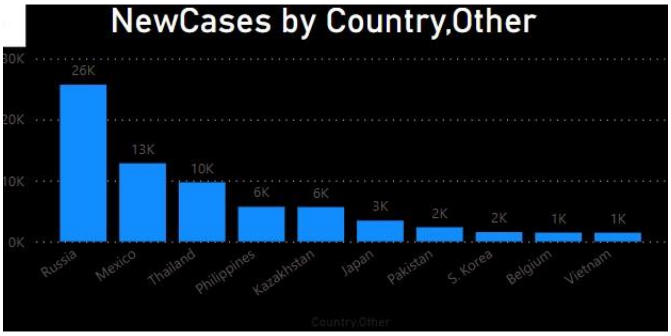


Table2: Code implementation for conversion of csv data:-

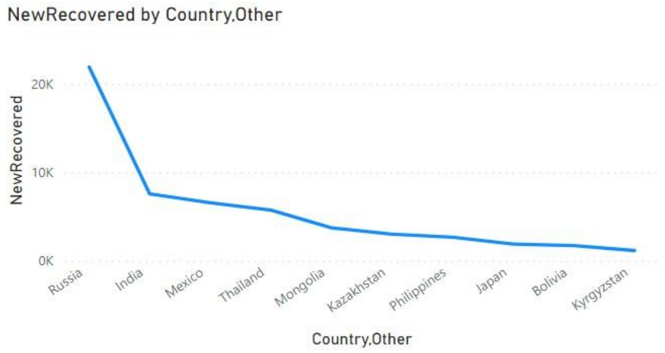


```
    var csvStr = JsonFields.join(",") + "\n";

    playerCtx?.filteredcommentsData?.forEach((element) => {
      let CommentText =
        element.comment_desc === null ? "" : element.comment_desc;
      var pattern = /\B#[A-Za-z0-9_./#&+-]+/gi;
      let mentionedUsers = CommentText.match(pattern);
      if (mentionedUsers != null) {
        mentionedUsers.forEach((element) => {
          try {
            if (
              element.startsWith("#") &&
              /^[0-9a-fA-F]{8}-[0-9a-fA-F]{4}-[0-9a-fA-F]{4}-[0-9a-fA-F]{4}-[0-9a-fA-F]{12}$/.test(
                element.replace("#", "")
              )
            ) {
              // test if is guid or not
              let UserId = element.replace("#", "");
              let UserName = _.findWhere(playerCtx.ProjectUserList, {
                user_id: element.replace("#", ""),
              }).user_name;
              CommentText = CommentText.replace(element, `@${UserName}`);
            }
          } catch (e) {}
        });
      }
```

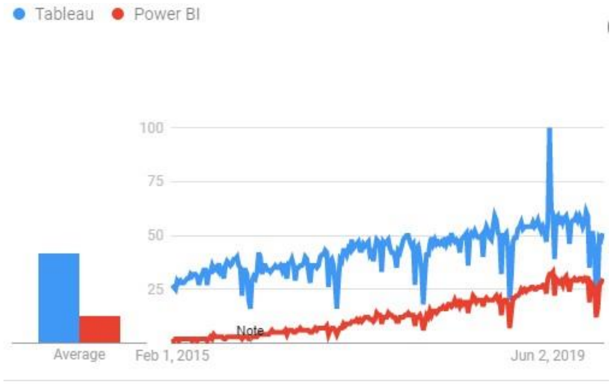
**Figure:9** Code for Statistical Result Of Covid Data



Graph:1 Country Wise Cases



Graph-2 New recovered cases



Graph: 3 Connections to various data sources

- Develop new columns of data
- Set-up, arrange and remove data
- Grouping data
- Transporting data
- Pivoting data
- Re-shaping data
- Manipulation of data by using formulae
- Publish data

## 9. Comparison of Power-BI and Tableau

| SN | Power-BI | Tableau |
|---|---|---|
| 1 | OS Support only Microsoft Windows | OS Support more for Mac, Microsoft Windows etc. |
| 2 | Low cost | More expensive |

|  |  |  |
|---|---|---|
| 3 | Components available are -<br>▪ Power-BI Service<br>▪ Power-BI Mobile App<br>▪ Power-BI Gateway<br>▪ Power-BI Reports<br>▪ Power-BI Desktop | Components available are -<br>▪ Tableau Visualize<br>▪ Tableau Server<br>▪ Tableau Reader<br>▪ Tableau Server<br>▪ Tableau Public |
| **4** | Query Editor Support | No any such query editor |
| **5** | Easy to embed the report in web portal | Difficult to embed the report in web portal |
| **6** | Support limited amount of data | Support big data |
| **7** | Easy to use | Required technical knowledge to use |
| **8** | Faster and good performance when small data | Faster and good performance even when big data |

Table 2 Comparative study

Power-BI and Tableau both are capable of creating visual dashboards that can analyse data visualizations from numerous reports. Power-BI and Tableau has been popular for its visually attractive dashboards that are simple to made using drag and drop options. With the help of Power-BI's visible metrics or tiles which connect directly to reports and datasets, creating dashboards is equally easy and simple. End-users can quickly pin tiles from a specific report to a dashboard to get required information for decision making.

## 10. Results and Discussion

a. **Data Mining:** Business Intelligence(BI) analytics tools are extremely well-suited for influential data mining. Data mining is the process of looking for patterns in data in order to recognize trends and present insights. It can be fragmented into five steps: gathering, warehousing and storing, grouping, analysis and visualisation etc. Some BI environments and platforms can achieve all of these steps as per business needs, although others need support from business analytics tools, Big Data analytics applications and data warehousing environments and platforms.

b. **Visualize Significant Information :**The core assistances of BI software tool is that it deliver effective and efficient data visualization competences, enabling users to generate instinctive data visuals that are simple to understand and easy to interpret

c. **Valuable Insights:** There are numerous types of MIS reports you can generate with the help of BI tools. All BI tools and platforms will deliver pre-formatted reporting competences that gather data for general business and process KPIs.

## 11. Performance Management

Business Intelligence (BI) assists to find, accomplish and implement performance objectives of a business. With BI tools business could provide database objectives like sales goals, quality goals, risk control goals, target delivery time, and then track progress on a regular basis as and when needed. This tracking technique is called as performance management, and it provides the most effective, efficient and easily implementation of management strategy to achieve high business productivity.

## 12. Conclusions

Business Intelligent tools play a vital role in taking business sentiments. As far as concern with Power-BI and Tableau, both Power-BI and Tableau has its own features strength and weakness for data capturing, storing, processing, analysing and visualizing etc. It all depends upon the business needs, conditions and strategigies.If new user wants to produce dashboard so they can analyse and visualize the data as per their requirements. The analysis done through this study clearly indicates if data contain small dataset then we can use Power-BI and if dataset is large then we should use Tableau.

## 13. References:

1. Jignesh Shah," 8 Major benefits of Microsoft Power BI you must know",Saviant,July,2018 ,https://www.saviantconsulting.com/blog/8-major-benefits-of-microsoft- power-BI.aspx

2. Michael Diamond , Angela Mattia "Data visualization: an exploratory study into the software tools used by businesses", Journal of Instructional Pedagogies,Vol.18,2018 ,https://files.eric.ed.gov/fulltext/EJ1151731.pdf

3. Vijay Krishnan,S Bharanidharan,G Krishnamoorthy, 11th International CALIBER-2017 ,"Research Data Analysis with Power BI" ,https://ir.inflibnet.ac.in/bitstream/1944/2116/1/24.pdf

4. Luís Miguel Ramos Silva ,"Industrial Reporting Using Power BI", Uporto,FEUP Facudade Engenharia,Universidade Do Porto,July,2021 ,https://repositorioaberto.up.pt/bitstream/10216/135436/2/486940.pdf

5. InterviewBit,"Power BI vs Tableau: Full Comparison",July,2022 ,https://www.interviewbit.com/blog/power-bi-vs-tableau/

6. Podeschi, R.J.: Experiential learning using QlikView business intelligence software. In: 2014 Proceedings of the Information Systems Educators Conference Baltimore, Maryland, USA.ISSN: 2167-1435

7. Shukla, A., Dhir, S.: Tools for data visualization in business intelligence: case study using the tool Qlikview. In: Information Systems Design and Intelligent Applications, vol. 434, pp. 319–326. AIC, Feb. 2016
8. MicroStraregy Inc.: MicroStrategy 9: Basic Reporting Guide. MicroStrategy
9. Negash, S. “Business Intelligence.” Communications of the Association for Information Systems 13:177-195, 2004.
10. "Announcing Power BI General Availability Coming July 24th." PowerBI. N.p., 10 July 2015.
11. Eigner, W. "Current Work Practice and Users' Perspectives on Visualization and Interactivity in Business Intelligence." 2013 17th International Conference on Information Visualisation. 2013.
12. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H. "Big Data: The next Frontier for Innovation, Competition, and Productivity." Big Data: The next Frontier for Innovation, Competition, and Productivity. May 2011.
13. Gounder, M.S., Iyer, V.V., Al Mazyad, A.: A survey on business intelligence tools for university dashboards development. In: 3rd MEC International Conference on Big Data and Smart City (2016)
14. Golfarelli, M.: Open source BI platforms: functional and architectural comparison. In: International Conference on Data Warehousing and Knowledge Discovery. LCNS, vol. 5691 (2009)
15. Watson, H.J., Wixom, B.H.: The current state of business intelligence. IEEE, vol. 40, issue 9, Sept. 2007 504 C. S. Reddy et al.
16. Power BI- Guided Learning material,(https://powerbi.microsoft.com/en-us/guided-learning/). 17. Power BI- Webinars (https://powerbi.microsoft.com/enus/documentation/powerbi-webinars/)
18. Spago BI, a fully open source business intelligence tool,(http://www.spagobi.org/).
19. Presentations of SpagoBI tools as slides ,(http://www.slideshare.net/spagoworld/ ).
20. Mobile development platform for SpagoBI ,(http://www.spagobi.org/homepage/product/mobile/).

Sinhgad Institute of Management PUNE 411 041.

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

Dr. D. Y. Patil Pratishthan's

Dr. D. Y. Patil Institute of Management Studies, Akurdi, Pune.

&

Dr. D. Y. Patil Institute of MCA and Management, Akurdi, Pune.

# 6TH ASIA -AFRICA DEVELOPMENT CONFERENCE, SUMMIT & AWARDS - 2022

# Certificate

*This is to Certify that*

*Dipali Prakash Patil*

*of Savitribai Phule Pune University*

*has participated & presented a research paper entitled*

*Detecting Skin Diseases Using Image Processing and Machine Learning algorithm: Review*

*at 6th Asia-Africa Development Conference, Summit and Awards 2022 on 1st and 2nd December 2022, jointly organized by Asia-Africa Development Council managed by the Council For Sustainable Peace & Development and Dr. D. Y. Patil Institute of Management Studies (DYPIMS), Pune.*

Prof. (Dr.) Kuldip Charak<br>
Director, DYPIMS

Prof. (Dr.) Ripu R. Sinha<br>
Director, CSPD

* 1 & 2, DECEMBER - 2022. *<br>
THE ORCHID HOTEL<br>
BALEWADI, PUNE, INDIA.

Dr. D. Y. Patil Pratishthan ᵗs

Dr. D. Y. Pat-il Institute of Management Studies, Akurdi, Pune.

Dr. D. Y. Patil Institute of MCA and Management, Akurdi, Pune.

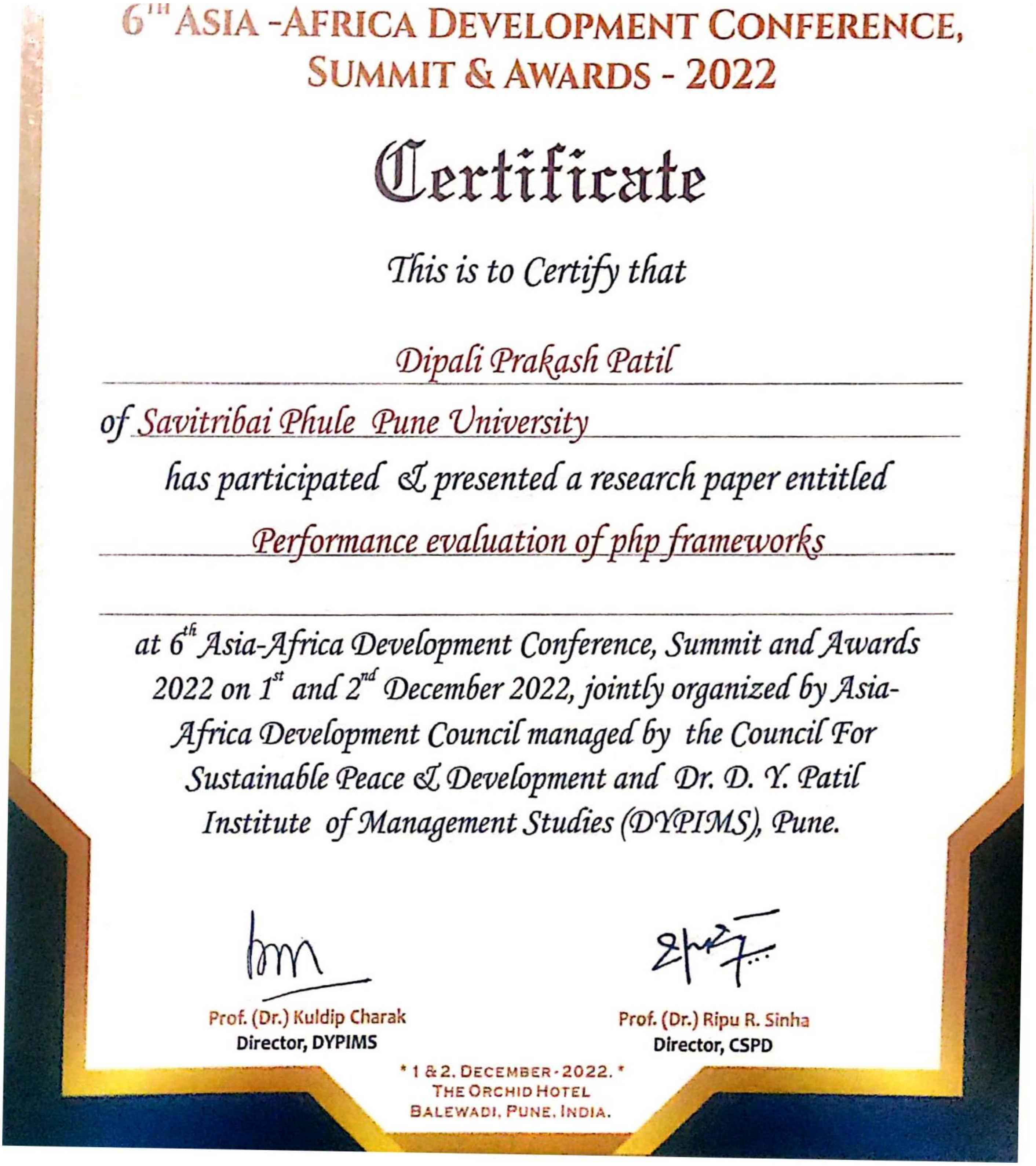
OYPIMS

110 N e

DYP

D.V, PATtL

r..srrrt.'JE0F



# 6TH ASIA -AFRICA DEVELOPMENT CONFERENCE, SUMMIT & AWARDS - 2022

# Certificate

*This is to Certify that*

*Dipali Prakash Patil*

*of Savitribai Phule Pune University*

*has participated & presented a research paper entitled*

*Performance evaluation of php frameworks*

*at 6th Asia-Africa Development Conference, Summit and Awards 2022 on 1st and 2nd December 2022, jointly organized by Asia-Africa Development Council managed by the Council For Sustainable Peace & Development and Dr. D. Y. Patil Institute of Management Studies (DYPIMS), Pune.*

Prof. (Dr.) Kuldip Charak
Director, DYPIMS

Prof. (Dr.) Ripu R. Sinha
Director, CSPD

* 1 & 2. DECEMBER - 2022. *
THE ORCHID HOTEL
BALEWADI, PUNE, INDIA.

Scanned with CamScanner

Sinhgad Institute of Management
PUNE
411 041.

Singh

Dr.Chandrani Singh ,Director –MCA,SIOM

Dr. D. Y. Patil Pratishthan's

Dr. D. Y. Patil Institute of Management Studies, Akurdi, Pune.

&

Dr. D. Y. Patil Institute of MCA and Management, Akurdi, Pune.

DYP
DR. D. Y. PATIL INSTITUTE OF ENGINEERING, MANAGEMENT AND RESEARCH

# 6TH ASIA -AFRICA DEVELOPMENT CONFERENCE, SUMMIT & AWARDS - 2022

## Certificate

*This is to Certify that*

*Dipali Prakash Patil*

*of Savitribai Phule Pune University*

*has participated & presented a research paper entitled*

*Detecting Skin Diseases Using Image Processing and Machine Learning algorithm: Review*

*at 6th Asia-Africa Development Conference, Summit and Awards 2022 on 1st and 2nd December 2022, jointly organized by Asia-Africa Development Council managed by the Council For Sustainable Peace & Development and Dr. D. Y. Patil Institute of Management Studies (DYPIMS), Pune.*

Prof. (Dr.) Kuldip Charak
Director, DYPIMS

Prof. (Dr.) Ripu R. Sinha
Director, CSPD

* 1 & 2, DECEMBER - 2022. *
THE ORCHID HOTEL
BALEWADI, PUNE, INDIA.

Dr. D. Y. Patil Pratishthan $^{t}$s

Dr. D. Y. Pat-il Institute of Management Studies, Akurdi, Pune.

Dr. D. Y. Patil Institute of MCA and Management, Akurdi, Pune.

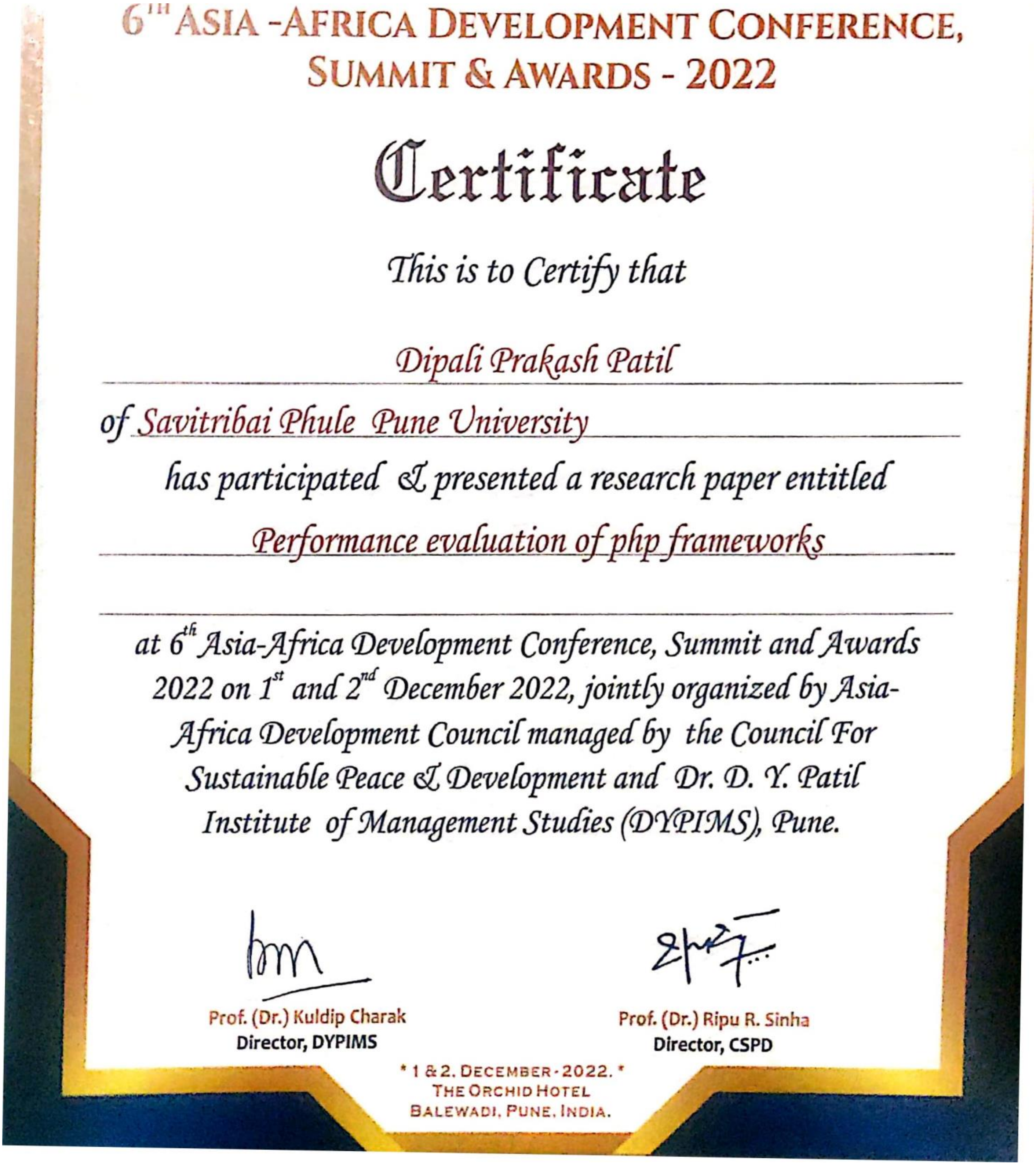
OYPIMS

110 N e

DYP

D.V, PATtL

r..srrrt.'JE0F



## 6TH ASIA -AFRICA DEVELOPMENT CONFERENCE, SUMMIT & AWARDS - 2022

# Certificate

*This is to Certify that*

*Dipali Prakash Patil*

*of Savitribai Phule Pune University*

*has participated & presented a research paper entitled*

*Performance evaluation of php frameworks*

*at 6th Asia-Africa Development Conference, Summit and Awards 2022 on 1st and 2nd December 2022, jointly organized by Asia-Africa Development Council managed by the Council For Sustainable Peace & Development and Dr. D. Y. Patil Institute of Management Studies (DYPIMS), Pune.*

Prof. (Dr.) Kuldip Charak
Director, DYPIMS

Prof. (Dr.) Ripu R. Sinha
Director, CSPD

* 1 & 2. DECEMBER - 2022. *
THE ORCHID HOTEL
BALEWADI, PUNE, INDIA.

Scanned with CamScanner

Sinhgad Institute of Management
PUNE
411 041.

Dr.Chandrani Singh ,Director –MCA,SIOM

# “Automatic Gas Booking System using IoT”

Prof. Shobha Sachendra Mishra[1], Dr.Ramesh D Jadhav[2] , Dr. Chandrani Singh[3]
[1]shobhamishra@sinhgad.edu , [2]ramesh.jadhav@sinhgad.edu , [3]directormca_siom@sinhgad.edu
*[1,2,3]Department of Computer Application,, Sinhgad Institute of Management,*
*Vadgaon Budruk(BK), Pune, India*

***Abstract –***Liquefied Petroleum Gas (LPG) is the major cooking fuel in India and other countries. LPG cook stoves are comparatively portable, clean, and highly efficient and requires less maintenance. However, these LPG cook stoves are 60-65% efficient and pollution from the cook stoves which are beyond the regulation standard of World Health Organization. The project's need is to save time when booking gas. Your order may not be logged, or your phone may not connect when you contact your gas dealer. This is all a waste of human effort. If you don't notice that you're out of gas, you would order it in black for an additional fee. In this project, you will continuously monitor the gas level and receive an alert when the supply of gas is depleted. With this article, we'll look at a microcontroller-based system that uses a weight sensor and a load cell to calculate the weight of gas in a cylinder.This block is linked to the alarm block and provides an audible or visual signal when the LPG cylinder is empty. At a reasonable price, the sensor provides adequate sensitivity and a quick response time. When a gas blockage is detected, a message is sent to family member to use the cellular network known as GSM in the normal way. It also has the ability to calculate the weight of an LPG cylinder and display that value on an LCD. Gas under 10 kg per cylinder is automatically reserved by texting the supplier. In addition, if the cylinder weighs less than 0.5 kg, a notification will be sent to notify family members to refill the cylinder.

***Keywords –***LPG,GSM,MQ6gas sensor,Weight sensor.

## I. INTRODUCTION

India is a large and developing country, and everyone is busy with their daily activities as most men and women work to support their families. Cooking is one of the tasks we have to do every day, using LPG. India's current population is 1,407,236,169, of which 92% use LPG for cooking. Political awareness of the need for sustainable cuisine has increased political efforts in many developing countries. India has a long history of subsidizing the cost of filling LPG cylinders for home cooking. Women who primarily manage the biomass production chain will benefit from the sustainable use of LPG, which provides more flexibility and ease of use and reduces indoor air pollution. As a major determinant of household spending, women will use high-quality energy sources that save time, provide better health and more free time.

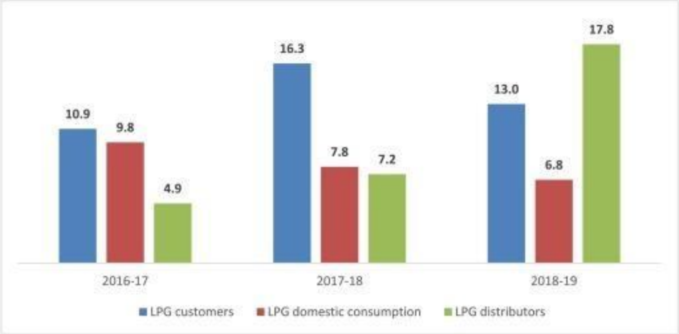


Figure 1. Annual growth rate (%) of domestic LPG sector

### 1.1 Topic

The project's need is to save time when booking gas. Your request may not be recorded, or your call may not connect when you call your gas dealer. This is all a waste of human time. If you don't realize that you're out of gas, you can order it in black for an additional fee. In this project, you will continuously check the gas level and receive an alert when the supply of gas is depleted.

### 1.2 Theory

LPG, initially manufactured in 1910 with the assistance of Dr. Walter Snelling, is a blend of commercial butane and commercial propane that contains both saturated and unsaturated hydrocarbons. LPG is rising rapidly day by day due to its flexible character, which is widely used in numerous fields alongside domestic gas and commercial fuel. Text messages are now used to reserve LPG tanks. The IVRS customer approach has been adopted by oil corporations as a customer-friendly service. As a result, an efficient device for weighing and showing the amount of LPG is required.

## II. STATISTICAL ANALYSIS

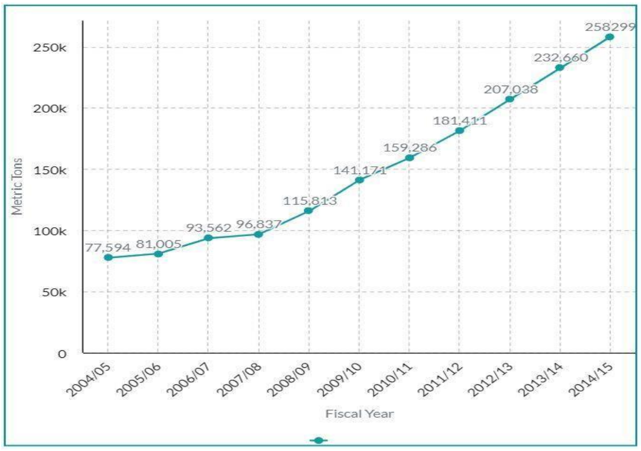


Figure 2. Historical LPG use trends

Firewood was a key source of cooking fuel in 2014-2015, according to the Annual Consumer Survey Report, feeding more than 59.3 percent of all households. LPG consumption has skyrocketed in recent years, with 25.8 percent of homes currently using it. It is the most often utilized cooking fuel in metropolitan areas, accounting for around 58.5 percent.

LPG is now the country's second-most popular cooking fuel. Approximately 77,594 tons of LPG were imported from India in fiscal year 2004-2005. Over a ten-year period, demand for LPG increased rapidly, with imports reaching 258,299 tons in fiscal year 2014-2015.

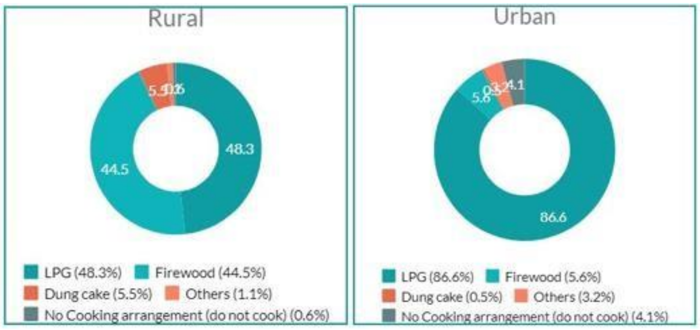


Figure 3.Percentage of homes using various forms of fuel (%)

Intriguingly, compared to 5.6% of dwellings in cities in 2018, 44.5% of homes in villages continued to used firewood, crop refuse, and chip for cooking.According to the NSO’s data, just 48.3% of rural homes used LPG, compared to86.6 percentage points in metropolitan regions.

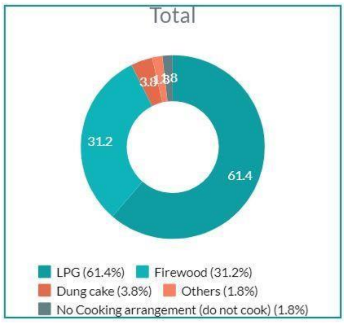


Figure 4. Total Percentage of homes using various forms of fuel (%)

As per NSO’s latest 2018 data, the overall consumption of LPG in country is 61.4% and firewood still accounts for 31.2%. The deep insights also shows that there is a big contrast between rural and urban where majority of our population lives in rural areas.

## III. SYSTEM OVERVIEW

It is made up of the components depicted. It has an ATMEGA 16A microcontroller, a weight sensor (Load Mobile-L6D), a gas sensor, a GSM module (SIMCOM300), and a display.

### 2.1 Micro-Controller

A prospective and fast-running controller is planned to detect LPG fuel consumption and the output of the stage (weight) sensor on a continuous basis. In this sense, the device must be capable of storing some data that can be processed. The microcontroller is at the heart of the device, as defined in Definition 1. Features such as 16KB of internal memory, which generates a clean garage of all code on the microcontroller itself, and a control loop execution charge of 1MIPS per MHz, which is a more desirable average overall device performance. The LCD module labelled as port b of the ATMega16A is used to display predetermined messages in four-part mode.

The output of the load cell module is coupled to a pin on a port that is used to continually sense the fuel level via the transmission circuit.

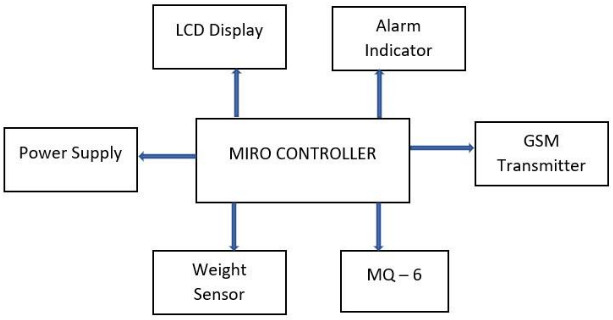


Figure 5: Microcontroller Block Diagram

### 2.2 Weight Sensor Module

To order a cylinder from a wholesaler, the volume of gas in the cylinder must be known. As a result, it is required to continuously determine the amount of gas in the cylinder. Because it provides the requisite weighing capability for domestic cylinders, the weight sensor module is used in combination with the load cell for assessment reasons. A L6D weight sensor module is inserted in the system. The load cell yield drives a transfer circuit that generates two logical pulses (for <= 10 kg and <= 0.5 kg), which are correspondingly connected to micro-controller port pins in order to detect the level of the gas.

### 2.3 GSM Module

The weight sensor indicates the amount of gas in the cylinder, and the microprocessor makes the appropriate action. The cylinder status must be reported to the device's rightful owner or home partner through the LCD display and GSM module. The GSM module is used for transmitting and receiving messages based on AT commands. Connect the modem to the microcontroller using this instruction to control it. In this case, SIMCOM 300 is used. It is powered by a 12-volt adaptor. Sending, in particular, consumes less memory.

### 2.4 LCD Display

The device performs control and management functions. In addition, a display should be built into the system that shows a number of texts, including gas weight, the cylinder's spare number when filling, and the display behavior. Messages are displayed on the microcontroller's 16X2 character LCD display, which operates in 4-bit mode and is powered by +5 volts. By combining the ATMegaL6D with simple computer code, it is possible to make the system user-friendly and easy to use.

### 2.5 MQ-6 Sensors

LPG is made up of propane, propylene, butane, and butylenes. A reliable, rapid, and robust gas sensor which can only identify LPG particles and it is less reactive to other gases is required (such as cooking exhaust gas, cigarettes, etc.) Tin (IV) Oxide (SNO2) is the MQ-6 gas sensor's delicate substance; it has low conductivity in pure air and its impact ability improves with gas concentration; also, it keeps a strategic distance from gases such as cooking vapour. In accordance with the vaporous condition, it demands a low and secure voltage of 0-5 volts. This sensor measures gas levels and, if they surpass a predetermined limit, activates the alarm, exhaust fan, and microcontroller.

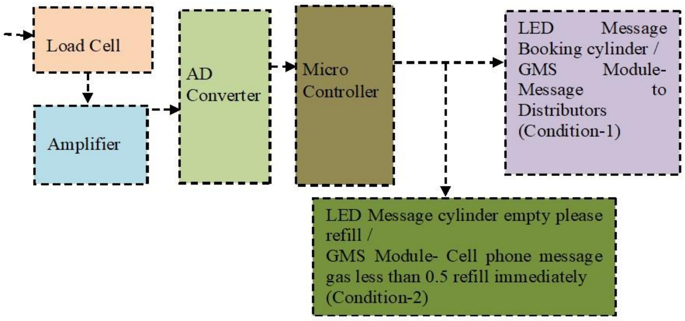

Load Cell
Amplifier
AD Converter
Micro Controller
LED Message Booking cylinder / GMS Module- Message to Distributors (Condition-1)
LED Message cylinder empty please refill / GMS Module- Cell phone message gas less than 0.5 refill immediately (Condition-2)

Figure 6. Block Diagram of Gas booking System using IoT

## IV. SYSTEM OPERATION

The following is an explanation of the operation's procedure from the glide chart for automated gas reservations:

The automatic gas booking device's L6D routinely measures the cylinder's weight and shows it on seven segment displays. A microcontroller port pin receives a logic pulse when the gas weight is less than or equal to 5 kg.

The microcontroller will send a reserving message to the distributor as soon as this pin goes high. Additionally, the message may also be read as "RESERVING CYLINDER" on the liquid crystal display at the same time. Any additional logic pulse is supplied to another port on the microcontroller when the gas weight is less than 0.5 kg.

When this port pin goes too high, the microcontroller will send the message "fuel last just 0.5 Kg. Through a GSM module, members are instructed to "immediately replace the cylinder" and are given the option of online payment or cash on delivery. The use of an alert and a MQ-6 sensor is also included in this project for fuel leak detection. With the help of a manual reset transfer, we can reset. Additionally, an interrupt of common sense (+5 v) is provided to the ATMega16 A microcontroller's int0 pin. "EMERGENCY ALERT: LPG gas leaking found in the residence," the microcontroller transmits.

## V. CONCLUSION

This study designs and implements a cost-effective system for monitoring gas levels and booking it immediately when the gas is nearing completion The proposed system satisfies the approach for efficiently reserving gas. Because of its ability to measure the weight of the LPG cylinder and show the value on an LCD, this device can be used in factories and other settings to assess how much gas is still in a cylinder. When compared to the cost of existing fuel detectors on the market, the costs of building this system is far less and, overall, significantly cheaper.

## VI. REFERENCES

[1] ATMega-16 Datasheet;www.atmel.com

[2] MQ-6 Technical Data; https Sensors/Biometric/MQ-6. pdf

[3]L6Dweightsensorspecifications;http://www.zemic.com.cn/e/showproduction.asp?num=33

[4] P Bala Gopal, K Hari Kishore, R.R Kalyan Venkatesh, P HarinathMandalapu "An FPGA Implementation of On Chip UART Testing with BIST Techniques", International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 10, Number 14 , pp. 34047-34051, August 2015.

[5]Meka Bharadwaj, Hari Kishore "Enhanced Launch-Off-Capture Testing Using BIST Designs" Journal of Engineering and Applied Sciences, ISSN No: 1816-949X, Vol No.12, Issue No.3, page: 636-643, April 2017.

[6] M.S. Kasar, Rupali Dhaygude, Snehal Godse and Sneha Gurgule, "Automatic LPG Gas Booking and Detection System", International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, vol. 5, no. 3, pp. 1250-1253, March 2016, ISSN 2278-8875.

[7] L. K. S. Rohan Chandra Pandey, Manish Verma, "Internet of things (IOT) based gas leakage monitoring and alerting system with MQ-2 sensor," International Journal of Engineering Development and Research, Vol. 5, 2017

[8] Ravindra R. Hiwase, Priya K. Kewate, Sushmita P. Tajane, JitendraWaghmare "Automatic LPG Cylinder Booking and Leakage Detection using Arduino UNO" IJESC.

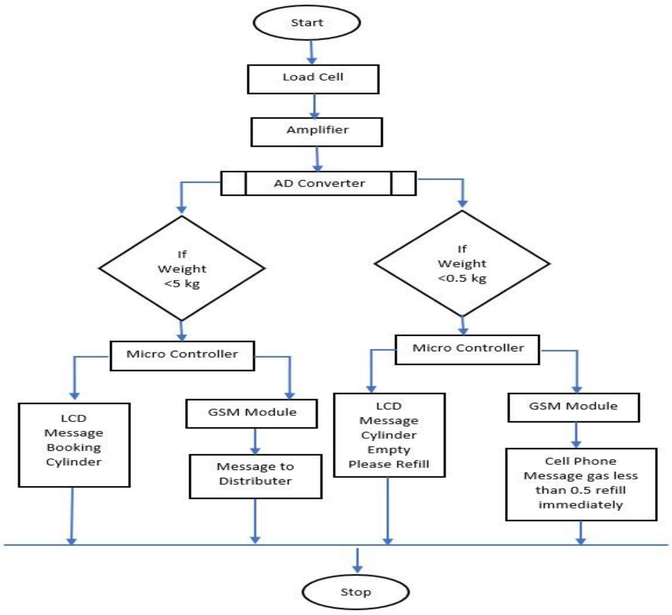


Figure 7. Automatic Gas Booking Flow Chart

Sinhgad Institute of Management PUNE 411 041. *

Dr.Chandrani Singh ,Director –MCA,SIOM

# Predictive analysis the study of different characteristics of Palmer penguins using R-programming

Dr. Ramesh D Jadhav[1,] Miss. Vaishali Bhujbal[2]

Sinhgad Institute of Management, Pune-41, (India)

## Abstract:

In 50th Century the statistics of chinstrap penguins in Antarctica have down up to 77%. Environment transformation is melting ice was adverse impact on krill which the penguins, jumbos and seals complete eat. The environment extremity was become risky on Antarctica's chinstrap penguins. Transformers on a Greenpeace way to Antarctica set up that the penguins' facts were dropping with one colony falling up to 77% in approximately in 50th century. This was predominately surprising due to the chinstrap penguin has been measured a species of less concern through the 'International Union for Conservation of Nature'(IUCN) as per the CN.

Penguin associations in almost corridor of the Antarctic have dropped through more than 75 percent completed the once partial era, mostly as an outcome of environment transformation, experimenters around. Researchers exposed that associations of chinstrap penguins furthermore recognized such as ringed or faced penguins have fallen histrionically subsequently they were previous plotted approximately 50 centuries back. Each collection plotted on Elephant Island, a significant penguin niche northeast of the Antarctic Peninsula, endured a people fall, according to self-determining experimenters who combined a Greenpeace passage towards the province.

On the previous check cutting-edge 1971, there be situated, 122,550 dyads of penguins crossways completely associations on Elephant Island. On the other hand, the latest total discovered just,786 dyads a drip of nearly 60 percent. The dimension of the people transformation diverse since group to group arranged Elephant Island. The major disaster 77 percent was documented by a group recognized such as Chinstrap Camp. Environment transformation has directed to reduced ocean snow and heater abysses, which takes destined lower krill, the foremost element of the penguin's food. In this paper highlighted global warming impact has been on Antarctica of the chinstrap penguins and its different characteristics analysis.

**Keyword**: Antarctica, Penguin, island, Predictive analysis, R-programming

## 1. Introduction:

Penguins be present a cluster of submarine earth-bound catcalls. They living nearly simply cutting-edge the Southern Hemisphere lone single classes, the Galápagos penguin, is set up northern of the Equator. largely acclimated for natural life in the seawater, penguins take athwart shadowed black also white plumage and members used for swimming. utmost penguins diet continuously krill, catch fish, squid and further classifications of ocean lifecycle which they hook using their fliers and gulp it complete though swimming. A penguin takes a nasty lingo & important lips to grip greasy prey.

Penguin Species: There are 17 penguin species on the earth, but the eight most iconic live in Antarctica, its near islets, and the sub-Antarctic archipelagos of South Georgia and the Falklands. For now, we're shifting our focus to substantially 3 type of penguins ie Adelie, Gentoo & Chinstrap Penguins The 'Adélie penguin' (Pygoscelis-adeliae) is a classes of penguin collective lengthways the whole seacoast of the 'Antarctic' mainland, which is the individual place anywhere it's set up. This one the widest penguin classes, & along using the king penguin, stands the greatest southerly scattered of entirely penguins. It's named afterwards 'Adélie Land', in turn so-called for 'Adèle Dumontd' Urville, who remained wedded to French discoverer 'Jules Dumontd' Urville, who initial exposed this penguin in 1840. 'Adélie' penguins gain their diet thru mutually predation and rustling, using a food of substantially krill & fish.

The chinstrap penguin ('Pygoscelis-antarcticus') is a types of penguin that populates a variation of islets & props in the 'Southern-Pacific' & the 'Antarctic' abysses. This one label stalks starting the thin black crowd less than its skull, which kinds it appears as if the situation were wearisome a black hat, building it informal to categorize. Additional mutual names exclude ringed penguin, unshaven penguin, and gravestone cracker-penguin, due to its lurid, strict song.

The Gentoo penguin (Pygoscelis-papua) is a penguin types present the rubric Pygoscelis, greatest nearly associated to the 'Adélie-penguin' (P. 'adeliae') & the chinstrap-penguin (P. 'antarcticus'). The foremost systematic explanation remained complete in 1781 by 'Johann-Reinhold-Forster' thru a category position in the 'Falkland' islets. The classes sound in a diversity of behaviors, but the greatest regularly received is a lurid announcing, which the raspberry secretes thru its skull thrown back. [1], [7], [8], [9] & [10]. See in the figure types of penguins.

CHINSTRAP!
GENTOO!
ADĒLIE!

Figure no.1: Chinstrap, Gentoo and adelie

2. **Problem statement**: Environment transformation directed to the drop in Chinstrap Penguin Types Thru Over 75% in the Previous 50 Centuries. Antarctic penguin associations in approximately quantities of the Antarctic have dropped by extra than 75% finished the latest 50 centuries.

3. **Objective of Study:** the following some objective are

To identify reason of global warming impact on Antarctica
To identify the chinstrap penguins different characterises
To identify problem over the different kinds of the penguins in Antarctica

4. **Research Design and Methodology:** In this paper we have used exploratory research design and its details given below

### 4.1 Data Collection

The datasets of Palmer Penguins were collected from Kaggle that spanned years from 2007 to 2009.The data was validated and up-to date and the codes were written and executed in RStudio.

### 4.2 Data Exploration:

The Penguin dataset consisted of 345 observations across 7 attributes as listed below:

| Sr.no | Attribute |
|---|---|
| 1 | Species |
| 2 | Island |
| 3 | culmen_length_mm |
| 4 | culmen_depth_mm |
| 5 | flipper_length_mm |
| 6 | body_mass_g |
| 7 | Sex |

Table no.:1: Attribute list

- **Penguins_size.csv**: Streamlined records from unique penguin records sets. Contains variables:

○ species: penguin species (Chinstrap, Adélie, or Gentoo)

○ culmen_length_mm: culmen length (mm)

○ culmen_depth_mm: culmen depth (mm)

○ flipper_length_mm: flipper length (mm)

○ body_mass_g: body mass (g)

○ island: island name (Dream, Torgersen, or Biscoe) in the Palmer Archipelago (Antarctica)

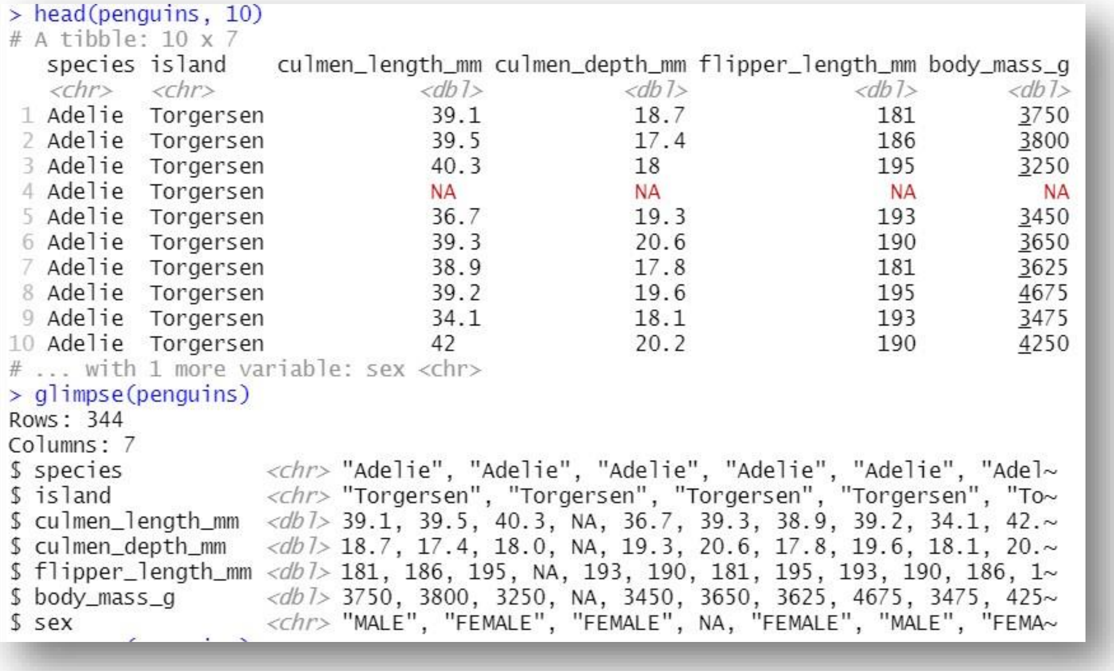
○ sex: penguin sex

### 4.3 Data Analysis:

In order to have a robust data analysis and visualization in representing the results, we installed some packages by using the following command and then running libraries as shown below:

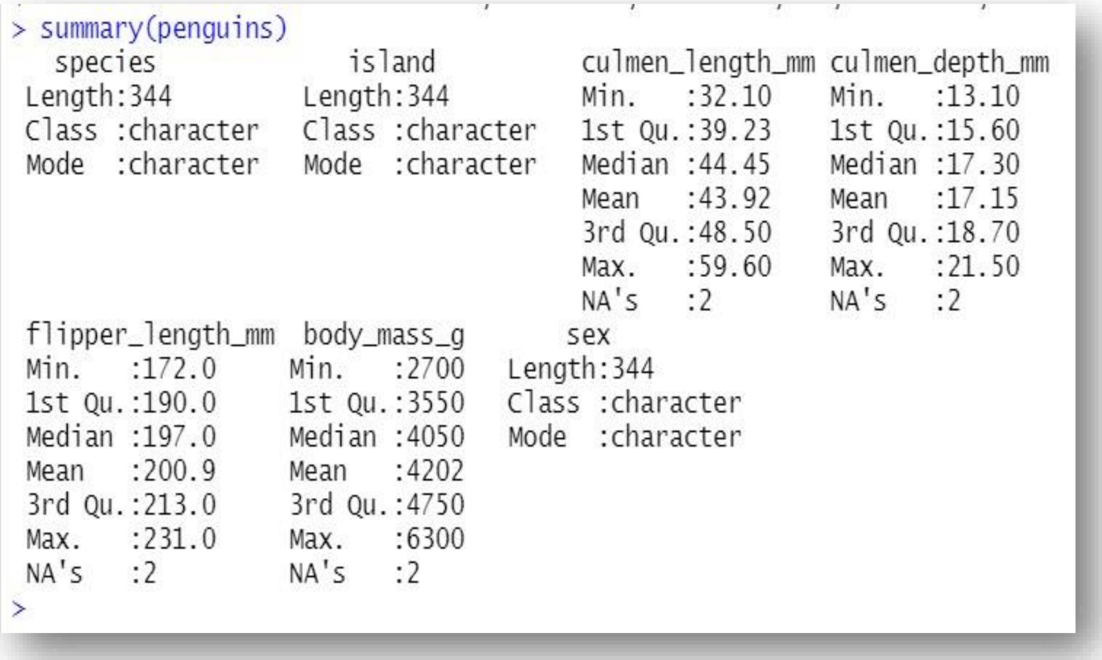
```
○   install.packages('tidyverse')
○   install.packages(‘palmerpenguins’)
○   install.packages(‘tinytex’)
○   install.packages(‘dplyr’)
○   library(tidyverse)
○   library(palmerpenguins)
○   library(tinytex)
○   library(dplyr)
○   library(ggplot2)
○   theme_set(theme_minimal())
```

The requisite lines of code then helped create a data frame by reading and binding the csv files which was then inspected with the head(), glimpse(), & summary() functions.



```
> head(penguins, 10)
# A tibble: 10 x 7
   species island    culmen_length_mm culmen_depth_mm flipper_length_mm body_mass_g
   <chr>   <chr>                <dbl>           <dbl>             <dbl>       <dbl>
 1 Adelie  Torgersen             39.1            18.7               181        3750
 2 Adelie  Torgersen             39.5            17.4               186        3800
 3 Adelie  Torgersen             40.3            18                 195        3250
 4 Adelie  Torgersen             NA              NA                  NA          NA
 5 Adelie  Torgersen             36.7            19.3               193        3450
 6 Adelie  Torgersen             39.3            20.6               190        3650
 7 Adelie  Torgersen             38.9            17.8               181        3625
 8 Adelie  Torgersen             39.2            19.6               195        4675
 9 Adelie  Torgersen             34.1            18.1               193        3475
10 Adelie  Torgersen             42              20.2               190        4250
# ... with 1 more variable: sex <chr>
> glimpse(penguins)
Rows: 344
Columns: 7
$ species           <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie", "Adel~
$ island            <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen", "To~
$ culmen_length_mm  <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.~
$ culmen_depth_mm   <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.~
$ flipper_length_mm <dbl> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 1~
$ body_mass_g       <dbl> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 425~
$ sex               <chr> "MALE", "FEMALE", "FEMALE", NA, "FEMALE", "MALE", "FEMA~
```

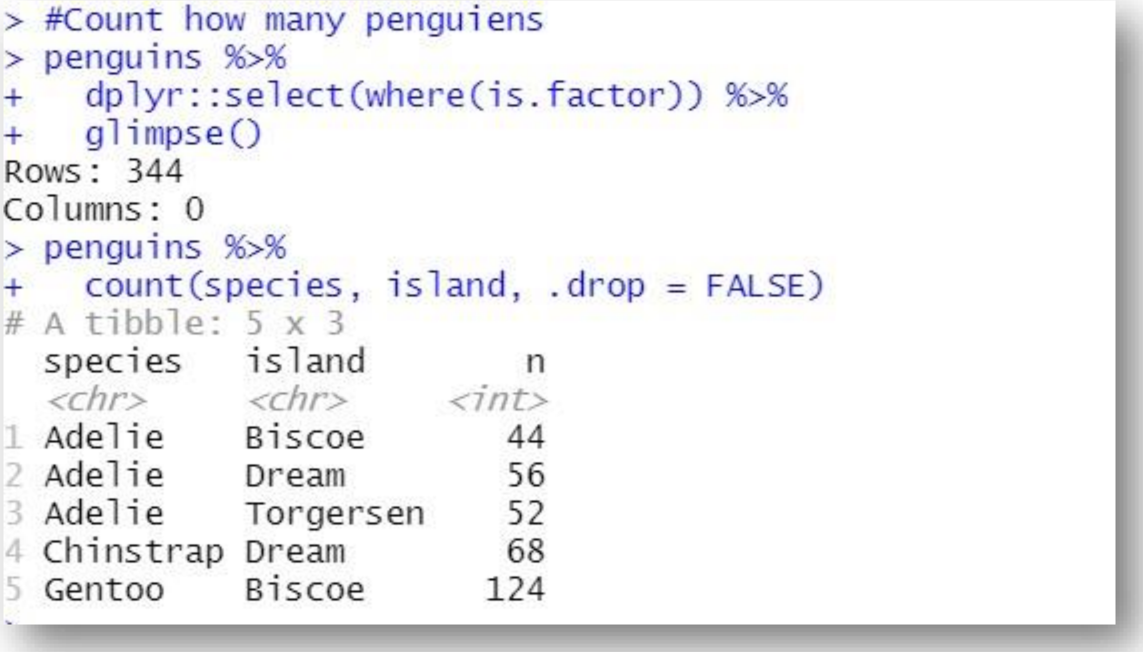
Table no. 2: Penguins-species length, flipper & body-mass.



```
> summary(penguins)
   species             island          culmen_length_mm culmen_depth_mm
 Length:344         Length:344         Min.   :32.10    Min.   :13.10  
 Class :character   Class :character   1st Qu.:39.23    1st Qu.:15.60  
 Mode  :character   Mode  :character   Median :44.45    Median :17.30  
                                       Mean   :43.92    Mean   :17.15  
                                       3rd Qu.:48.50    3rd Qu.:18.70  
                                       Max.   :59.60    Max.   :21.50  
                                       NA's   :2        NA's   :2      
 flipper_length_mm  body_mass_g       sex           
 Min.   :172.0     Min.   :2700   Length:344        
 1st Qu.:190.0     1st Qu.:3550   Class :character  
 Median :197.0     Median :4050   Mode  :character  
 Mean   :200.9     Mean   :4202                     
 3rd Qu.:213.0     3rd Qu.:4750                     
 Max.   :231.0     Max.   :6300                     
 NA's   :2         NA's   :2                        
>
```

Table no.3: Summary (Penguins)

The code and the output data is shown in table 2 and table 3. On this data set the analysis and visualizations are performed of penguins with different characteristics.



```
> #Count how many penguiens
> penguins %>%
+   dplyr::select(where(is.factor)) %>%
+   glimpse()
Rows: 344
Columns: 0
> penguins %>%
+   count(species, island, .drop = FALSE)
# A tibble: 5 x 3
  species   island        n
  <chr>     <chr>     <int>
1 Adelie    Biscoe       44
2 Adelie    Dream        56
3 Adelie    Torgersen    52
4 Chinstrap Dream        68
5 Gentoo    Biscoe      124
```

Table no. 4: Count of Penguiens types wise

**Analysis**: On this table 4, shows the number and different species of penguins on the three islands.

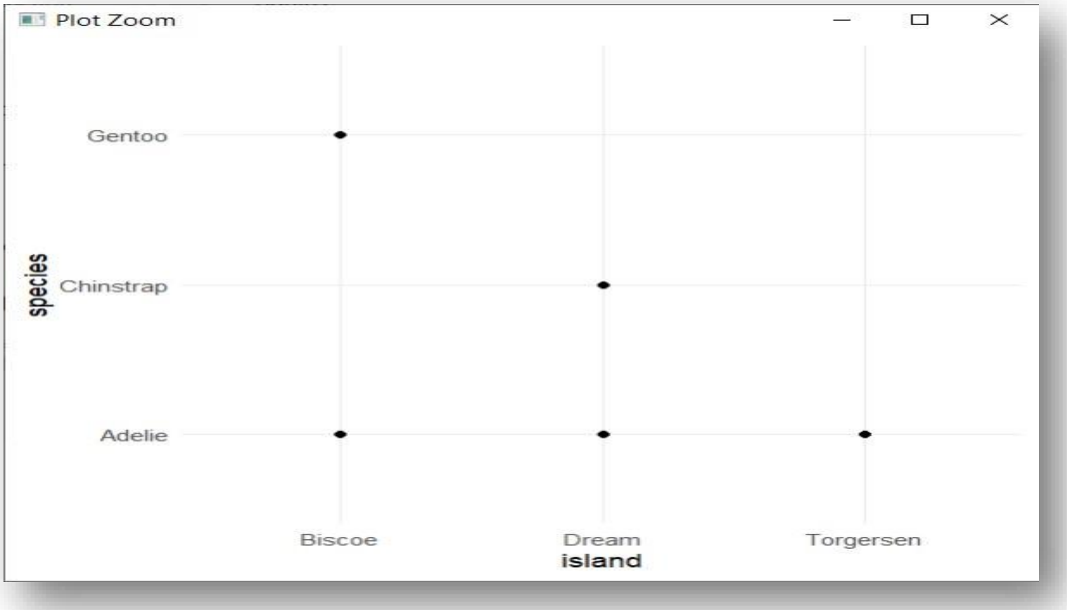


Figure 2 Species VS island

**Analysis**: On this Figure no.2: It shows that species of the penguins on the located island.

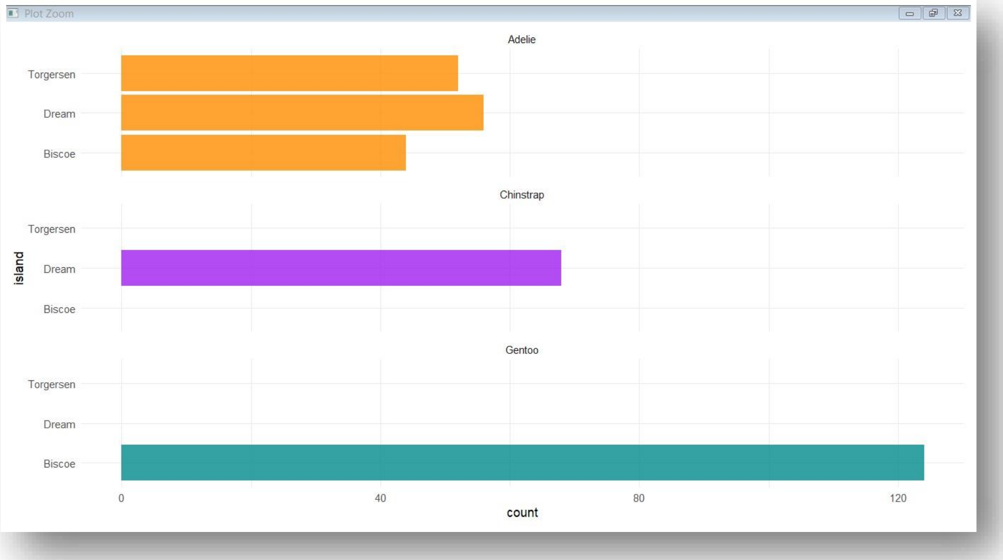
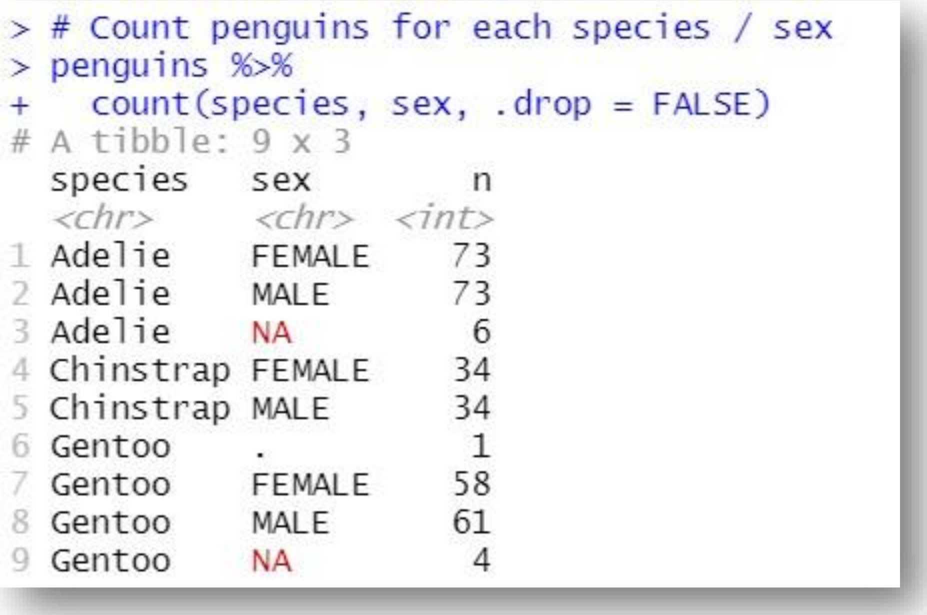


Figure no. 3: Island VS Species wise count

**Analysis**: On this figure no. 3, shows that how many Penguins of which species are on which island.



```
> # Count penguins for each species / sex
> penguins %>%
+   count(species, sex, .drop = FALSE)
# A tibble: 9 x 3
  species   sex        n
  <chr>     <chr>  <int>
1 Adelie    FEMALE    73
2 Adelie    MALE      73
3 Adelie    NA         6
4 Chinstrap FEMALE    34
5 Chinstrap MALE      34
6 Gentoo    .          1
7 Gentoo    FEMALE    58
8 Gentoo    MALE      61
9 Gentoo    NA         4
```

Table no. 5: Sex wise Penguins count

**Analysis**: Table 5, it shows that which species have how many male and female details.

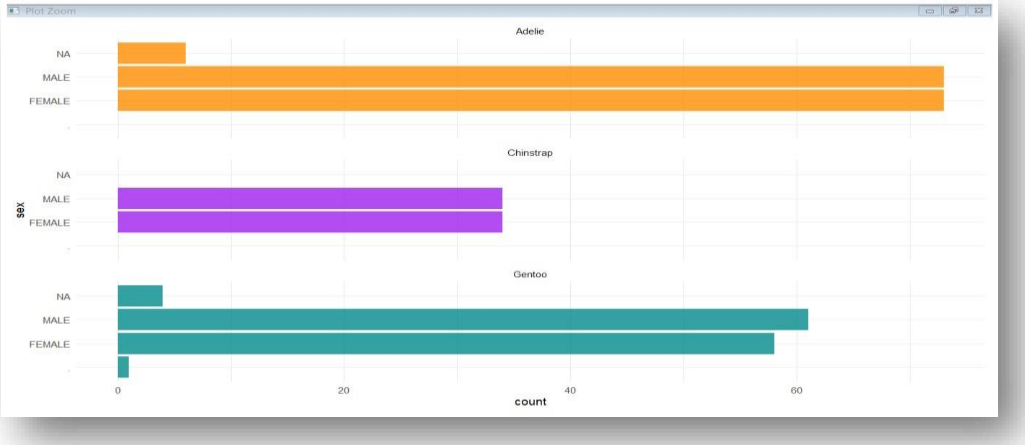
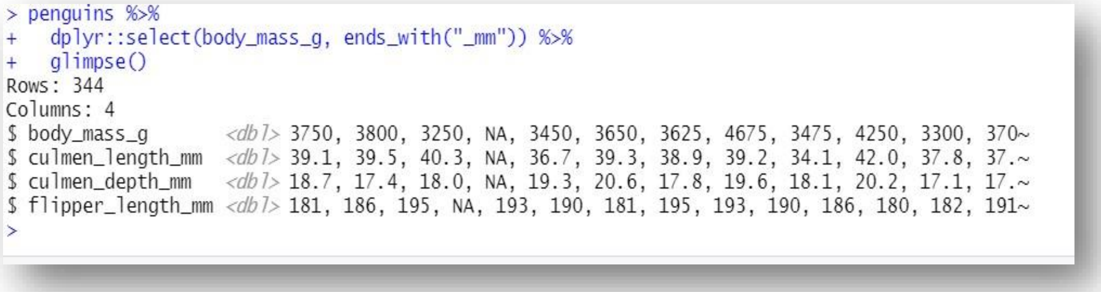


Figure no. 4: Species wise count VS Sex

**Analysis**: On this figure no. 4, shows Sex-wise Penguins of which species are on which island.



```
> penguins %>%
+   dplyr::select(body_mass_g, ends_with("_mm")) %>%
+   glimpse()
Rows: 344
Columns: 4
$ body_mass_g       <dbl> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, 4250, 3300, 370~
$ culmen_length_mm  <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, 42.0, 37.8, 37.~
$ culmen_depth_mm   <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, 20.2, 17.1, 17.~
$ flipper_length_mm <dbl> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186, 180, 182, 191~
>
```

Table no. 6: Penguins body_mass, depth, length

**Analysis**: Table no. 6, it shows that different characteristics of Penguins like as Body mass in g ,Culmen length in mm, Culmen depth in mm, flipper length in mm and Sex.

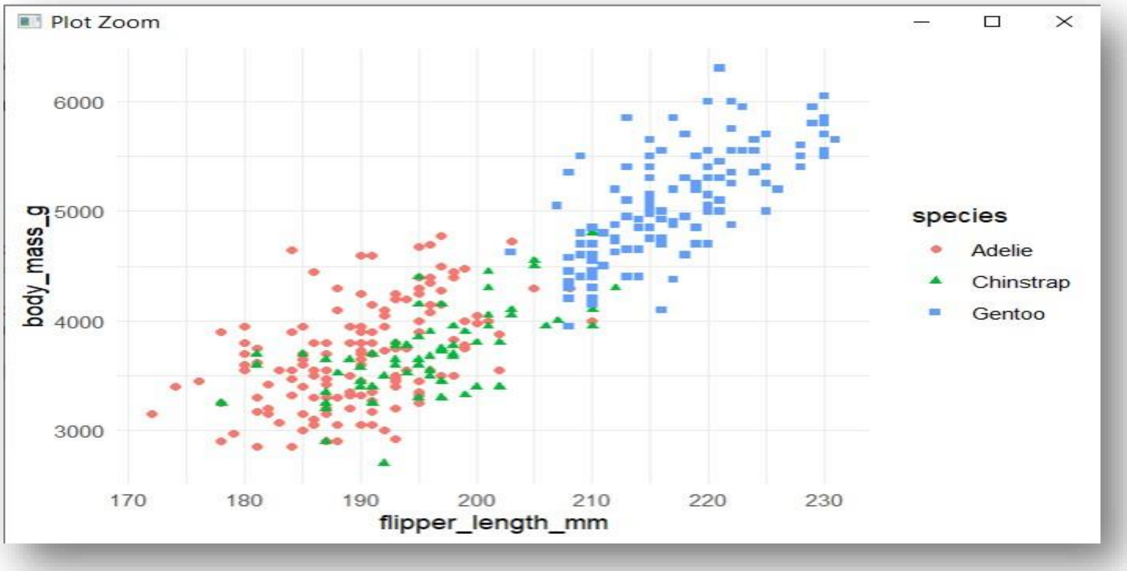


Figure no.5: body_mass_g VS Flipper_length _mm

**Analysis**: On this Figure no. 5, shows that body mass and flipper length of penguins. red diamond represents Adelie Penguin species, green triangle represents Chinstrap Penguin species and blue square represent Gentoo Penguin species. Now, with different color and shape for different species, it is easily understandable.

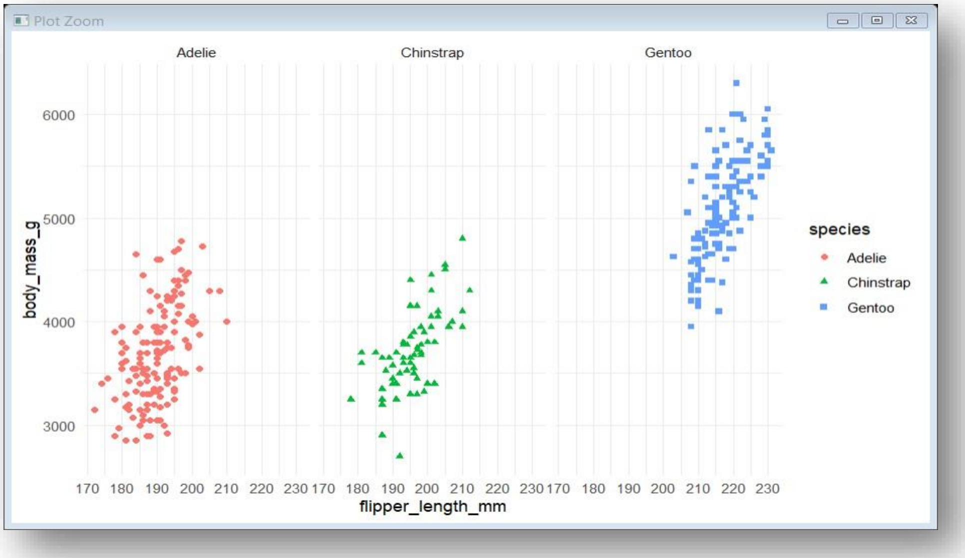


Figure no. 6: body_mass_g VS Flipper_length _mm

**Analysis**: In order to make it even more clear, the plot can be divided into three sub plots corresponding to each species. Finally, adding labels and annotations to the plot to make the plot more informative. Adelie appearances small in size associating with other classes.

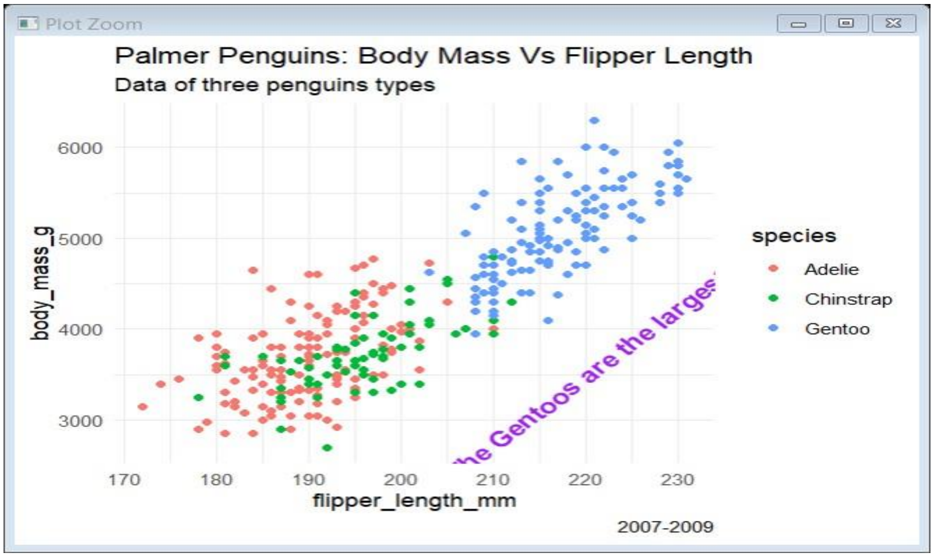


Figure no.7: body_mass_g VS Flipper_length _mm

**Analysis**:  Clearly shows that the Gentoo Penguins species are the Largest penguin species.

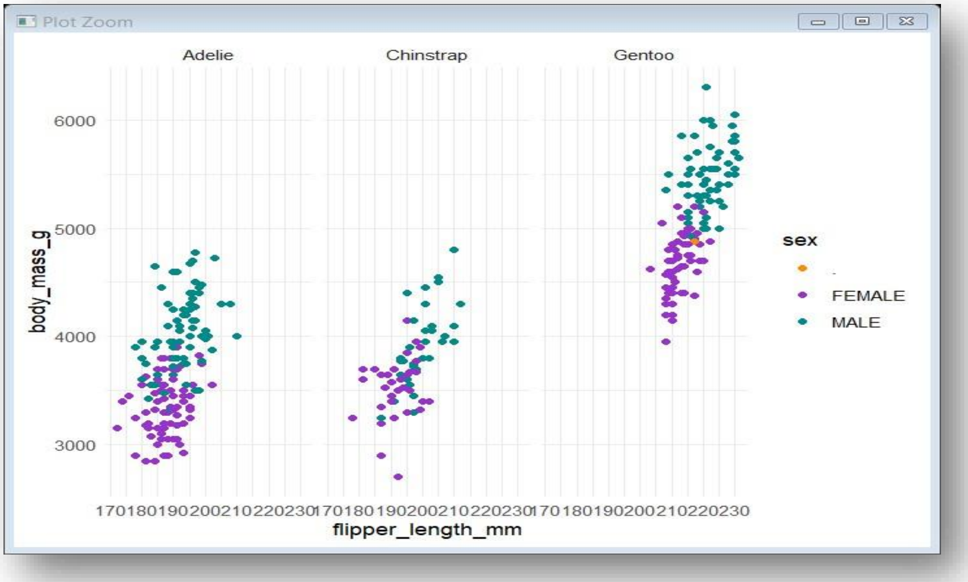


Figure no.8: body_mass_g VS Flipper_length _mm

**Analysis**: The above graph plot shows species wise male and female body mass and flipper length. By using above graph we analyzed male penguins have large body mass and flipper length of all species.

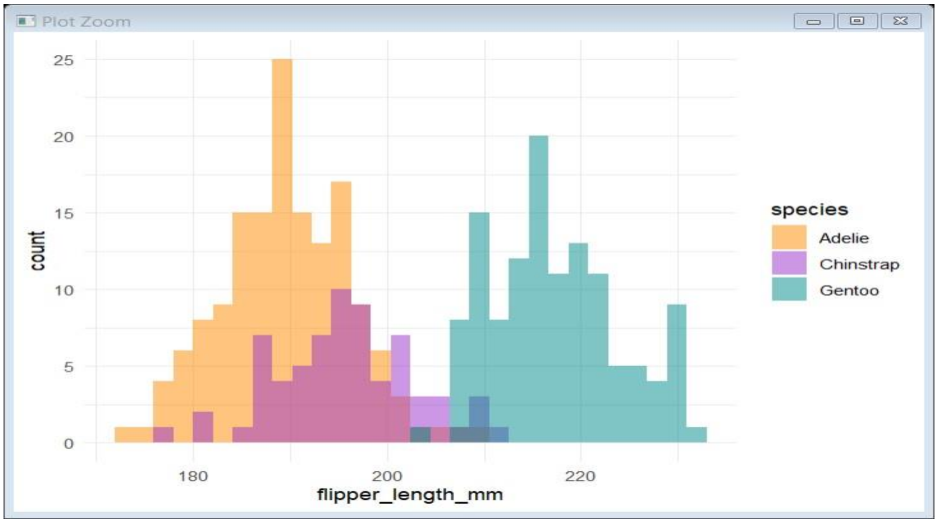


Figure no. 9: Count VS Flipper_length _mm

**Analysis**: Above figure no.9 it shows that count wise flipper length in mm. By using above histogram we clearly say that large no of Adelie have flipper length is between 180-200 mm. And the Gentoo penguin have largest flipper which is 220 mm.

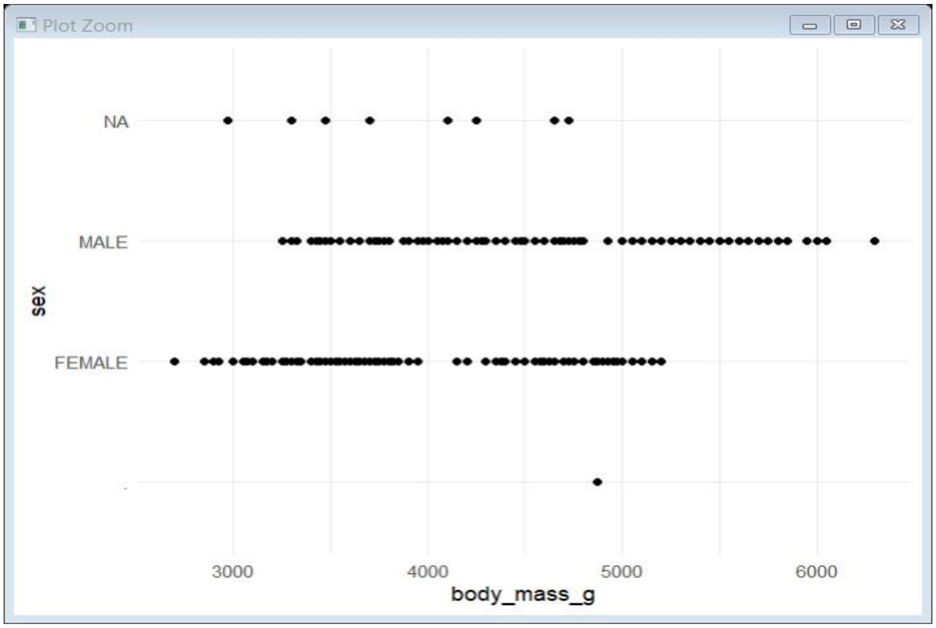


Figure no .10: Sex VS body_mass_g

**Analysis**: Above figure no.10 shows that sex wise body mass in gram. By using above graph we clearly say that most of male penguins have body mass between 3000-5000 grams, and female have body mass between 3000-4000 grams, that why we conclude that male penguins are bigger than female penguins.

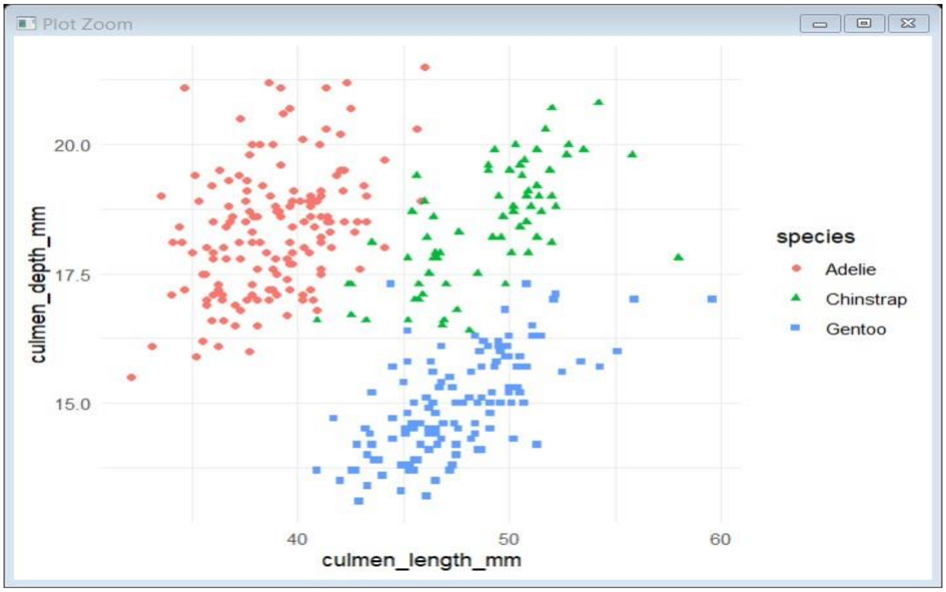


Figure no.11: Culmen_depth_mm VS Culmen_length _mm

**Analysis**: On this Figure no.11, shows that culmen depth and culmen length in mm of penguins.

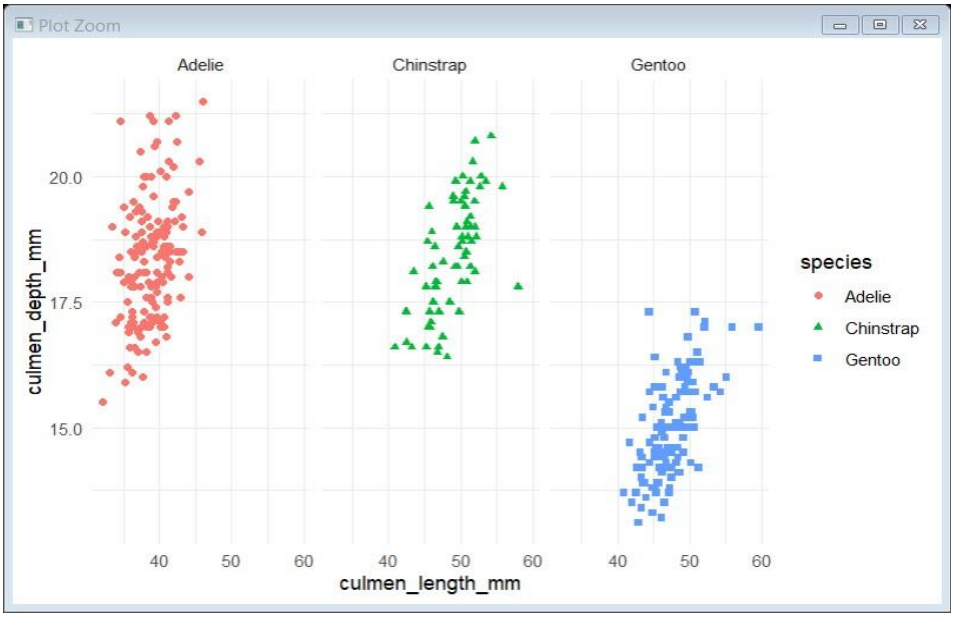


Figure no.12: Culmen_depth_mm VS Culmen_length _mm

**Analysis**: Above graph plot shows Adelie penguins culmen depth is large but Gentoo penguins culmen length is large.

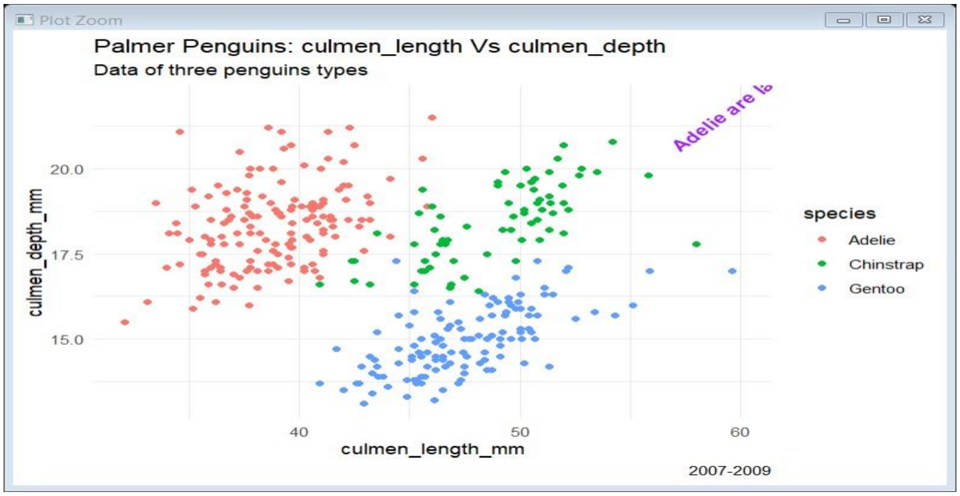


Figure no.13: Culmen_depth_mm VS Culmen_length _mm

**Analysis**: From the above graph we conclude that Adelie penguins culmen depth is large but Gentoo penguins culmen length is large.

## 5. Observations:

- It is observed that Body mass and Flipper length are positively correlated, showing that Gentoos have a longer Flipper length and also heavier body mass.
- Culmen length and Flipper also positively correlated, adelies have shorter flippers and shorter culmen lengths.

## 6. Conclusions:

A penguin to be a Gentoo, it must a has body mass relatively heavier than all species and a longer Flipper length larger but a shorter culmen depth and also found in the Biscoe islands only. Also a Adelie penguins, more probable has comparatively shorter Culmen-length and and longer culmen depth and can be found in all Islands.

For Chinstrap, a relatively longer length of both culmen length and culmen depth and only found on the Dream Islands. Adelie lives in all three islands whereas Gentoo lives in Biscoe and Chinstrap lives in Torgersen. maximum of Gentoo has distinctive flipper length, body mass, culmen depth. most of adelie culmen length is less than 40mm, most of the eggs observed in November month, Population of adelie and gentoo increased from 2007 to 2011, Generally, for the reason that of climate transformation. The heating up of the Antarctic-Peninsula is affecting alterations to the physical and living atmosphere of Antarctica. The scattering of penguin associations has transformed such as the oceanic snow environments change. Melting of continuing ice and snow shelters has resulted in augmented colonization by vegetation.

## 7. References

1. Gorman KB, Williams TD, Fraser WR (2014). Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus *Pygoscelis*). PLoS ONE 9(3):e90081.
2. Palmer Station Antarctica LTER and K. Gorman, 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Adélie penguins, Gentoo penguin & Chinstrap penguin (*Pygoscelis adeliae*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 5 and 6. Environmental Data Initiative.
3. Steven D. Emslie , William R. Fraser (1998) ,Abandoned penguin colonies and environmental change in the Palmer Station area, Anvers Island, Antarctic Peninsula
4. Megan A. Cimino,☒ 1 Donna L. Patterson-Fraser, (2019) The interaction between island geomorphology and environmental parameters drives Adélie penguin breeding phenology on neighboring islands near Palmer Station, Antarctica
5. Schuyler C. Nardelli,Megan A. Cimino, (2021),Krill availability in adjacent Adélie and gentoo penguin foraging regions near Palmer Station, Antarctica
6. Retrieved web site 30 June 2020: https://www.kaggle.com/datasets/parulpandey/palmer-archipelagoantarctica-penguin-data.
7. Retrieved web site 30 June 2020: https://simple.wikipedia.org/wiki/Penguin
8. Retrieved web site 30 June 2020: https://en.wikipedia.org/wiki/Penguin#cite_note-pengsent-6

9. Retrieved web site 30 June 2020: https://simple.wikipedia.org/wiki/Chinstrap_penguin
10. Retrieved web site 30 June 2020: https://simple.wikipedia.org/wiki/Gentoo_penguin

Sinhgad Institute of Management
PUNE
411 041.
*

CSingh

Dr.Chandrani Singh ,Director –MCA,SIOM

मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ.शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई - ४०० ०९८.

Estd. 1971

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Tel. No. - 022 2652 7082

Website : mu.ac.in/distance-open-learning    e-mail : director@idol.mu.ac.in

"सुवर्ण महोत्सवी वर्ष २०२०–२०२१"    "Golden Jubilee Year 2020 – 2021"

IDOL/SMU/2066 of 2021
Date: 2021

To
Ms Kumudini Manwar
Assistant Professor
Lalit, F-1104, Nanded City, NRD - 2,
Pune 411041

To
Mr. Kiran Pandurang Patil
Assistant Professor
At-Pandewadi, Post-Sawarde, Taluka-
Radhanagari, Dist-Kolhapur, PIN 416212

To
Ms Tejàswini Nadgauda
Associate Professor
Janaki .Plot.No.24 Sagar Society , Sahakar
Nagar -2 Pune 411009

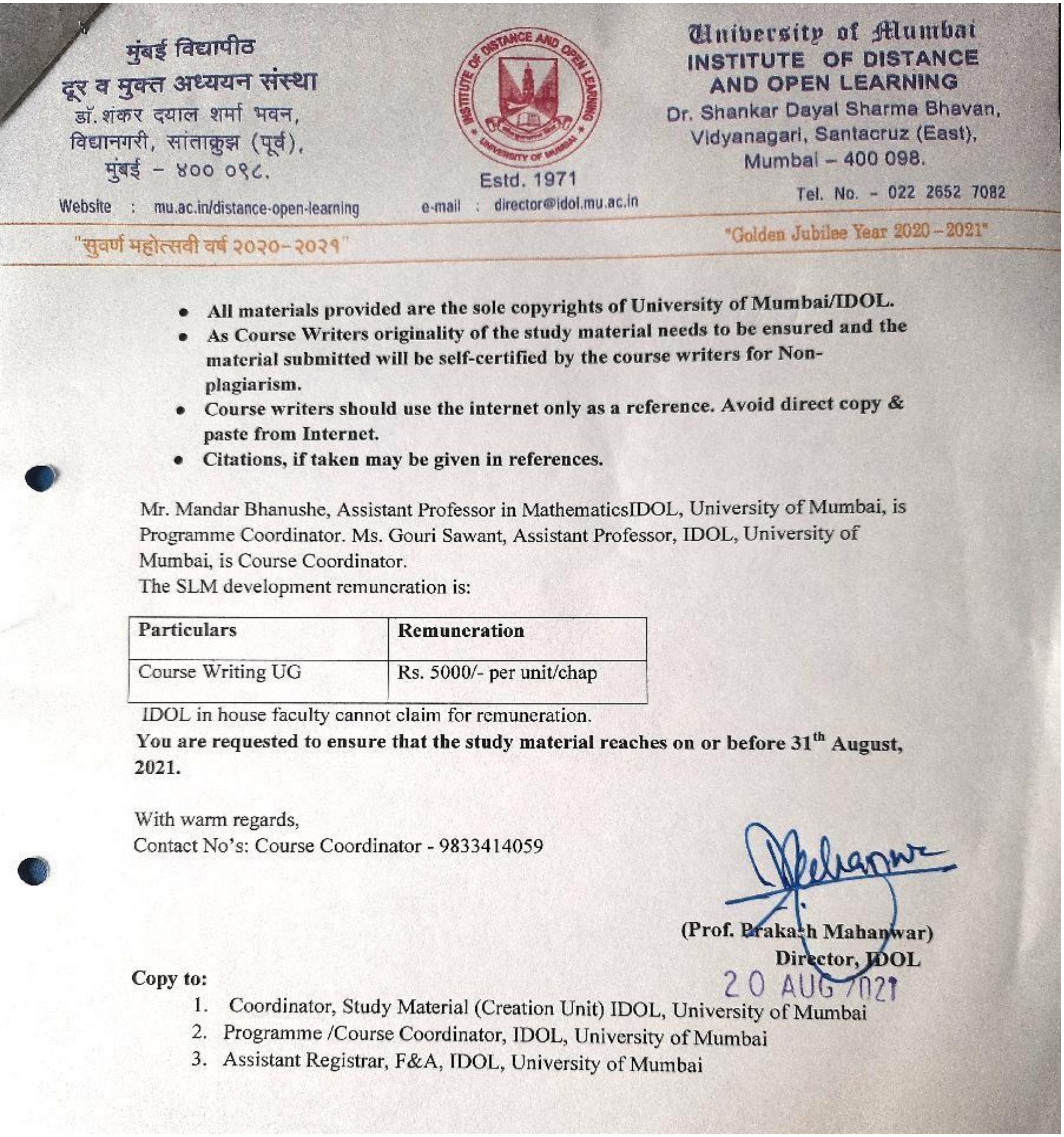
To
Mr Nimesh Punjani
Assistant Professor
7, Ahuja Bhavan, B.P Cross Road NO. 4,
Mulund West, Mumbai 400080

It's my great pleasure to inform that you have been appointed as course writer for preparing Self Learning Material (SLM) in the Course **B.Sc.I.T** for **Computer Oriented Statistical Technique , Semester IV.**

Individual faculty is requested to prepare the study material as under: -

| Module / Unit No. | No of Chapters | Name of the course writer |
|---|---|---|
| Unit - I | 3 | Ms Kumudini Manwar |
| Unit - II | 3 | Ms Tejàswini Nadgauda |
| Unit - III | 3 | Mr. Kiran Pandurang Patil |
| Unit - IV | 2 | Mr Nimesh Punjani |
| Unit – V | 2 | Mr Nimesh Punjani |

Detailed syllabus and Format of the respective units is enclosed for information. Further entire syllabus should not exceed 8 sub- units/chapters and same will be of 3500 to 5000 words. Course writers are requested to prepare their respective units duly taking into consideration the standard guidelines of University of Mumbai with regard to Plagiarism. Guiding principles are given below for reference.



मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Estd. 1971

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Website : mu.ac.in/distance-open-learning    e-mail : director@idol.mu.ac.in    Tel. No. - 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०–२०२१"    "Golden Jubilee Year 2020 – 2021"

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As Course Writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Course writers should use the internet only as a reference. Avoid direct copy & paste from Internet.**
- **Citations, if taken may be given in references.**

Mr. Mandar Bhanushe, Assistant Professor in MathematicsIDOL, University of Mumbai, is Programme Coordinator. Ms. Gouri Sawant, Assistant Professor, IDOL, University of Mumbai, is Course Coordinator.
The SLM development remuneration is:

| **Particulars** | **Remuneration** |
|---|---|
| Course Writing UG | Rs. 5000/- per unit/chap |

IDOL in house faculty cannot claim for remuneration.
**You are requested to ensure that the study material reaches on or before 31th August, 2021.**

With warm regards,
Contact No's: Course Coordinator - 9833414059

Mahanwar

**(Prof. Prakash Mahanwar)**
**Director, IDOL**

20 AUG 2021

**Copy to:**

1. Coordinator, Study Material (Creation Unit) IDOL, University of Mumbai
2. Programme /Course Coordinator, IDOL, University of Mumbai
3. Assistant Registrar, F&A, IDOL, University of Mumbai

**UNIVERSITY OF MUMBAI**
**INSTITUTE OF DISTANCE AND OPEN LEARNING**
**VIDYANAGARI**

**No. IDOL/SMU/05** 10/51/**of 2021.**
**Date :** 10/03/ **2021.**

**To,**
Prof. Santosh S. Deshmukh
SARTHAK BEAULIEU, sr. no. 32/5a/2,
Flat No – B 603, Near Balaji Hotel,
Pisoli, Pune – 411 028.

Prof. Rahul Borate
Flat no. 8 ganesh shrushti phase – I,
Bhintadenagar Ambegaon Bk,
Pune – 411 046.

Prof. Rahul Navale
Flat no. 13, Phase – 6B,
Chandrangan Associates,
Ambegaon Bk Pune – 411 046.

**I am pleased to inform that you have been appointed as course writer for the purpose to prepare the study material in Self Learning Material (SLM) format, in the Web Programming for F.Y.BSc. I.T. Semester II, is segregated in Units and individual faculty member is requested to prepare the study material as under:-**

| **Chapters** | **Name of the course writers** |
|---|---|
| Unit 1,Chapters (1) | Prof.Santosh S. Deshmuk |
| Unit 2,Chapters (4) | Prof. Rahul Borate |
| Unit 4,Chapters (4) | Prof. Rahul Navale |
| Unit 5,Chapters ~~(1)~~ (2) | Prof. Santosh S. Deshmukh |

Detailed syllabus and format of the respective units are enclosed for information. Further, entire syllabus should not exceed 16 sub- units/chapters and same will be of 3500 to 5000 words. Course writers are requested to prepare their respective units duly taking into consideration the standard guidelines of University of Mumbai with regard to Plagiarism guiding principles are given below for reference.

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As course writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Citations, if taken may be given in references.**
- **Course Writers Should use the internet only as a reference. Avoid direct copy & paste from internet.**

- **The faculty head of Science & Technology is Mr. Mandar Bhanushe Assistant Professor (M.Sc. Mathematics) in IDOL, University of Mumbai.**

मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई - ४०० ०९८.

Website : mu.ac.in/distance-open-learning

Estd. 1971
e-mail : director@idol.mu.ac.in

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०-२०२१" "Golden Jubilee Year 2020-2021"

IDOL/SMU/967 of 2021
Date: 29/06/ 2021

To
Mr Ankush Kudale
Assistant Professor,
301, Aura Elegance, Behind Hotel Dawat,
Vadgaon Bk, Off Sinhgad Road, Pune
411041

To
Dr. Milind Godase
Assistant Professor,
Flat. No. 04, Vitthal Chaya, 19/8, Raykar
Nagar, Dhayari, Pune 411041

To
Ms. Pradnya Patil
Assistant Professor,
A 303, Park Dew CHS, Plot No 73, Sector
20, Kharghar, Navi Mumbai-410210

To
Ms Jyotika D. Chheda
12/13 2A Wing, Matrushraddha CHS,
Near Tilak School, Tilak Nagar, Dombivli
East-421201

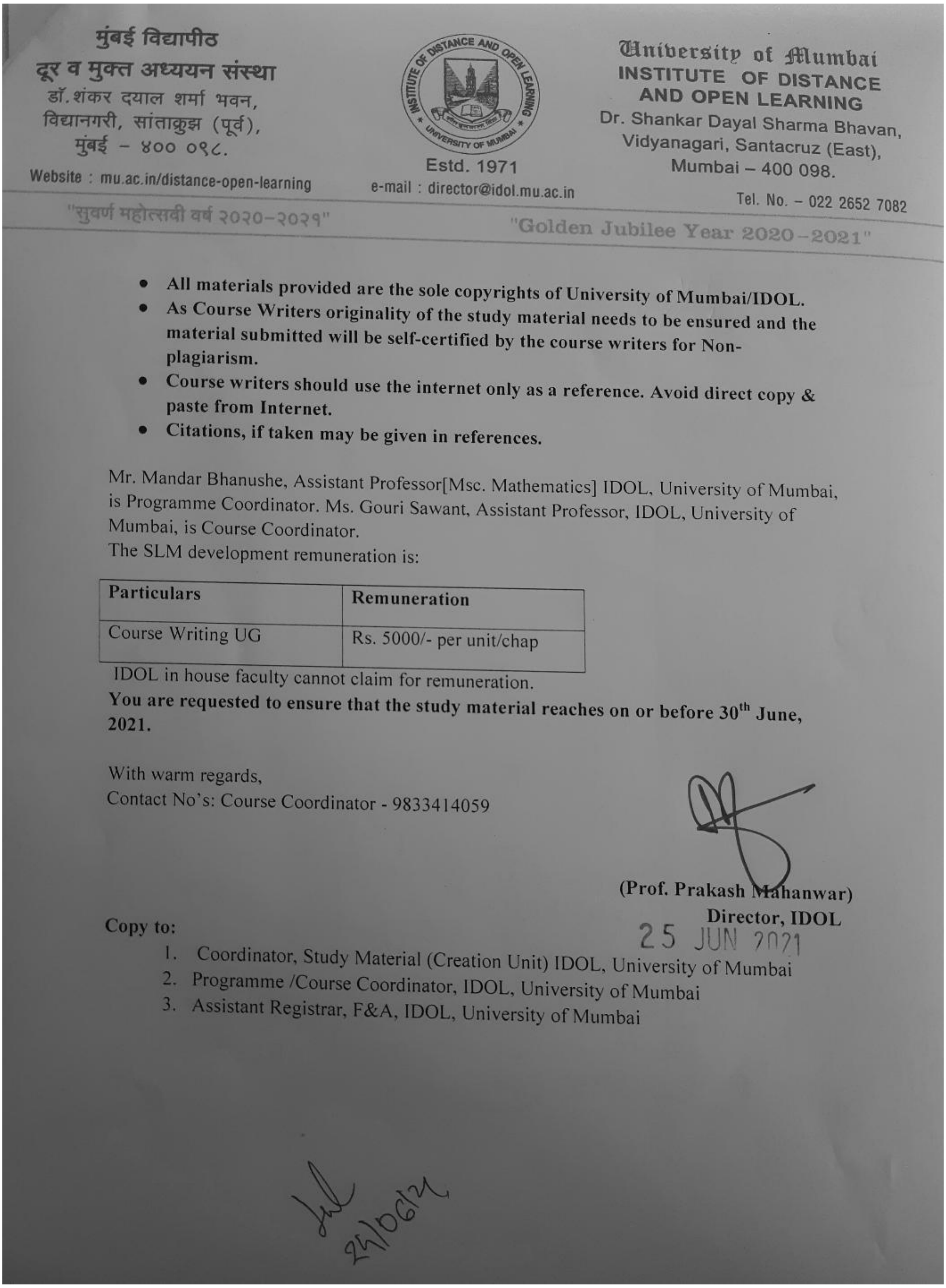
It's my great pleasure to inform that you have been appointed as course writer for preparing Self Learning Material (SLM) in the Course **B.Sc.I.T** for **Computer Networks, Semester III.**

Individual faculty is requested to prepare the study material as under:-

| Module / Unit No. | Name of the course writer |
|---|---|
| Unit - I | Mr Ankush Kudale |
| Unit - II | Dr Milind Godase |
| Unit - III | Ms Pradnya Patil |
| Unit - IV | Mr Jyotika Chedda |
| Unit – V | Mr Milind Godase |

Detailed syllabus and Format of the respective units is enclosed for information. Further entire syllabus should not exceed 8 sub- units/chapters and same will be of 3500 to 5000 words.

Course writers are requested to prepare their respective units duly taking into consideration the standard guidelines of University of Mumbai with regard to Plagiarism. Guiding principles are given below for reference.



मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.
Website : mu.ac.in/distance-open-learning

INSTITUTE OF DISTANCE AND OPEN LEARNING · UNIVERSITY OF MUMBAI
Estd. 1971
e-mail : director@idol.mu.ac.in

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.
Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०–२०२१" "Golden Jubilee Year 2020–2021"

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As Course Writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Course writers should use the internet only as a reference. Avoid direct copy & paste from Internet.**
- **Citations, if taken may be given in references.**

Mr. Mandar Bhanushe, Assistant Professor[Msc. Mathematics] IDOL, University of Mumbai, is Programme Coordinator. Ms. Gouri Sawant, Assistant Professor, IDOL, University of Mumbai, is Course Coordinator.
The SLM development remuneration is:

| **Particulars** | **Remuneration** |
|---|---|
| Course Writing UG | Rs. 5000/- per unit/chap |

IDOL in house faculty cannot claim for remuneration.
**You are requested to ensure that the study material reaches on or before 30th June, 2021.**

With warm regards,
Contact No's: Course Coordinator - 9833414059

**(Prof. Prakash Mahanwar)**
**Director, IDOL**

25 JUN 2021

**Copy to:**

1. Coordinator, Study Material (Creation Unit) IDOL, University of Mumbai
2. Programme /Course Coordinator, IDOL, University of Mumbai
3. Assistant Registrar, F&A, IDOL, University of Mumbai

25/06/21

Sinhgad Institute of Management
PUNE 411 041.

मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ.शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Estd. 1971

University of Mumbai
INSTITUTE OF DISTANCE AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Website : mu.ac.in/distance-open-learning    e-mail : director@idol.mu.ac.in    Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०-२०२१"    "Golden Jubilee Year 2020–2021"

**IDOL/SMU/1034 of 2021**
**Date: 5/07/ 2021**

To
Mr.Rahul Navale
Assistant Professor
Flat No.13 Phase 6B, Chandrangan
Associates S.No 15/5B, Ambegaon(Bk)
Pune-411046

To
Mr Sunil khilari
Assistant Professor,
STES,Rajgad 1ldg.,Sinhgad Boy's Hostel,
Sinhgad Road,Vadon-Bk,Pune-411041

To
Ms Aarti Sahitya
Assistant Professor,
Flat no 402, Melbourne B12, opp
yogidham autostand, yogidham kalyan
west 421301

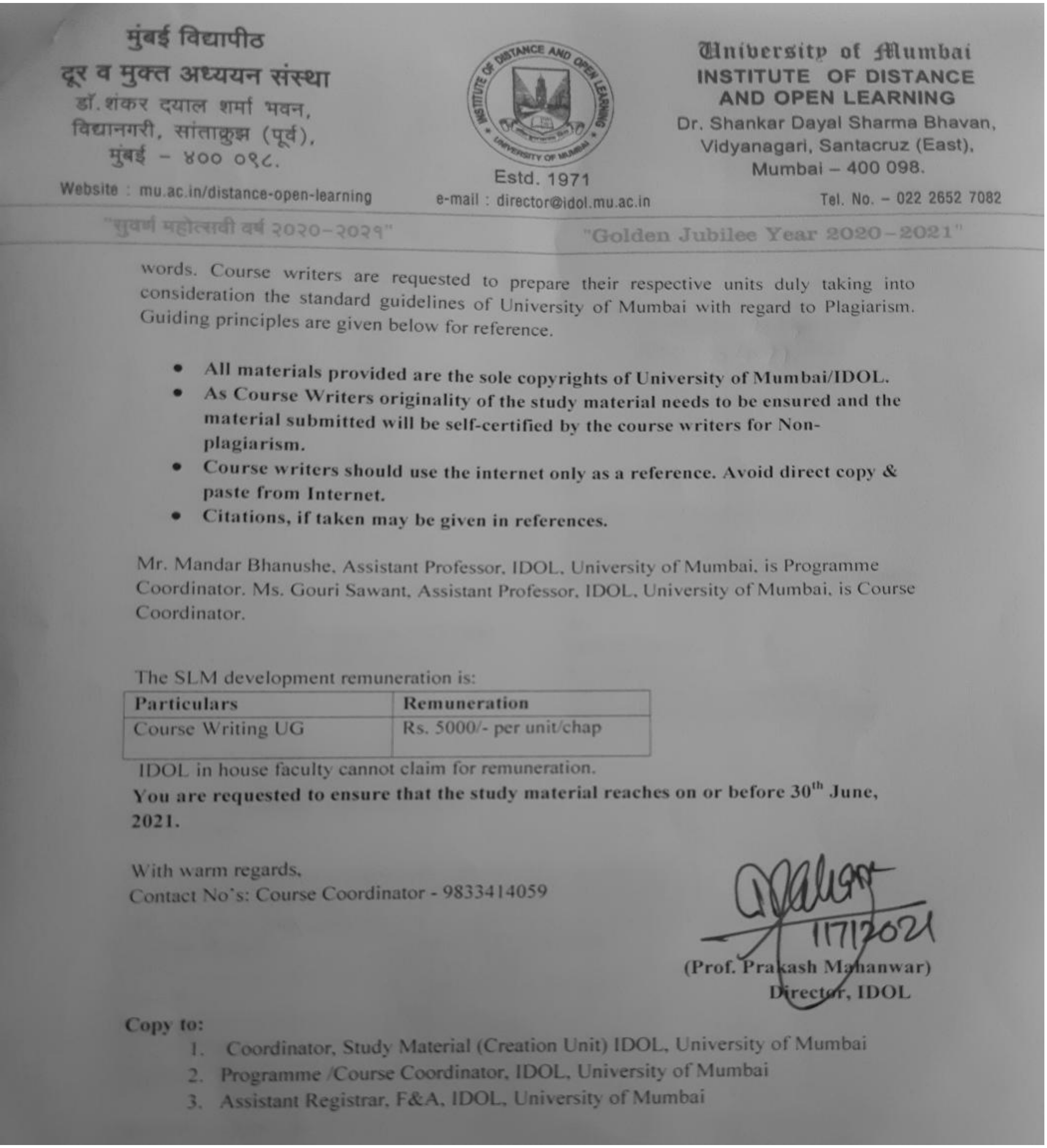
To
Inumarthi Veerabhadra Srinivas
Assistant Professor,
Flat No 203, FE1, Kasturi CHS, Moraj
Residency, Sector 16, Sanpada, Navi Mumbai

To
Mr Rahul Borate
Assistant Professor,
Flat No.13 Phase 6B, Chandrangan
Associates Ambegaon(Bk) Pune-411046

It's my great pleasure to inform that you have been appointed as course writer for preparing Self Learning Material (SLM) in the Course **B.Sc.I.T** for **Python Programming, Semester III.** Individual faculty is requested to prepare the study material as under:-

| Module / Unit No. | Name of the course writer |
|---|---|
| Unit I | Mr Rahul Navale |
| Unit II | Mr Innumarthi veerabhadra srinivas |
| Unit III | Mr Sunil Khilari |
| Unit IV | Mr Rahul Borate |
| Unit V | Ms Aarti Sahitya |

Detailed syllabus and Format of the respective units is enclosed for information. Further entire syllabus should not exceed 8 sub- units/chapters and same will be of 3500 to 5000



मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Website : mu.ac.in/distance-open-learning

INSTITUTE OF DISTANCE AND OPEN LEARNING
UNIVERSITY OF MUMBAI
Estd. 1971
e-mail : director@idol.mu.ac.in

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०-२०२१" "Golden Jubilee Year 2020-2021"

words. Course writers are requested to prepare their respective units duly taking into consideration the standard guidelines of University of Mumbai with regard to Plagiarism. Guiding principles are given below for reference.

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As Course Writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Course writers should use the internet only as a reference. Avoid direct copy & paste from Internet.**
- **Citations, if taken may be given in references.**

Mr. Mandar Bhanushe, Assistant Professor, IDOL, University of Mumbai, is Programme Coordinator. Ms. Gouri Sawant, Assistant Professor, IDOL, University of Mumbai, is Course Coordinator.

The SLM development remuneration is:

| **Particulars** | **Remuneration** |
|---|---|
| Course Writing UG | Rs. 5000/- per unit/chap |

IDOL in house faculty cannot claim for remuneration.

**You are requested to ensure that the study material reaches on or before 30th June, 2021.**

With warm regards,
Contact No's: Course Coordinator - 9833414059

11/7/2021
**(Prof. Prakash Mahanwar)**
**Director, IDOL**

**Copy to:**

1. Coordinator, Study Material (Creation Unit) IDOL, University of Mumbai
2. Programme /Course Coordinator, IDOL, University of Mumbai
3. Assistant Registrar, F&A, IDOL, University of Mumbai

मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Estd. 1971

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Website : mu.ac.in/distance-open-learning
e-mail : director@idol.mu.ac.in
Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०–२०२१"
"Golden Jubilee Year 2020–2021"

**IDOL/SMU/968 of 2021**
**Date:** 29/06/**2021**

To
Ms Kumudini Manwar
Assistant Professor,
Lalit F-1104, Nanded City, NRD 2,
Pune –411041

To
Dr Chandrani Singh
Director MCA,
A-102 Indraprastha Apartment Rd-7
Central Avenue Kayaninagar Pune-411006

To
Mr Amar A. Shinde
Academic Counsellor
S.No. 53/2/A , Dhareshwar Nagar,
Dhayari, Off. Sinhgad Road ,
Pune- 411041

To
Dr. Manisha Kumbhar
Assistant Professor,
Sai-Krupa,Plot No.29, Shreedhar Col,
Karvenagar,Pune-411052

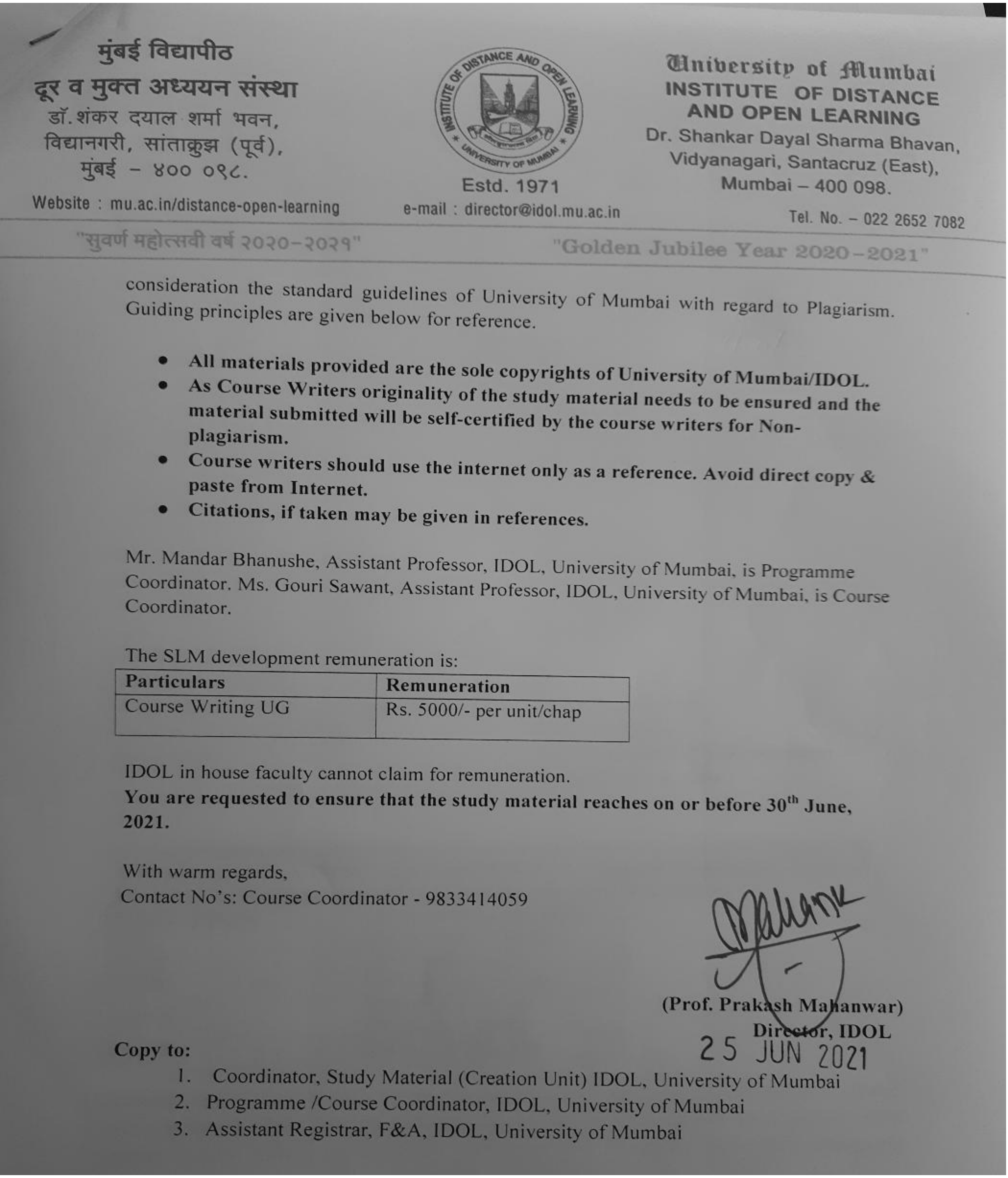
To
Mrs Priti A. Shinde
Assistant Professor,
S.No. 53/2/A ,Dhareshwar Nagar, Dhayari,
Off.Sinhgad-Road,
Pune-411041

To
Mr Ravikant D. Kale
Assistant Professor
Visawa Bunglow,Bhairavnath-nagar ,
Kusgaon (Bk) Lonavla

It's my great pleasure to inform that you have been appointed as course writer for preparing Self Learning Material (SLM) in the Course **B.Sc.I.T** for **Applied Mathematics, Semester III.** Individual faculty is requested to prepare the study material as under:-

| **Module / Unit No.** | **Name of the course writer** |
|---|---|
| Unit I | Ms. Kumudini Manwar |
| Unit II | Dr Chandrani Singh |
| Unit III | Dr. Manisha Kumbhar |
| Unit IV | Mrs. Priti A. Shinde<br>Mr. Amar A. Shinde |
| Unit V | Mr. Ravikant D. Kale |

Detailed syllabus and Format of the respective units is enclosed for information. Further entire syllabus should not exceed 8 sub- units/chapters and same will be of 3500 to 5000 words. Course writers are requested to prepare their respective units duly taking into



मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Website : mu.ac.in/distance-open-learning

Estd. 1971
e-mail : director@idol.mu.ac.in

University of Mumbai
INSTITUTE OF DISTANCE AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०-२०२१"  "Golden Jubilee Year 2020-2021"

consideration the standard guidelines of University of Mumbai with regard to Plagiarism. Guiding principles are given below for reference.

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As Course Writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Course writers should use the internet only as a reference. Avoid direct copy & paste from Internet.**
- **Citations, if taken may be given in references.**

Mr. Mandar Bhanushe, Assistant Professor, IDOL, University of Mumbai, is Programme Coordinator. Ms. Gouri Sawant, Assistant Professor, IDOL, University of Mumbai, is Course Coordinator.

The SLM development remuneration is:

| **Particulars** | **Remuneration** |
|---|---|
| Course Writing UG | Rs. 5000/- per unit/chap |

IDOL in house faculty cannot claim for remuneration.

**You are requested to ensure that the study material reaches on or before 30th June, 2021.**

With warm regards,
Contact No's: Course Coordinator - 9833414059

**(Prof. Prakash Mahanwar)**
**Director, IDOL**

25 JUN 2021

**Copy to:**

1. Coordinator, Study Material (Creation Unit) IDOL, University of Mumbai
2. Programme /Course Coordinator, IDOL, University of Mumbai
3. Assistant Registrar, F&A, IDOL, University of Mumbai

मुंबई विद्यापीठ
दूर व मुक्त अध्ययन संस्था
डॉ. शंकर दयाल शर्मा भवन,
विद्यानगरी, सांताक्रुझ (पूर्व),
मुंबई – ४०० ०९८.

Estd. 1971

University of Mumbai
INSTITUTE OF DISTANCE
AND OPEN LEARNING
Dr. Shankar Dayal Sharma Bhavan,
Vidyanagari, Santacruz (East),
Mumbai – 400 098.

Website : mu.ac.in/distance-open-learning
e-mail : director@idol.mu.ac.in
Tel. No. – 022 2652 7082

"सुवर्ण महोत्सवी वर्ष २०२०-२०२१"
"Golden Jubilee Year 2020-2021"

**IDOL/SMU/966 of 2021**
**Date: 29/06/2021**

To
Ms Aarti sahitya
Assistant Professor,
Flat no 402, Melbourne B12, opp
Yogidham autostand, Yogidham Kalyan
west 421301

To
Dr Vidya Gavekar
Associate Professor,
C-1,Bhagyashree Complex,Raikar
Nagar,Near Solapur Janta Shakari
Bank,Dhayri,Pune-411041

To
Mr Sumit Mali
Assistant Professor,
205, 2bhk ambegaon staff quarter, besides
NBN sinhgad school of engineering,
Back side of Kaveri boy's hostel,
near bank of india ,Singad institute,
ambegaon (bk.), pune 41

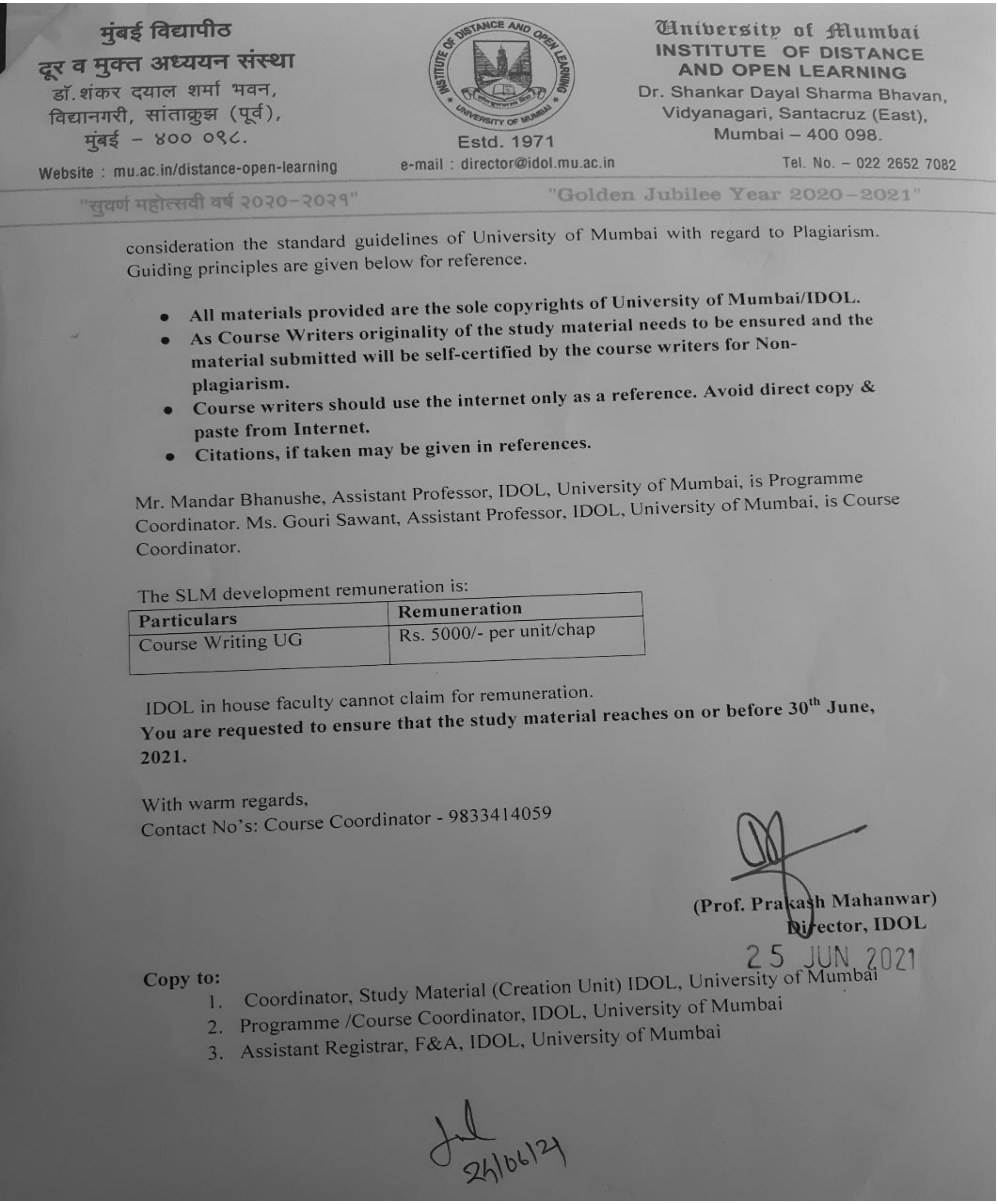
To
Mr Zahirabbas Jainuddin Mulani
Assistant Professor,
kh2/18/202,sec16 Kharghar,navi mumbai
410210

To
Mr. Abhijeet Pawaskar
3/2, Shivsadan, sitanagar,
Bandrekarwadi, Near Shree Siddhivinayak
Mandir,
Jogeshwari East , Mumbai 400060

It's my great pleasure to inform that you have been appointed as course writer for preparing Self Learning Material (SLM) in the Course **B.Sc.I.T** for **Data Structures, Semester III.** Individual faculty is requested to prepare the study material as under:-

| **Module / Unit No.** | **Name of the course writer** |
|---|---|
| Unit - I | Ms Aarti sahitya |
| Unit - II | Dr Vidya Gavekar |
| Unit - III | Mr Sumit Mali |
| Unit - IV | Mr Zahirabbas Jainuddin Mulani |
| Unit – V | Mr. Abhijeet Pawaskar |

Detailed syllabus and Format of the respective units is enclosed for information. Further entire syllabus should not exceed 8 sub- units/chapters and same will be of 3500 to 5000 words. Course writers are requested to prepare their respective units duly taking into



consideration the standard guidelines of University of Mumbai with regard to Plagiarism. Guiding principles are given below for reference.

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As Course Writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Course writers should use the internet only as a reference. Avoid direct copy & paste from Internet.**
- **Citations, if taken may be given in references.**

Mr. Mandar Bhanushe, Assistant Professor, IDOL, University of Mumbai, is Programme Coordinator. Ms. Gouri Sawant, Assistant Professor, IDOL, University of Mumbai, is Course Coordinator.

The SLM development remuneration is:

| **Particulars** | **Remuneration** |
|---|---|
| Course Writing UG | Rs. 5000/- per unit/chap |

IDOL in house faculty cannot claim for remuneration.

**You are requested to ensure that the study material reaches on or before 30th June, 2021.**

With warm regards,
Contact No's: Course Coordinator - 9833414059

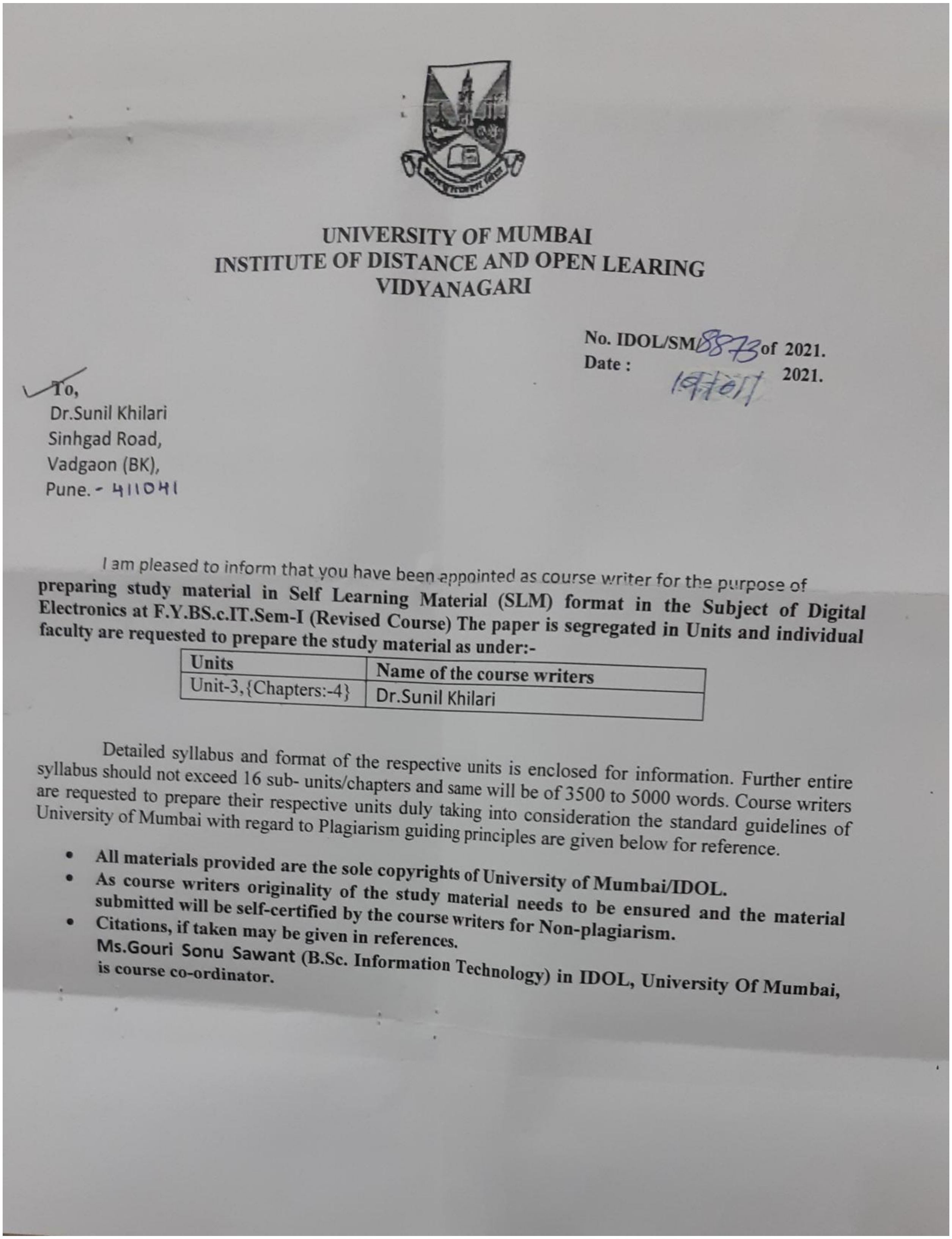
**(Prof. Prakash Mahanwar)**
**Director, IDOL**

25 JUN 2021

**Copy to:**

1. Coordinator, Study Material (Creation Unit) IDOL, University of Mumbai
2. Programme /Course Coordinator, IDOL, University of Mumbai
3. Assistant Registrar, F&A, IDOL, University of Mumbai

24/06/21



# UNIVERSITY OF MUMBAI
# INSTITUTE OF DISTANCE AND OPEN LEARING
# VIDYANAGARI

No. IDOL/SM/8873 of 2021.
Date : 19.7.2021 2021.

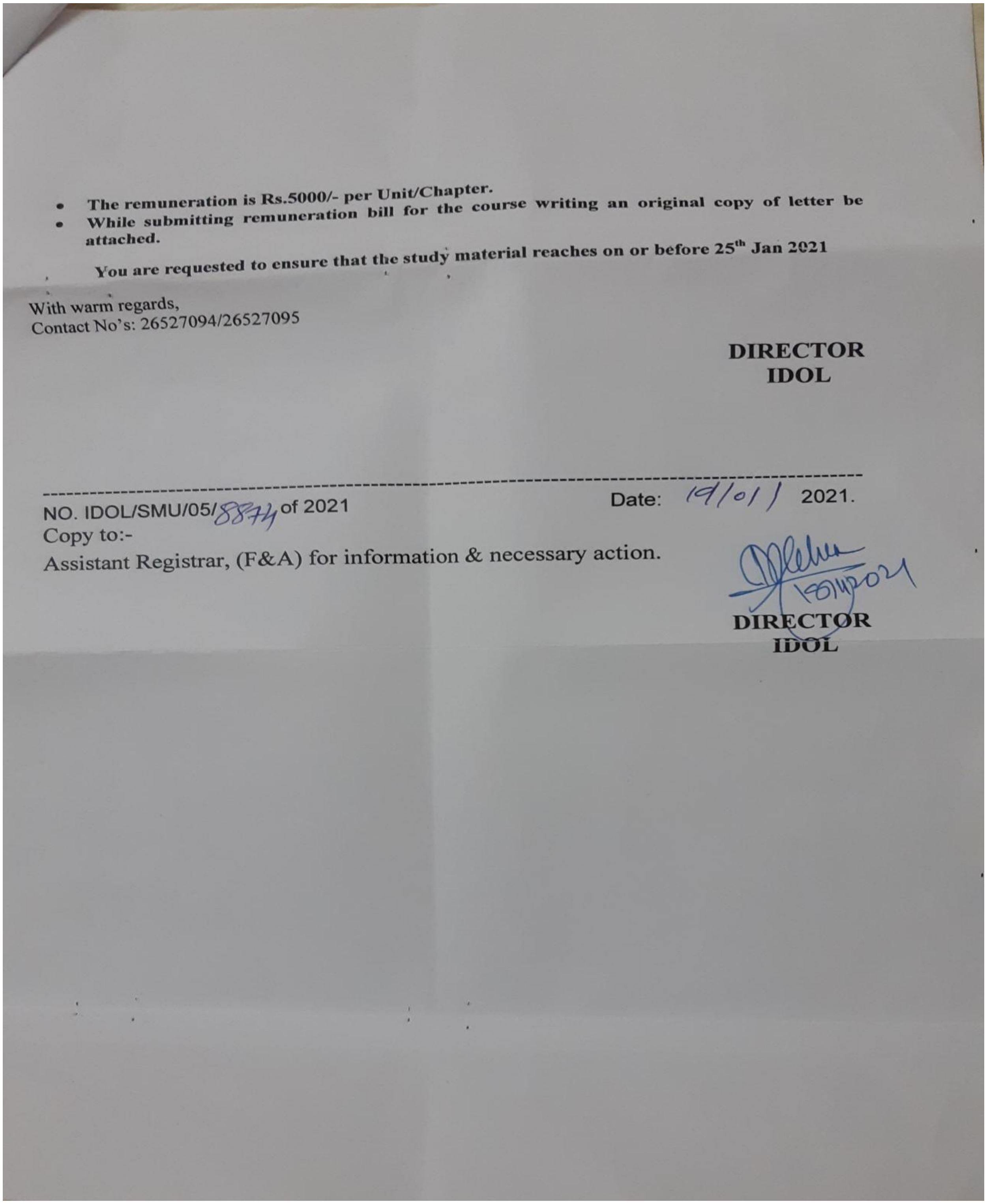
To,
Dr.Sunil Khilari
Sinhgad Road,
Vadgaon (BK),
Pune. - 411041

I am pleased to inform that you have been appointed as course writer for the purpose of **preparing study material in Self Learning Material (SLM) format in the Subject of Digital Electronics at F.Y.BS.c.IT.Sem-I (Revised Course) The paper is segregated in Units and individual faculty are requested to prepare the study material as under:-**

| Units | Name of the course writers |
|---|---|
| Unit-3,{Chapters:-4} | Dr.Sunil Khilari |

Detailed syllabus and format of the respective units is enclosed for information. Further entire syllabus should not exceed 16 sub- units/chapters and same will be of 3500 to 5000 words. Course writers are requested to prepare their respective units duly taking into consideration the standard guidelines of University of Mumbai with regard to Plagiarism guiding principles are given below for reference.

- **All materials provided are the sole copyrights of University of Mumbai/IDOL.**
- **As course writers originality of the study material needs to be ensured and the material submitted will be self-certified by the course writers for Non-plagiarism.**
- **Citations, if taken may be given in references.**
  **Ms.Gouri Sonu Sawant (B.Sc. Information Technology) in IDOL, University Of Mumbai, is course co-ordinator.**



- **The remuneration is Rs.5000/- per Unit/Chapter.**
- **While submitting remuneration bill for the course writing an original copy of letter be attached.**

**You are requested to ensure that the study material reaches on or before 25th Jan 2021**

With warm regards,
Contact No's: 26527094/26527095

**DIRECTOR**
**IDOL**

---

NO. IDOL/SMU/05/8874 of 2021 Date: 19/01/ 2021.

Copy to:-

Assistant Registrar, (F&A) for information & necessary action.

19/01/2021

**DIRECTOR**
**IDOL**

# *Clinical Data Analysis in Healthcare Using Clustering Algorithms: A Review*

*Pradnya Bhambre† , Dr. Mrs. Nusrat Khan**

***ABSTRACT***

*In the healthcare domain, we are dealing with very large volumes of data. This data is not well organized and may not provide the actual scenario of the prevalence of a particular disease. To analyze this data, it is important to organize the large volume clinical data into meaningful information. Because of which Healthcare Professionals are facing many difficulties in analyzing various parameters of the disease with respect to the outbreak of that disease. Researchers use different tools and techniques of data mining to analyze the data and get the required information. Cluster analysis is an important unsupervised machine learning tool which discovers hidden patterns of data, gains new insights into structures of data, explores the correlation between different variables to extract useful information from huge data. This paper explains various applications of data analysis in the healthcare data and explores previously analyzed clinical data analysis work of different researchers which uses various clustering and classification techniques. Also, it explains advantages, disadvantages and applications of frequently used clustering techniques in the analysis of different disease datasets. This research also discovers the challenges faced in obtaining effective analysis of this clinical data to get the required knowledge. This work will be helpful to discover new criteria for analysis of disease datasets and to understand the possible implementation of these criteria by applying an appropriate clustering algorithm to these disease datasets. These analysis criteria may reveal the trends in biomedical, clinical and environmental factors in the disease1 datasets and may prove very useful in getting significant information which can be crucial and effective for prevention, treatment and cure of the diseases.*

*Keywords clustering algorithm, disease dataset, clinical data analysis, healthcare, clustering techniques*

## 1. Introduction

Data mining is a subfield of machine learning which uses unsupervised learning methods for exploratory data analysis. Data mining is the most important step in the knowledge discovery process. It consists of data pre- processing, post- processing and visualizing it to get the required information. To form the exact data analysis strategy in the knowledge discovery process (KDD), it is required to consider the exact difference between classification of data and clustering of data.

Cluster analysis is much better than usual classification techniques, in terms of providing the required information by analyzing data or to get new insights into the large volume of data. This paper focuses on the varied applications of clustering algorithms for analyzing different types of data in the healthcare domain. Therefore, literature of various applications of clustering algorithms in healthcare is reviewed and summarized.

## 2. Various applications of data analysis in healthcare

| *Application* | *Uses* |
|---|---|
| 1. Executives Information System[2] | ● *clinical decision making*[1]<br>● *higher quality services*[1,2]<br>● *more accurate and reliable decisions*[1]<br>● *early identification of high risk patients*[1]<br>● *avoid resubmission of insurance claims*[1]<br>● *reduce healthcare cost*[1,2]<br>● *disease prevention*[1]<br>● *reduce adverse drug effects*[2] |
| 2. Genetics | ● *the effects of genetics on different diseases can be investigated at micro-level*[1] |
| 3. Public Health Informatics[2] | ● *health policies and administration*[1]<br>● *E-governance in healthcare*[2]<br>● *identify the causes, trends and patterns of disease spreading in the population*[1] |
| 4. Forecasting Treatment[2] | ● *prediction of treatment cost*[2]<br>● *estimate demand for resources*[2]<br>● *patient's future behavior* |
| 5. Health insurance[2] | ● *detect insurance fraud*[1]<br>● *identify high-cost patients*[1]<br>● *prevent significant costs by early stage care*[1] |

**Table 1. Applications and uses of data analysis in healthcare**

## 3. Literature review of previously analyzed clinical data using clustering techniques

### I. Heart disease data

Heart Disease dataset from machine learning repository is analyzed by using various classification algorithms and the performance of KEEL tool is compared for time taken to build the model using two different distance functions ( Heterogeneous Value Difference Metric and Euclidean Distance) and three different pre- processing techniques (Generational Genetic Algorithm, Steady-state Genetic Algorithm, CHC Adaptive Search Algorithm) in three validation modes (5-Fold cross validation, K-Fold Validation and without validation). There are many factors which influence the performance analysis like the nature of the dataset, encompassing validation mode, distance function. This study demonstrates that Lazy Learning- k-nearest neighbor algorithm from artificial intelligence is efficient for prediction of the performance using without validation mode for the heart disease dataset. Advantage of lazy learning systems is that it can handle changes in the problem domain successfully and can solve multiple problems simultaneously. Lazy classifiers are most useful to analyze large datasets with fewer attributes. But it has some disadvantages also, like large space requirement, these methods are slower to evaluate.2

## II. Diabetics, thyroid and cancer data

There are some genetic diseases which are inherited from one generation to others due to the change in regular food habits and physical activities of the human beings. The most common of these hereditary diseases that stay for lifetime are diabetics, thyroid and cancer. Prediction of such hereditary diseases should be done at an early stage. Here, the advanced Prediction modeling is implemented in three phases. In the first phase, the issue is defined and data collection is completed. In the second phase a model is selected to perform training and testing. And in the third phase, the model is applied in the real-world. This is a crucial task to have immediate disease diagnosis in the medical field. Such automatic healthcare prediction systems could implement modern Artificial Intelligent technology to develop an easy way to identify the existence of the diseases. This research examines the diseases through some disease parameters and classified them using various intense classification algorithms such as Decision tree, Support Vector Machine, K-nearest neighbor, Logistic Regression, Naive Bayes etc. The proposed classification algorithms estimate the accurate prediction of the disease by measuring the diseases using the disease datasets. This experimental analysis has been carried out for three disease datasets which are Diabetics data set, Thyroid dataset and Cancer dataset. Classification techniques can be used to provide automatic predictions and quick treatment of patients. Here five different classification techniques were implemented to predict the diseases. These datasets were tested using classification algorithms in a Python environment. Support Vector Machine Classification Algorithm gives the best accuracy results as compared to other techniques. Following bar chart is obtained in analysis of this research data.

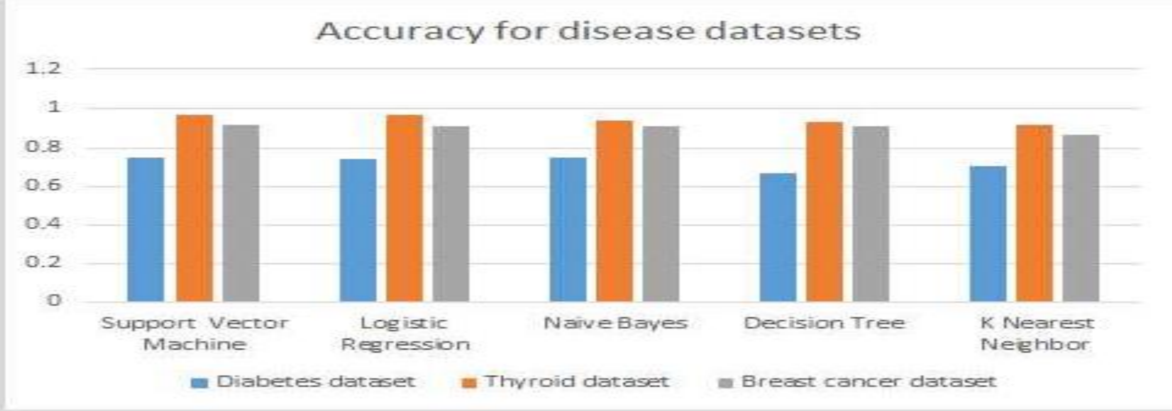


**Fig. 1. Accuracy of classification algorithms for different disease datasets**

## III. Liver and thyroid disorders Data

For liver and thyroid datasets K-means, Fuzzy c-means and Fuzzy Possibilistic c-Means clustering algorithms are analyzed to check the performance, clustering output efficiency as well as the percentage of correctness. The classification performance of FPCM is highest. K-Means and FCM have similar classification performance for the liver dataset. FCM and FPCM have similar and highest classification performance for the thyroid dataset.

Therefore, it is concluded that FPCM is the best in terms of high percentage of correctness and classification performance.

## IV. Alzheimer's disease data

The previous studies on clustering methods applied to Alzheimer's Disease dataset are summarized and concluded that Hierarchical agglomerative clustering algorithm was frequently used for this disease dataset followed by k-Means, multi-layer clustering and then k-Means-Mode. This shows the Hierarchical agglomerative clustering algorithm is more suitable and appropriate in the analysis and interpretation of useful results from Alzheimer's Disease dataset.

## V. Diabetic data

New clustering method for Clinical data to predict the likelihood of diabetic disease by combining k-means and k-mode algorithms and incorporating medical background knowledge is proposed. This method clusters numerical and categorical data efficiently and allows the user to specify constraints on selected attributes to participate in the clustering process. Following graph shows the clustering accuracy results. It is concluded that neither the medical BK nor hybrid-clustering algorithm performs very well, but combination of both produces excellent results.

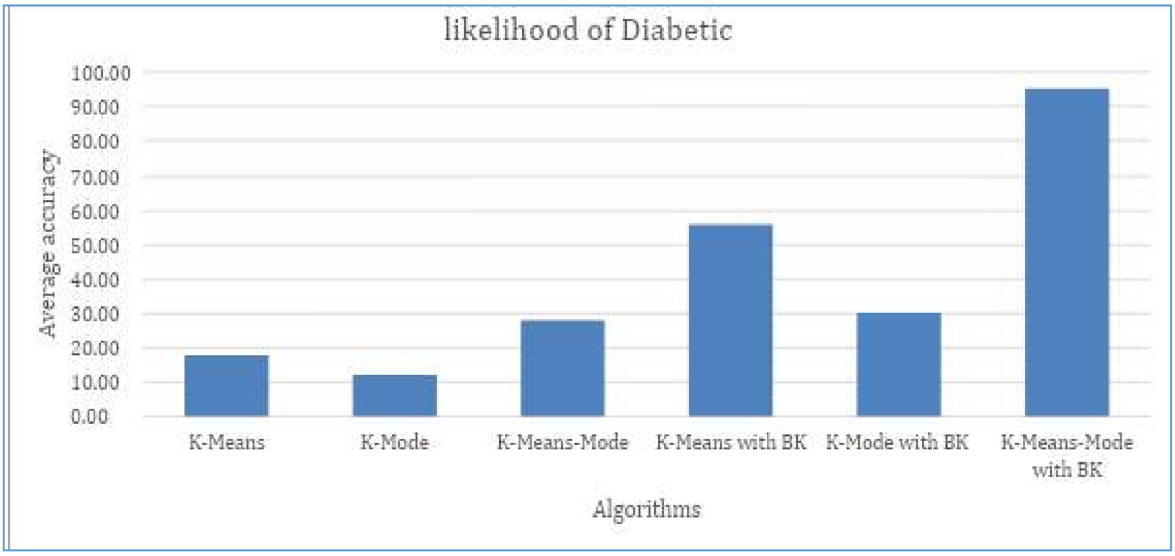


**Fig.2. Average accuracy of different algorithms to calculate Likelihood of diabetic data**

## VI. Hemogram blood test data

A new weight- based k-means clustering algorithm was developed for identifying the diseases namely inflammatory disease, leukemia, HIV infection, viral infection, and pernicious anaemia from the hemogram blood test samples data set. The performances of this algorithm are compared with K-means and fuzzy c means clustering algorithms in terms of execution time, clustering accuracy and error rate. The proposed algorithm provides the highest accuracy in less time.

## 4. Categories of clustering techniques

Following table illustrates advantages, disadvantages and applications of frequently used clustering techniques in the analysis of disease data.

| *Algorithm* | *Advantages* | *Disadvantages* | *Applications for disease data* |
|---|---|---|---|
| *1. Hierarchical Clustering* | *embedded flexibility, ease of handling of any kinds of similarity or distance, applicability to any attribute types* | *Vagueness of termination criteria* | *predict the severity of the Rheumatoid Arthritis* |
| *2. Partition Methods* | *simple clustering approach and efficient* | *number of clusters required in advance and could not discover the cluster with non-convex shape* | *grouping of persons according to high BP and cholesterol level into low risk and high risk of heart disease, detection of the recurrence of breast cancer* |
| *3. Density Based Clustering* | *no need to specify the number of clusters in advance, easily handle cluster with arbitrary shape* | *not handling the data points with varying densities and results depend on the distance measure* | *discovers the area of homogeneous color in biomedical images, separates the wound from healthy skin and discovers the sub regions of spotted part inside the unhealthy skin* |
| *4. K-Nearest-Neighbor* | *easy to implement* | *large database required, sensitive to noise, testing is slow* | *used with adaptive fuzzy K-NN approach for Parkinson disease* |

## 5. Challenges in effective analysis of clinical data

i. Data quality - Clinical data is usually collected from many different sources which could make the data dirty, more complex and different coding standards. But quality of data is important for achieving useful and reliable information using the data analysis techniques.

ii. Missing values - Most of the time clinical data contains missing values of some variables. So the results obtained from this incomplete data can be misleading. Even probability formulas are applied to find missing values, these results are not accurate and results may be incorrect.

iii. Mixed data types - Clinical data contains variables of mixed data types like numerical, character and categorical. Therefore, it is difficult to apply some formulas and algorithms.

iv. Extraction of comprehensible knowledge - Clinical data contains lots of variables and constraints. Therefore, it is very difficult to extract comprehensible knowledge from this data.

v. Variations in laboratory test parameters - There are differences in pathology test readings according to age, gender, health conditions. Even more, these readings can vary with time of the day, vary with underlying diseases, vary with days of the month for females due to menstruation, and may vary with time period in the populations. So, it is difficult to apply the same algorithm or analyzing techniques every time, they need to be updated.

vi. Sensitivity of diagnostic tests- Some lab tests have very sensitive methods of obtaining results. Therefore, proper care should be taken when implementing these tests otherwise because of minor errors also, wrong results are generated.

vii. Data sharing and privacy issues - Sharing of health data may prone the risk of threats to the privacy of patients. Even more, preparing a secure infrastructure for gathering data from different sources is expensive and time consuming.

viii. Relying on predictive models - In case of clinical decision- making models, it would be dangerous to completely rely on the predictive models only when making critical decisions because it may be vital to the life of the patient.

ix. Variety of methods and complex maths - Almost all data mining techniques involve somewhat complex mathematics, thus health administrators usually prefer to continue work with traditional methods and avoid implementing new algorithms and techniques.

## 6. Conclusion

This paper explores various applications of different clustering algorithms for the data analysis in the healthcare domain like decision support, forecasting, prediction and estimation.2 Many different disease datasets whose clinical data is analyzed using clustering algorithms are also enlisted here. Reviewing this research work, it is possible to implement various data analysis criteria with the required adaptations, in any particular disease dataset for the data analysis in the onset and prevalence of this disease. This data analysis can provide effective information about details, specifics and even minorities 1 of different parameters of the disease under study. Furthermore, these data analysis solutions obtained can be modified at any required stage by just making simple modifications in the clustering algorithms used. Therefore, by considering required accuracy and specifications in the required solution data, an appropriate clustering algorithm can be selected, tailored as per requirement, and applied to the particular disease dataset to effectively analyze it to provide more exact and quick information. The aim of this review paper is to encourage new researchers, scientists in the healthcare domain to investigate new views and approaches in the analysis of disease data under study. This would be proven beneficial for enhancing the quality of research work by discovering new aspects in the data analysis of the clinical data.

## References

1. M. H. Tekieh and B. Raahemi, Importance of Data Mining in Healthcare: A Survey, Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, pp. 1057-1062, doi: 10.1145/2808797.2809367

2. D. Kaur and A. Paul, Performance Analysis of Different Data mining Techniques over Heart Disease dataset, International Journal of Current Engineering and Technology, 2014, 4(1), http://inpressco.com/category/ijcet

3. E. L. Lydia, N. Sharmil, K. Shankar and A. Maseleno, Analysing the Performance of Classification Algorithms on Diseases Datasets, International Journal on Emerging Technologies, 2019, 10(3), pp. 224-230.

4. B. Venkataramana, L. Padmasree, M. S. Rao, G. Ganesan and K. R. Krishna, Implementation of Clustering Algorithms for real datasets in Medical Diagnostics using MATLAB. Journal of Soft Computing and Applications, 2017(1), pp. 53-66.

5. H. Alashwal, M. El Halaby, J. J. Crouse, A. Abdalla and A. A. Moustafa, The Application of Unsupervised Clustering Methods to Alzheimer’sDisease, Front. Comput. Neurosci., Vol. 13:31. doi: 10.3389/fncom.2019.00031

6. R. Paul and Abu Sayed Md. Latiful Hoque, Clustering Medical Data to Predict the Likelihood of Diseases, Fifth International Conference on Digital Information Management: ICDIM 2010: July 5-8, 2010, Lakehead University, Thunder Bay, Canada. doi: 978-1-4244-7571-1

7. S. Vijayarani and S. Sudha, An efficient clustering algorithm for predicting diseases from hemogram blood test samples, Indian Journal of Science and Technology, 2015, 8(17). doi: 10.17485/ijst/2015/v8i17/52123

8. N.S. Nithya, K. Duraiswamy and P. Gomathy, A Survey on Clustering Techniques in Medical Diagnosis. International Journal of Computer Science Trends and Technology (IJCST) –2013, 1(2), available at www.ijcstjournal.org

# Prevalence and Awareness of Thyroid Disorders in Pune and Pimpri- Chinchwad

*Pradnya Bhambre*[1] *Dr. Mrs. Nusrat Khan*[2]

## ABSTRACT

Thyroid disorders are the most common endocrine disorders worldwide. In India the population suffering from these disorders is increasing day by day. Thus, there is a need to create awareness about these disorders and help people to take care of themselves, properly consume the medicine, be familiar with the long- term consequences of these disorders and aid in regaining a normal healthy life. There are several factors which may cause outbreak of these disorders including heredity, age, gender and other biological and/or environmental factors. Therefore, a survey has been conducted by authors to understand the prevalence and awareness of thyroid disorders in Pune and Pimpri-Chinchwad geographical region. And the data collected is analyzed to interpret the results with respect to different parameters like heredity, age and gender and health status of thyroid patients etc.

**Keywords** thyroid disorders, hypothyroid, hyperthyroid, endocrine disorder, health status

## I Introduction

There are many diseases like diabetes, cancer etc. which are influencing population more than the past few years. This is because of lifestyle changes, changes in food habits, changes in food quality, sedentary lifestyle, exposure to chemicals and pollution etc.

Along with these diseases, thyroid disorders are also influencing the population to a greater extent from last few years. So, there is a need to aware people about this increasing hazard. A survey has been conducted by the authors of this paper in Pune and Pimpri Chinchwad and responses from 43 thyroid patients were collected. Then by analyzing this data, prevalence of thyroid disorders is evaluated in terms of age and gender.

For this a vast literature survey has been conducted to check different parameters of the thyroid disorders. Accordingly, the questionnaire had been formulated and responses had been collected. Information is collected related to thyroid patient's age group, gender, type of thyroid disorders, genetic factor, medicine, vitamin/ minerals deficiency, thyroid antibodies etc. The categories of age groups considered here are according to different stages in human life span as Childhood - (0-12 years), Adolescence - (13-18 years), Adulthood - (19-59 years) and Elderly - (60 years and above). Using these responses, the data is analyzed and results are interpreted.

## II Literature Study

There is a significant burden of Thyroid Disorders in India. It has been estimated from various studies that around 42 million people were suffering from thyroid disorders in India. Common thyroid disorders in India are: (1) hypothyroidism (2) hyperthyroidism (3) goiter or iodine deficiency disorders (4) Hashimoto's thyroiditis (5) thyroid cancer. [1]

Population studies have interpreted that about 16.7% of thyroid patients possess anti-thyroid peroxidase (TPO or AMA) antibodies and about 12.1% have anti-thyroglobulin (ATG) antibodies. [1]

Autoimmune Thyroid Disorders are complex diseases which may be caused by the combined effects of genetic factors and/or environmental triggers. The interaction

1. Research Scholar, Savitribai Phule Pune University, Pune

2. Associate Professor, Sinhgad Institutes, Pune

between susceptibility genes and environmental factors results in the breakdown of self-tolerance ability and may lead to Autoimmune Thyroid Disorders. [2]

Even if the value of TSH is within limits, women with hypothyroidism represent a poorer quality of life as compared to the women without hypothyroidism. Therefore, there is a need to assess the Health- Related Quality of Life of women with Hypothyroid disorders. [3]

Quality of Life was impaired in patients receiving Levothyroxine treatment, irrespective of the hormonal status. Therefore, management of comorbid diseases and patients' health status should be taken into consideration to achieve an optimal treatment. So, the integration of health- related quality of life assessment is highly recommended in primary health care systems. [4]

It can be suggested from a population- based study that Health Related Quality of Life scores of patients with suppressed TSH values or markedly elevated TSH values were generally not significantly lower than those of patients with normal or slightly elevated TSH values. [5]

Health Related Quality of Life is impaired in patients with thyroid disorders, both in the untreated patients and patients suffering from these disorders in the long- term. [6]

Vit-B12 and vit-D deficiency are associated with autoimmune hypothyroid disorders, and there is a negative correlation between vit-B12 and vit-D levels and anti-TPO antibodies in these patients. In patients with autoimmune hypothyroid disorders, vit-D and vit-B12 deficiency should be evaluated during the treatment and periodic follow-ups. [7]

## III Research Methodology

An explorative study has been conducted by authors to understand the Prevalence of Thyroid Disorders in Pune and Pimpri- Chinchwad geographical region. Therefore, Thyroid patients from Pune and Pimpri- Chinchwad were considered as respondents for this study. Survey method with random sampling is used to collect the primary data. Primary data has been collected from 43 valid respondents out of total 47 respondents. Data is collected by using Questionnaire Technique. This questionnaire is designed by considering many factors like gender, age group, type of thyroid disorder, dosage of medicine, suffering from how many years, regularity of medicine, periodic checking TSH level, regular exercise etc. then it is extended to check the awareness of vitamins and minerals deficiencies and also about thyroid antibodies. Samples are selected from Pune and Pimpri- Chinchwad area is as follows:

**Table 1: Selected sample in Pune and Pimpri-Chinchwad area**

| Area | Type of thyroid patient | Sample Number | Percentage |
|---|---|---|---|
| Pune | Hypothyroid Patients | 13 | 30.23% |
| | Hyperthyroid Patients | 4 | 9.30% |
| Pimpri-Chinchwad | Hypothyroid Patients | 23 | 53.48% |
| | Hyperthyroid Patients | 3 | 6.97% |

**Objective of the Study:** To understand the prevalence of Thyroid Disorders and awareness about it in the people.

**Sub - objectives:**

1. To understand the prevalence of different types of Thyroid Disorders according to different age groups.
2. To understand the prevalence of different types of Thyroid Disorders according to gender.
3. To check awareness in people about Thyroid Disorders related information and patient’s health status.

## IV Data analysis

Data is collected through questionnaires by interacting with Thyroid Patients in Pune and Pimpri- Chinchwad. This primary data is analyzed to understand the prevalence of different types of Thyroid Disorders according to different age groups and gender. Forty- three valid responses were received from thyroid patients of all age groups. Results are summarized as follows:

- The bar graph Fig. 1 shows that, prevalence of Hypothyroid Disorders is maximum i.e. almost 44% is in the adulthood i.e. age group of 19 to 59 years. Whereas in the childhood i.e. age group of 0 to 12 years the prevalence is minimum i.e. approximately 9%.
- The bar graph Fig. 2 represents that, Hypothyroid Disorders with respect to gender are more prevalent in Females i.e. 33% whereas in males only 28%. But Hyperthyroid Disorders are more prevalent in Males i.e. approximately 5% and in Females it is approximately 2%.

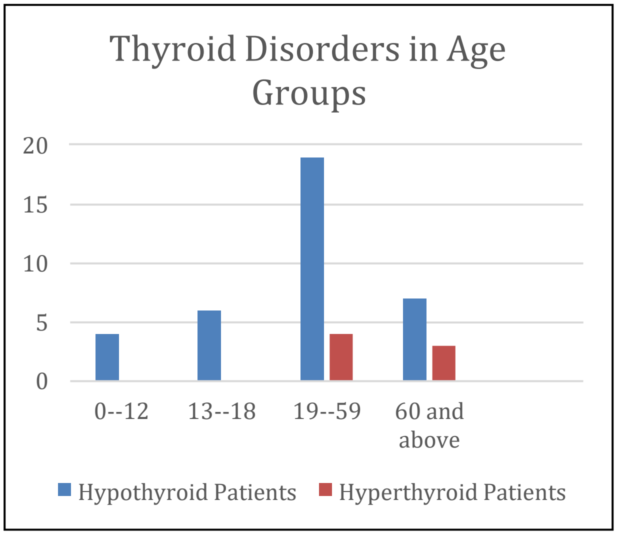


**Fig. 1 Prevalence of Thyroid Disorders in different age groups**

- The presence of Heredity Factor is evaluated as 39.5% approximately.
- The thyroid patients continuing treatment are 83% whereas the remaining almost 17% had discontinued treatment regimen.

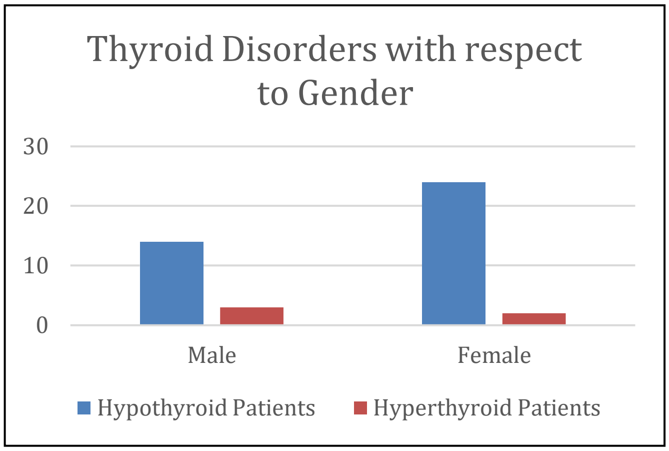


**Fig. 2 Prevalence of Thyroid Disorders with respect to Gender**

- The thyroid patients who know the medicine and dosage they are consuming are around 67% and the remaining 33% patients even don't know the name and dosage of the medicine they are consuming.
- Approximately 72% of patients test their TSH regularly while remaining 28% do not bother about it.
- Only 53% patients are exercising regularly while the remaining 47% are not exercising because of either time constraint, willingness, muscle pain or they don't know that exercise is must for thyroid disease management.
- Around only 21% thyroid patients are aware of the fact that they have some vitamins or minerals deficiencies. Even though 37% of patients are consuming vitamins and minerals supplements.
- Only 13% thyroid patients possess some knowledge about thyroid antibodies while others are not.

## V Results

- Prevalence of thyroid disorders is maximum in the adulthood i.e. age group of 19 to 59 years and Hypothyroidism is more prevalent in this age group. In the childhood i.e. age group of 0 to 12 years the prevalence is minimum. This implication directs towards the fact that prevalence of thyroid disorders is increasing with age. Therefore, people should have been more conscious about thyroid function in adulthood.

- Hypothyroid disorders are more prevalent in Females, but Hyperthyroid disorders are more prevalent in Males.
- Most of the Thyroid patients have very little information about the disease like they don't know name of the medicine, dosage, importance of continuing treatment, necessity of checking TSH regularly and also do not aware of the health problems associated with these disorders.

## VI Conclusion

In India, Thyroid Disorders are on the rise. [1, 8] Approximately, 1 in 10 adults suffer from hypothyroidism. In Pune, 17.85 per cent prevalence of Thyroid Disorders is evaluated in a study conducted in 2016. [8] But Thyroid Patients in Pune and Pimpri-Chinchwad have very little awareness about different types of Thyroid Disorders. There are many health and psychological problems that are associated with Thyroid disorders. [3, 4, 5, 6] Patient should be aware of these problems and should know how to cope with them. Healthcare Professionals should consider these problems when treating the patient for optimizing the treatment and enhancing the health- related quality of life of the patient. [3, 4] Following these changes, Thyroid Patients can be a step ahead in improving and maintaining their health status.

## References

*[1]A. G. Unnikrishnan and U. V. Menon, "Thyroid disorders in India: An epidemiological perspective", Indian Journal of Endocrinology and Metabolism, vol. 15, suppl. no. 2, p. S78- S81, 2011. Available: http://www.ijem.in [Accessed: March 24, 2021].*

*[2]Y. Tomar, "Genetic susceptibility to autoimmune thyroid disease: past, present, and future", Thyroid: official journal of the American Thyroid Association, vol. 20, no. 7, p. 715-725, May, 2010. DOI: 10.1089/thy.2010.1644*

*[3]B. Romero-Gómez, P. Guerrero-Alonso, J. Carmona-Torres, D. Pozuelo-Carrascosa, J. Laredo-Aguilera, A. Cobo-Cuenca, "Health-Related Quality of Life in Levothyroxine-Treated Hypothyroid Women and Women without Hypothyroidism: A Case–Control Study", journal of Clinical Medicine, vol. 9, no. 12, 3864, 2020. DOI: 10.3390/jcm9123864*

*[4]T. Al Quran, Z. Bataineh, A. Al-Mistarehi, A. Okour, O. Beni Yonis, A. Khassawneh, R. AbuAwwad, A. Al qura'an, "Quality of life among patients on levothyroxine: A cross-sectional study", JOURNAL ARTICLE, Annals of Medicine and Surgery 60, p. 182–187, October, 2020. DOI: 10.1016/j.amsu.2020.10.030*

*[5]E. Klaver, H. Van Loon, R. Stienstra, T. Links, J. Keers, I. Kema, A. Kobold, M. Van Der Klauw, B. Wolffenbuttel, "Thyroid hormone status and health-related quality of life in the lifeLines cohort study", Thyroid, vol. 23, no. 9, p. 1066-1073, 2013. DOI: 10.1089/thy.2013.0017*

*[6]T. Watt, M. Groenvold, A. Rasmussen, S. Bonnema, L. Hegedüs, J. Bjorner, U. Feldt-Rasmussen, "Quality of life in patients with benign thyroid disorders. A review", European Journal of Endocrinology, vol. 154, no. 4, p. 501-510, 2006. DOI: 10.1530/eje.1.02124*

*[7]H. Aktaş, "Vitamin B12 and Vitamin D Levels in Patients with Autoimmune Hypothyroidism and Their Correlation with Anti-Thyroid Peroxidase Antibodies", Medical principles and practice: international journal of the Kuwait University, Health Science Centre, vol. 29, no. 4, p. 364-370, 2020. DOI: 10.1159/000505094*

*[8]THE WEEK MAGAZINE, "Thyroid Disorders on the rise in India", by Pooja Biraia Jaiswal, July 23, 2019 22:32 IST Accessed at https://www.theweek.in/news/health/2019/07/23/thyroid-disorders-rise-india.html*

Sinhgad Institutes

VALENCE

AICTE Sponsored
[Ref. No. 34-67 I 1.L5 / FDC / FDP / P -t I 2019 -20]
Faculty Development Programme (FDP): Online Mode

On

# MACHINE LEARNING, DATA SCIENCE & DEEP LEARNING WITH PYTHON

(1st July – 16th July, 2021)

Organized By
Sinhgad Institute of Management, Pune-41
(Platinum Category Institute by AICTE-CII Survey 4th Consecutive year)
(Accredited by National Assessment and Accreditation Council)
(Affiliated to Savitribai Phule Pune University and Approved by AICTE)
Venue
Thin Client Lab/T1-Lab-Online Mode
Vadgaon (Bk.), Pune - 411041
Tel. 020-24358360
Website: www.sinhgad.edu

Sinhgad Institute of Management
PUNE
411 041.

AICTE Sponsored

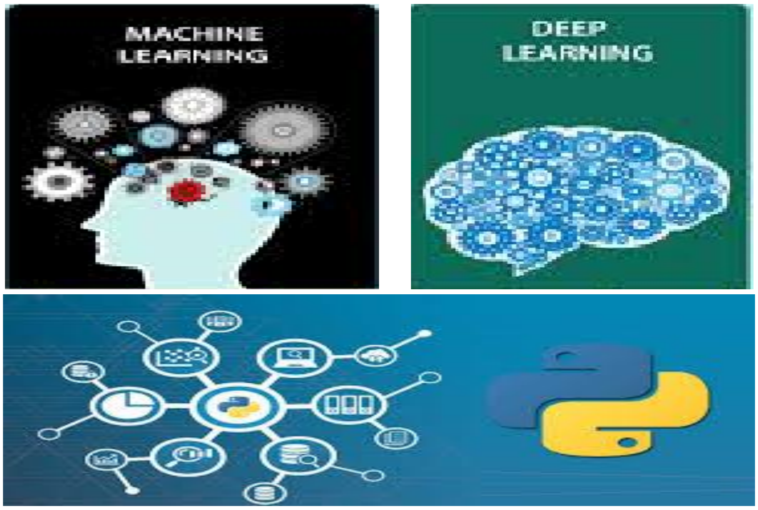
[Ref. No. 34-67 I 1.L5 / FDC / FDP / P -t I 2019 -20]

Faculty Development Programme- Online Mode

On

# MACHINE LEARNING, DATA SCIENCE & DEEP LEARNING WITH PYTHON

(1st July – 16th July, 2021)



Organized By

Sinhgad Institute of Management, Pune-41

(Platinum Category Institute by AICTE-CII Survey 4th Consecutive year)

(Accredited by National Assessment and Accreditation Council)

(Affiliated to Savitribai Phule Pune University and Approved by AICTE)

Venue

Thin Client Lab/T1-Lab-Online Mode

Vadgaon (Bk.), Pune - 411041

Tel. 020-24358360

**Website:** www.sinhgad.edu

All the printing & distribution process of the book is powered by

**24by7 Publishing**
13 New Road, Kolkata - 51, India
https://www.24by7Publishing.com
mail@24by7publishing.com
+91 9831 470 133
+91 9433 444 334

First Published in August, 2021

Version 1.00

**ISBN: 978-93-90979-11-0**

Powered by

**24**by**7**Publishing.com

**Disclaimer**

Institute, AICTE and the editorial team and publishing agency will not take any responsibility for any mistake/error incorporated in the articles of aauthors, either in settings, in grammatical way, in formatting according to the contents given, in printing especially images or in any contents of topics. Opinions expressed and figures provided in the Proceedings' MACHINE LEARNING, DATA SCIENCE & DEEP LEARNING WITH PYTHON are the sole responsibility of the authors.

**Copyright**

**Printed and published by**
Sinhgad Technical Education Society's, Sinhgad Institute of Management
S. No. 44/1, Vadgaon (Bk), Off Sinhgad Road, Pune-411041, www.sinhgad.edu

**Editorial Team**
Dr. Chandrani Singh, Prof.Monalisa Bhinge, Dr Vidya Gavekar, Dr. Milind Godase, Dr. Sunil Khilari

**Compilation team**
Kumudini Manwar, Rahul Navale, Rahul Dwivedi, Ankush Kudale, Rahul Borate and
Dr. Manisha Kumbhar

**Resource persons**

- Dr. Amlan Chakrabarti, Professor & Director, A.K. Choudhury School of IT, University of Calcutta
- Dr. Jyoti Gautam, Associate Professor, JSS Academy of Technical Education, Noida
- Ms. Nitima Malsa, Assistant Professor, JSS Academy of Technical Education, Noida
- Mr. Anchal Koshta, CEO & Founder, ABCROB Technologies Pvt. Ltd
- Ms. Priyanka Bhalere, Senior Analyst, Accenture
- Mr. Vivek Nikam, Data Scientist, CTO, Code Nautics Technologies
- Mr. Satheesh Kumar, Ex. Principal Consultant, Wipro Ltd, Bangalore
- Dr. Shailesh Kasande , CEO & Group Director, Suryadatta Group of Institutes
- Dr. Chandrani Singh , Director-MCA, Sinhgad Institute of Management

# Founder President Message, STES

STES' Sinhgad Institute of Management welcomes you all for the Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python" to be held in online mode in the first fortnight of July 2021. The theme of the Faculty Development Program centres around the latest topics for learning and research in the field of AI & Machine Learning, Data Science, Big Data Management and Analytics', Science of Cloud Data and Computing.The FDP promises of outstanding resources and tools those which will help render knowledge and in-depth insights.

The FDP will be a platform boasting of eminent educationists, entrepreneurs and technology experts across the globe. No one will leave without learning best practices, making new contacts and creating strong bonds of support. This FDP is being hosted by STES's SIOM Vadgaon, and is being sponsored by AICTE.

I look forward to welcome you and to be a part of productive and fulfilling series of sessions. I am sure the aspirants will gain real knowledge from informative skill-building sessions and hands on training that would be imparted through online mode.

I hope that all participants will enjoy and leave with pleasant memories at the end of the program.

All the Best!!!

**Prof. M. N. Navale**

**Founder President, STES**

# Message from Founder Secretary, STES

Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python" is an amalgamation of IT and Management tracks where all the aspirant learners will come together and discuss and exchange knowledge on various subjects that have immense market value in today's era. It is noteworthy that, academicians, technologists and IT professionals all over the globe are coming to be facilitated through the course. This FDP provides a platform to impart knowledge in multidisciplinary fields such as AI and Machine Learning, Data Science and Deep Learning.

Sinhgad Institutes have always worked from academic and research perspective in diverse fields of IT and Management which helps in culminating the academic brilliance across all the dimensions.

I express my gratitude to AICTE for being the main sponsor for the Faculty Development Program in this important knowledge sharing endeavor.

I extend my warm wishes to all the participants and faculty members who worked hard to make this FDP a grand success.

**Dr. (Mrs.) Sunanda M. Navale**

**Founder Secretary, STES**

# Message from Vice President (HR), STES

Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python" is a true reflection of progress that Sinhgad Institutes is making by aligning to current trends. The FDP is meant for faculty researchers, practitioners, management consultants, students, industry leaders and other experts to improvise on their analytical vision through the learning and suggest measures for meeting the evolving challenges.

The exchange will hopefully benefit the aspiring learners in the relevant field. STES' Sinhgad Institutes is one of the largest educational conglomerates in western India with a vision towards creating excellence in all the spheres.

Sinhgad Institutes have always endeavored towards providing quality education along with overall development of the individuals through Faculty Development Programs, Research Conferences and many other value-added programs.

I congratulate everyone for attending this FDP and derive concrete outcomes!

**Mr. Rohit M. Navale**

**Vice President (HR), STES**

# Message from Vice President (Admin.), STES

Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python" is a medium and stage for people from relevant fields to interact, discuss and deliberate on the recent trends in the above disciplines and to share their expertise and insights. The new revolution will harness mental and cognitive ability brought in by Artificial Intelligence and its associated branches and it is imperative for professionals to up-skill and re-skill themselves.

Sinhgad Institutes is a national leader in imparting education and is recognized globally. It is an organization which is raising its global prominence through research and other allied activities and ensures that it stays ahead of time.

To enhance its position, it has made an agreement with several international research bodies for strengthening its research base and embracing diversity.

I wish you all the very best to showcase your talents and acquire tremendous knowledge from this Faculty Development Program.

**Mrs. Rachana Navale Ashtekar**

**Vice President (Admin.), STES**

# Message from Sr. Director, Management Programme, STES

Dear All,

Sinhgad Institutes have been instrumental in imparting Quality Education with the help of its state-of-the-art infrastructure and dedicated faculty members. Our success in the field of Research Projects, Industry tie-ups, Scholarships and Placements are well known in the academic circles. I believe there is always scope for improvement in whatever we do.

As part of this initiative, a Faculty Development Programme (FDP) on "Data Science, Machine Learning and Deep learning using Python" is being conducted for faculty, researchers, post graduate scholars in academia and industry. The purpose of conducting this 15 days' workshop is to provide a learning and hands-on-experience on applications using AI & ML techniques. Academicians and Industry experts are participating in a major way to deliver sessions that promote experiential learning and concrete course outcomes. I extend my support and best wishes to all the participants and hope that through this faculty development program there is adequate knowledge transfer and productive exchange in alignment to the current market trends.

I welcome all the participants for the Faculty Development Program and wish them Happy Learning.

**Mr. G.K. Shahani,**
**Sr. Director,**
**Management Programme, STES**

# Message from Director, STES Vadgaon (BK)

Dear Participants,

The Faculty Development Programme on 'Data Science, Machine Learning and Deep Learning using Python' will help to disseminate the knowledge in the domain of data science and Machine learning. It empowers the participants to understand how data science can be used to innovate and improve the business processes. Machine Learning is a fast-growing field of Artificial Intelligence concerned with the study and design of computer algorithms for learning good representations of data, at multiple levels of abstraction. Since data is overwhelming, organizations are struggling to extract the powerful insights they need to make smarter business decisions. The participants will be trained using hands-on approach in order to have an in-depth insight into the domain of Data Sciences and expose them to the future scope.

Wishing All the participants Happy Learning.

**Dr. A. V. Deshpande**
**Director,**
**STES Vadgaon (BK), Pune, India**

# Message from Director, SIOM

It is my great pleasure to welcome you all to the Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python" which is taking place in the beautiful city of Pune in Maharashtra, India in an online mode from 1 July -16 July 2021. The FDP aims to bring together the professionals, industrialists, researchers, to discuss and learn the latest advents in the area of Data science and Analytics.

The data science related areas such as machine learning, data analytics, AI, Cloud computing are trending and provide many new opportunities perspective to the aspirants. In the age where educational institutes are under growing pressure to reduce costs and increase efficiency/productivity analytics promises to be the important lens through which to view and plan for the change at the institution level.

Wishing you "All the Best" for the FDP.

**Dr. Daniel Penkar**
**Director, SIOM**

# Message from Director, SIOM-MCA

Dear Faculty Members,

I take great pride in welcoming all the attendees for the Faculty Development Program on "Machine Learning, Data Science and Deep Learning with Python". This FDP aims to provide an open forum for discussions on the themes as AI and Machine Learning, Deep Learning, Data Science, RPA and IoT.

It gives opportunity to exchange the knowledge between the academics and various industry experts in the said field. The objective of this FDP is also to nurture research under the vertical of Data Science. Data Analytics has proved, since its beginnings, its importance in creating new information and in revealing hidden insights across different domains. This FDP provides a national forum for academia and industry to exchange and share their experiences, research results, and new ideas on hot and emerging topics such as data science and data analytics.

On behalf of Sinhgad Institute of Management, I welcome you, and wish that the FDP could enrich you in the Data Science domain and associated branches.

**Dr. Chandrani Singh**
**Director, SIOM-MCA**

# PREFACE

**Dear All,**

The Organizing Committee warmly welcomes our distinguished delegates and guests to the Faculty Development Program on “Machine Learning, Data Science and Deep Learning with Python” to be held from 1st July - 16th July 2021 in Pune, Maharashtra, India. The FDP is sponsored by AICTE. This FDP aims on discussing present and future technologies in data science, machine learning and deep learning.

This FDP is organized for creating avenues in learning through collaborations in the Data Science vertical and associated disciplines with faculty members and industry professionals from around the nation so that they can learn the leading-edge technologies, thereby expanding community’s knowledge and insights for undertaking significant challenges that are currently being addressed. This proceeding comprises of contents prepared by the speakers and the faculty members. The FDP themes around Machine Learning, Data Science, Deep learning, RPA, Industrial IOT, and Strategic Mmanagement and Team building. The main goal of this event is to create a national scientific forum for exchange of new ideas in number of fields, interact through discussions with peers and constructive collaboration.

All the submitted contents in the proceedings have been reviewed by the editorial board depending on the subject matter. Reviewing and initial selection were undertaken electronically. After the rigorous review process, the submitted contents were selected on the basis of originality, significance, and clarity for the purpose of the FDP. The FDP is extremely rich, featuring high-impact presentations. We are grateful to all those who have contributed to the success of this FDP.

We hope that all participants and other interested readers benefit from the proceedings and FDP.

With best wishes from

**Editorial Board FDP 2021**

# Table of Contents

| Sr. No | Contents | Authors |
|---|---|---|
| 1 | Machine Learning | Dr. Vidya Gavekar, Prof. Rahul Navale |
| 2 | Deep Learning | Mr. Vivek Nikam, Prof. Monalisa Bhinge, Prof. Ankush Kudale |
| 3 | Data Science | Dr. Amlan Chakrabarti, Dr. Jyoti Gautam, Prof. Nitima Malsa, Prof. Rahul Borate |
| 4 | RPA | Ms. Priyanka Bhalere, Dr Sunil Khilari, Mr. Shripad Kulkarni, Prof. Rahul Dwivedi |
| 5 | Internet of Things (IoT) | Mr. Anchal Koshta, Dr. Milind Godase |
| 6 | Blended Learning | Dr. Chandrani Singh, Prof. Archana Nair<br>Dr. Manisha Kumbhar |
| 7 | National Education Policy (NEP-2020) | Dr. Shailesh Kasande, Dr. Chandrani Singh |
| 8 | Python Assignments | Prof. Nitima Malsa |

# Machine learning

Dr. Vidya Gavekar

Prof. Rahul Navale

## 1. Introduction:

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more.

Machine learning is a kind of artificial intelligence (AI) that provides computers the ability to understand and learn without any need for explicit programming. Machine learning comprises of algorithms to be trained using a given set of data, and utilize this training to predict the characteristics of any given data. The primary focus of Machine learning is on the development of computer pprograms that tend to change when exposed to any new data.

Machine Learning is a computer algorithm that is able to adjust its own internal parameters using sample data, in order to be able to estimate/predict something useful for similar data.

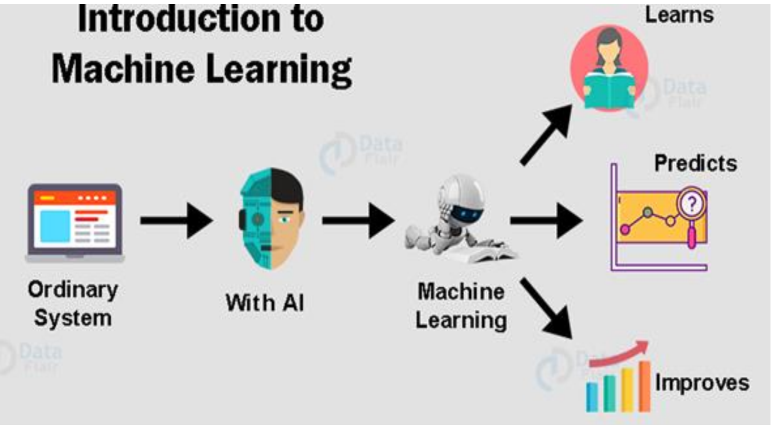


**Fig 1.1: Process of Machine Learning**

## 2. Need and Importance of Machine Learning:

Machine Learning today has all the attention it needs. Machine Learning can automate many tasks, especially the ones that only humans can perform with their innate intelligence. Replicating this intelligence can be achieved only with the help of machine learning.

With the help of Machine Learning, businesses can automate routine tasks. It also helps in automating and quickly create models for data analysis. Various industries depend on vast quantities of data to optimize their operations and make intelligent decisions. Machine Learning helps in creating models that can process and analyze large amounts of complex data to deliver accurate results. These models are precise and scalable and function with less turnaround time. By building such precise Machine Learning models, businesses can leverage profitable opportunities and avoid unknown risks that are being witnessed on account of fourth industrial revolution.

In order to derive meaningful insights from this data and learn from the way in which people and the system interface with the data, we need computational algorithms that can churn the data and provide us with results that would benefit us in various ways.

Examples of Machine Learning are as follows: Search Engines as Google that are able to provide with appropriate search results based on browsing habits. Similarly, Netflix is capable of recommending the films or shows that one would want to watch based on the machine learning algorithms that perform predictions based on the history.

Furthermore, machine learning has facilitated the automation of redundant tasks that have taken away the need for manual labor. All of this is possible due to the massive amount of data that one can generate on a daily basis.

Machine Learning facilitates several methodologies to make sense of this data and provide an individual with steadfast and accurate results.

Image recognition, text generation, and many other use-cases are finding applications in the real world. This is increasing the scope for machine learning experts to shine as a much sought-after professional.

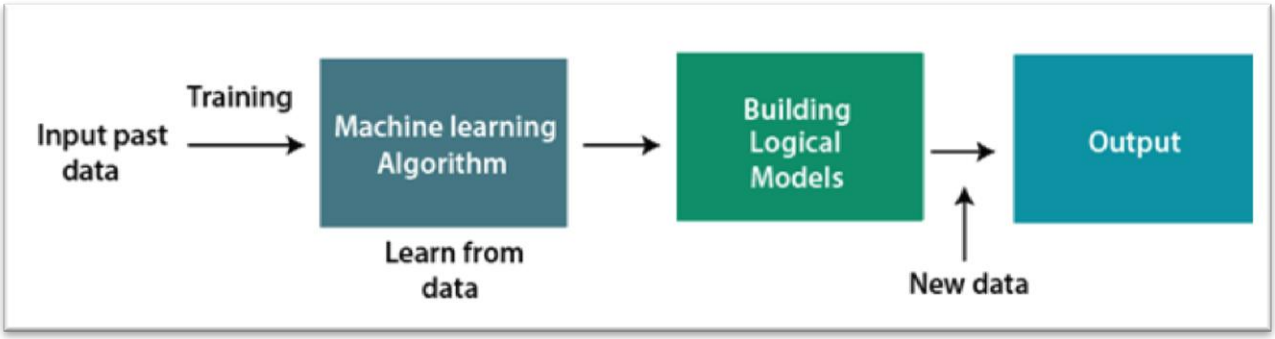


**Fig 1.2: Working of Machine Learning Algorithm**

## 3. Working of Machine Learning:

A machine learning model learns from the historical data fed to it and then builds prediction algorithms to predict the output for the new set of data that comes in as input to the system. The accuracy of these models depends on the quality and amount of input data. A large amount of data helps build a better model which predicts the output more accurately.

Suppose there is a complex problem at hand that requires to perform certain predictions. Now, instead of writing a code, this problem could be solved by feeding the given data to generic machine learning algorithms. With the help of these algorithms, the machines develop the logic and predict the output. Machine learning has transformed the way one approaches the business and social problems. Below is a diagram that briefly explains the working of a machine learning model/ algorithm. The below block diagram explains the working of Machine Learning algorithm:

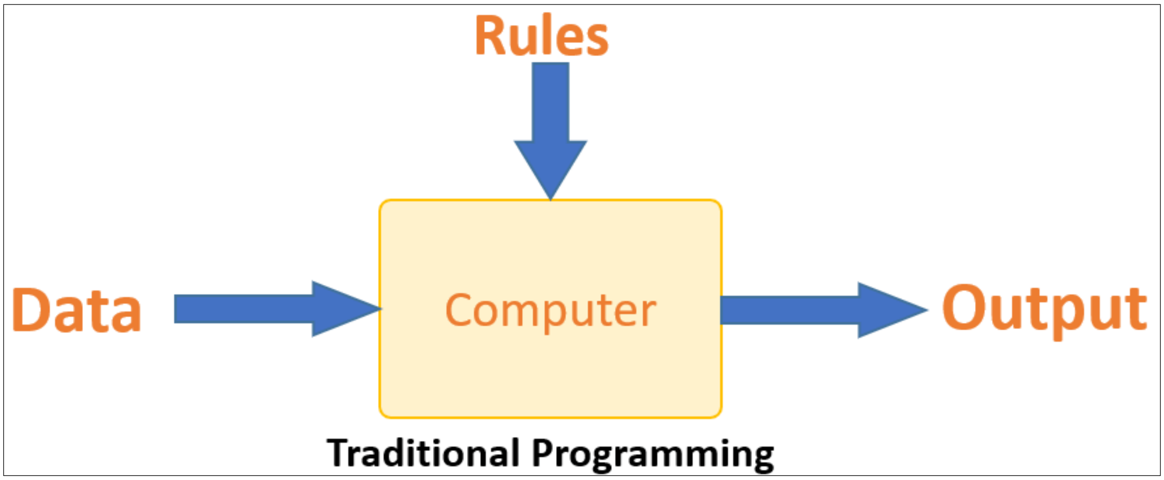


**Fig.1.3: Logical Process**

## 4. Comparison between Machine Learning and Traditional Programming:

Traditional programming differs significantly from machine learning. In traditional programming, a programmer codes all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time when there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.

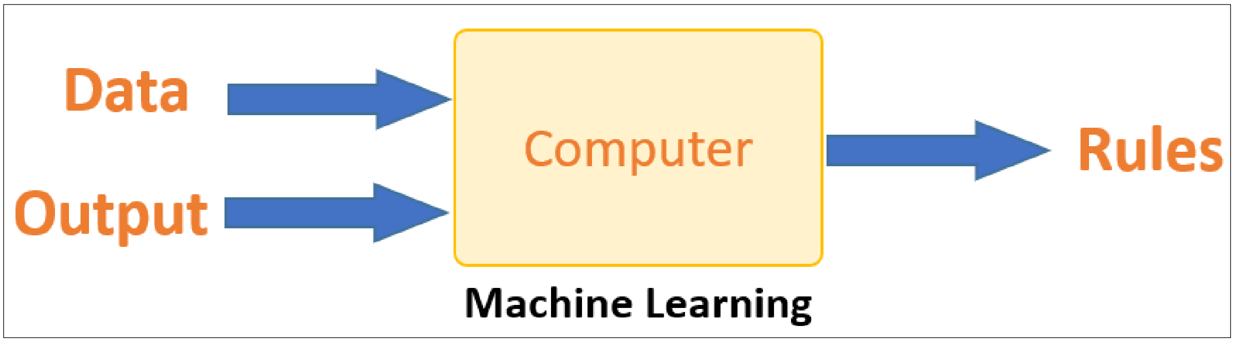


**Fig 1.4: Processing of input and output**

## 5. Steps of Machine Learning Algorithm:

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more is known, the more easily it can be predicted. By analogy, when an unknown situation is faced, the likelihood of success is lower than the known situation. Machines are trained on the same. To make an accurate prediction, the machine sees an example when the machine is given a similar example, it can figure out the outcome. However, like a human, if its feed is a previously unseen example, the machine has difficulties to predict.

The core objective of machine learning is the learning and inference. First of all, the machine learns through the discovery of patterns. One crucial aspect is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a feature vector. One can think of a feature vector as a subset of data that is used to tackle a problem.

The machine uses some visualization algorithms to simplify the reality and transform this discovery into a model. The learning stage is used to describe the data and summarize it into a model.

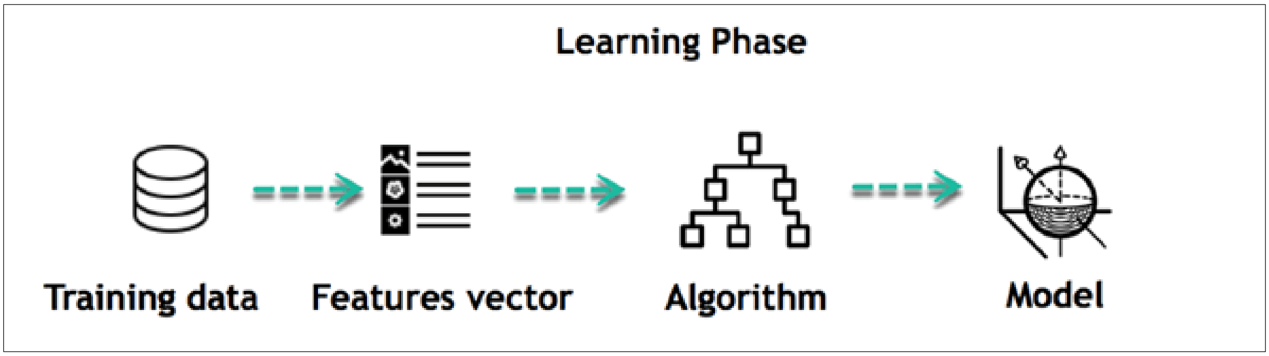


**Fig 1.5: Learning Phase Model**

For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant:

Inference

When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a feature vector, subjected to the model and is generated prediction is there is no need to update the rules or train again the model. One can use the model previously trained to make inference on the new data.

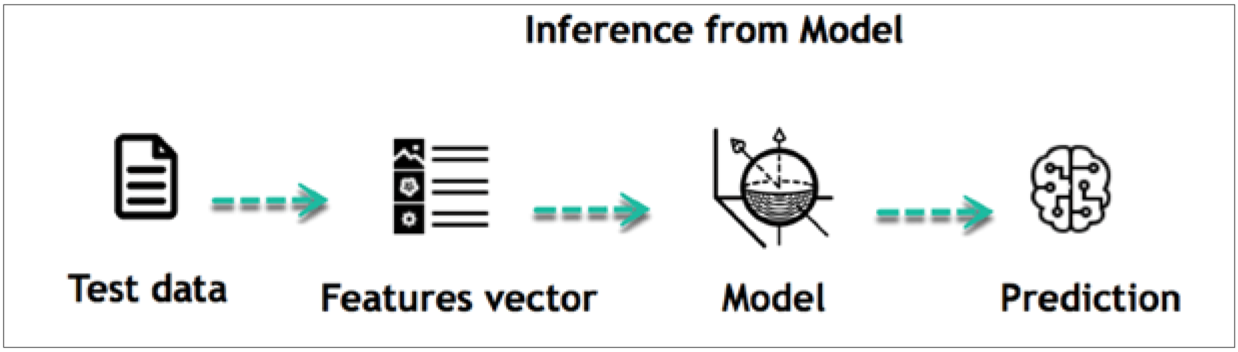


**Fig 1.6: Prediction Model**

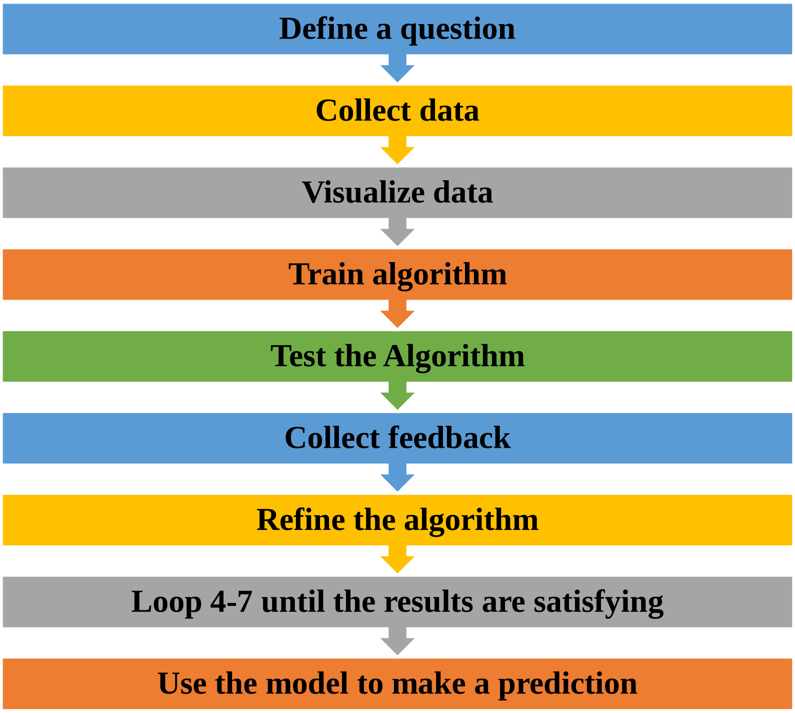


**Fig 1.7: Steps of Machine Learning Model**

Once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data.

## 6. Data-set in Machine Learning:

Using the right data-set while studying for machine learning or data science developments is a relatively tough job. And, to develop correct models, one needs large quantity of data. Following are some machine learning datasets that one can use to develop few projects.

**1. Mall Customer Data-set**

In the Mall customers data-set holds data about persons going in the mall. The data-set contains gender, customer id, age, yearly income, and expenditure score. It assembles perceptions from the statistics and group peoples based on their performance's.

**2. Iris Data-set**

The iris data-set is a well-known and user-friendly data-set that comprises information about the floret petal and sepal sizes. The data-set has 3 categories with 50 examples in each category, then it includes 150 tuples with only 4 features.

**3. MNIST Data-set**

It includes a catalogue of handwritten numbers. It includes 60,000 training pictures and 10,000 testing pictures. This is a good dataset for applying picture sorting where one can categorize a number from 0 to 9.

**4. Titanic Data-set**

In this dataset Titanic boat descended and slaughtered 1502 travelers out of 2224. The dataset covers information like name, age, sex, number of siblings aboard, etc of about 891 passengers in the training set and 418 passengers in the testing set.

**5. Uber Pickup Data-set**

The data-set has data of 4.5 million uber cartridges in New York City from April 2014 to September 2014 and 14million more from January 2015 to June 2015. Users can make data investigation and collect perceptions from the information.

## 7. Data pre-processing in machine learning:

- Facts pre-processing in Machine Learning is a vital step that helps improve the value of data to encourage the abstraction of evocative visions from the data.
- Data pre-processing in Machine Learning denote the method of making (cleaning and organizing) the data to make it appropriate for construction and training Machine Learning models.
- Classically, practical data is imperfect, unpredictable, imprecise (contains errors or outliers), and frequently lacks specific attribute values or trends. This is where data pre-processing enters the scenario – it helps to clean, format, and organize the raw data, thereby making it ready-to-go for Machine Learning models.

**Steps in Data Pre-Processing in Machine Learning:**

There are seven significant steps in data pre-processing in Machine Learning is as follow

**1. Acquire the data-set**

- To figure and grow Machine Learning models, one must primarily obtain the relevant data-set. This data-set will comprise of data collected from manifold and dissimilar sources which are then mutual in a proper format to form a dataset.
- Data-set arrangements differ according to use cases. For instance, a business data-set will be entirely different from a medical data-set. While a business data-set will contain relevant industry and business data, a medical data-set will include healthcare-related data.
- There are several online sources from where datasets can be downloaded like https://www.kaggle.com/uciml/datasets and https://archive.ics.uci.edu/ml/index.php.
- Also, one can also create a data-set by collecting data via different Python APIs. Once the dataset is ready, you must put it in a CSV, or HTML, or XLSX file formats.

**2. Import the crucial libraries**

- The predefined Python libraries can achieve exact data pre-processing works. The three essential Python libraries used for this data pre-processing in Machine Learning are:
- **NumPy** – NumPy is the important package for procedural calculation in Python. Hence, it is

used for introducing any type of mathematical process in the code. Using NumPy, you can also add large multidimensional arrays and matrices in the code.

- **Pandas** – Pandas is an excellent open-source Python library for data manipulation and analysis. It is extensively used for importing and managing the datasets. It packs in high-performance, easy-to-use data structures and data analysis tools for Python.

- **Matplotlib** – Matplotlib is a Python 2D plotting library that is used to plot any type of charts in Python. It can deliver publication-quality figures in numerous hard copy formats and interactive environments across platforms (IPython shells, Jupyter notebook, web application servers, etc.).

**3. Identifying and handling the missing values**

In data pre-processing, it is pivotal to identify and correctly handle the missing values, failing to do this, one might draw inaccurate and faulty conclusions and inferences from the data. Basically, there are two ways to handle missing data.

- **Deleting a particular row**

In this method, remove a specific row that has a null value for a feature or a particular column where more than 75% of the values are missing. However, this method is not 100% efficient, and it is recommended that it is used only when the dataset has adequate samples.

- **Calculating the mean, median and mode**

This method is useful for features having numeric data like age, salary, year, etc. Here, one can calculate the mean, median, or mode of a particular feature or column or row that contains a missing value and replace the result for the missing value.

- **Encoding the categorical data**

Categorical data refers to the information that has specific categories within the data-set. In the dataset for example, there are two categorical variables – country and purchased using one-hot-encoding technique that convert categorical data into numeric data.

**4. Splitting the data-set**

- Every data-set for Machine Learning model must be split into two separate sets-training set and test set.

- Training set denotes the subset of a dataset that is used for training the machine learning model.

- Usually, the dataset is split into 70:30 ratio or 80:20 ratio. This means that you either take 70% or 80% of the data for training the model while leaving out the rest 30% or 20%. The splitting process varies according to the shape and size of the data-set.

**5. Feature scaling**

Feature scaling marks the end of the data pre-processing in Machine Learning. It is a method to standardize the independent variables of a dataset within a specific range. In other words, feature scaling limits the range of variables so that one can compare them on common grounds.

## 8. Types of Machine Learning Algorithms

There some variations of how to define the types of Machine Learning Algorithms but commonly they can be divided into categories according to their purpose and the main categories are the following:

Supervised learning

Unsupervised Learning

### 1. Supervised Learning

- Supervised Learning is a technique contains principles using categorized previous information and the algorithm shall forecast the tag for hidden or forthcoming data.
- A supervised machine learning discoveries the underlying designs that yield the expected output to within acceptable degree of correctness.
- In other words, using these prior known outputs, the machine learning algorithm studies from the past data and then produces an equation for the label or the value. This stage is called the training stage.

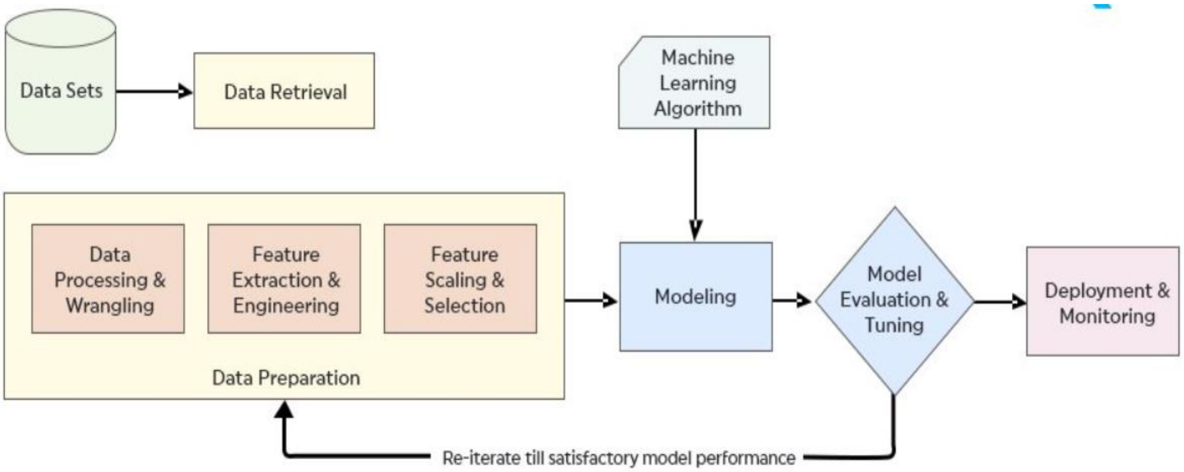

Data Sets
Data Retrieval
Machine Learning Algorithm
Data Processing & Wrangling
Feature Extraction & Engineering
Feature Scaling & Selection
Data Preparation
Modeling
Model Evaluation & Tuning
Deployment & Monitoring
Re-iterate till satisfactory model performance

**Fig 1.8: Phases of Supervised Learning**

### Phases of Supervised Learning

- A Supervised Learning algorithm has the following set of responsibilities – information gathering, data preparation, modelling, model assessment, deployment, and monitoring.
- Information gathering or collection at relevant data essential for the supervised learning algorithm. This data can be invented via activities like – transactions, demographics, inspections, etc.
- Data Preparation is where we adapt and alter the data using the essential steps. It is highly vital to eliminate unsolicited data points and fill-in the contradictions in the data. This step confirms correctness.
- Modeling or training stage where the association between label and other variables are recognized.
- Deployment and monitoring happen on unseen data, a stage where the model is applied and outputs

are produced.

**Commonly used Supervised Learning Algorithms:**

### 1. Linear Regression

Linear Regression is a Machine Learning procedure that charts numeric involvements to numeric productions, by fitting a line into the data points. Simply put, Linear Regression is a way of model the association among one or more self-determining variables in a way that they come together to form a driving force for the reliant on numerical variable. It is typically identified by the linear equation:

$$y = mx + c$$

```
import pandas as pd

import matplotlib. pyplot as plt

%matplotlib inline

df = pd. read_csv('Salary_Data_reg.csv')

df. tail (10)

df = pd. read_csv('Salary_Data_reg.csv')

df. tail(10)

plt.scatter(df['YearsExperience'], df['Salary'])

plt.ylim([0,130000])
```

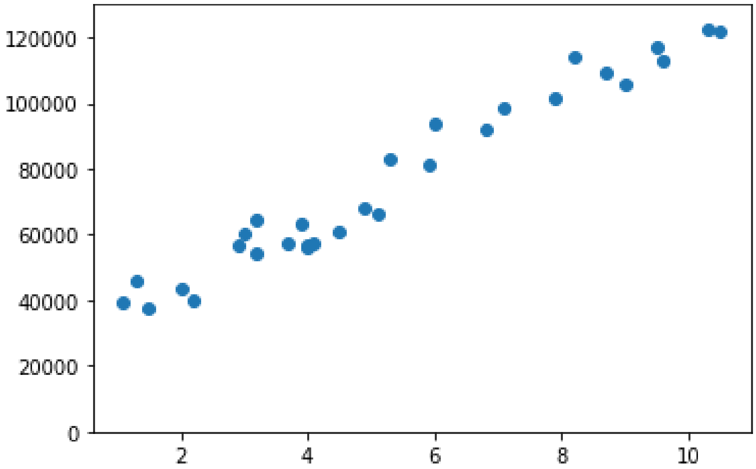


**Fig 1.9: Linear Regression**

```
from sklearn.linear_model import LinearRegression

model = LinearRegression ()

model.fit(df[['YearsExperience']], df['Salary'])

model.intercept_ , model.coef_

type(df['YearsExperience'])
```

```
# How to predict salary for given experience?

model.predict([[15],[20]])

'''x1, x2 = 0, 11

y1,y2 = model.predict([[x1],[x2]])

y1,y2'''

y1 = model.predict([[x1]])[0]

y1

x1,x2,y1,y2

plt.plot([x1,x2],[y1,y2])

plt.scatter(df['YearsExperience'], df['Salary'])

plt.ylim([0,130000])
```

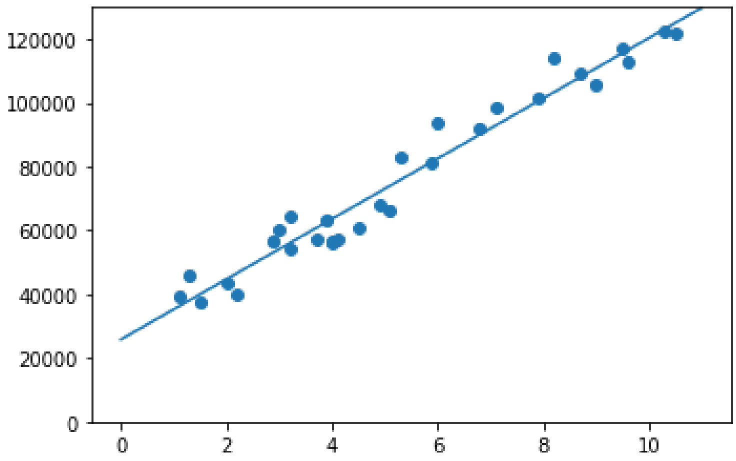


**Fig 1.10: Logistic Regression**

```
model.predict([[5]])[0]

73042.01180594409

model.predict([[x1],[x2]])

array ([ 25792.20019867, 129741.78573467])
```

**2. Logistic Regression**

The Logistic Regression algorithm classifies a connection among variables and a class. It is classically uses to forecast an occasion class, wherever we take a predefined and known group of actions. The dependent variable is certainly a clear-cut variable but the inner working of the Logistic regression algorithm really converts the variable by making usage of a logit function, which calculates the log odds ratio for the events and hence construction a linear equation for the same.

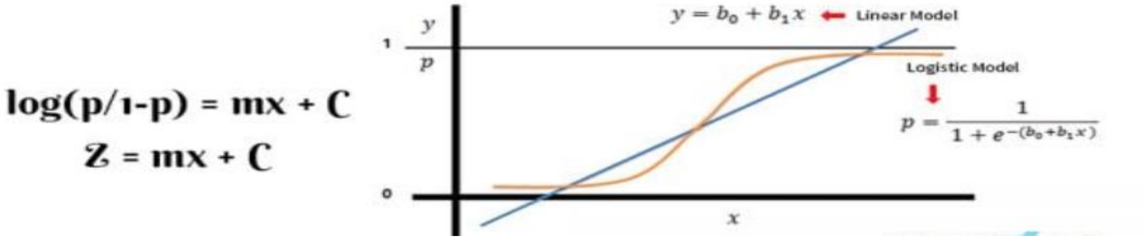


**Fig 1.11: Logistic Regression**

**Algorithm:**

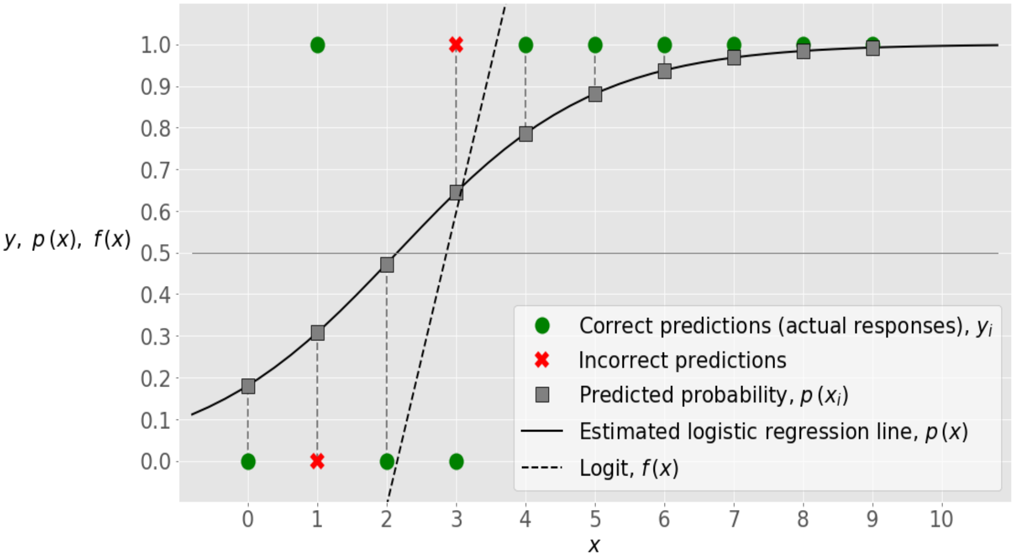
```
import numpy as np

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import classification_report, confusion_matrix

# Get data

x = np.arange(10).reshape(-1, 1)

y = np.array([0, 1, 0, 0, 1, 1, 1, 1, 1, 1])

# Create a model and train it

model = LogisticRegression(solver='liblinear', C=10.0, random_state=0)

model.fit(x, y)

#  Evaluate the model

p_pred = model.predict_proba(x)

y_pred = model.predict(x)

score_ = model.score(x, y)

conf_m = confusion_matrix(y, y_pred)

report = classification_report(y, y_pred)

>>> print('x:', x, sep='\n')

x:

[[0]

[1]

[2]

[3]

[4]

[5]
```

```
[6]
[7]
[8]
[9]]
>>> print('y:', y, sep='\n', end='\n\n')
y:
[0 1 0 0 1 1 1 1 1 1]
>>> print('intercept:', model.intercept_)
intercept: [-1.51632619]
>>> print('coef:', model.coef_, end='\n\n')
coef: [[0.703457]]
>>> print('p_pred:', p_pred, sep='\n', end='\n\n')
p_pred:
[[0.81999686 0.18000314]
[0.69272057 0.30727943]
[0.52732579 0.47267421]
[0.35570732 0.64429268]
[0.21458576 0.78541424]
[0.11910229 0.88089771]
[0.06271329 0.93728671]
[0.03205032 0.96794968]
[0.0161218  0.9838782 ]
[0.00804372 0.99195628]]
>>> print('y_pred:', y_pred, end='\n\n')
y_pred: [0 0 0 1 1 1 1 1 1 1]
>>> print('score_:', score_, end='\n\n')
score_: 0.8
>>> print('conf_m:', conf_m, sep='\n', end='\n\n')
conf_m:
[[2 1]
```

```
[1 6]]
>>> print('report:', report, sep='\n')
report:
precision    recall  f1-score   support
0       0.67      0.67      0.67         3
1       0.86      0.86      0.86         7
accuracy                         0.80        10
macro avg       0.76      0.76      0.76        10
weighted avg       0.80      0.80      0.80        10
```

In this case, the score (or accuracy) is 0.8. There are two observations classified incorrectly. One of them is a false negative, while the other is a false positive.

The figure below illustrates this example with eight correct and two incorrect predictions:



This figure reveals one important characteristic of this example. Unlike the previous one, this problem is not linearly separable. That means onecan't find a value of y and draw a straight line to separate the observations with y=0 and those with y=1. There is no such line.

### 3. Support Vector Machines (SVM)

This algorithm is mainly used for classification but can also be used for regression tasks. In this algorithm, each data item is plotted as a point in n-dimensional space, where n denotes the number of features one have, with the value of each feature as the value of a particular coordinate.

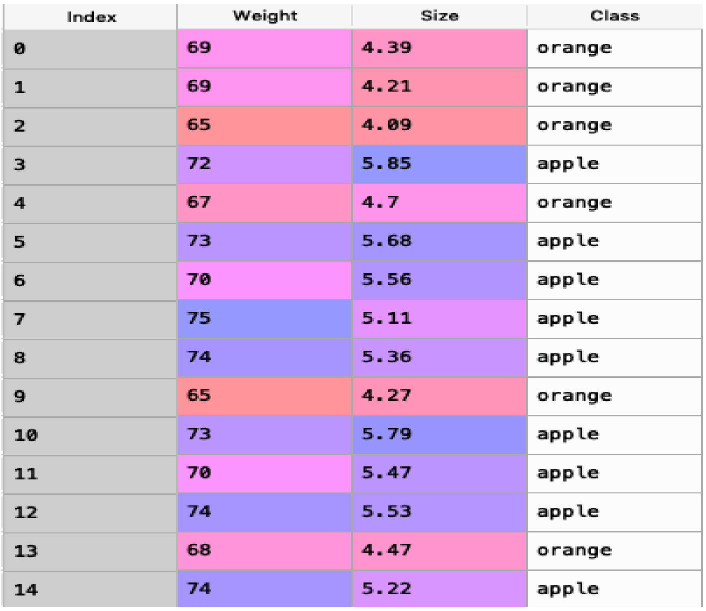
The objective of SVM is to draw a line that best separates the two classes of data points.

SVM generates a line that can cleanly separate the two classes. There are many possible ways of drawing a line that separates the two classes, however, in SVM, it is determined by the **margins** and the **support vectors**.

Let’s use the same dataset of apples and oranges. We will consider the Weights and Size for 20 each.

Importing the dataset

```
import pandas as pd
data = pd.read_csv("apples_and_oranges.csv")
```



| Index | Weight | Size | Class |
|---|---|---|---|
| 0 | 69 | 4.39 | orange |
| 1 | 69 | 4.21 | orange |
| 2 | 65 | 4.09 | orange |
| 3 | 72 | 5.85 | apple |
| 4 | 67 | 4.7 | orange |
| 5 | 73 | 5.68 | apple |
| 6 | 70 | 5.56 | apple |
| 7 | 75 | 5.11 | apple |
| 8 | 74 | 5.36 | apple |
| 9 | 65 | 4.27 | orange |
| 10 | 73 | 5.79 | apple |
| 11 | 70 | 5.47 | apple |
| 12 | 74 | 5.53 | apple |
| 13 | 68 | 4.47 | orange |
| 14 | 74 | 5.22 | apple |

**Fig. 1.12: Splitting the dataset into training and test samples**

```
from sklearn.model_selection import train_test_split
training_set, test_set = train_test_split(data, test_size = 0.2, random_state = 1)
```

Classifying the predictors and target

```
X_train = training_set.iloc[:,0:2].values
Y_train = training_set.iloc[:,2].values
X_test = test_set.iloc[:,0:2].values
Y_test = test_set.iloc[:,2].values
```

Initializing Support Vector Machine and fitting the training data

```
from sklearn.svm import SVC
classifier = SVC(kernel='rbf', random_state = 1)
classifier.fit(X_train,Y_train)
```

Predicting the classes for test set

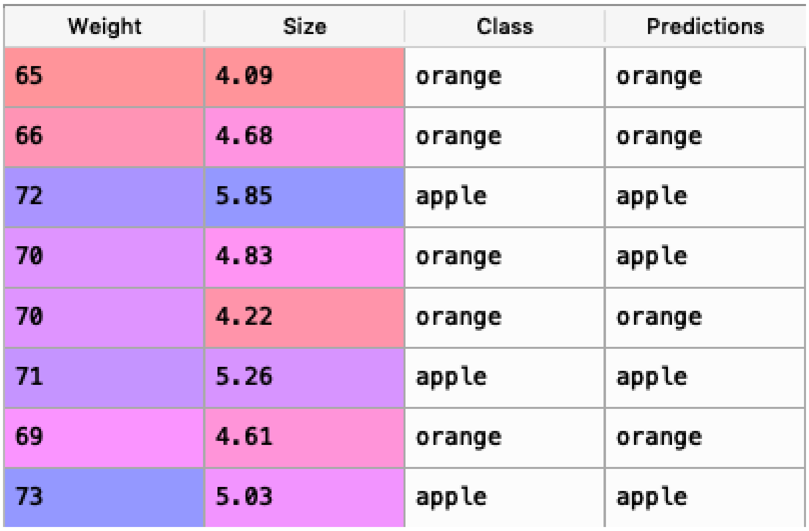
```
Y_pred = classifier.predict(X_test)
```

Attaching the predictions to test set for comparing

```
test_set["Predictions"] = Y_pred
```

Comparing the actual classes and predictions

Let’s have a look at the test_set:



| Weight | Size | Class | Predictions |
|---|---|---|---|
| 65 | 4.09 | orange | orange |
| 66 | 4.68 | orange | orange |
| 72 | 5.85 | apple | apple |
| 70 | 4.83 | orange | apple |
| 70 | 4.22 | orange | orange |
| 71 | 5.26 | apple | apple |
| 69 | 4.61 | orange | orange |
| 73 | 5.03 | apple | apple |

Comparing the ‘Class’ and ‘Predictions’ column we find that only one of the 8 predictions has gone wrong.

Calculating the accuracy of the predictions

We will calculate the accuracy using the confusion matrix as follows :

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(Y_test,Y_pred)
accuracy = float(cm.diagonal().sum())/len(Y_test)
print("\nAccuracy Of SVM For The Given Dataset : ", accuracy)
```

**Output:**

```
Accuracy Of SVM For The Given Dataset :  0.875
```

Visualizing the classifier

Before we visualize we might need to encode the classes ‘apple’ and ‘orange’ into numerical. We can achieve that using the label encoder.

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
Y_train = le.fit_transform(Y_train)
```

After encoding , fit the encoded data to the SVM

```
from sklearn.svm import SVC
classifier = SVC(kernel='rbf', random_state = 1)
classifier.fit(X_train,Y_train)
```

Let’s Visualize!

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
```

```
plt.figure(figsize = (7,7))
X_set, y_set = X_train, Y_train
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape), alpha =
0.75, cmap = ListedColormap(('black', 'white')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1], c = ListedColormap(('red', 'orange'))(i), label = j)
plt.title('Apples Vs Oranges')
plt.xlabel('Weight In Grams')
plt.ylabel('Size in cm')
plt.legend()
plt.show()
```

Output :

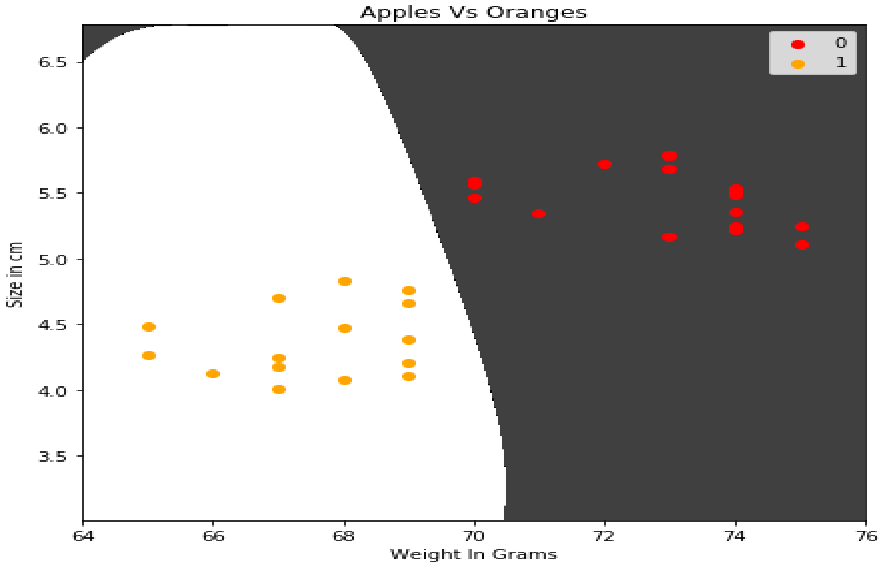


**Fig 1.13: SVM**

The above image shows the plotting of the training set after fitting the training data to the classifier. The border that separates both the white and black colours represent the Maximum Margin Hyperplane or Line in this case.

According to the SVM classifier, any new data point that falls within the white region is classified as oranges (denoted in orange colour) and any data point that falls in black region is classified as apples(denoted in red colour).

Visualizing the predictions

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
```

```
plt.figure(figsize = (7,7))
X_set, y_set = X_test, Y_test
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step =
0.01),np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),alpha =
0.75, cmap = ListedColormap(('black', 'white')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],c = ListedColormap(('red', 'orange'))(i), label = j)
plt.title('Apples Vs Oranges Predictions')
plt.xlabel('Weight In Grams')
plt.ylabel('Size in cm')
plt.legend()
plt.show()
```

Output:

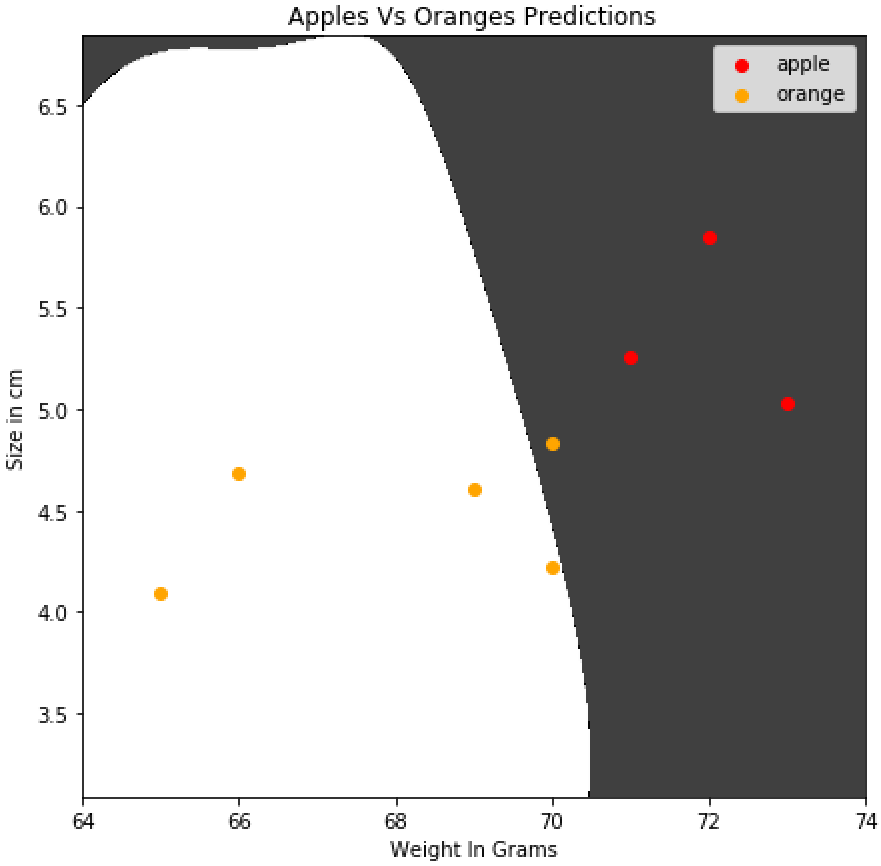


**Fig 1.14: Apples Vs Oranges prediction**

In the above image we can see that one of the orange data points is lying outside of the white region. This represents the false prediction that we saw earlier while comparing the actual test set classes and the predicted classes.

## 4. Decision Trees

It includes non-parametric supervised learning method that can be used for both Sorting and Regression problems, by recognizing appropriate approaches to divided information based on numerous circumstances into a tree-like assembly. The conclusion box is to forecast an occurrence or a value by leveraging the circumstances.

The tree-like construction is really a graph where the nodes represent an underlying question about an attribute, the edges which typically contain the answers and the leaves represent the output which, can be a value or a class. Thus, allowing us to predict values and events. The procedure usually follows a top-down method, by choosing a variable at each step which can split the next set of data items and usually represented by a metric such as GINI impurity, Information Gain, Variance Reduction, etc. to measure the best approach for splitting.

Is a Person Fit?

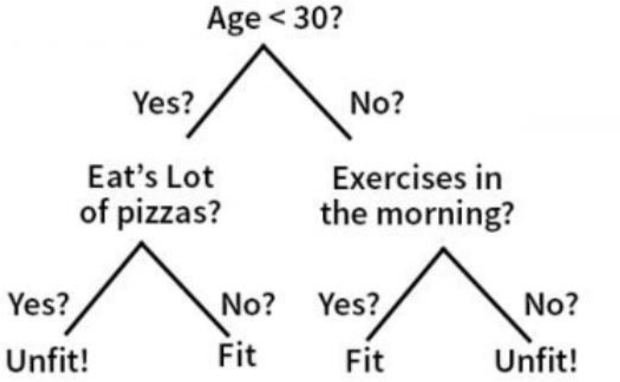


**Fig 1.15: Decision Tree**

```
import pandas as pd

import matplotlib. pyplot as plt

df = pd. read_csv('Social_Net_class.csv')

df.head()
```

| | User ID | Gender | Age | Estimated Salary | Purchased |
|---|---|---|---|---|---|
| 0 | 15624510 | Male | 19 | 19000 | 0 |
| 1 | 15810944 | Male | 35 | 20000 | 0 |
| 2 | 15668575 | Female | 26 | 43000 | 0 |
| 3 | 15603246 | Female | 27 | 57000 | 0 |
| 4 | 15804002 | Male | 19 | 76000 | 0 |

```
X = df[['Age','EstimatedSalary']]

y = df['Purchased']

from sklearn.model_selection import train_test_split

X_train,X_test, y_train,y_test=train_test_split(X,y,test_size=0.25, random_state=91)

X_test.shape

(100, 2)
```

```
from sklearn. tree import DecisionTreeClassifier
model = DecisionTreeClassifier ()
model.fit(X_train,y_train)
model. score(X_test,y_test), model.score(X_train,y_train)
(0.84, 0.9933333333333333)
plt.scatter(X_test['Age'],X_test['EstimatedSalary'])
```

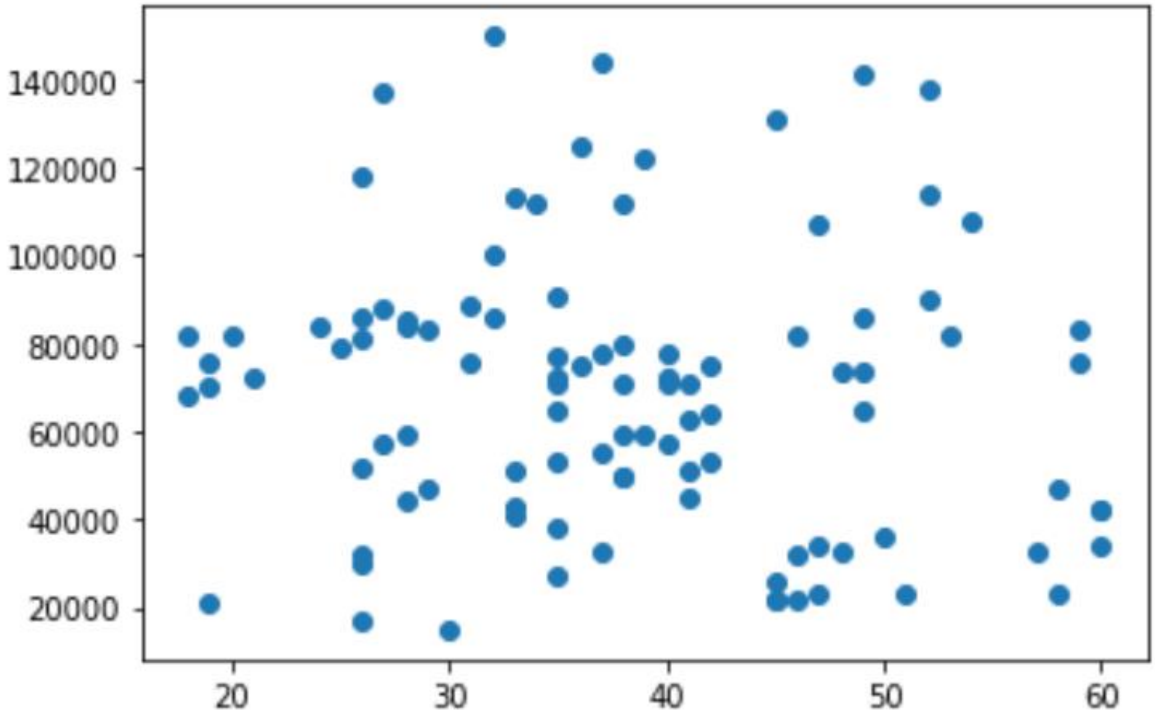


**Fig 1.16: Scatter Plot**

```
import numpy as np
plot_data = []
for x in range(X_test['Age'].min()*10,X_test['Age'].max()*10,5):
    for y in range(X_test['EstimatedSalary'].min(),X_test['EstimatedSalary'].max(),1350):
        plot_data.append([x/10,y])
plot_data = np.array(plot_data)
(X_test['EstimatedSalary'].max() - X_test['EstimatedSalary'].min())/100
plot_data[50:150]
y_pre=model.predict(plot_data)
y_pre
```

```
array([0, 0, 0, ..., 1, 1, 1], dtype=int64)
class_0 = plot_data[y_pre==0]
class_1 = plot_data[y_pre==1]
plt.scatter(class_0[:,0],class_0[:,1],c='red')
plt.scatter(class_1[:,0],class_1[:,1],c='blue')
```

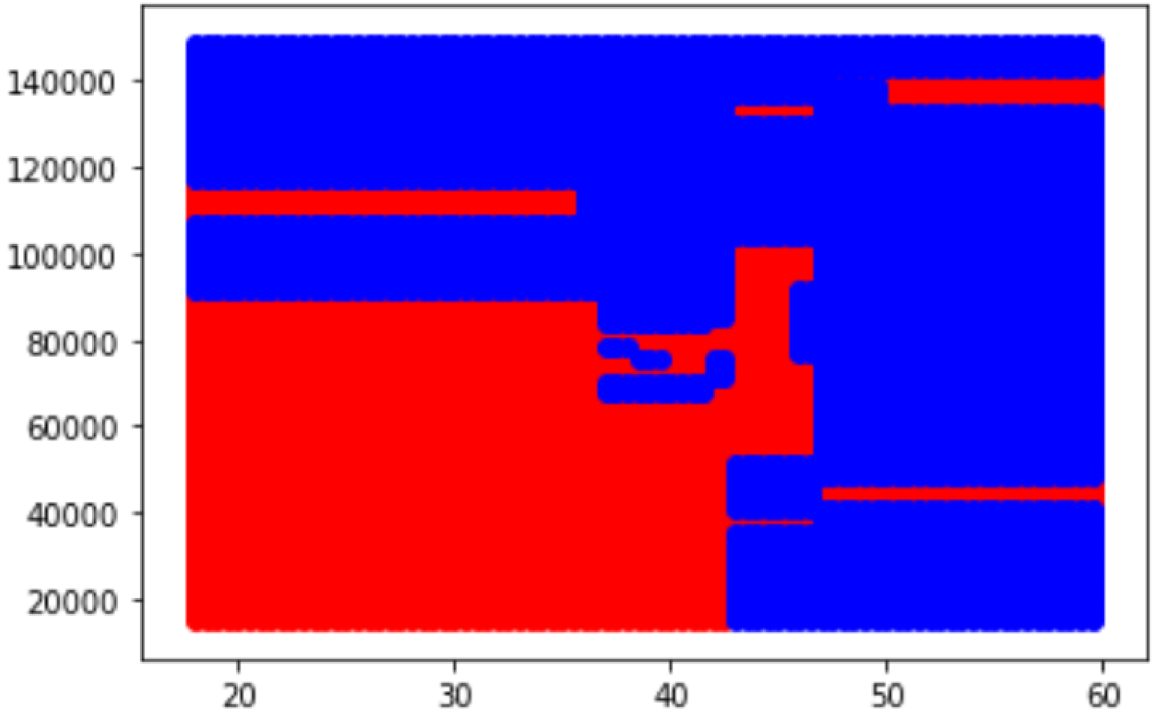
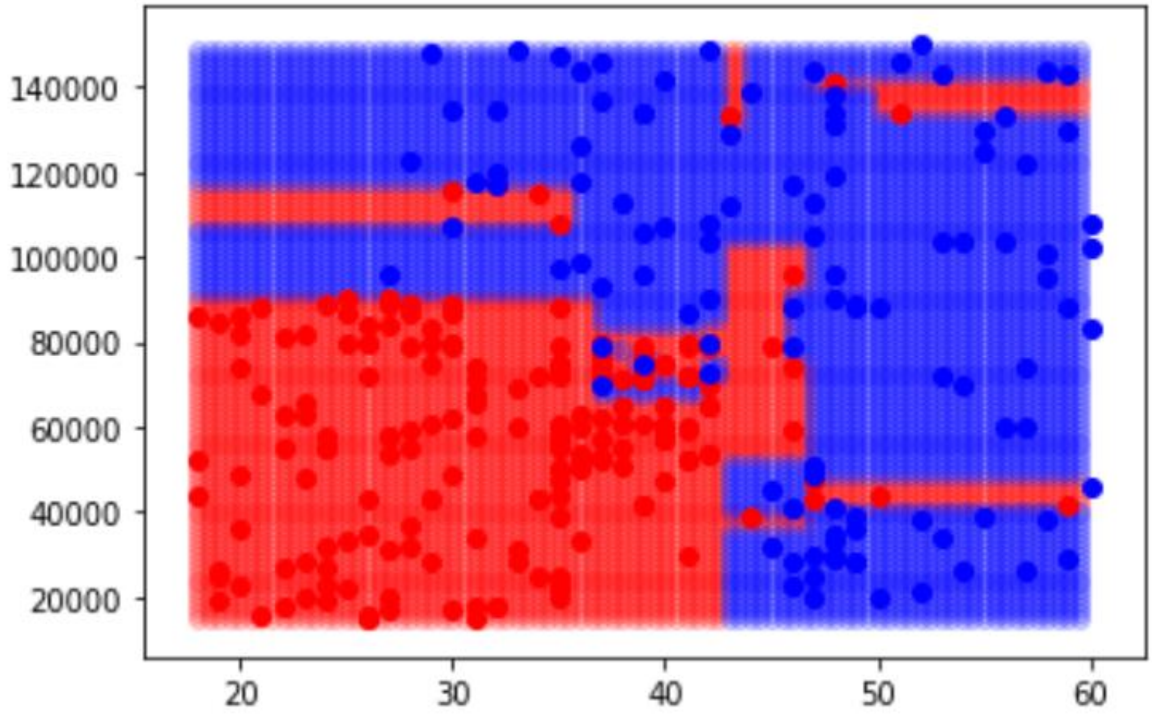


**Fig 1.17: Plot Data**

```
class_0 = plot_data[y_pre==0]
class_1 = plot_data[y_pre==1]
plt.scatter(class_0[:,0],class_0[:,1],c='red',alpha=0.2)
plt.scatter(class_1[:,0],class_1[:,1],c='blue',alpha=0.2)
class_0_tr = X_train[y_train==0]
class_1_tr = X_train[y_train==1]
plt.scatter(class_0_tr['Age'],class_0_tr['EstimatedSalary'],c='red')
plt.scatter(class_1_tr['Age'],class_1_tr['EstimatedSalary'],c='blue')
```



model. score(X_train,y_train)

0.9933333333333333

from sklearn.metrics import confusion_matrix

confusion_matrix(y_test,model.predict(X_test))

model.predict([[40,40000],[55,65000]])

model_1 = DecisionTreeClassifier (max_depth=4, min_samples_leaf=4,random_state=10)

model_1.fit(X_train,y_train)

model_1.score(X_test,y_test), model_1.score(X_train,y_train)

(0.88, 0.9233333333333333)

**2. Unsupervised Learning:**

Unsupervised culture is a kind of Machine Learning algorithm that includes drill a machine classically using unlabeled figures, and that procedures the main point of alteration with supervised machine learning procedures which classically use labeled data. In this form of machine learning, we let the algorithm to self-discover the fundamental designs, similarities, equations, and relations in the data without adding any bias from the users' end. Although the end result of these is totally random and cannot be controlled, Unsupervised Learning discoveries its residence is advanced exploratory data analysis and particularly, Cluster Analysis.

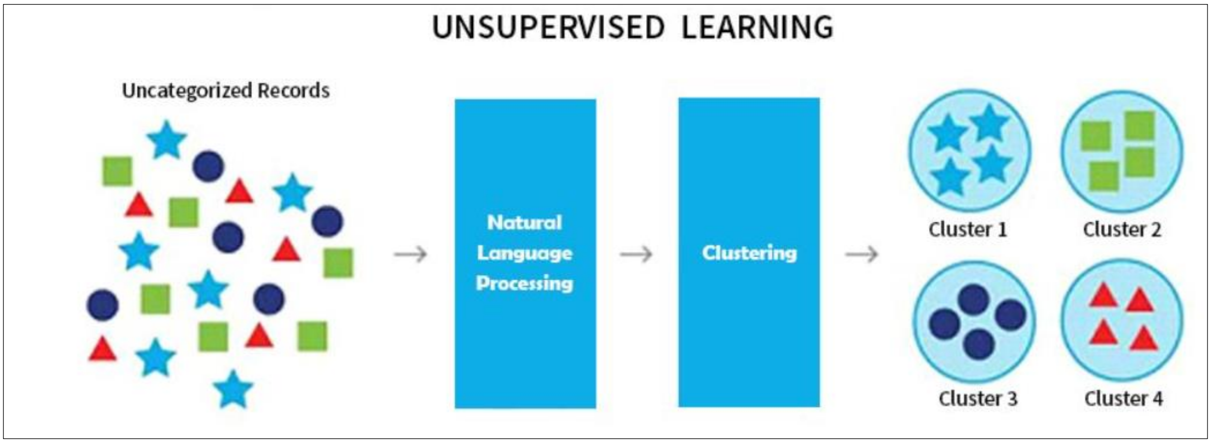


**Fig 1.18: Unsupervised Learning**

**Frequently used Unsupervised Learning Algorithms:**

**1. The k-means Clustering**

The k-means clustering is a development that aids to divide the data points or explanations into k unknown clusters in such a method that each remark definitely fits to a cluster. This cluster associativity is strongminded by the nearness of that data point with the adjacent mean, otherwise known as cluster centroid. Due to the connection of nearness measure in the data, various distance algorithms are used in the process to measure the familiarity of data to the cluster center.

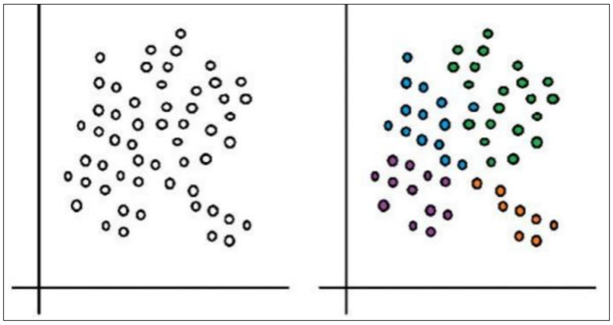


**Fig 1.19: k-means Clustering**

The only and main disadvantage of xk-means is the detail that the algorithm cannot start without the user specifying the required number of clusters apriori. Added to that, there is no mathematical or technical method to control the optimal number of clusters. The application typically happens based on the trial and error method, where a set of “K” values are measured originally and the best one is chosen according to the empirical information.

Contempt that disadvantage, k-means is one of the most simpler, yet powerful algorithms that has been functional to many business cases, like customer Segmentation, image Segmentation, text segmentation etc.

```
import pandas as pd

import matplotlib. pyplot as plt

df = pd. read_csv('../../Downloads/Mall_Customers.csv')
```

```
df.head()
```

| | CustomerID | Genre | Age | Annual Income (k$) | Spending Score (1-100) |
|---|---|---|---|---|---|
| 0 | 1 | Male | 19 | 15 | 39 |
| 1 | 2 | Male | 21 | 15 | 81 |
| 2 | 3 | Female | 20 | 16 | 6 |
| 3 | 4 | Female | 23 | 16 | 77 |
| 4 | 5 | Female | 31 | 17 | |

```
X = df[['Annual Income (k$)','Spending Score (1-100)']]

X

plt. scatter (X['Annual Income (k$)'], X['Spending Score (1-100)'])
```

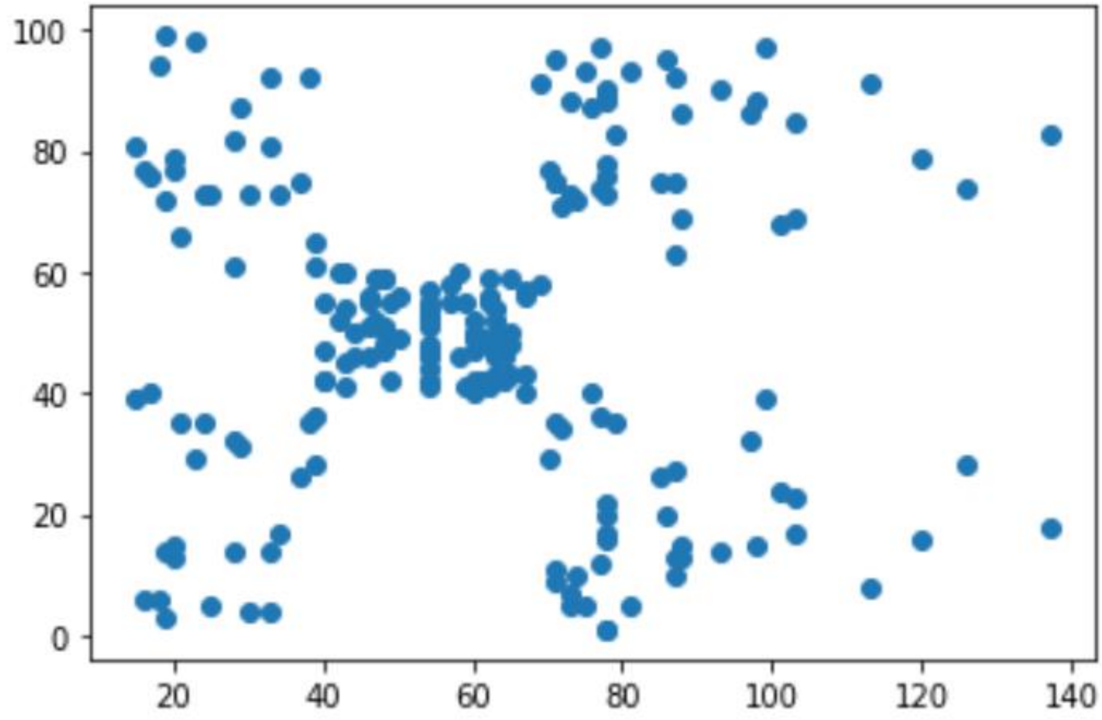
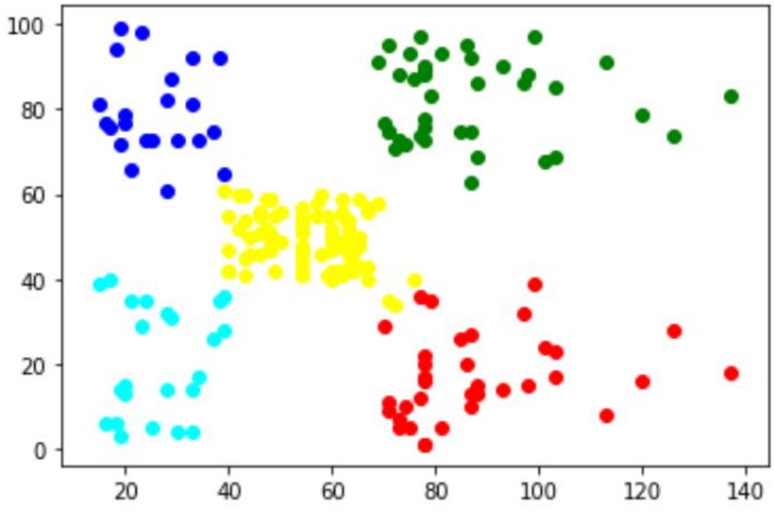


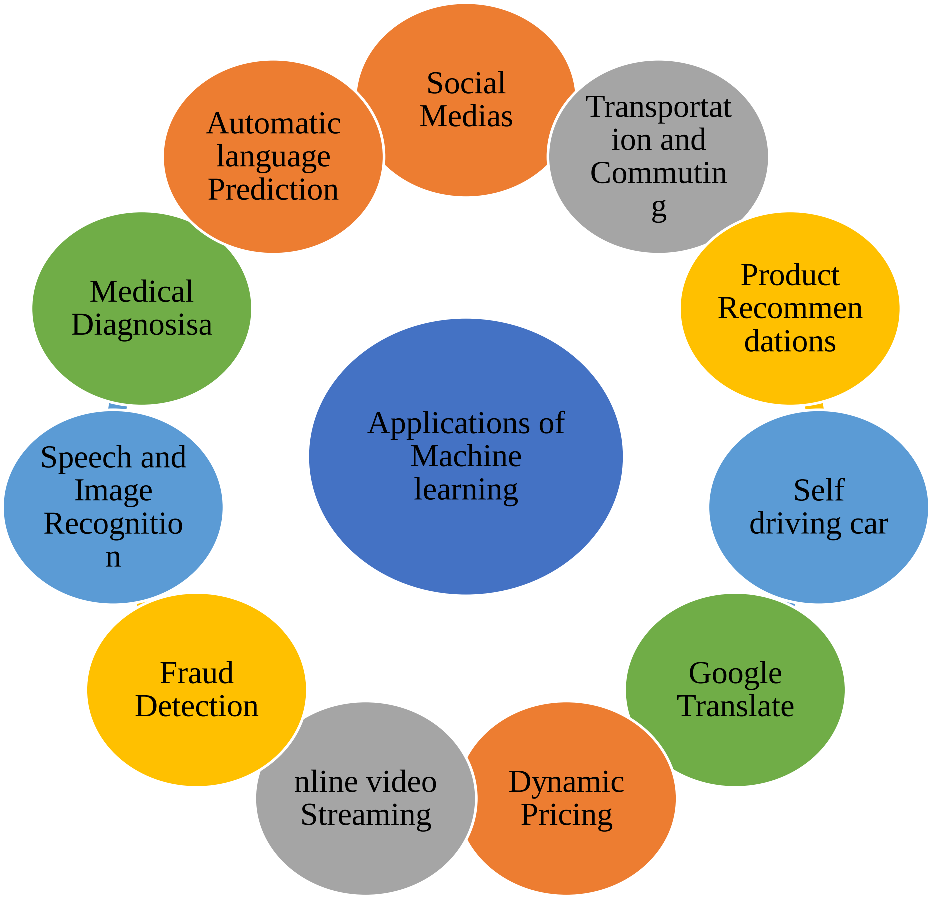
**Fig 1.22 Applications of Machine Learning**

```
from sklearn. cluster import KMeans

model = KMeans(n_clusters=5)

model.fit(X)

model.cluster_centers_

array([[88.2       , 17.11428571],

       [25.72727273, 79.36363636],

       [55.2962963 , 49.51851852],

       [26.30434783, 20.91304348],
       [86.53846154, 82.12820513]])
cluster_number = model. predict(X)
len(cluster_number)
200
len(X)
200
c0 = X[cluster_number==0]
c1 = X[cluster_number==1]
c2 = X[cluster_number==2]
c3 = X[cluster_number==3]
c4 = X[cluster_number==4]
plt.scatter(c0['Annual Income (k$)'], c0['Spending Score (1-100)'],c='red')
plt.scatter(c1['Annual Income (k$)'], c1['Spending Score (1-100)'],c='blue')
plt.scatter(c2['Annual Income (k$)'], c2['Spending Score (1-100)'],c='yellow')
plt.scatter(c3['Annual Income (k$)'], c3['Spending Score (1-100)'],c='cyan')
plt.scatter(c4['Annual Income (k$)'], c4['Spending Score (1-100)'],c='green')
```



## 9. Applications of Machine Learning:

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:


Applications of Machine learning
Social Medias
Transportat ion and Commutin g
Product Recommen dations
Self driving car
Google Translate
Dynamic Pricing
nline video Streaming
Fraud Detection
Speech and Image Recognitio n
Medical Diagnosisa
Automatic language Prediction

# Deep Learning

Mr. Vivek Nikam

Prof. Monalisa Bhinge

Prof. Ankush Kudale

## 2.1 Introduction:

Data has become much more pervasive. We are living in the age of big data and the algorithms need huge amounts of data to succeed.  Intelligence is the ability to process information to inform future decisions. Artificial Intelligence is the field which focuses on building algorithms required to process the information. Machine Learning is a subset of AI specifically that focuses on actually teaching an algo how to do this without being explicitly programmed to do the task at hand. Data Learning is a subset of ML which takes this idea even a step further and says how we can automatically extract the useful pieces of information needed to inform those future predictions or make a decision. It is the way to extract useful pattern from the data in an automated way with a little human effort required.

### 2.2 History of Deep learning Ideas and Milestones

1943: McCulloch Pitts Neuron (Neural networks) – Beginning

1957: Frank Rosenblatt creates Perceptron

1974-86: Backpropagation, RBM, RNN

1989-98: CNN, MNIST, LSTM, Bidirectional RNN

2006: “Deep Learning”, DBN

2009: ImageNet

2012: AlexNet, Dropout

2014: GANs

2014: DeepFace

2016: AlphaGo

2017: AlphaZero, Capsule Networks

2018: BERT

### 2.3 History of Deep learning Tools

1960: Mark 1 Perceptron

2002: Torch

2007: CUDA

2008: Theano

2014: Caffe

2011: DistBelief

2015: TensorFlow 0.1

2017: PyTorch 0.1

2017: TensorFlow 1.0

2017: PyTorch 1.0

2019: TensorFlow 2.0

If we want to understand why Deep learning, then we need to understand traditional ML. ML algorithms are defined as set of rules or features in the data.

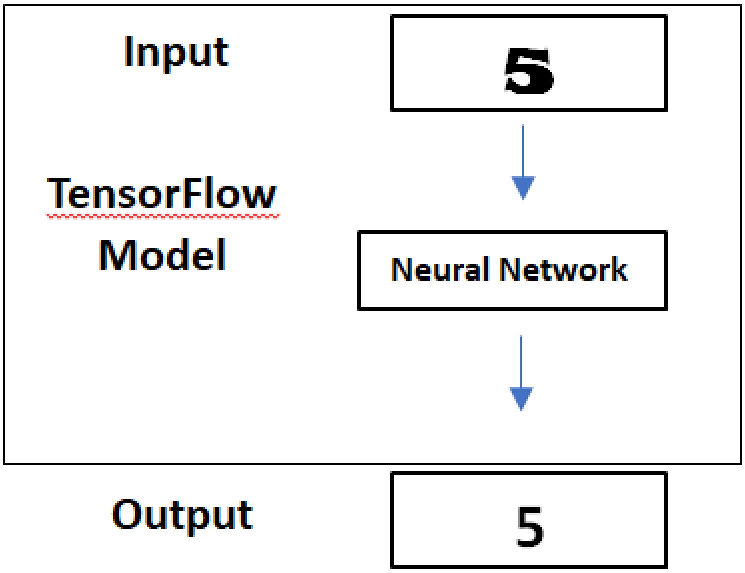


**Fig 2.1 Example of a DIGIT**

Let us consider a simple example, handwritten digit **5** as the input to which is given to machine learning system or neural network system. In the diagram the hand written number is the input & the output is the number that is the same digit. In the above example we have used TensorFlow a popular deep learning tool box to implement the algorithms. TensorFlow is an automated tool which extracts useful patterns from data with little human intervention. So, with small 6 little pieces of code, we can train a neural network or a machine learning system that can understand what the image is. Steps are as follows:

- Import the library TensorFlow (Deep Learning Library)
- Import the data set MNIST
- Stack on top of each other all the neural network layer as a hidden layer, an input layer and the output layer
- Training the model as model fit
- Evaluate the model with the help of a testing data set

- Predict what is in the image

### Deep Learning is Representation Learning

Deep Learning as the meaning goes is the ability to form higher level of abstractions of representations in data and raw patterns. Higher & higher levels of understanding of patterns and those representations are extremely important and effective for being able to interpret data. Under certain representations data is trivial to understand.

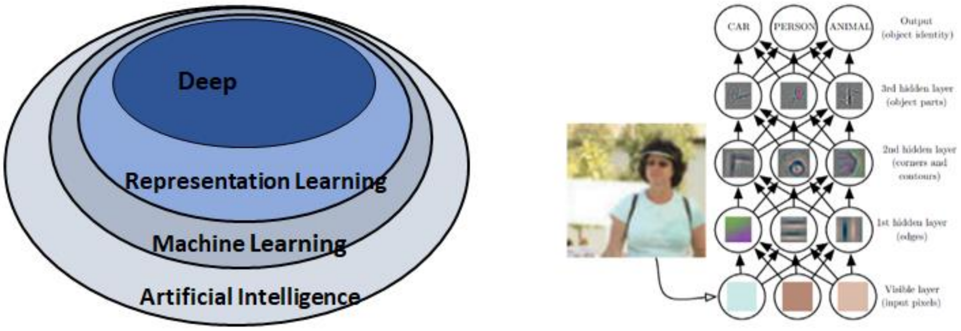


**Fig 2.1: Trivial example of representation**

So, in the below example the task of drawing a line under polar coordinates is trivial, under Cartesian coordinates is very difficult or almost impossible to do accurately. So, DL with ML in general is forming representations that map the topology. So, whatever the topology is, the problem is map it in such a way that the final representation is trivial to work with, trivial to classify, trivial to perform regression, trivial to generate new samples of that data. Higher levels of representation are the vision of artificial intelligence.

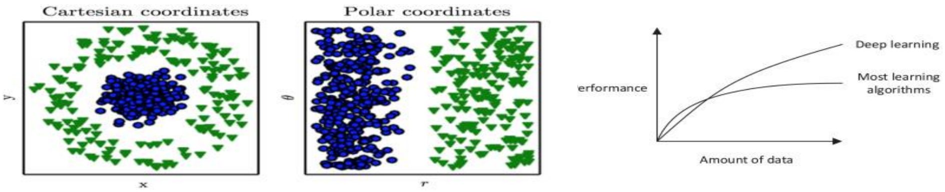


**Fig 2.2: Representation Matters**

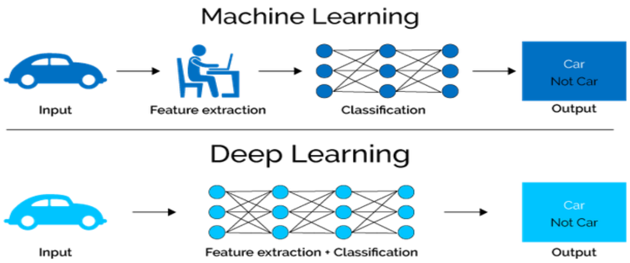


**Fig 2.3 Why Deep Learning? Scalable with Machine Learning and Artificial Intelligence**

Deep Learning is revolutionizing across fields from robotics to medicine. Deep Learning is the ability to remove the efforts of human expert. DL automates much of the extractions from the raw data and form representations from the raw data without the need of human involvement. The automated extraction of the features enables us to work with larger data sets without human experts.

## 2.4 Logical Computations with Neurons:

Warren McCulloch and Walter Pitts proposed a very simple model of the biological neuron, which later became known as an artificial neuron: it has one or more binary (on/off) inputs and one binary output. The artificial neuron simply activates its output when more than a certain number of its inputs are active. McCulloch and Pitts showed that even with such a simplified model it is possible to build a network of artificial neurons that computes any logical proposition you want. For example, let's build a few ANNs that perform various logical computations (see Fig.3), assuming that a neuron is activated when at least two of its inputs are active.

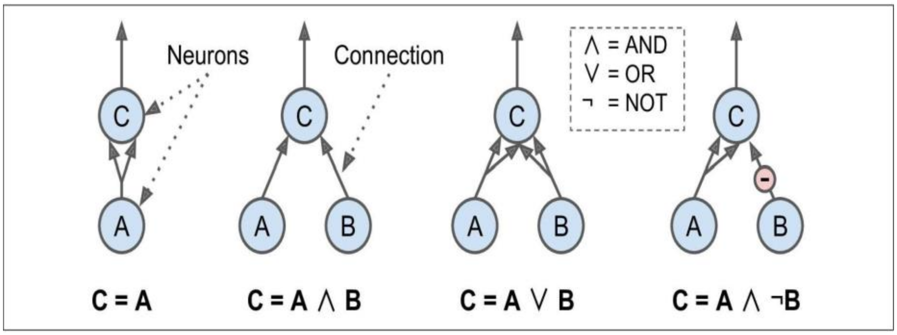


**Fig 2.4: ANNs performing simple logical computations**

- The first network on the left is simply the identity function: if neuron A is activated, then neuron C gets activated as well (since it receives two input signals from neuron A), but if neuron A is off, then neuron C is off as well.

- The second network performs a logical AND: neuron C is activated only when both neurons A and B are activated (a single input signal is not enough to activate neuron C).

- The third network performs a logical OR: neuron C gets activated if either neuron A or neuron B is activated (or both).

- Finally, if we suppose that an input connection can inhibit the neuron's activity (which is the case with biological neurons), then the fourth network computes a slightly more complex logical proposition: neuron C is activated only if neuron A is active and if neuron B is off. If neuron A is active all the time, then you get a logical NOT: neuron C is active when neuron B is off, and vice versa.

You can easily imagine how these networks can be combined to compute complex logical expressions (see the exercises at the end of the chapter).

## 2.5 The Perceptron:

The Perceptron is one of the simplest ANN architectures, invented in 1957 by Frank Rosenblatt. It is based on a slightly different artificial neuron (see Fig.4) called a threshold logic unit (TLU), or sometimes a linear threshold unit (LTU): the inputs and output are now numbers (instead of binary on/off values) and each input connection is associated with a weight. The TLU computes a weighted sum of its inputs ($z = w, x, + w_2 x_2 + \bullet + w_n x_r, = xT w$), then applies a step function to that sum and outputs the result: $h_w(x) = step(z)$, where $z = x^T W$.

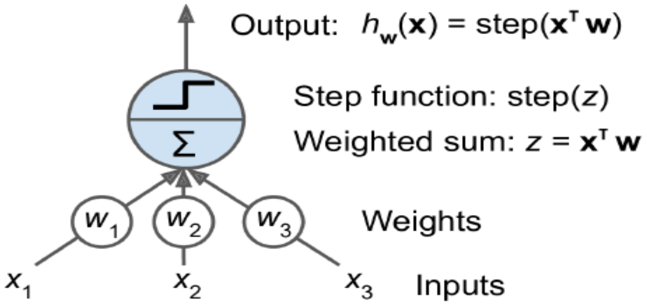

Output: $h_{\mathbf{w}}(\mathbf{x}) = \text{step}(\mathbf{x}^T \mathbf{w})$
Step function: step($z$)
Weighted sum: $z = \mathbf{x}^T \mathbf{w}$
Σ
$w_1$
$w_2$
$w_3$
Weights
$x_1$
$x_2$
$x_3$
Inputs

**Fig 2.5: Threshold logic unit**

The most common step function used in Perceptron's is the Heaviside step function (see Equation. 1). Sometimes the sign function is used instead.

Equation. 1 Common step functions used in Perceptrons

$$\text{heaviside}(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases} \qquad \text{sgn}(z) = \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ +1 & \text{if } z > 0 \end{cases}$$

A single TLU can be used for simple linear binary classification. It computes a linear combination of the inputs and if the result exceeds a threshold, it outputs the positive class or else outputs the negative class (just like a Logistic Regression classifier or a linear SVM). For example, you could use a single TLU to classify iris flowers based on the petal length and width (also adding an extra bias feature $x_o = 1$, just like we did in previous chapters). Training a TLU in this case means finding the right values for $w_o$, $w_1$, and $w_2$ (the training algorithm is discussed shortly).

A Perceptron is simply composed of a single layer of TLUs,[6] with each TLU connected to all the inputs. When all the neurons in a layer are connected to every neuron in the previous layer (i.e., its input neurons), it is called a fully connected layer or a dense layer. To represent the fact that each input is sent to every TLU, it is common to draw special passthrough neurons called input neurons: they just output whatever input they are fed. All the input neurons form the input layer. Moreover, an extra bias feature is generally added ($x_0 = 1$): it is typically represented using a special type of neuron called a bias neuron, which just outputs 1 all the time. A Perceptron with two inputs and three outputs is represented in Figure 10-5. This Perceptron can classify instances simultaneously into three different binary classes, which makes it a multi-

output classifier.

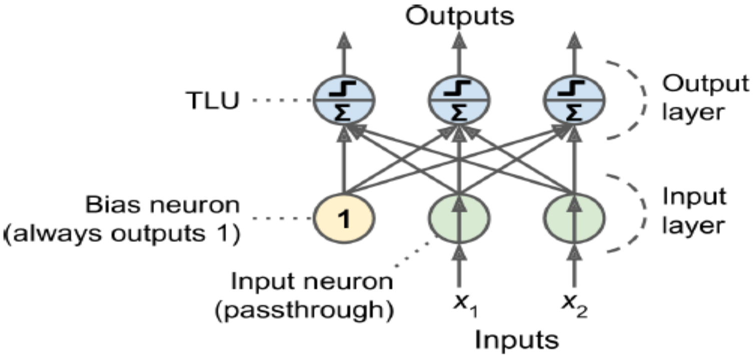


**Fig 2.6: Perceptron diagram**

Thanks to the magic of linear algebra, it is possible to efficiently compute the outputs of a layer of artificial neurons for several instances at once, by using Equation. 2:

Equation. 2 Computing the outputs of a fully connected layer

$$h_{w\,b}(X) = \Phi(XW + b)$$

- As always, X represents the matrix of input features. It has one row per instance, one column per feature.
- The weight matrix W contains all the connection weights except for the ones from the bias neuron. It has one row per input neuron and one column per artificial neuron in the layer.
- The bias vector b contains all the connection weights between the bias neuron and the artificial neurons. It has one bias term per artificial neuron.
- The function 0 is called the activation function: when the artificial neurons are TLUs, it is a step function (but we will discuss other activation functions shortly).

So how is a Perceptron trained? The Perceptron training algorithm proposed by Frank Rosenblatt was largely inspired by Hebb's rule. In his book The Organization of Behavior, published in 1949, Donald Hebb suggested that when a biological neuron often triggers another neuron, the connection between these two neurons grows stronger. This idea was later summarized by Siegrid Lowel in this catchy phrase: "Cells that fire together, wire together:' This rule later became known as Hebb's rule (or Hebbian learning); that is, the connection weight between two neurons is increased whenever they have the same output. Perceptrons are trained using a variant of this rule that takes into account the error made by the network; it reinforces connections that help reduce the error. More specifically, the Perceptron is fed one training instance at a time, and for each instance it makes its predictions. For every output neuron that produced a wrong prediction, it reinforces the connection weights from the inputs that would have contributed to the correct prediction. The rule is shown in Equation. 3.

Equation. 3. Perceptron learning rule (weight update)

$$w_{i,\,j}^{(\text{next step})} = w_{i,\,j} + \eta\left(y_j - \hat{y}_j\right)x_i$$

- $w_{i,,}$ is the connection weight between the ith input neuron and the $J^{11}$ output neuron.
- $x_i$ is the $i^{th}$ input value of the current training instance.
- $Y_i$ is the output of the $j^{th}$ output neuron for the current training instance.
- $y_j$ is the target output of the $f^{h}$ output neuron for the current training instance.
- ri is the learning rate.

The decision boundary of each output neuron is linear, so Perceptrons are incapable of learning complex patterns (just like Logistic Regression classifiers). However, if the training instances are linearly separable, Rosenblatt demonstrated that this algorithm would converge to a solution.' This is called the Perceptron convergence theorem.

Scikit-Learn provides a Perceptron class that implements a single TLU network. It can be used pretty much as you would expect—for example, on the iris dataset

```
import numpy as np

from sklearn.datasets import load_iris

from sklearn.linear_model import Perceptron

iris = load_iris()

X = iris data[:, (2, 3)] # petal length, petal width
y = (iris target == 0).astype(npAnt) # Iris Setosa?

per_clf = Perceptron ()

per_clf.fit(X, y)


y_pred = per_clf.predicta[2, 0.5]])
```

You may have noticed the fact that the Perceptron learning algorithm strongly resembles Stochastic Gradient Descent. In fact, Scikit-Learn's Perceptron class is equivalent to using an SGDClassifier with the following hyperparameters: loss="perceptron", learning_rate="constant", eta0=1 (the learning rate), and penalty=None (no regularization).

Note that contrary to Logistic Regression classifiers, Perceptrons do not output a class probability; rather, they just make predictions based on a hard threshold. This is one of the good reasons to prefer Logistic Regression over Perceptrons.

In their 1969 monograph titled Perceptrons, Marvin Minsky and Seymour Papert highlighted a number of serious weaknesses of Perceptrons, in particular the fact that they are incapable of solving some trivial problems (e.g., the Exclusive OR (XOR) classification problem; see the left side of Figure. 6). Of course this is true of any other linear classification model as well (such as Logistic Regression classifiers), but researchers had expected much more from Perceptrons, and their disappointment was great, and many researchers dropped neural networks altogether in favor of higher-level problems such as logic, problem solving, and search.

However, it turns out that some of the limitations of Perceptrons can be eliminated by stacking multiple Perceptrons. The resulting ANN is called a Multi-Layer Perceptron (MLP). In particular, an MLP can solve the XOR problem, as you can verify by computing the output of the MLP represented on the right of Figure. 6: with inputs (0, 0) or (1, 1) the network outputs 0, and with inputs (0, 1) or (1, 0) it outputs 1. All connections have a weight equal to 1, except the four connections where the weight is shown. Try verifying that this network indeed solves the XOR problem!

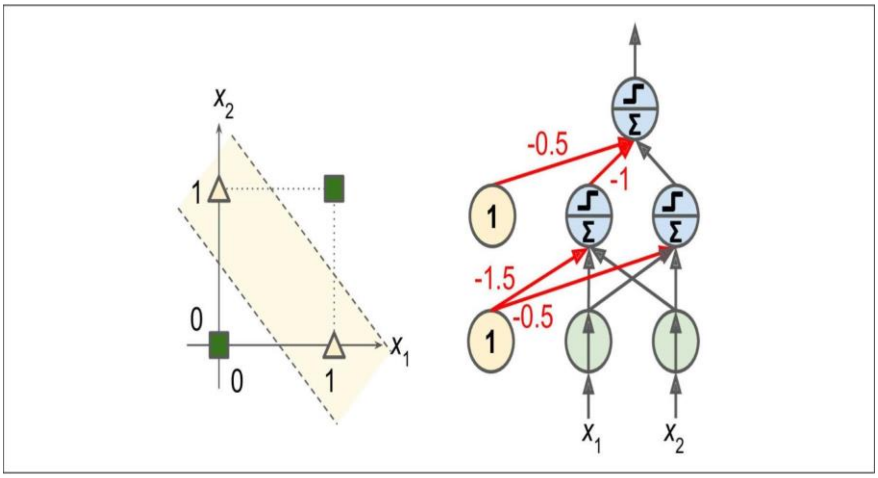


**Fig 2.7: XOR classification problem and an MLP that solves it**

## 2.6 Multi-Layer Perceptron and Backpropagation:

An MLP is composed of one (pass through) input layer, one or more layers of TLUs, called hidden layers, and one final layer of TLUs called the output layer (see Figure 10-7). The layers close to the input layer are usually called the lower layers, and the ones close to the outputs are usually called the upper layers. Every layer except the output layer includes a bias neuron and is fully connected to the next layer.

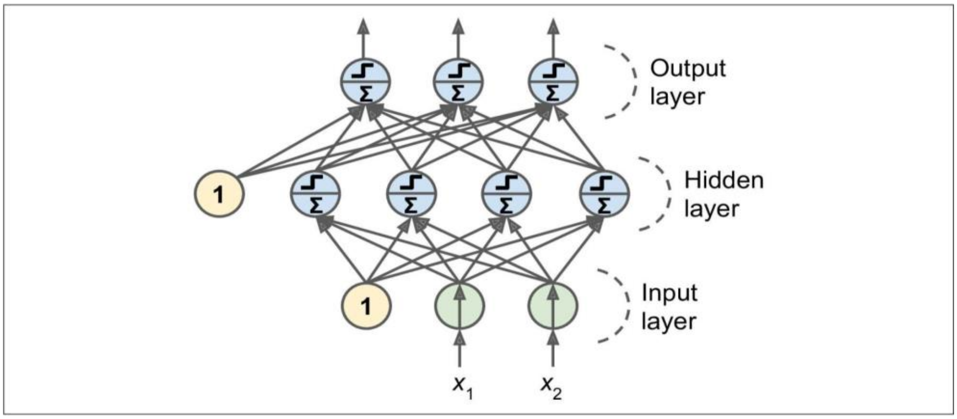


**Fig 2.8: Multi-Layer Perceptron**

The signal flows only in one direction (from the inputs to the outputs), so this architecture is an example of a feed forward neural network (FNN). When an ANN contains a deep stack of hidden layers[8], it is called a deep neural network (DNN). The field of Deep Learning studies DNNs, and more generally models containing deep stacks of computations. However, many people talk about Deep Learning whenever neural networks are involved (even shallow ones).

For many years researchers struggled to find a way to train MLPs, without success. But in 1986, David Rumelhart, Geoffrey Hinton and Ronald Williams published a groundbreaking paper[9] introducing the backpropagation training algorithm, which is still used today. In short, it is simply Gradient using an efficient technique for computing the gradients automatically[10]: in just two passes through the network (one forward, one backward), the backpropagation algorithm is able to compute the gradient of the network's error with regards to every single model parameter. In other words, it can find out how each connection weight and each bias term should be tweaked in order to reduce the error. Once it has these gradients, it just performs a regular Gradient Descent step, and the whole process is repeated until the network converges to the solution.

> Automatically computing gradients is called automatic differentiation, or autodiff. There are various autodiff techniques, with different pros and cons. The one used by backpropagation is called reverse-mode autodiff. It is fast and precise, and is well suited when the function to differentiate has many variables (e.g., connection weights) and few outputs (e.g., one loss). If you want to learn more about autodiff, check out ???.

Let's run through this algorithm in a bit more detail:

- It handles one mini-batch at a time (for example containing 32 instances each), and it goes through the full training set multiple times. Each pass is called an epoch, as we saw in
- Each mini-batch is passed to the network's input layer, which just sends it to the first hidden layer. The algorithm then computes the output of all the neurons in this layer (for every instance in the mini-batch). The result is passed on to the next layer, its output is computed and passed to the next layer, and so on until we get the output of the last layer, the output layer. This is the forward pass: it is exactly like making predictions, except all intermediate results are preserved since they are needed for the backward pass.
- Next, the algorithm measures the network's output error (i.e., it uses a loss function that compares the desired output and the actual output of the network, and returns some measure of the error).
- Then it computes how much each output connection contributed to the error. This is done analytically by simply applying the chain rule (perhaps the most fundamental rule in calculus), which makes this step fast and precise.
- The algorithm then measures how much of these error contributions came from each connection in the layer below, again using the chain rule—and so on until the algorithm reaches the input layer. As we explained earlier, this reverse pass efficiently measures the error gradient across all the connection weights in the network by propagating the error gradient backward through the network (hence the name of the algorithm).
- Finally, the algorithm performs a Gradient Descent step to tweak all the connection weights in the network, using the error gradients it just computed.

This algorithm is so important, it's worth summarizing it again: for each training instance the backpropagation algorithm first makes a prediction (forward pass), measures the error, then goes through each layer in reverse to measure the error contribution from each connection (reverse pass), and finally slightly tweaks the connection weights to reduce the error (Gradient Descent step).

It is important to initialize all the hidden layers' connection weights randomly, or else training will fail. For example, if you initialize all weights and biases to zero, then all neurons in a given layer will be perfectly identical, and thus backpropagation will affect them in exactly the same way, so they will remain identical. In other words, despite having hundreds of neurons per layer, your model will act as if it had only one neuron per layer: it won't be too smart. If instead you randomly initialize the weights, you break the symmetry and allow backpropagation to train a diverse team of neurons.

In order for this algorithm to work properly, the authors made a key change to the MLP's architecture: they replaced the step function with the logistic function, u(z) = 1 / (1 + exp(-z)). This was essential because the step function contains only flat segments, so there is no gradient to work with (Gradient Descent cannot move on a flat surface), while the logistic function has a well-defined nonzero derivative everywhere, allowing Gradient Descent to make some progress at every step. In fact, the backpropagation algorithm works well with many other activation functions, not just the logistic function. Two other popular activation functions are:

The hyperbolic tangent function tanh(z) = 2σ(2z) - 1

Just like the logistic function it is S-shaped, continuous, and differentiable, but its output value ranges from -1 to 1 (instead of 0 to 1 in the case of the logistic function), which tends to make each layer's output more or less centered around 0 at the beginning of training. This often helps speed up convergence.

The Rectified Linear Unit function: ReLU(z) = max(0, z)

It is continuous but unfortunately not differentiable at z = 0 (the slope changes abruptly, which can make Gradient Descent bounce around), and its derivative is 0 for z < 0. However, in practice it works very well and has the advantage of being fast to compute". Most importantly, the fact that it does not have a maximum output value also helps reduce some issues during Gradient Descent (we will come back to this in Chapter 11).

These popular activation functions and their derivatives are represented in Figure 10-8. But wait! Why do we need activation functions in the first place? Well, if you chain several linear transformations, all you get is a linear transformation. For example, say f(x) = 2 x + 3 and g(x) = 5 x - 1, then chaining these two linear functions gives you another linear function: f(g(x)) = 2(5 x - 1) + 3 = 10 x + 1. So if you don't have some non-linearity between layers, then even a deep stack of layers is equivalent to a single layer: you cannot solve very complex problems with that.

Okay! So now you know where neural nets came from, what their architecture is and how to compute their outputs, and you also learned about the backpropagation algorithm. But what exactly can you do with them?

### 2.7 Regression MLPs:

First, MLPs can be used for regression tasks. If you want to predict a single value (e.g., the price of a house given many of its features), then you just need a single output neuron: its output is the predicted value. For multivariate regression (i.e., to predict multiple values at once), you need one output neuron per output dimension. For example, to locate the center of an object on an image, you need to predict 2D coordinates, so you need two output neurons. If you also want to place a bounding box around the object, then you need two more numbers: the width and the height of the object. So you end up with 4 output neurons. In general, when building an MLP for regression, you do not want to use any activation function for the output neurons, so they are free to output any range of values. However, if you want to guarantee that the output will always be positive, then you can use the ReLU activation function, or the softplus activation function in the output layer. Finally, if you want to guarantee that the predictions will fall within a given range of values, then you can use the logistic function or the hyperbolic tangent, and scale the labels to the appropriate range: 0 to 1 for the logistic function, or -1 to 1 for the hyperbolic tangent.

The loss function to use during training is typically the mean squared error, but if you have a lot of outliers in the training set, you may prefer to use the mean absolute error instead. Alternatively, you can use the Huber loss, which is a combination of both.

The Huber loss is quadratic when the error is smaller than a threshold 6 (typically 1), but linear when the error is larger than 6. This makes it less sensitive to outliers than the mean squared error, and it is often more precise and converges faster than the mean absolute error.

Table 2.1 summarizes the typical architecture of a regression MLP.

Table 1 Typical Regression MLP Architecture

| Hyperparameter | Typical Value |
|---|---|
| # Input neurons | One per input feature(e.g. 28*28= 784 for MNIST) |
| # hidden layers | Depends on the problem typically 1 to 5 |
| # neurons per hidden layers | Depends on the problem typically 1 to 100 |
| # output neurons | 1 per prediction dimension |
| Hidden activation | ReLU(or SELU) |
| Output activation | None or ReLU/softplus(if positive outputs) or Logistic/Tanh(if bounded outputs) |
| Loss Function | MSE or MAE/Huber(if outlieers) |

### 2.8 Classification of MLPs:

MLPs can also be used for classification tasks. For a binary classification problem, you just need a single output neuron using the logistic activation function: the output will be a number between 0 and 1, which you can interpret as the estimated probability of the positive class. Obviously, the estimated probability of the negative class is equal to one minus that number.

MLPs can also easily handle multilabel binary classification tasks. For example, you could have an email classification system that predicts whether each incoming email is ham or spam, and simultaneously predicts whether it is an urgent or non-urgent email. In this case, you would need two output neurons, both using the logistic activation function: the first would output the probability that the email is spam and the second would output the probability that it is urgent. More generally, you would dedicate one output neuron for each positive class. Note that the output probabilities do not necessarily add up to one. This lets the model output any combination of labels: you can have non-urgent ham, urgent ham, non-urgent spam, and perhaps even urgent spam (although that would probably be an error).

If each instance can belong only to a single class, out of 3 or more possible classes (e.g., classes 0 through 9 for digit image classification), then you need to have one output neuron per class, and you should use the softmax activation function for the whole output layer (see Figure 9). The softmax function will ensure that all the estimated probabilities are between 0 and 1 and that they add up to one (which is required if the classes are exclusive). This is called multiclass classification.

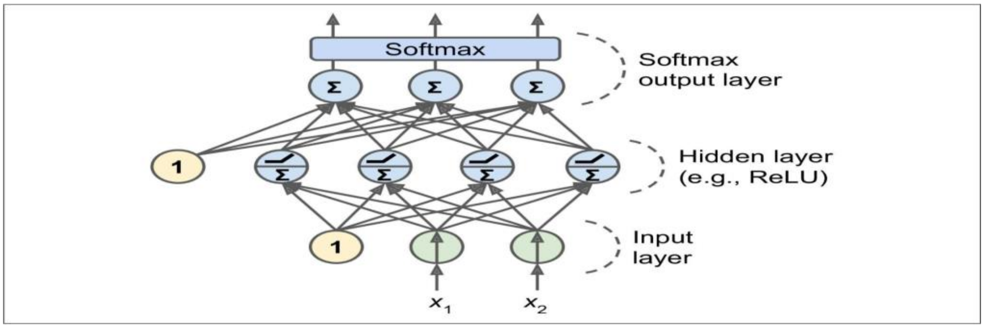


**Fig 2.9: A modern MPL (including ReLU and softmax) for classification**

Regarding the loss function, since we are predicting probability distributions, the cross-entropy (also called the log loss, see Chapter 4) is generally a good choice.

Table 2.2 summarizes the typical architecture of a classification MLP.

Table 2.2: Typical Classification MLP Architecture

| Hyperparameter | Binary classification | Multilabel binary classification | Multiclass classification |
|---|---|---|---|
| Input and hidden layer | Same as regression | Same as regression | Same as regression |
| # output neurons | 1 | 1 per label | 1 per class |
| Output layer activation | Logistic | Logistic | Softmax |

### 2.9 Deep Computer Vision Using Convolution Neural Networks:

Although IBM's Deep Blue supercomputer beat the chess world champion Garry Kasparov back in 1996, it wasn't until fairly recently that computers were able to reliably perform seemingly trivial tasks such as detecting a puppy in a picture or recognizing spoken words. Why are these tasks so effortless to us humans? The answer lies in the fact that perception largely takes place outside the realm of our consciousness, within specialized visual, auditory, and other sensory modules in our brains. By the time sensory information reaches our consciousness, it is already adorned with high-level features; for example, when you look at a picture of a cute puppy, you cannot choose not to see the puppy, or not to notice its cuteness. Nor can you explain how you recognize a cute puppy; it's just obvious to you. Thus, we cannot trust our subjective experience: perception is not trivial at all, and to understand it we must look at how the sensory modules work.

Convolutional neural networks (CNNs) emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, thanks to the increase in computational power, the amount of available training data, and the tricks presented in Chapter 11 for training deep nets, CNNs have managed to achieve superhuman performance on some complex visual tasks. They power image search services, self-driving cars, automatic video classification systems, and more. Moreover, CNNs are not restricted to visual perception: they are also successful at many other tasks, such as voice recognition or natural language processing (NLP); however, we will focus on visual applications for now

In this chapter we will present where CNNs came from, what their building blocks look like, and how to implement them using TensorFlow and Keras. Then we will discuss some of the best CNN architectures, and discuss other visual tasks, including object detection (classifying multiple objects in an image and placing bounding boxes around them) and semantic segmentation (classifying each pixel according to the class of the object it belongs to).

## 2.10 The Architecture of the Visual Cortex:

David H. Hubel and Torsten Wiesel performed a series of experiments on cats in 1958' and 1959[2] (and a few years later on monkeys[3]), giving crucial insights on the structure of the visual cortex (the authors received the Nobel Prize in Physiology or Medicine in 1981 for their work). In particular, they showed that many neurons in the visual cortex have a small local receptive field, meaning they react only to visual stimuli located in a limited region of the visual field (see Figure 10, in which the local receptive fields of five neurons are represented by dashed circles). The receptive fields of different neurons may overlap, and together they tile the whole visual field. Moreover, the authors showed that some neurons react only to images of horizontal lines, while others react only to lines with different orientations (two neurons may have the same receptive field but react to different line orientations). They also noticed that some neurons have larger receptive fields, and they react to more complex patterns that are combinations of the lower-level patterns. These observations led to the idea that the higher-level neurons are based on the outputs of neighboring lower-level neurons (in Figure. 10, notice that each neuron is connected only to a few neurons from the previous layer). This powerful architecture is able to detect all sorts of complex patterns in any area of the visual field.

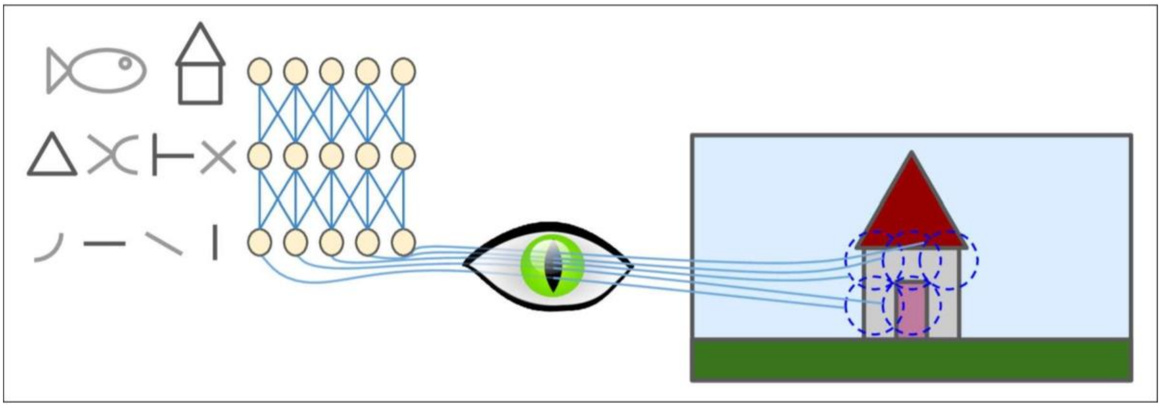


**Fig 2.10: Local receptive fields in the visual cortex**

These studies of the visual cortex inspired the neocognitron, introduced in 1980,[4] which gradually evolved into what we now call convolutional neural networks. An important milestone was a 1998 papers by Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner, which introduced the famous LeNet-5 architecture, widely used to recognize handwritten check numbers. This architecture has some building blocks that you already know, such as fully connected layers and sigmoid activation functions, but it also introduces two new building blocks: convolutional layers and pooling layers. Let's look at them now

Why not simply use a regular deep neural network with fully connected layers for image recognition tasks? Unfortunately, although this works fine for small images (e.g., MNIST), it breaks down for larger images because of the huge number of parameters it requires. For example, a 100 x 100 image has 10,000 pixels, and if the first layer has just 1,000 neurons (which already severely restricts the amount of information transmitted to the next layer), this means a total of 10 million connections. And that's just the first layer.

CNNs solve this problem using partially connected layers and weight sharing.

## 2.11 Convolutional Layer:

The most important building block of a CNN is the convolutional layer:[6] neurons in the first convolutional layer are not connected to every single pixel in the input image (like they were in previous chapters), but only to pixels in their receptive fields (see Figure 11). In turn, each neuron in the second convolutional layer is connected only to neurons located within a small rectangle in the first layer. This architecture allows the network to concentrate on small low-level features in the first hidden layer, then assemble them into larger higher-level features in the next hidden layer, and so on. This hierarchical structure is common in real-world images, which is one of the reasons why CNNs work so well for image recognition.

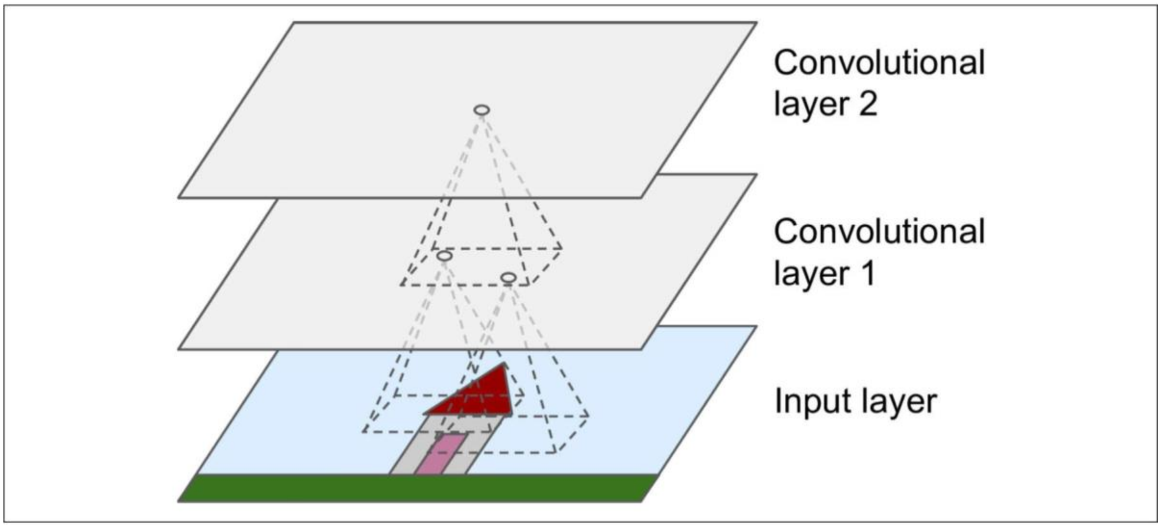


**Fig 2.11: CNN layers with rectangular local receptive fields**

Until now, all multilayer neural networks we looked at had layers composed of a long line of neurons, and we had to flatten input images to 1D before feeding them to the neural network. Now each layer is represented in 2D, which makes it easier to match neurons with their corresponding inputs.

A neuron located in row i, column j of a given layer is connected to the outputs of the neurons in the previous layer located in rows i to i + $f_h$ - 1, columns j to j + f,,, - 1, where $f_h$ and f,,, are the height and width of the receptive field (see Figure 14-3). In order for a layer to have the same height and width as the previous layer, it is common to add zeros around the inputs, as shown in the diagram. This is called zero padding.

It is also possible to connect a large input layer to a much smaller layer by spacing out the receptive fields, as shown in Figure 13. The shift from one receptive field to the next is called the stride. In the diagram, a 5 x 7 input layer (plus zero padding) is connected to a 3 x 4 layer, using 3 x 3 receptive fields and a stride of 2 (in this example the stride is the same in both directions, but it does not have to be so). A neuron located in row i, column j in the upper layer is connected to the outputs of the neurons in the previous layer located in rows i x $s_h$ to i x $s_h$ + $f_h$ — 1, columns j x s,, to j x $s_w$ + f, —1, where $s_h$ and $s_w$ are the vertical and horizontal strides.

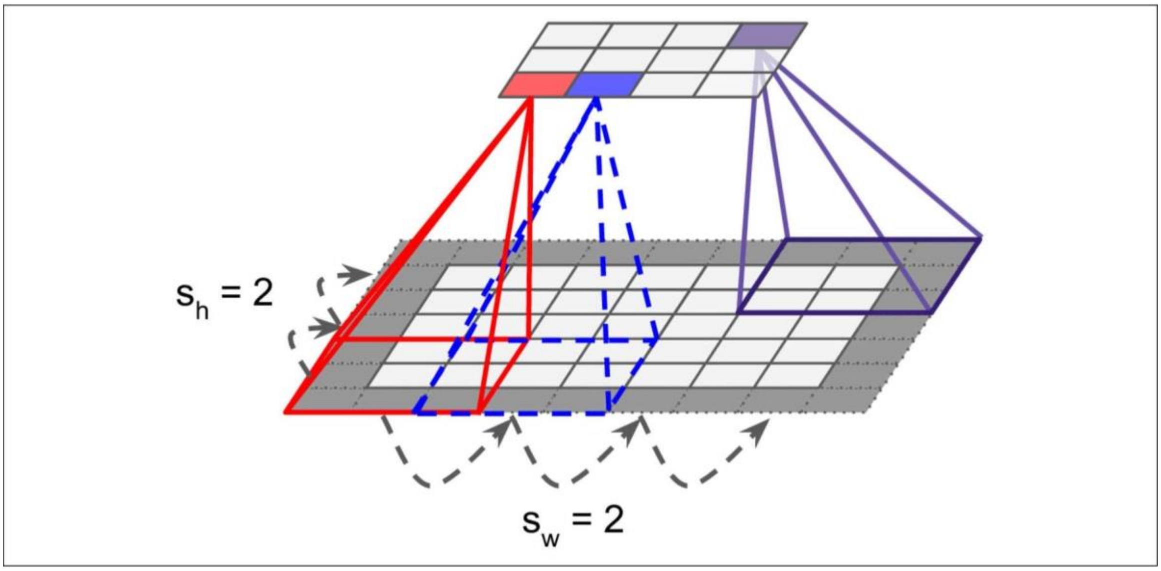


**Fig 2.12 Reducing dimensionality using a stride of 2**

### 2.12 Filters:

A neuron's weights can be represented as a small image the size of the receptive field. For example, Figure 14-5 shows two possible sets of weights, called filters (or convolution kernels). The first one is represented as a black square with a vertical white line in the middle (it is a 7 x 7 matrix full of Os except for the central column, which is full of 1s); neurons using these weights will ignore everything in their receptive field except for the central vertical line (since all inputs will get multiplied by 0, except for the ones located in the central vertical line). The second filter is a black square with a horizontal white line in the middle. Once again, neurons using these weights will ignore everything in their receptive field except for the central horizontal line.

Now if all neurons in a layer use the same vertical line filter (and the same bias term), and you feed the network the input image shown in Figure 14-5 (bottom image), the layer will output the top-left image. Notice that the vertical white lines get enhanced while the rest gets blurred. Similarly, the upper-right image is what you get if all neurons use the same horizontal line filter; notice that the horizontal white lines get enhanced while the rest is blurred out. Thus, a layer full of neurons using the same filter outputs a feature map, which highlights the areas in an image that activate the filter the most. Of course you do not have to define the filters manually: instead, during training the convolutional layer will automatically learn the most useful filters for its task, and the layers above will learn to combine them into more complex patterns.

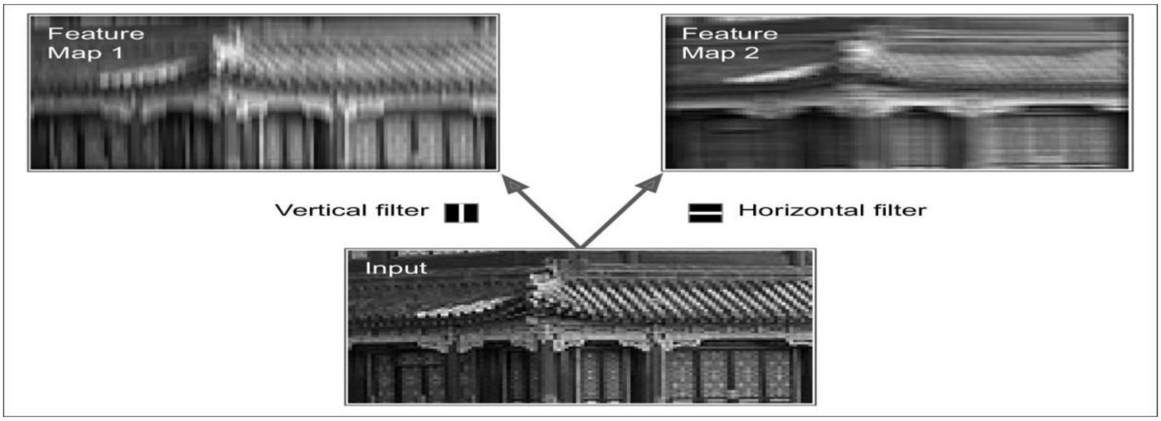


**Fig 2.13 Applying two different filters to get two feature maps**

### 2.13 Stacking Multiple Feature Maps:

Up to now, for simplicity, I have represented the output of each convolutional layer as a thin 2D layer, but in reality a convolutional layer has multiple filters (you decide how many), and it outputs one feature map per filter, so it is more accurately represented in 3D (see Figure 14-6). To do so, it has one neuron per pixel in each feature map, and all neurons within a given feature map share the same parameters (i.e., the same weights and bias term). However, neurons in different feature maps use different parameters. A neuron's receptive field is the same as described earlier, but it extends across all the previous layers' feature maps. In short, a convolutional layer simultaneously applies multiple trainable filters to its inputs, making it capable of detecting multiple features anywhere in its inputs.

The fact that all neurons in a feature map share the same parameter dramatically reduces the number of parameters in the model. Moreover, once th CNN has learned to recognize a pattern in one location, it can recognize it ir any other location. In contrast, once a regular DNN has learned to recognize pattern in one location, it can recognize it only in that particular location.

Moreover, input images are also composed of multiple sublayers: one per color channel. There are typically three: red, green, and blue (RGB). Grayscale images have just one channel, but some images may have much more—for example, satellite images.

Specifically, a neuron located in row i, column j of the feature map k in a given convolutional layer 1 is connected to the outputs of the neurons in the previous layer 1 - 1, located in rows i x $s_h$, to i x $s_h$ + $f_h$ - 1 and columns j x $s_i$, to j x + $f_W$ - 1, across all feature maps (in layer 1 - 1). Note that all neurons located in the same row i and column j but in different feature maps are connected to the outputs of the exact same neurons in the previous layer.

Equation 4 summarizes the preceding explanations in one big mathematical equation: it shows how to compute the output of a given neuron in a convolutional layer.

It is a bit ugly due to all the different indices, but all it does is calculate the weighted sum of all the inputs, plus the bias term.

Equation 4 Computing the output of a neuron in a convolutional layer $f_h$- 1 f $_w$— 1 f $_{n'}$ —1

1·/ —— i X Sh+ 14

$Z_{i, j, k} = b_k + 1$ x x.1,. •i, with

u = ,:i v 0 k' = 0 I k' u, v, k', k I = i X $s_w$ + v

- k is the output of the neuron located in row i, column j in feature map k of the convolutional layer (layer 1).

- As explained earlier, $s_h$ and s,,, are the vertical and horizontal strides, $f_h$ and $L_i$ are the height and width of the receptive field, and $f_n$i is the number of feature maps in the previous layer (layer / - 1).

- xi,1r, le is the output of the neuron located in layer 1 - 1, row i', column j', feature map k' (or channel k' if the previous layer is the input layer).

- $b_k$ is the bias term for feature map k (in layer 1). You can think of it as a knob that tweaks the overall brightness of the feature map k.

- Wu, v, k1 ,k is the connection weight between any neuron in feature map k of the layer / and its input located at row u, column v (relative to the neuron's receptive field), and feature map k'.

### 2.14 TensorFlow Implementation:

In TensorFlow, each input image is typically represented as a 3D tensor of shape [height, width, channels]. A mini-batch is represented as a 4D tensor of shape [mini-batch size, height, width, channels]. The weights of a convolutional layer are represented as a 4D tensor of shape [fp f,,, $f_{i1}$l, $f_n$]. The bias terms of a convolutional layer are simply represented as a 1D tensor of shape [ $f_n$].

Let's look at a simple example. The following code loads two sample images, using Scikit-Learn's load_sample_images() (which loads two color images, one of a Chinese temple, and the other of a flower). The pixel intensities (for each color channel) is represented as a byte from 0 to 255, so we scale these features simply by dividing by 255, to get floats ranging from 0 to 1. Then we create two 7 x 7 filters (one with a vertical white line in the middle, and the other with a horizontal white line in the middle), and we apply them to both images using the tf.nn.conv2d() function, which is part of TensorFlow's low-level Deep Learning API. In this example, we use zero padding (padding="SAME") and a stride of 2. Finally, we plot one of the resulting feature maps (similar to the top-right image in Figure 14).

```
    from sklearn.datasets import load_sample_image

    # Load sample images

china = load_sample_image("china.jpg") / 255 flower = load_sample_image("flower.jpg") / 255
images = np.array([china, flower])

    batch_size, height, width, channels = images.shape

    # Create 2 filters

    filters = np.zeros(shape=(7, 7, channels, 2), dtype=np.float32)

outputs = tf.nn.conv2d(images, filters, strides=1, padding="SAME") plt.imshow(outputs[0, :, :, 1],
cmap="gray") # plot 1st image's 2nd feature map plt.show()
```

Most of this code is self-explanatory, but the tf . nn . conv2d( ) line deserves a bit of explanation:

- images is the input mini-batch (a 4D tensor, as explained earlier).
- filters is the set of filters to apply (also a 4D tensor, as explained earlier).
- strides is equal to 1, but it could also be a 1D array with 4 elements, where the two central elements are the vertical and horizontal strides ($s_h$ and $s_{i,}$). The first and last elements must currently be equal to 1. They may one day be used to specify a batch stride (to skip some instances) and a channel stride (to skip some of the previous layer's feature maps or channels).
- padding must be either **"VALID"** or **"SAME":**

  If set to "VALID", the convolutional layer does *not* use zero padding, and may ignore some rows and columns at the bottom and right of the input image, depending on the stride, as shown in Figure 14-7 (for simplicity, only the horizontal dimension is shown here, but of course the same logic applies to the vertical dimension).

— If set to "SAME", the convolutional layer uses zero padding if necessary. In this case, the number of output neurons is equal to the number of input neurons divided by the stride, rounded up (in this example, 13 / 5 = 2.6, rounded up to 3). Then zeros are added as evenly as possible around the inputs. In this example, we manually defined the filters, but in a real CNN you would normally define filters as trainable variables, so the neural net can learn which filters work best, as explained earlier. Instead of manually creating the variables, however, you can simply use the ke r as . layers . Conv2D layer:

```
cony = keras.layers.Conv2D(filters=32, kernel_size=3, strides=1,

                          padding="SAME", activation="relu")
```

This code creates a Conv2D layer with 32 filters, each 3 x 3, using a stride of 1 (both horizontally and vertically), SAME padding, and applying the ReLU activation function to its outputs. As you can see, convolutional layers have quite a few hyperparameters: you must choose the number of filters, their height and width, the strides, and the padding type. As always, you can use cross-validation to find the right hyperparameter values, but this is very time-consuming. We will discuss common CNN architectures later, to give you some idea of what hyperparameter values work best in practice.

### 2.15: Memory Requirements:

Another problem with CNNs is that the convolutional layers require a huge amount of RAM. This is especially true during training, because the reverse pass of backpropagation requires all the intermediate values computed during the forward pass.

For example, consider a convolutional layer with 5 x 5 filters, outputting 200 feature maps of size 150 x 100, with stride 1 and SAME padding. If the input is a 150 x 100 RGB image (three channels), then the number of parameters is (5 x 5 x 3 + 1) x 200 = 15,200 (the +1 corresponds to the bias terms), which is fairly small compared to a fully connected layer.' However, each of the 200 feature maps contains 150 x 100 neurons, and each of these neurons needs to compute a weighted sum of its 5 x 5 x 3 = 75 inputs: that's a total of 225 million float multiplications. Not as bad as a fully connected layer, but still quite computationally intensive. Moreover, if the feature maps are represented using 32-bit floats, then the convolutional layer's output will occupy 200 x 150 x 100 x 32 = 96 million bits (12 MB) of RAM.[8] And that's just for one instance! If a training batch contains 100 instances, then this layer will use up 1.2 GB of RAM!

During inference (i.e., when making a prediction for a new instance) the RAM occupied by one layer can be released as soon as the next layer has been computed, so you only need as much RAM as required by two consecutive layers. But during training everything computed during the forward pass needs to be preserved for the reverse pass, so the amount of RAM needed is (at least) the total amount of RAM required by all layers.

If training crashes because of an out-of-memory error, you can try reducing the mini-batch size. Alternatively, you can try reducing dimensionality using a stride, or removing a few layers. Or you can try using 16-bit floats instead of 32-bit floats. Or you could distribute the CNN across multiple devices.

Now let's look at the second common building block of CNNs: the *pooling layer.*

**2.16: Pooling Layer:**

Once you understand how convolutional layers work, the pooling layers are quite easy to grasp. Their goal is to *subsample* (i.e., shrink) the input image in order to reduce the computational load, the memory usage, and the number of parameters (thereby limiting the risk of overfitting).

Just like in convolutional layers, each neuron in a pooling layer is connected to the outputs of a limited number of neurons in the previous layer, located within a small rectangular receptive field. You must define its size, the stride, and the padding type, just like before. However, a pooling neuron has no weights; all it does is aggregate the inputs using an aggregation function such as the max or mean. Figure 14-8 shows a *max pooling layer,* which is the most common type of pooling layer. In this example,

we use a 2 x 2 _pooling kerne12, with a stride of 2, and no padding. Only the max input value in each receptive field makes it to the next layer, while the other inputs are dropped. For example, in the lower left receptive field in Figure 14-8, the input values are 1, 5, 3, 2, so only the max value, 5, is propagated to the next layer. Because of the stride of 2, the output image has half the height and half the width of the input image (rounded down since we use no padding).

A pooling layer typically works on every input channel independently, so the output depth is the same as the input depth.

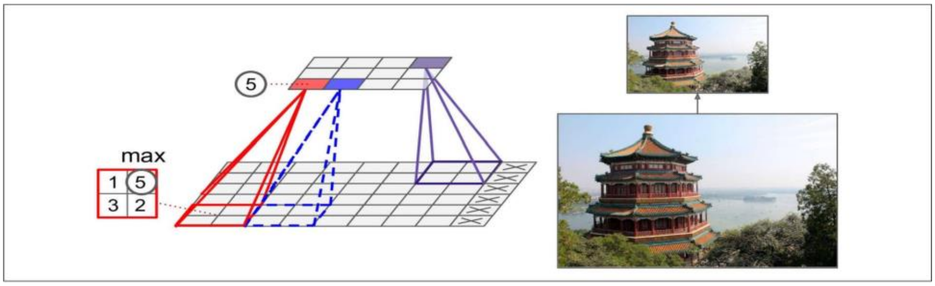


**Fig 2.14: Max pooling layer (2 x 2 pooling kernel, stride 2, no padding)**

Other than reducing computations, memory usage and the number of parameters, a max pooling layer also introduces some level of *invariance* to small translations, as shown in Figure 14-9. Here we assume that the bright pixels have a lower value than dark pixels, and we consider 3 images (A, B, C) going through a max pooling layer with a 2 x 2 kernel and stride 2. Images B and C are the same as image A, but shifted by one and two pixels to the right. As you can see, the outputs of the max pooling layer for images A and B are identical. This is what translation invariance means. However, for image C, the output is different: it is shifted by one pixel to the right (but there is still 75% invariance). By inserting a max pooling layer every few layers in a CNN, it is possible to get some level of translation invariance at a larger scale. Moreover, max pooling also offers a small amount of rotational invariance and a slight scale invariance. Such invariance (even if it is limited) can be useful in cases where the predictions should not depend on these details, such as in classification tasks.

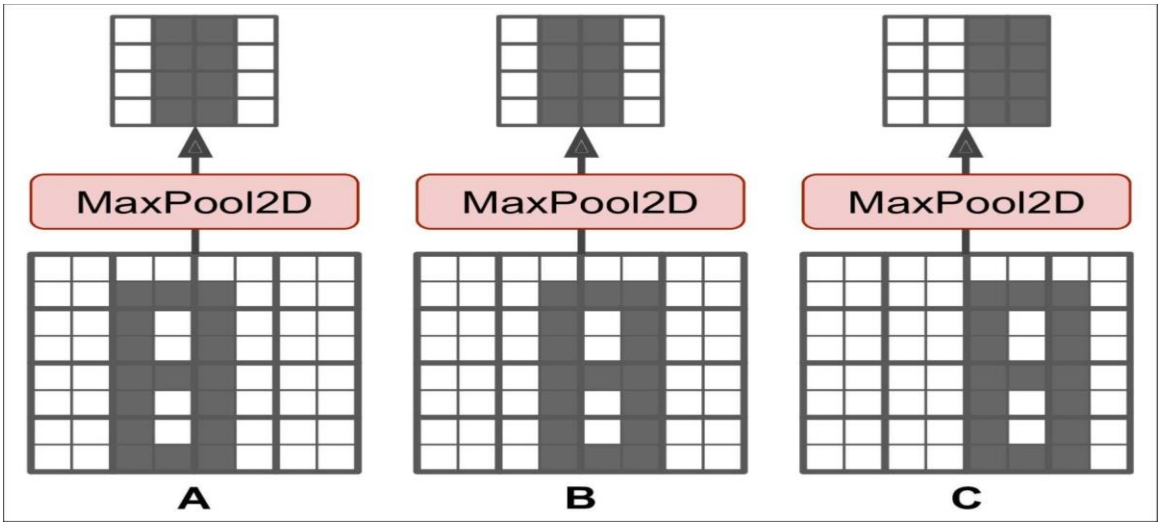


**Fig 2.15: Invariance to small translations**

But max pooling has some downsides: firstly, it is obviously very destructive: even with a tiny 2 x 2 kernel and a stride of 2, the output will be two times smaller in both directions (so its area will be four times smaller), simply dropping 75% of the input values. And in some applications, invariance is not desirable, for example for *semantic segmentation:* this is the task of classifying each pixel in an image depending on the object that pixel belongs to: obviously, if the input image is translated by 1 pixel to the right, the output should also be translated by 1 pixel to the right. The goal in this case is *equivariance,* not invariance: a small change to the inputs should lead to a corresponding small change in the output.

### 2.17 TensorFlow Implementation:

Implementing a max pooling layer in TensorFlow is quite easy. The following code creates a max pooling layer using a 2 x 2 kernel. The strides default to the kernel size, so this layer will use a stride of 2 (both horizontally and vertically). By default, it uses VALID padding (i.e., no padding at all):

```
max_pool = keras.layers.MaxPool2D(pool_size=2)
```

To create an *average pooling layer,* just use Avg Pool2D instead of MaxPool2D. As you might expect, it works exactly like a max pooling layer, except it computes the mean rather then the max. Average pooling layers used to be very popular, but people mostly use max pooling layers now, as they generally perform better. This may seem surprising, since computing the mean generally loses less information than computing the max. But on the other hand, max pooling preserves only the strongest feature, getting rid of all the meaningless ones, so the next layers get a cleaner signal to work with. Moreover, max pooling offers stronger translation invariance than average pooling.

Note that max pooling and average pooling can be performed along the depth dimension rather than the spatial dimensions, although this is not as common. This can allow the CNN to learn to be invariant to various features. For example, it could learn multiple filters, each detecting a different rotation of the same pattern, such as handwritten digits (see Figure 14-10), and the depth-wise max pooling layer would ensure that the output is the same regardless of the rotation. The CNN could similarly learn to be invariant to anything else: thickness, brightness, skew, color, and so on.

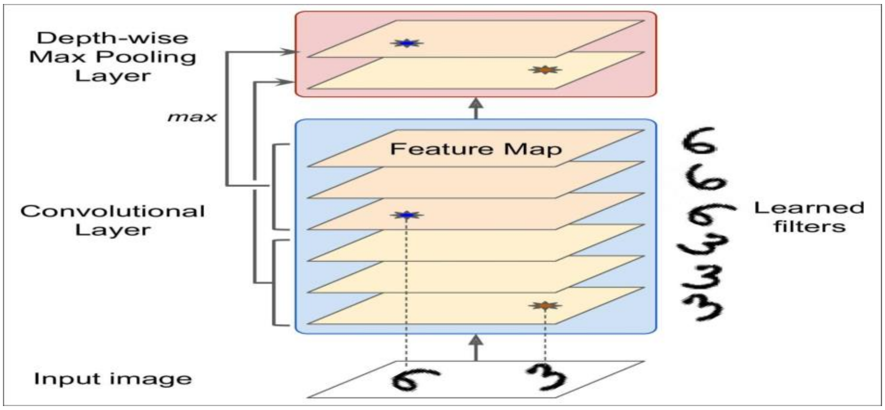


**Fig 2.16 Depth-wise max pooling can help the CNN learn any invariance**

Keras does not include a depth-wise max pooling layer, but TensorFlow's low-level Deep Learning API does: just use the tf nn max_pool( ) function, and specify the kernel size and strides as 4-tuples. The first three values of each should be 1: this indicates that the kernel size and stride along the batch, height and width dimensions shoud be 1. The last value should be whatever kernel size and stride you want along the depth dimension, for example 3 (this must be a divisor of the input depth; for example, it will not work if the previous layer outputs 20 feature maps, since 20 is not a multiple of 3):

```
output = tf.nn,max_pool(images, ksize=(1, 1, 1, 3), strides=(1, 1, 1, 3), padding="VALID")
```

If you want to include this as a layer in your Keras models, you can simply wrap it in a Lambda layer (or create a custom Keras layer):

```
depth_pool = keras.layers.Lambda(lambda X: tf.nn.max_pool(X, ksize.(1, 1, 1, 3), strides=(1, 1, 1, 3),
padding="VALID"))
```

One last type of pooling layer that you will often see in modern architectures is the global average pooling layer. It works very differently: all it does is compute the mean of each entire feature map (it's like an average pooling layer using a pooling kernel with the same spatial dimensions as the inputs). This means that it just outputs a single number per feature map and per instance. Although this is of course extremely destructive (most of the information in the feature map is lost), it can be useful as the output layer, as we will see later in this chapter. To create such a layer, simply use the keras .layers.GlobalAvgPool2D class:

```
    global_avg_pool = keras.layers.GlobalAvgPool2D()
```

It is actually equivalent to this simple Lamba layer, which computes the mean over the spatial dimensions (height and width):

global_avg_pool = keras.layers.Lambda(lambda X: tf.reduce_mean(X, axis=[1, 2])) Now you know all the building blocks to create a convolutional neural network. Let's see how to assemble them.

## 2.18 CNN Architectures:

Typical CNN architectures stack a few convolutional layers (each one generally followed by a ReLU layer), then a pooling layer, then another few convolutional layers (+ReLU), then another pooling layer, and so on. The image gets smaller and smaller as it progresses through the network, but it also typically gets deeper and deeper (i.e., with more feature maps) thanks to the convolutional layers (see Figure 14-11). At the top of the stack, a regular feedforward neural network is added, composed of a few fully connected layers (+ReLUs), and the final layer outputs the prediction (e.g., a softmax layer that outputs estimated class probabilities).

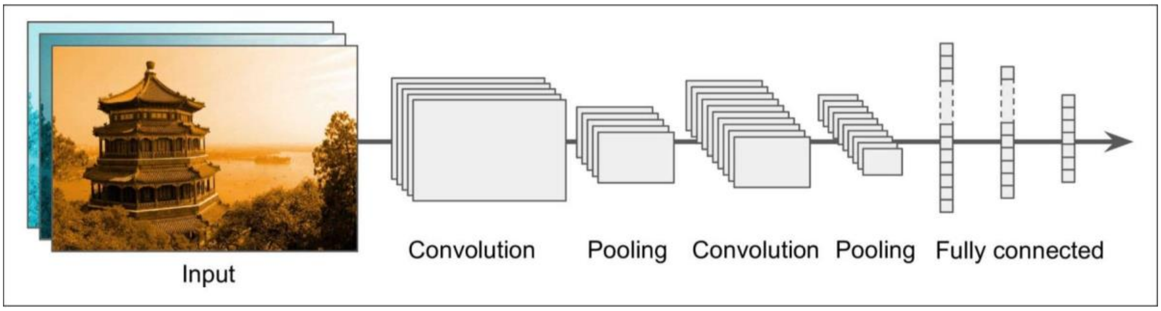


**Fig 2.18: Typical CNN architecture**

A common mistake is to use convolution kernels that are too large. For example, instead of using a convolutional layer with a 5 x 5 kernel, it is generally preferable to stack two layers with 3 x 3 kernels: it will use less parameters and require less computations, and it will usually perform better. One exception to this recommendation is for the first convolutional layer: it can typically have a large kernel (e.g., 5 x 5), usually with stride of 2 or more: this will reduce the spatial dimension of the image without losing too much information, and since the input image only has 3 channels in general, it will not be too costly.

Here is how you can implement a simple CNN to tackle the fashion MNIST dataset (introduced in Chapter 10):

```
from functools import partial

DefaultConv2D = partial(keras.layers.Conv2D,

                        kernel_size=3, activation='relu', padding="SAME")

model = keras.models.Sequential([

    DefaultConv2D(filters=64, kernel_size=7, input_shape=[28, 28, 1]),
    keras.layers.MaxPooling2D(pool_size=2),

   DefaultConv2D(filters=128),                              DefaultConv2D(filters=128),
   keras.layers.MaxPooling2D(pool_size=2),

   DefaultConv2D(filters=256),                              DefaultConv2D(filters=256),
   keras.layers.MaxPooling2D(pool_size=2),
    keras.layers.Flatten(), keras.layers.Dense(units=128, activation='relu'),
    keras.layers.Dropout(0.5), keras.layers.Dense(units=64, activation='relu'),
    keras.layers.Dropout(0.5), keras.layers.Dense(units=10, activation='softmax'),
])
```

- In this code, we start by using the partial( ) function to define a thin wrapper around the Conv2D class, called DefaultConv2D: it simply avoids having to repeat the same hyperparameter values over and over again.
- The first layer uses a large kernel size, but no stride because the input images are not very large. It also sets input_shape=[28, 28, 1], which means the images are 28 x 28 pixels, with a single color channel (i.e., grayscale).
- Next, we have a max pooling layer, which divides each spatial dimension by a factor of two (since pool_size=2).
- Then we repeat the same structure twice: two convolutional layers followed by a max pooling layer. For larger images, we could repeat this structure several times (the number of repetitions is a hyperparameter you can tune).
- Note that the number of filters grows as we climb up the CNN towards the output layer (it is initially 64, then 128, then 256): it makes sense for it to grow, since the number of low level features is often fairly low (e.g., small circles, horizontal lines, etc.), but there are many different ways to combine them into higher level features. It is a common practice to double the number of filters after each pooling layer: since a pooling layer divides each spatial dimension by a factor of 2, we can afford doubling the number of feature maps in the next layer, without fear of exploding the number of parameters, memory usage, or computational load.
- Next is the fully connected network, composed of 2 hidden dense layers and a dense output layer. Note that we must flatten its inputs, since a dense network expects a 1D array of features for each instance. We also add two dropout layers, with a dropout rate of 50% each, to reduce overfitting.

This CNN reaches over 92% accuracy on the test set. It's not the state of the art, but it is pretty good, and clearly much better than what we achieved with dense networks in Chapter 10.

Over the years, variants of this fundamental architecture have been developed, leading to amazing advances in the field. A good measure of this progress is the error rate in competitions such as the ILSVRC ImageNet challenge. In this competition the top-5 error rate for image classification fell from over 26% to less than 2.3% in just six years. The top-five error rate is the number of test images for which the system's top 5 predictions did not include the correct answer. The images are large (256 pixels high) and there are 1,000 classes, some of which are really subtle (try distinguishing 120 dog breeds). Looking at the evolution of the winning entries is a good way to understand how CNNs work.

We will first look at the classical LeNet-5 architecture (1998), then three of the winners of the ILSVRC challenge: AlexNet (2012), GoogLeNet (2014), and ResNet (2015).

### 2.19: LeNet-5:

The LeNet-5 architecture" is perhaps the most widely known CNN architecture. As mentioned earlier, it was created by Yann LeCun in 1998 and widely used for handwritten digit recognition (MNIST). It is composed of the layers shown in Table 14-1.

**Table 2.3: LeNet-5 architecture**

| Layer | Type | Maps | Size | Kernel size | Stride | Activation |
|---|---|---|---|---|---|---|
| Out | Fully Connected | — | 10 | — | — | RBF |
| F6 | Fully Connected | — | 84 | — | — | tanh |
| C5 | Convolution | 120 | 1 x 1 | 5 x 5 | 1 | tanh |
| S4 | Avg Pooling | 16 | 5 x 5 | 2 x 2 | 2 | tanh |
| C3 | Convolution | 16 | 10 x 10 | 5 x 5 | 1 | tanh |
| S2 | Avg Pooling | 6 | 14 x 14 | 2 x 2 | 2 | tanh |
| Cl | Convolution | 6 | 28 x 28 | 5 x 5 | 1 | tanh |
| In | Input | 1 | 32 x 32 | | | |

There are a few extra details to be noted:

- MNIST images are 28 x 28 pixels, but they are zero-padded to 32 x 32 pixels and normalized before being fed to the network. The rest of the network does not use any padding, which is why the size keeps shrinking as the image progresses through the network.
- The average pooling layers are slightly more complex than usual: each neuron computes the mean of its inputs, then multiplies the result by a learnable coefficient (one per map) and adds a learnable bias term (again, one per map), then finally applies the activation function.
- Most neurons in C3 maps are connected to neurons in only three or four S2 maps (instead of all six S2 maps). See table 1 (page 8) in the original paper" for details.

The output layer is a bit special: instead of computing the matrix multiplication of the inputs and the weight vector, each neuron outputs the square of the Euclidian distance between its input vector and its weight vector. Each output measures how much the image belongs to a particular digit class. The cross entropy cost function is now preferred, as it penalizes bad predictions much more, producing larger gradients and converging faster.

Yann LeCun's website ("LENET" section) features great demos of LeNet-5 classifying digits.

### 2.19: AlexNet:

The *AlexNet* CNN architecturell won the 2012 ImageNet ILSVRC challenge by a large margin: it achieved 17% top-5 error rate while the second best achieved only 26%! It was developed by Alex Krizhevsky (hence the name), Ilya Sutskever, and Geoffrey Hinton. It is quite similar to LeNet-5, only much larger and deeper, and it was the first to stack convolutional layers directly on top of each other, instead of stacking a pooling layer on top of each convolutional layer. Table 14-2 presents this architecture.

**Table 2.4: AlexNet architecture**

| LayerType | | Maps | Size | Kernel size | Stride | Padding | Activation |
|---|---|---|---|---|---|---|---|
| Out | Fully Connected | | 1,000 | | — | — | Softmax |
| F9 | Fully Connected | — | 4,096 | — | — | — | ReLU |
| F8 | Fully Connected | — | 4,096 | — | — | — | ReLU |
| C7 | Convolution | 256 | 13 x 13 | 3 x 3 | 1 | SAME | ReLU |
| C6 | Convolution | 384 | 13 x 13 | 3 x 3 | 1 | SAME | ReLU |
| C5 | Convolution | 384 | 13 x 13 | 3 x 3 | 1 | SAME | ReLU |
| S4 | Max Pooling | 256 | 13 x 13 | 3 x 3 | 2 | VALID | |
| C3 | Convolution | 256 | 27 x 27 | 5 x 5 | 1 | SAME | ReLU |
| S2 | Max Pooling | 96 | 27 x 27 | 3 x 3 | 2 | VALID | — |
| C1 | Convolution | 96 | 55 x 55 | 11 x 11 | 4 | VALID | ReLU |
| In | Input | 3 (RGB) | 227 x 227 | — | — | — | — |

To reduce overfitting, the authors used two regularization techniques: first they applied dropout (introduced in Chapter 11) with a 50% dropout rate during training to the outputs of layers F8 and F9. Second, they performed *data augmentation* by randomly shifting the training images by various offsets, flipping them horizontally, and changing the lighting conditions.

### 2.20: Data Augmentation:

Data augmentation artificially increases the size of the training set by generating many realistic variants of each training instance. This reduces overfitting, making this ideally, given an image from the augmented training set, a human should not be able to tell whether it was augmented or not. Moreover, simply adding white noise will not help; the modifications should be learnable (white noise is not).

For example, you can slightly shift, rotate, and resize every picture in the training set by various amounts and add the resulting pictures to the training set (see Figure 14-12). This forces the model to be more tolerant to variations in the position, orientation, and size of the objects in the pictures. If you want the model to be more tolerant to different lighting conditions, you can similarly generate many images with various contrasts. In general, you can also flip the pictures horizontally (except for text, and other non-symmetrical objects). By combining these transformations, you can greatly increase the size of your training set.

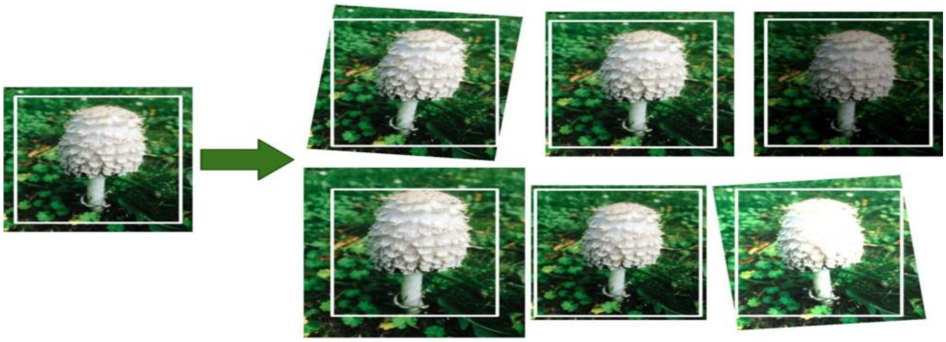


**Fig 2.19: Generating new training instances from existing**

AlexNet also uses a competitive normalization step immediately after the ReLU step of layers C1 and C3, called *local response normalization*. The most strongly activated neurons inhibit other neurons located at the same position in neighboring feature maps (such competitive activation has been observed in biological neurons). This encourages different feature maps to specialize, pushing them apart and forcing them to explore a wider range of features, ultimately improving generalization. Equation 14-2 shows how to apply LRN.

'high

)high = min $(i + {}_{2}—^{r}, {}_{f11—} 1)$

—P

max $(0, i — —^{r}$

with

$b_i = a_i k + aa.^2$

'low =

2

-'*low*

**Equation 14-2. Local response normalization**

- *b,* is the normalized output of the neuron located in feature map *i,* at some row *u* and column v (note that in this equation we consider only neurons located at this row and column, so *u* and v are not shown).
- $a_t$ is the activation of that neuron after the ReLU step, but before normalization.
- *k, a, /3,* and r are hyperparameters. *k* is called the *bias,* and r is called the *depth radius.*
- *f,* is the number of feature maps.

For example, if r = 2 and a neuron has a strong activation, it will inhibit the activation of the neurons located in the feature maps immediately above and below its own.

In AlexNet, the hyperparameters are set as follows: r = 2, *a* = 0.00002, *j3* = 0.75, and *k* = 1. This step can be implemented using the tf nn local_response_normali z a tion ( ) function (which you can wrap in a Lambda layer if you want to use it in a Keras model).

A variant of AlexNet called *ZF Net* was developed by Matthew Zeiler and Rob Fergus and won the 2013 ILSVRC challenge. It is essentially AlexNet with a few tweaked hyperparameters (number of feature maps, kernel size, stride, etc.).

### 2.21 GoogLeNet:

The GoogLeNet architecture was developed by Christian Szegedy et al. from Google Research,[12] and it won the ILSVRC 2014 challenge by pushing the top-5 error rate below 7%. This great performance came in large part from the fact that the network was much deeper than previous CNNs (see Figure 14-14). This was made possible by sub-networks called *inception modules,*[13] which allow GoogLeNet to use parameters much more efficiently than previous architectures: GoogLeNet actually has 10 times fewer parameters than AlexNet (roughly 6 million instead of 60 million).

Figure 14-13 shows the architecture of an inception module. The notation "3 x 3 + 1(S)" means that the layer uses a 3 x 3 kernel, stride 1, and SAME padding. The input signal is first copied and fed to four different layers. All convolutional layers use the ReLU activation function. Note that the second set of convolutional layers uses different kernel sizes (1 x 1, 3 x 3, and 5 x 5), allowing them to capture patterns at different scales. Also note that every single layer uses a stride of 1 and SAME padding (even the max pooling layer), so their outputs all have the same height and width as their inputs. This makes it possible to concatenate all the outputs along the depth dimension in the final *depth concat layer* (i.e., stack the feature maps from all four top convolutional layers). This concatenation layer can be implemented in TensorFlow using the tf .concat() operation, with axis=3 (axis 3 is the depth).

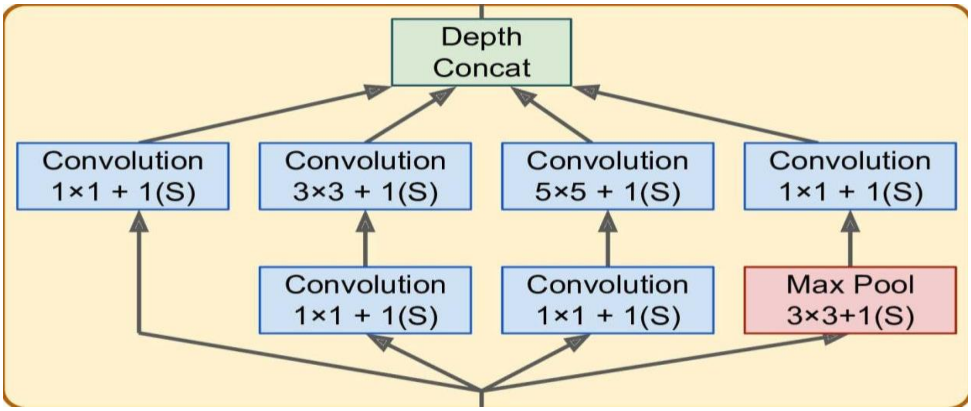


**Fig 2.20 Inception module**

You may wonder why inception modules have convolutional layers with 1 x 1 kernels. Surely these layers cannot capture any features since they look at only one pixel at a time? In fact, these layers serve three purposes:

- First, although they cannot capture spatial patterns, they can capture patterns along the depth dimension.
- Second, they are configured to output fewer feature maps than their inputs, so they serve as *bottleneck layers,* meaning they reduce dimensionality. This cuts the computational cost and the number of parameters, speeding up training and improving generalization.
- Lastly, each pair of convolutional layers ([1 x 1, 3 x 3] and [1 x 1, 5 x 5]) acts like a single, powerful convolutional layer, capable of capturing more complex patterns. Indeed, instead of sweeping a simple linear classifier across the image (as a single convolutional layer does), this pair of convolutional layers sweeps a two-layer neural network across the image as a single convolutional layer does), this pair of convolutional layers sweeps a two-layer neural network across the image.

In short, you can think of the whole inception module as a convolutional layer on steroids, able to output feature maps that capture complex patterns at various scales.

The number of convolutional kernels for each convolutional layer is a hyperparameter. Unfortunately, this means that you have six more hyperparameters to tweak for every inception layer you add.

Now let's look at the architecture of the GoogLeNet CNN (see Figure 14-14). The number of feature maps output by each convolutional layer and each pooling layer is shown before the kernel size. The architecture is so deep that it has to be represented in three columns, but GoogLeNet is actually one tall stack, including nine inception modules (the boxes with the spinning tops). The six numbers in the inception modules represent the number of feature maps output by each convolutional layer in the module (in the same order as in Figure 14-13). Note that all convolution layers use the ReLU activation function through this network:

- The first two layers divide the image's height and width by 4 (so its area is divided by 16), to reduce the computational load. The first layer uses a large kernel size, so that much of the information is still preserved.
- Then the local response normalization layer ensures that the previous layers learn a wide variety of features (as discussed earlier).
- Two convolutional layers follow, where the first acts like a *bottleneck layer*. As explained earlier, you can think of this pair as a single smarter convolutional layer.
- Again, a local response normalization layer ensures that the previous layers capture a wide variety of patterns.
- Next a max pooling layer reduces the image height and width by 2, again to speed up computations.
- Then comes the tall stack of nine inception modules, interleaved with a couple max pooling layers to reduce dimensionality and speed up the net.
- Next, the global average pooling layer simply outputs the mean of each feature map: this drops any remaining spatial information, which is fine since there was not much spatial information left at that point. Indeed, GoogLeNet input images are typically expected to be 224 x 224 pixels, so after 5 max pooling layers, each dividing the height and width by 2, the feature maps are down to 7 x 7. Moreover, it is a classification task, not localization, so it does not matter where the object is. Thanks to the dimensionality reduction brought by this layer, there is no need to have several fully connected layers at the top of the CNN (like in AlexNet), and this considerably reduces the number of parameters in the network and limits the risk of overfitting.
- The last layers are self-explanatory: dropout for regularization, then a fully connected layer with 1,000 units, since there are a 1,000 classes, and a softmax activation function to output estimated class probabilities.

This diagram is slightly simplified: the original GoogLeNet architecture also included two auxiliary classifiers plugged on top of the third and sixth inception modules. They were both composed of one average pooling layer, one convolutional layer, two fully connected layers, and a softmax activation layer. During training, their loss (scaled down by 70%) was added to the overall loss. The goal was to fight the vanishing gradients problem and regularize the network. However, it was later shown that their effect was relatively minor.

Several variants of the GoogLeNet architecture were later proposed by Google researchers, including Inception-v3 and Inception-v4, using slightly different inception modules, and reaching even better performance.

### 2.22 VGGNet:

The runner up in the ILSVRC 2014 challenge was VGGNetN, developed by K. Simonyan and A. Zisserman. It had a very simple and classical architecture, with 2 or 3 convolutional layers, a pooling layer, then again 2 or 3 convolutional layers, a pooling layer, and so on (with a total of just 16 convolutional layers), plus a final dense network with 2 hidden layers and the output layer. It used only 3 x 3 filters, but many filters.

### 2.23 ResNet:

The ILSVRC 2015 challenge was won using a *Residual Network* (or *ResNet),* developed by Kaiming He et al.,'[5] which delivered an astounding top-5 error rate under 3.6%, using an extremely deep CNN composed of 152 layers. It confirmed the general trend: models are getting deeper and deeper, with fewer and fewer parameters. The key to being able to train such a deep network is to use *skip connections* (also called *shortcut connections):* the signal feeding into a layer is also added to the output of a layer located a bit higher up the stack. Let's see why this is useful.

When training a neural network, the goal is to make it model a target function h(x). If you add the input x to the output of the network (i.e., you add a skip connection), then the network will be forced to model f(x) = h(x) - x rather than h(x). This is called *residual learning* (see figure 20)

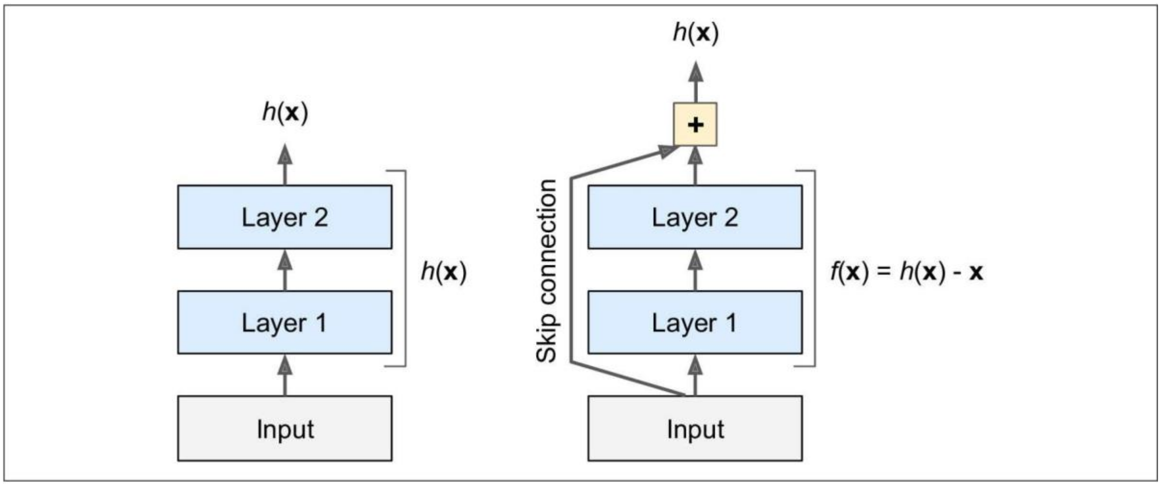


**Fig 2.21: Residual learning**

When you initialize a regular neural network, its weights are close to zero, so the network just outputs value close to zero. If you add a skip connection, the resulting network just outputs a copy of its inputs; in other words, it initially models the identity function. If the target function is fairly close to the identity function (which is often the case), this will speed up training considerably.

Moreover, if you add many skip connections, the network can start making progress even if several layers have not started learning yet (see Figure 14-16). Thanks to skip connections, the signal can easily make its way across the whole network. The deep residual network can be seen as a stack of *residual units,* where each residual unit is a small neural network with a skip connection.

Now let's look at ResNet's architecture (see Figure 14-17). It is actually surprisingly simple. It starts and ends exactly like GoogLeNet (except without a dropout layer), and in between is just a very deep stack of simple residual units. Each residual unit is composed of two convolutional layers (and no pooling layer!), with Batch Normalization (BN) and ReLU activation, using 3 x 3 kernels and preserving spatial dimensions (stride 1, SAME padding).

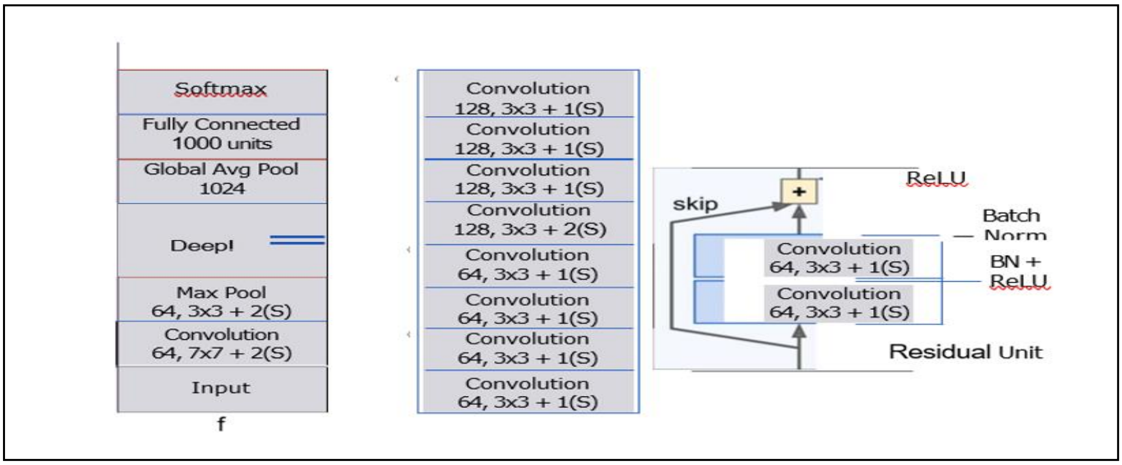


**Fig 2.22: ResNet architecture**

Note that the number of feature maps is doubled every few residual units, at the same time as their height and width are halved (using a convolutional layer with stride 2). When this happens the inputs cannot be added directly to the outputs of the residual unit since they don't have the same shape (for example, this problem affects the skip connection represented by the dashed arrow in Figure 14-17). To solve this problem, the inputs are passed through a 1 x 1 convolutional layer with stride 2 and the right number of output feature maps (see Figure 14-18).

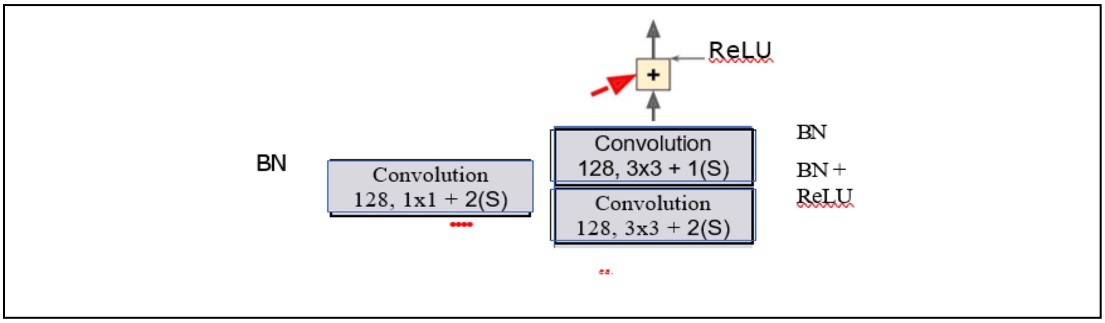


**Fig 2.23: Skip connection when changing feature map size and depth**

ResNet-34 is the ResNet with 34 layers (only counting the convolutional layers and the fully connected layer) containing three residual units that output 64 feature maps, 4 RUs with 128 maps, 6 RUs with 256 maps, and 3 RUs with 512 maps. We will implement this architecture later in this chapter.

ResNets deeper than that, such as ResNet-152, use slightly different residual units. Instead of two 3 x 3 convolutional layers with (say) 256 feature maps, they use three convolutional layers: first a 1 x 1 convolutional layer with just 64 feature maps (4 times less), which acts as a bottleneck layer (as discussed already), then a 3 x 3 layer with 64 feature maps, and finally another 1 x 1 convolutional layer with 256 feature maps (4 times 64) that restores the original depth. ResNet-152 contains three such RUs that output 256 maps, then 8 RUs with 512 maps, a whopping 36 RUs with 1,024 maps, and finally 3 RUs with 2,048 maps.

Google's Inception-v4['6] architecture merged the ideas of GoogLeNet and ResNet and achieved close to 3% top-5 error rate on ImageNet classification.

Xception

Another variant of the GoogLeNet architecture is also worth noting: Xception" (which stands for *Extreme Inception)* was proposed in 2016 by Francois Chollet (the author of Keras), and it significantly outperformed Inception-v3 on a huge vision task (350 million images and 17,000 classes). Just like Inception-v4, it also merges the ideas of GoogLeNet and ResNet, but it replaces the inception modules with a special type of layer called a *depthwise separable convolution* (or *separable convolution* for short"). These layers had been used before in some CNN architectures, but they were not as central as in the Xception architecture. While a regular convolutional layer uses filters that try to simultaneously capture spatial patterns (e.g., an oval) and cross-channel patterns (e.g., mouth + nose + eyes = face), a separable convolutional layer makes the strong assumption that spatial patterns and cross-channel patterns can be modeled separately (see Figure 14-19). Thus, it is composed of two parts: the first part applies a single spatial filter for each input feature map, then the second part looks exclusively for cross-channel patterns—it is just a regular convolutional layer with 1 x 1 filters.

Since separable convolutional layers only have one spatial filter per input channel, you should avoid using them after layers that have too few channels, such as the input layer (granted, that's what Figure 14-19 represents, but it is just for illustration purposes). For this reason, the Xception architecture starts with 2 regular convolutional layers, but then the rest of the architecture uses only separable convolutions (34 in all), plus a few max pooling layers and the usual final layers (a global average pooling layer, and a dense output layer).

You might wonder why Xception is considered a variant of GoogLeNet, since it contains no inception module at all? Well, as we discussed earlier, an Inception module contains convolutional layers with 1 x 1 filters: these look exclusively for cross-channel patterns. However, the convolution layers that sit on top of them are regular convolutional layers that look both for spatial and cross-channel patterns. So you can think of an Inception module as an intermediate between a regular convolutional layer (which considers spatial patterns and cross-channel patterns jointly) and a separable convolutional layer (which considers them separately). In practice, it seems that separable convolutions generally perform better.

Separable convolutions use less parameters, less memory and less computations than regular convolutional layers, and in general they even perform better, so you should consider using them by default (except after layers with few channels).

The ILSVRC 2016 challenge was won by the CUlmage team from the Chinese University of Hong Kong. They used an ensemble of many different techniques, including a sophisticated object-detection system called GBD-Net[19], to achieve a top-5 error rate below 3%. Although this result is unquestionably impressive, the complexity of the solution contrasted with the simplicity of ResNets. Moreover, one year later another fairly simple architecture performed even better, as we will see now

### SENet

The winning architecture in the ILSVRC 2017 challenge was the Squeeze-andExcitation Network (SENet)[20]. This architecture extends existing architectures such as inception networks or ResNets, and boosts their performance. This allowed SENet to win the competition with an astonishing 2.25% top-5 error rate! The extended versions of inception networks and ResNet are called *SE-Inception* and *SE-ResNet* respectively. The boost comes from the fact that a SENet adds a small neural network, called a *SE Block,* to every unit in the original architecture (i.e., every inception module or every residual unit), as shown in Figure 14-20.

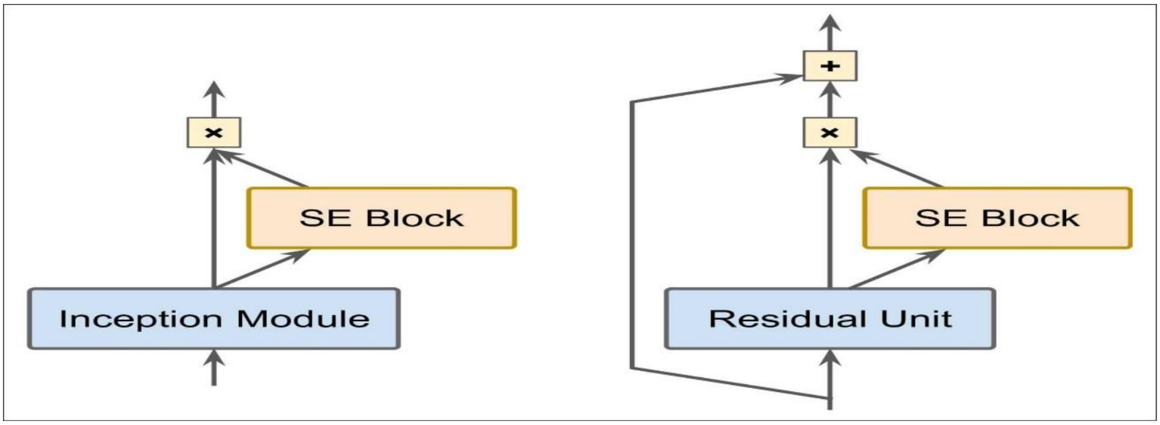


**Fig 2.24: SE-Inception Module (left) and SE-ResNet Unit (right)**

A SE Block analyzes the output of the unit it is attached to, focusing exclusively on the depth dimension (it does not look for any spatial pattern), and it learns which features are usually most active together. It then uses this information to recalibrate the feature maps, as shown in Figure 14-21. For example, a SE Block may learn that mouths, noses and eyes usually appear together in pictures: if you see a mouth and a nose, you should expect to see eyes as well. So if a SE Block sees a strong activation in the mouth and nose feature maps, but only mild activation in the eye feature map, it will boost the eye feature map (more accurately, it will reduce irrelevant feature maps). If the eyes were somewhat confused with something else, this feature map recalibration will help resolve the ambiguity.

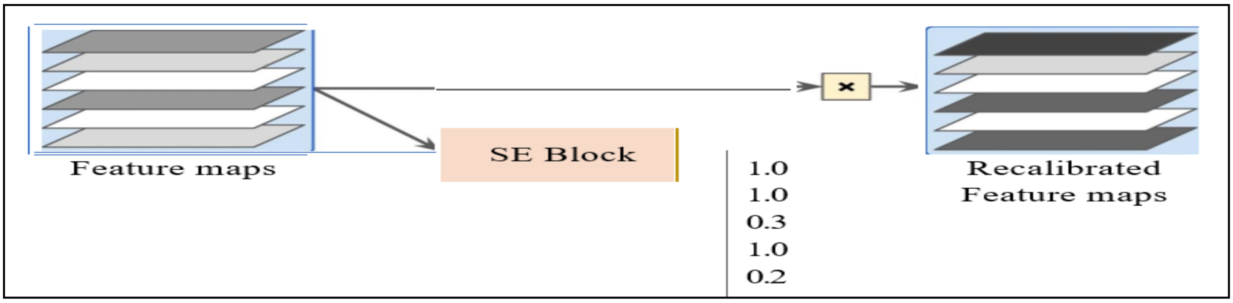


**Fig 2.25 An SE Block Performs Feature Map Recalibration**

A SE Block is composed of just 3 layers: a global average pooling layer, a hidden dense layer using the ReLU activation function, and a dense output layer using the sigmoid activation function.

**2.24 Conclusion:**

We have learned the artificial neural networks with ANN architecture. The ANNs are at the very core of Deep Learning. They are versatile, powerful, and scalable, making them ideal to tackle large and highly complex Machine Learning tasks, such as classifying billions of images, powering speech recognition services, recommending the best videos to watch to hundreds of millions of users every day, or learning to beat the world champion at the game of Go by examining millions of past games and then playing against itself.

The biological neuron which is also known as artificial neuron. The artificial neuron simply activates its output when more than a certain number of its inputs are active. McCulloch and Pitts showed that even with such a simplified model it is possible to build a network of artificial neurons that computes any logical proposition you want.

The Perceptron is one of the simplest ANN architectures It is based on a slightly different artificial neuron the inputs and output are now numbers (instead of binary on/off values) and each input connection is associated with a weight.

An Multi-Layer Perceptron is composed of one (passthrough) input layer, one or more layers of LTUs, called hidden layers, and one final layer of LTUs called the output layer. Biological neurons seem to implement a roughly sigmoid (S-shaped) activation function, so researchers stuck to sigmoid functions for a very long time. But it turns out that the ReLU activation function generally works better in ANNs. This is one of the cases where the biological analogy was misleading.

Class MLP Regressor implements a multi-layer perceptron (MLP) that trains using backpropagation with no activation function in the output layer, which can also be seen as using the identity function as activation function.

Class MLP Classifier implements a multi-layer perceptron (MLP) algorithm that trains using Backpropagation.

Convolutional neural networks apply a filter to an input to create a feature map that summarizes the presence of detected features in the input. Filters can be handcrafted, such as line detectors, but the innovation of convolutional neural networks is to learn the filters during training in the context of a specific prediction problem. How to calculate the feature map for one- and two-dimensional convolutional layers in a convolutional neural network.

The output from multiplying the filter with the input array one time is a single value. As the filter is applied multiple times to the input array, the result is a two-dimensional array of output values that represent a filtering of the input. As such, the two-dimensional output array from this operation is called a "feature map".

The depth of a filter in a CNN must match the depth of the input image. The number of color channels in the filter must remain the same as the input image. Different Conv2D filters are created for each of the three channels for a color image. Filters for each layer are randomly initialized based on either Normal or Gaussian distribution. Initial layers of a convolutional network extract high-level features from the image, so use fewer filters. As we build further deeper layers, we increase the number of filters to twice or thrice the size of the filter of the previous layer. Filters of the deeper layers learn more features but are computationally very intensive.

TensorFlow is the premier open-source deep learning framework developed and maintained by Google. Although using TensorFlow directly can be challenging, the modern tf.keras API beings the simplicity and ease of use of Keras to the TensorFlow project.

The difference between Keras and tf.keras and how to install and confirm TensorFlow is working. The 5-step life-cycle of tf.keras models and how to use the sequential and functional APIs. How to develop MLP, CNN, and RNN models with tf.keras for regression, classification, and time series forecasting.

How to use the advanced features of the tf.keras API to inspect and diagnose your model. How to improve the performance of your tf.keras model by reducing overfitting and accelerating training.

Pooling is required to down sample the detection of features in feature maps. How to calculate and implement average and maximum pooling in a convolutional neural network. How to use global pooling in a convolutional neural network.

The CNN performance is greatly influenced by hyper-parameter selection. Any small change in the hyper-parameter values will affect the general CNN performance. Therefore, careful parameter selection is an extremely significant issue that should be considered during optimization scheme development.

Impressive and robust hardware resources like GPUs are required for effective CNN training. Moreover, they are also required for exploring the efficiency of using CNN in smart and embedded systems.

**References**

1. https://semiengineering.com/deep-learning-spreads/
2. https://www.oreilly.com/library/view/neural-networks-and/9781492037354/ch01.html
3. https://books.google.co.in/books?id=khpYDgAAQBAJ&printsec=frontcover#v=onepage&q&f=false

# Data Science

Dr. Amlan Chakrabarti

Dr. Jyoti Gautam

Prof. Nitima Malsa

Prof. Rahul Borate

**3.1 Introduction to Data Science/ Analytics:**

## 1. WHAT IS DATA SCIENCE & WHY IS IT IMPORTANT?

Data is meaningless until its conversion into valuable information. Data Science involves mining large datasets containing structured and unstructured data and identifying hidden patterns to extract actionable insights. The importance of Data Science lies in its innumerable uses that range from daily activities like asking Siri or Alexa for recommendations to more complex applications like operating a self-driving car.

The interdisciplinary field of Data Science encompasses Computer Science, Statistics, Inference, Machine Learning algorithms, Predictive Analysis, and new technologies.

## 2. Why Data Matters

Data is the new electricity. We are living in the age of the fourth industrial revolution. This is the era of Artificial Intelligence and Big Data. There is a massive data explosion that has resulted in the culmination of new technologies and smarter products.

Around 2.5 exabytes of Data is created each day. The need for data has risen tremendously in the last decade. Many companies have centred their business on data. Data has created new sectors in the IT industry. However,

- Why do we need Data?
- Why do industries need Data?
- What makes data a precious commodity?

The answer to these questions lies in the way companies have sought to transform their products.

Below image is the Life Cycle of Data Science.

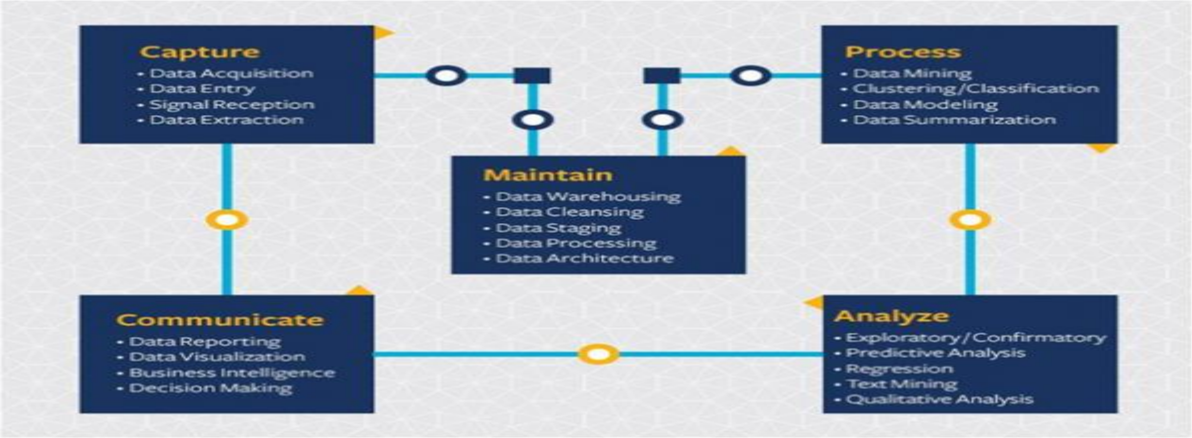


**Fig 3.1: Life Cycle of Data Science**

**3. Why Business Success depend on Data Science**

Data is important so decoding is easy. Millions of bytes of data is being generated and so its importance has gone beyond oil. The role of a data scientist is and will be crucial to peoples who are across many verticals.

Data needs to be analyzed on top. It is essential to maintain and analyze data quality and understand how to make data-driven discoveries.

For goods and products, data science can be used by machine learning which will enable companies to produce products that are liked by customers. For example, a great recommendation system for an ecommerce company can help them look for their shopping history and find their customers.

Data science is can be used **in healthcare**, technology and consumer goods etc.

## 3.2 OVERVIEW OF DATA SCIENCE

In 1962, John Tukey wrote about the convergence of Statistics and computers to devise measurable outputs in hours. In 1974, Peter Naur mentioned the term 'Data Science' multiple times in his review, Concise Survey of Computer Methods. In 1977, the International Association for Statistical Computing (IASC) was formed to link modern computer technology, traditional statistical methodology, and domain expertise to convert data into knowledge. In the same year, Tukey composed a paper, Exploratory Data Analysis, that briefed the importance of using data.

By 1994, organizations had started gathering tremendous individual data for new showcasing efforts. In 1999, Jacob Zahavi stressed the need for new devices to deal with the gigantic chunk of organizational data. In 2001, William S. Cleveland presented an activity plan depicting how to create a specialized understanding and scope of Data Scientists and indicated six regions of studies for offices and colleges.

In 2002, the International Council for Science published the Data Science Journal focusing on Data Science issues like data systems explanation, application, and more. In 2003, Columbia University published the Data Science Journal to set a platform for data teams. In the year 2005, the National Science Board published an existing collection of digital data, and in 2013, IBM revealed that 90% of the global data had been created in the past two years. By this time, organizations realized the importance of Data Science to convert huge data clusters into usable information to gain.

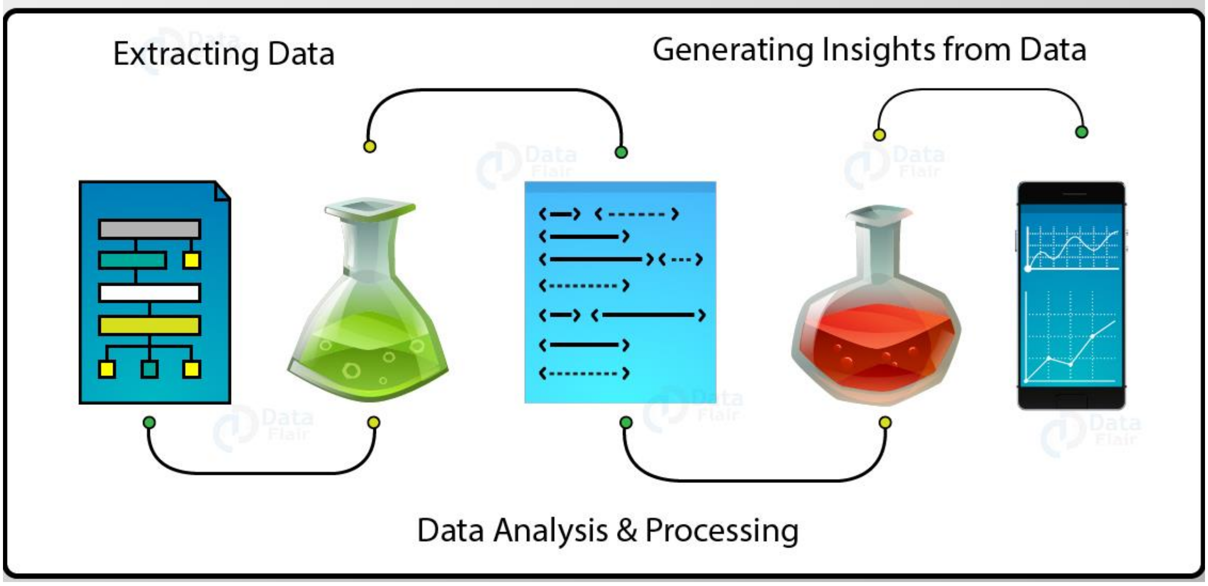


**Fig 3.2: Data Analysis and Processing**

### 1. Essential Math for Data Science

The key topics to master to become a better data scientist

Mathematics is the bedrock of any contemporary discipline of science. Almost all the techniques of modern data science, including machine learning, have a deep mathematical underpinning.

It goes without saying that you will absolutely need all the other pearls of knowledge—programming ability, some amount of business acumen, and your unique analytical and inquisitive mindset—about the data to function as a top data scientist. But it always pays to know the machinery under the hood, rather than just being the person behind the wheel with no knowledge about the car. Therefore, a solid understanding of the mathematical machinery behind the cool algorithms will give you an edge among your peers.

The knowledge of this essential math is particularly important for newcomers arriving at data science from other professions: hardware engineering, retail, the chemical process industry, medicine and health care, business management, etc. Although such fields may require experience with spreadsheets, numerical calculations, and projections, the math skills required in data science can be significantly different.

Consider a web developer or business analyst. They may be dealing with a lot of data and information on a daily basis, but there may not be an emphasis on rigorous modeling of that data. Often, the emphasis is on using the data for an immediate need and moving on, rather than on deep scientific exploration. Data science, on the other hand, should always be about the science (not the data). Following that thread, certain tools and techniques become indispensable. Most are the hallmarks of the sound scientific process:

- Modeling a process (physical or informational) by probing the underlying dynamics
- Constructing hypotheses

- Rigorously estimating the quality of the data source
- Quantifying the uncertainty around the data and predictions
- Identifying the hidden pattern from the stream of information
- Understanding the limitation of a model
- Understanding mathematical proof and the abstract logic behind it

Data science, by its very nature, is not tied to a particular subject area and may deal with phenomena as diverse as cancer diagnoses and social behavior analysis. This produces the possibility of a dizzying array of n-dimensional mathematical objects, statistical distributions, optimization objective functions, etc.

Here are my suggestions for the topics to study to be at the top of the game in data science.

**2. Functions, Variables, Equations, and Graphs**

This area of math covers the basics, from the equation of a line to the binomial theorem and everything in between:

- Logarithm, exponential, polynomial functions, rational numbers
- Basic geometry and theorems, trigonometric identities
- Real and complex numbers, basic properties
- Series, sums, inequalities
- Graphing and plotting, Cartesian and polar coordinates, conic sections
    - **uses**

If you want to understand how a search runs faster on a million-item database after you've sorted it, you will come across the concept of "binary search." To understand the dynamics of it, you need to understand logarithms and recurrence equations. Or, if you want to analyze a time series, you may come across concepts like "periodic functions" and "exponential decay."

**3. Statistics**

The importance of having a solid grasp over essential concepts of statistics and probability cannot be overstated. Many practitioners in the field actually consider classical (non-neural network) machine learning to be nothing but statistical learning. The subject is vast, and focused planning is critical to cover the most essential concepts:

- Data summaries and descriptive statistics, central tendency, variance, covariance, correlation
- Basic probability: basic idea, expectation, probability calculus, Bayes' theorem, conditional probability
- Probability distribution functions: uniform, normal, binomial, chi-square, Student's t-distribution, central limit theorem

- Sampling, measurement, error, random number generation
- Hypothesis testing, A/B testing, confidence intervals, p-values
- ANOVA, t-test
- Linear regression, regularization

### 4. Linear Algebra

This is an essential branch of mathematics for understanding how machine-learning algorithms work on a stream of data to create insight. Everything from friend suggestions on Facebook, to song recommendations on Spotify, to transferring your selfie to a Salvador Dali-style portrait using deep transfer learning involves matrices and matrix algebra. Here are the essential topics to learn:

- Basic properties of matrix and vectors: scalar multiplication, linear transformation, transpose, conjugate, rank, determinant
- Inner and outer products, matrix multiplication rule and various algorithms, matrix inverse
- Special matrices: square matrix, identity matrix, triangular matrix, idea about sparse and dense matrix, unit vectors, symmetric matrix, Hermitian, skew-Hermitian and unitary matrices
- Matrix factorization concept/LU decomposition, Gaussian/Gauss-Jordan elimination, solving Ax=b linear system of equation
- Vector space, basis, span, orthogonality, orthonormality, linear least square
- Eigenvalues, eigenvectors, diagonalization, singular value decomposition

### 5. Calculus

Whether you loved or hated it in college, calculus pops up in numerous places in data science and machine learning. It lurks behind the simple-looking analytical solution of an ordinary least squares problem in linear regression or embedded in every back-propagation your neural network makes to learn a new pattern. It is an extremely valuable skill to add to your repertoire. Here are the topics to learn:

- Functions of a single variable, limit, continuity, differentiability
- Mean value theorems, indeterminate forms, L'Hospital's rule
- Maxima and minima
- Product and chain rule
- Taylor's series, infinite series summation/integration concepts
- Fundamental and mean value-theorems of integral calculus, evaluation of definite and improper integrals
- Beta and gamma functions

- Functions of multiple variables, limit, continuity, partial derivatives
- Basics of ordinary and partial differential equations

### 6. Discrete Math

This area is not discussed as often in data science, but all modern data science is done with the help of computational systems, and discrete math is at the heart of such systems. A refresher in discrete math will include concepts critical to daily use of algorithms and data structures in analytics project:

- Sets, subsets, power sets
- Counting functions, combinatorics, countability
- Basic proof techniques: induction, proof by contradiction
- Basics of inductive, deductive, and propositional logic
- Basic data structures: stacks, queues, graphs, arrays, hash tables, trees
- Graph properties: connected components, degree, maximum flow/minimum cut concepts, graph coloring
- Recurrence relations and equations
- Growth of functions and O(n) notation concept

### 7. Optimization and Operation Research Topics

These topics are most relevant in specialized fields like theoretical computer science, control theory, or operation research. But a basic understanding of these powerful techniques can also be fruitful in the practice of machine learning. Virtually every machine-learning algorithm aims to minimize some kind of estimation error subject to various constraints—which is an optimization problem. Here are the topics to learn:

- Basics of optimization, how to formulate the problem
- Maxima, minima, convex function, global solution
- Linear programming, simplex algorithm
- Integer programming
- Constraint programming, knapsack problem
- Randomized optimization techniques: hill climbing, simulated annealing, genetic algorithms

## 3.3 Introductory Concepts of R

### 1. Introduction to R:

R, primarily an open-source programming language, provides an environment for performing statistical computing and graphics. It has a suite of software packages that can be used to accomplish a wide range of tasks such as data mining, time series analysis, machine learning, multivariate statistical analysis, analysis of spatial data, graphical plotting, etc.

- **Origin of R**

R is an alternate implementation of the statistical programming language called S. S-PLUS was developed post S as its commercial version. R was introduced later by Ross_Ihaka and Robert_Gentleman in 1991. Though R is independent of S-PLUS, much of its code works without any alteration for R too. The first official version of R was released in 1995 as an open-source software package under the GNU General Public License.

- **Fundamental operations and concepts**

Here, we explain in brief some basic yet essential concepts and functions an R programming beginner should know. Each of the further sub-topics has been demonstrated with a snippet of code implemented in R Console (RGui (32-bit)), which can be installed from here.

- Help-related functions

  - `help.start()`: opens R's official documentation for general help on available functionalities.

`help("sum")` or `?sum` : opens documentation for the sum() function

Note: If there is no function with the parameter name, a message will be displayed on the console informing the user that there is no documentation for it in the specified packages and libraries. E.g. help("add") gives the result:

`help.search("sum")` or `??sum` : searches the help system to find instances of the string "sum"

  - `apropos("sum", mode="function")` : lists all the available functions with "sum" string present in their name

`data()` : lists all the example datasets available in the currently loaded packages. (A new window named 'R data sets' gets opened in the console in which the output appears)

Some general purpose functions

  - `getwd()`: to know the current working directory

`setwd(PATH)` : sets the specified path as the current working directory (changes done can be verified using getwd())

  - `ls()` : lists the objects in the current workspace.

  - `rm(objects)`: removes the object(s) specified as parameters from the current workspace

The following snippet creates objects x,y, z and then use ls() to display the objects' names. On executing rm(x,y) removes objects x and y so again doing ls() gives only "z" as output.

history(num): opens a new window named 'R History', which contains names of 'num' number of last executed commands. If nothing is specified as an argument, last 25 commands are displayed by default.

e.g. > history(5)

Sample output:

savehistory("fname") saves the workspace history in a 'fname' named file which can be loaded into the current workspace using loadhistory("fname") command.

- save.image("my_workspace") saves the current workspace to a file named 'my_workspace' which can further be loaded using load("my_workspace") command.
- q() : a dialog box will ask if you want to save the current workspace and then the R console will be exited.

## 2. Packages in R

A package in R is a collection of data, functions and compiled code in a properly defined format. Several packages are stored in a library.

- .libPaths() command shows the path location where the library is located
- library() command displays the list of all the packages saved in the library.

Sample condensed output:

- Package installation: install.packages() command displays a list of CRAN mirror websites for installing a package.

Sample condensed output:

- update.packages() can be used to to get the changes/updates done to each package in the library
- installed.packages() displays the list of all the installed packages along with some additional information such as version number, dependencies etc.

Particular package can be loaded in the current session using library("package_name") command.

## 3. Objects in R

An object refers to anything that can be assigned to a variable. Each object has two attributes:

1. length: number of elements in the object
2. mode: denotes type of the object's data (numeric, character, complex or logical)

Note: 'numeric' data type in R by default means decimal value and not an integer. E.g if we assign x=10 and then check is.integer(x), it will return FALSE. It can be converted to integer type using as.integer() as follows:

There are six types of R objects as follows:

1. **Vector:** a 1D array which is a collection of fixed-sized cells having the same type of data.

Ways to create a vector:

- vector1 <- 1:10 (has elements from 1 to 10)
- Use 'seq' to create a vector of sequence

e.g. seq(from=1,to=10, by=2) (choose elements from 1 to 10 in step of 2)

- Use 'rep' to create vector having repeated element or another vector

e.g. rep("Hi",4)

- Use c() method where 'c' stands for 'combine'
- e.g. vector1 <- c(1,2,3,4,5)

Element(s) of a vector can be accessed using indexing as follows:

2. **Matrix:** It is a 2D vector with fixed-sized cells having the same type of data.

Matrix creation example:

Where, nrow and ncol denote the rows and columns respectively; byrow=TRUE means the matrix will be filled row-by-row.

Image Classification Using R

Ways to access element(s) of a matrix:

- M[n]: nth element of matrix M (counting occurs column-wise, with n=1 denoting the first element)
- M[n,] : nth row of matrix M (n=1 denotes first row)
- M[,n] : nth column of matrix M (n=1 denotes first column)
- M[x,y] : element at xth row and yth column
- M[,c(x,y)] : extract xth and yth columns at a time
- M[c(x,y),] : extract xth and yth rows at a time

**3.** **Array :** It is one or more dimensional array. So 1D array and 2D array are (almost) the same as a vector and a matrix respectively.

The one with 3 or more dimensions is said to be a multidimensional array.

4. **List :** It is a collection of elements which can be of different data types. Also, the size of a list can be expanded on the fly.

**5.** **Factor :** A factor in R is a data object which deals with categorical variables (i.e. those having some fixed possible values, e.g. 'gender' and 'months' variable). Each factor has a levels attribute that denotes the permitted values of the variable. The usefulness of a factor can be understood from the following short example.

e.g. Suppose, there is a list x1 having some of the months' names as its elements. We create a factor with the data of x1 and initialize the 'levels' attribute with a list named 'months' which contains names of all the months in a year. If we simply sort x1, it will be sorted in alphabetical order, but if the factor y1 with well-defined levels is sorted, we get the x1's elements sorted in the order in which those months occur in a year.

Now suppose there is a value in a list which does not match any of the 'levels' list, it will be converted to NA in the factor and the wrong element will be missing in the output if the factor is sorted.

If we miss defining the levels, explicitly, they will be taken as the data's values sorted in alphabetical order.

Levels of a factor can be known using levels() method by passing the factor's name as its argument.

6. **Data frame :** A data frame in R refers to a data table in which the columns can be of different types but each particular column holds the same type of data.

   Some inbuilt datasets such as the Iris Flower dataset can be loaded by loading the 'datasets' package and then loading the dataset

   using data.frame() as follows:

   We can also create a custom data frame as follows:

   Number of rows and columns of a dataframe can be known using nrow() and ncol() methods respectively.

**4. Machine Learning**

- **Predictive Analysis in R Programming**

Predictive analysis in R Language is a branch of analysis which uses statistics operations to analysed historical facts to make predict future events. It is a common term used in data mining and machine learning. Methods like time series analysis, non-linear least square, etc. are used in predictive analysis. Using predictive analytics can help many businesses as it finds out the relationship between the data collected and based on the relationship, the pattern is predicted. Thus, allowing businesses to create predictive intelligence.

In this article, we'll discuss the process, need and applications of predictive analysis with example codes.

I. *Process of Predictive Analysis*

Predictive analysis consists of 7 processes as follows:

- Define project: Defining the project, scope, objectives and result.
- Data collection: Data is collected through data mining providing a complete view of customer interactions.
- Data Analysis: It is the process of cleaning, inspecting, transforming and modelling the data.

- Statistics: This process enables validating the assumptions and testing the statistical models.
- Modelling: Predictive models are generated using statistics and the most optimized model is used for the deployment.
- Deployment: The predictive model is deployed to automate the production of everyday decision-making results.
- Model monitoring: Keep monitoring the model to review performance which ensures expected results.

*II. Need of Predictive Analysis*

- Understanding customer behavior: Predictive analysis uses data mining feature which extracts attributes and behaviour of customers. It also finds out the interests of the customers so that business can learn to represent those products which can increase the probability or likelihood of buying.
- Gain competition in the market: With predictive analysis, businesses or companies can make their way to grow fast and stand out as a competition to other businesses by finding out their weakness and strengths.
- Learn new opportunities to increase revenue: Companies can create new offers or discounts based on the pattern of the customers providing an increase in revenue.
- Find areas of weakening: Using these methods, companies can gain back their lost customers by finding out the past actions taken by the company which customers didn't like.

*III. Applications of Predictive Analysis*

- Health care: Predictive analysis can be used to determine the history of patient and thus, determining the risks.
- Financial modelling: Financial modelling is another aspect where predictive analysis plays a major role in finding out the trending stocks helping the business in decision making process.
- Customer Relationship Management: Predictive analysis helps firms in creating marketing campaigns and customer services based on the analysis produced by the predictive algorithms.
- Risk Analysis: While forecasting the campaigns, predictive analysis can show an estimation of profit and helps in evaluating the risks too.

**Example:**

Let us take an example of time analysis series which is a method of predictive analysis in R programming:

```
x <-c(580, 7813, 28266, 59287, 75700,

      87820, 95314, 126214, 218843, 471497,

      936851, 1508725, 2072113)
```

```
    # library required for decimal_date() function
library(lubridate)
 # output to be created as png file
png(file="predictiveAnalysis.png")
 # creating time series object
# from date 22 January, 2020
mts <-ts(x, start =decimal_date(ymd("2020-01-22")),
                        frequency =365.25/7)
 plotting the graph
plot(mts, xlab ="Weekly Data of sales",
         ylab ="Total Revenue",
         main ="Sales vs Revenue",
         col.main ="darkgreen")
    # saving the file
dev.off()
```

Output:

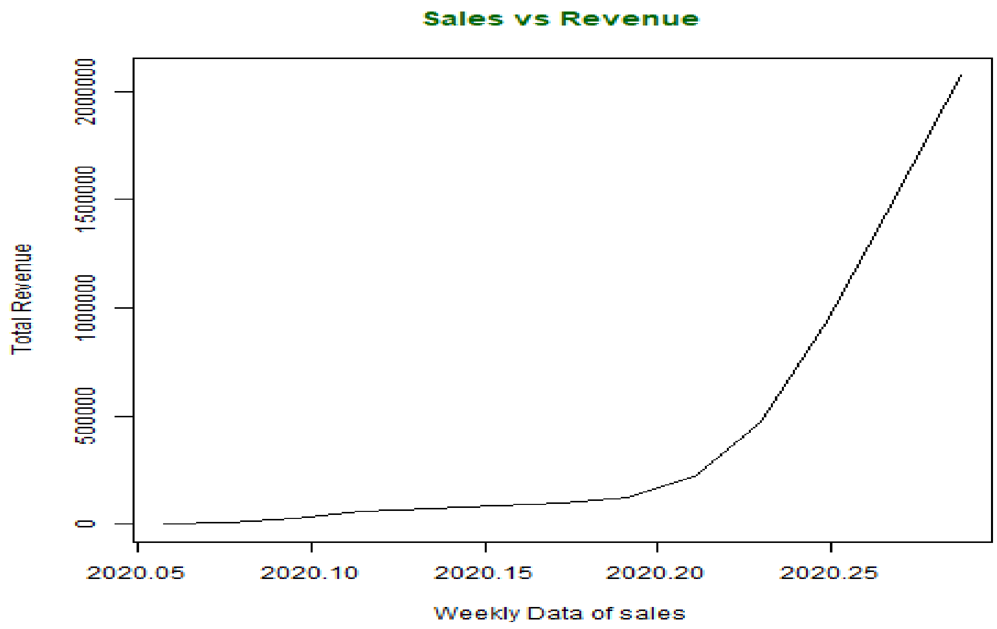


**Fig 3.3: Predictive Analysis**

## 5. Logistic Regression Analysis in R

- **Introduction**

In this article, you'll learn about Logistic Regression in detail. Believe me, Logistic Regression isn't easy to master. It does follow some assumptions like Linear Regression. But its method of calculating model fit and evaluation metrics is entirely different from Linear/Multiple regression.

- **What is Logistic Regression?**

Many a time, situations arise where the dependent variable isn't normally distributed; i.e., the assumption of normality is violated. For example, think of a problem when the dependent variable is binary (Male/Female). Will you still use Multiple Regression? Of course not! Why? We'll look at it below.

Logistic Regression belongs to the family of generalized linear models. It is a binary classification algorithm used when the response variable is dichotomous (1 or 0). Inherently, it returns the set of probabilities of target class. But, we can also obtain response labels using a probability threshold value. Following are the assumptions made by Logistic Regression:

1. The response variable must follow a binomial distribution.
2. Logistic Regression assumes a linear relationship between the independent variables and the link function (logit).
3. The dependent variable should have mutually exclusive and exhaustive categories.

In R, we use glm() function to apply Logistic Regression. In Python, we use sklearn.linear_model function to import and use Logistic Regression.

What are the types of Logistic Regression techniques?

Logistic Regression isn't just limited to solving binary classification problems. To solve problems that have multiple classes, we can use extensions of Logistic Regression, which includes Multinomial Logistic Regression and Ordinal Logistic Regression. Let's get their basic idea:

1. Multinomial Logistic Regression: Let's say our target variable has K = 4 classes. This technique handles the multi-class problem by fitting K-1 independent binary logistic classifier model. For doing this, it randomly chooses one target class as the reference class and fits K-1 regression models that compare each of the remaining classes to the reference class.

Due to its restrictive nature, it isn't used widely because it does not scale very well in the presence of a large number of target classes. In addition, since it builds K - 1 models, we would require a much larger data set to achieve reasonable accuracy.

2. Ordinal Logistic Regression: This technique is used when the target variable is ordinal in nature. Let's say, we want to predict years of work experience (1,2,3,4,5, etc). So, there exists an order in the value, i.e., 5>4>3>2>1. Unlike a multinomial model, when we train K -1 models, Ordinal Logistic Regression builds a single model with multiple threshold values.

If we have K classes, the model will require K -1 threshold or cutoff points. Also, it makes an imperative assumption of proportional odds. The assumption says that on a logit (S shape) scale, all of the thresholds lie on a straight line.

Note: Logistic Regression is not a great choice to solve multi-class problems. But, it's good to be aware of its types. In this tutorial we'll focus on Logistic Regression for binary classification task.

How does Logistic Regression work?

Now comes the interesting part!

As we know, Logistic Regression assumes that the dependent (or response) variable follows a binomial distribution. Now, you may wonder, what is binomial distribution? Binomial distribution can be identified by the following characteristics:

- There must be a fixed number of trials denoted by n, i.e. in the data set, there must be a fixed number of rows.
- Each trial can have only two outcomes; i.e., the response variable can have only two unique categories.
- The outcome of each trial must be independent of each other; i.e., the unique levels of the response variable must be independent of each other.
- The probability of success (p) and failure (q) should be the same for each trial.

## Data Visualization in R

### 3.4 Data Visualization in R:

In this article, we will create the following visualizations:

<u>Basic Visualization</u>

1. Histogram
2. Bar / Line Chart
3. Box plot
4. Scatter plot

- Basic graphs in R can be created quite easily. The **plot** command is the command to note.
- It takes in many parameters from x axis data , y axis data, x axis labels, y axis labels, color and title. To create line graphs, simply use the parameter, type=l.
- If you want a boxplot, you can use the word boxplot, and for barplot use the barplot function.

### 1. <u>Histogram</u>

Histogram is basically a plot that breaks the data into bins (or breaks) and shows frequency distribution of these bins. You can change the breaks also and see the effect it has data visualization in terms of understandability.

Let me give you an example.

Note: We have used par(mfrow=c(2,5)) command to fit multiple graphs in same page for sake of clarity(

see the code below).

The following commands show this in a better way. In the code below, the **main** option sets the Title of Graph and the **col** option calls in the color pallete from RColorBrewer to set the colors.

```
library(RColorBrewer)
```

```
data(VADeaths)

par(mfrow=c(2,3))

hist(VADeaths,breaks=10, col=brewer.pal(3,"Set3"),main="Set3 3 colors")

hist(VADeaths,breaks=3 ,col=brewer.pal(3,"Set2"),main="Set2 3 colors")

hist(VADeaths,breaks=7, col=brewer.pal(3,"Set1"),main="Set1 3 colors")

hist(VADeaths,,breaks= 2, col=brewer.pal(8,"Set3"),main="Set3 8 colors")

hist(VADeaths,col=brewer.pal(8,"Greys"),main="Greys 8 colors")

hist(VADeaths,col=brewer.pal(8,"Greens"),main="Greens 8 colors")
```

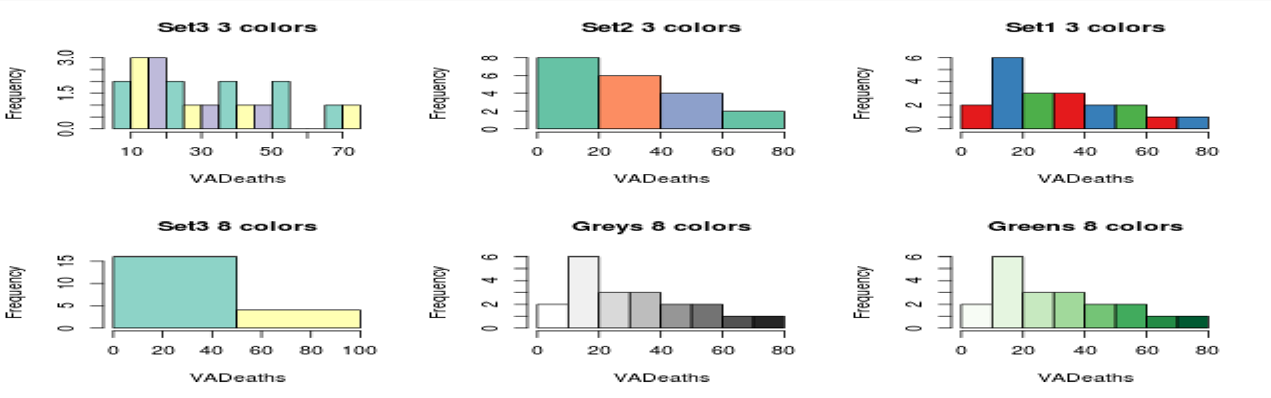


**Fig 3.4: Histogram**

Notice, if number of breaks is less than number of colors specified, the colors just go to extreme values as in the "Set 3 8 colors" graph. If number of breaks is more than number of colors, the colors start repeating as in the first row.

## 2. Bar/ Line Chart

- *Line Chart*

Below is the line chart showing the increase in air passengers over given time period. Line Charts are commonly preferred when we are to analyse a trend spread over a time period. Furthermore, line plot is also suitable to plots where we need to compare relative changes in quantities across some variable (like time). Below is the code:

```
plot(AirPassengers,type="l")  #Simple Line Plot
```

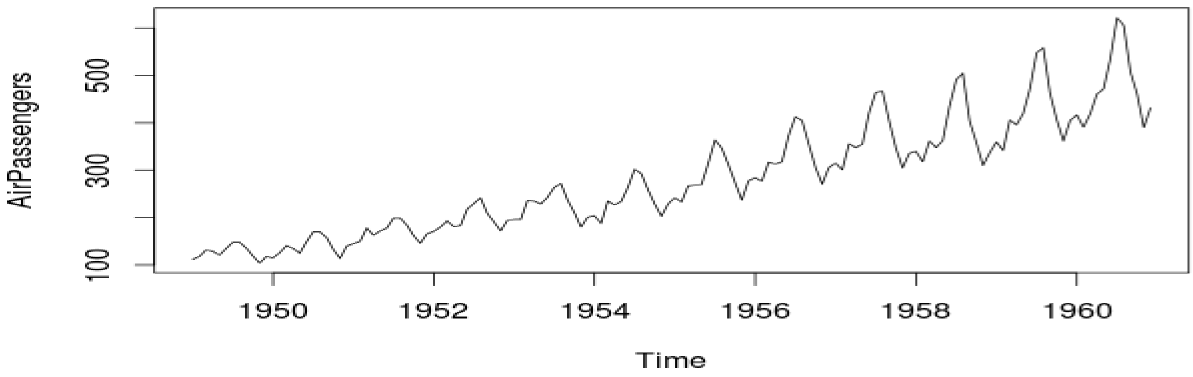


**Fig 3.5: Line Chart**

- *Bar Chart*

Bar Plots are suitable for showing comparison between cumulative totals across several groups. Stacked Plots are used for bar plots for various categories. Here's the code:

```
barplot(iris$Petal.Length) #Creating simple Bar Graph

barplot(iris$Sepal.Length,col  = brewer.pal(3,"Set1"))

barplot(table(iris$Species,iris$Sepal.Length),col  = brewer.pal(3,"Set1")) #Stacked Plot
```

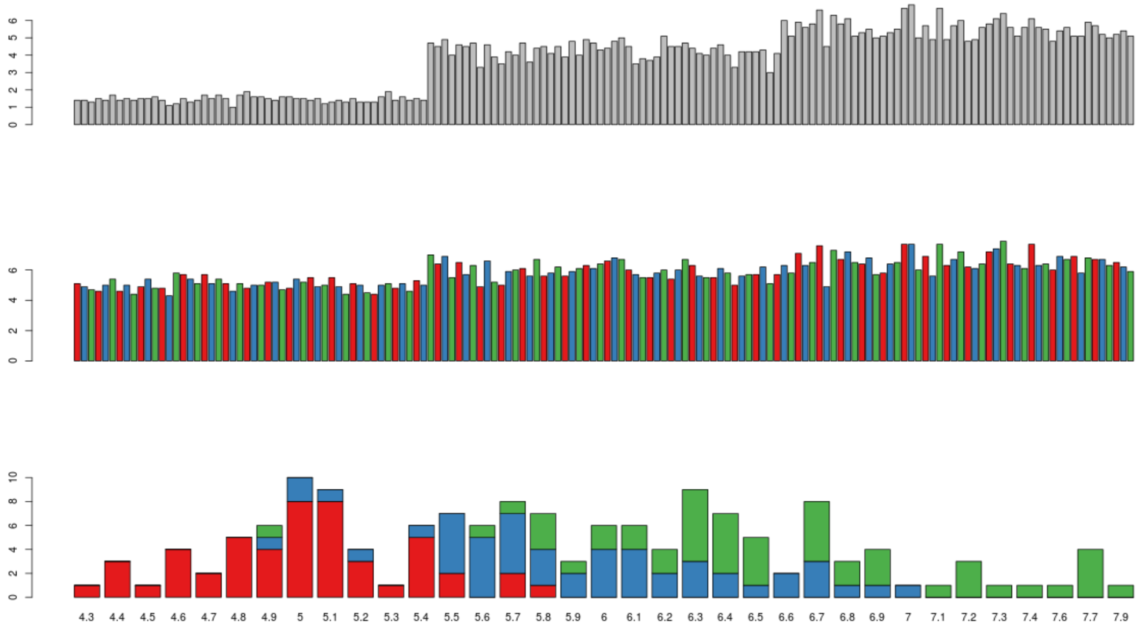


**Fig 3.6: Bar Chart**

3. **Box Plot**

Box Plot  shows 5 statistically significant numbers- the minimum, the 25th percentile, the median, the

75th percentile and the maximum. It is thus useful for visualizing the spread of the data is and deriving inferences accordingly. Here’s the basic code:

```
boxplot(iris$Petal.Length~iris$Species) #Creating Box Plot between two variable
```

Let’s understand the code below:

In the example below, I have made 4 graphs in one screen. By using the ~ sign, I can visualize how the spread (of Sepal Length) is across various categories ( of Species). In the last two graphs I have shown the example of color palettes. A color palette is a group of colors that is used to make the graph more appealing and helping create visual distinctions in the data.

```
data(iris)

par(mfrow=c(2,2))

boxplot(iris$Sepal.Length,col="red")

boxplot(iris$Sepal.Length~iris$Species,col="red")

oxplot(iris$Sepal.Length~iris$Species,col=heat.colors(3))

boxplot(iris$Sepal.Length~iris$Species,col=topo.colors(3))
```

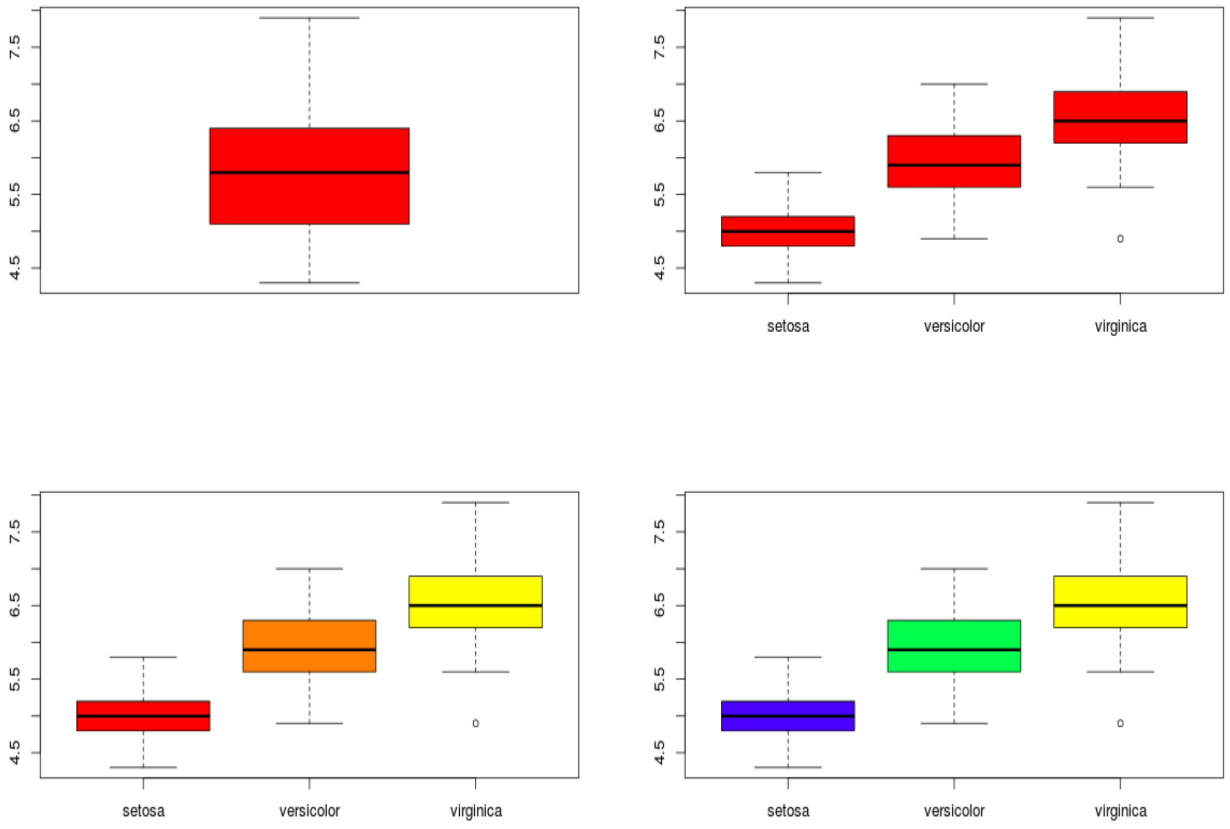


**Fig 3.7: Box Plot**

## 1. Data Science/Analytics Fundamentals

**Learning Outcomes:**

After attending the course, the participants understand and apply the fundamental concepts and techniques in data science

1 Key concepts in data science, including tools, approaches, and application scenarios

2 Topics in Supervised and Unsupervised Machine Learning Techniques

3 Topics in Ensemble Learning

4 Key Concepts in Fine Tuning of the models using Regularization and Feature Engineering.

5 Topics in Time Series Analysis

6 Understanding Artificial Neural Networks and Convolution Neural Network

7 Solve real-world data-science problems in the domains of Computer Vision and Internet of Things

**Pre-requisite:**

- Basic knowledge of Mathematics

**Duration: 40 Hours**

**Software/ Hardware Requirements:**

**Personal Computers with Internet connectivity**

**Google Colab Environment Setup Instructions**

Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs.

1. Setting up your drive

1.1 Create a folder for your notebooks

You can create your folder by going to your Google Drive and clicking **"New"** and then creating a new folder.

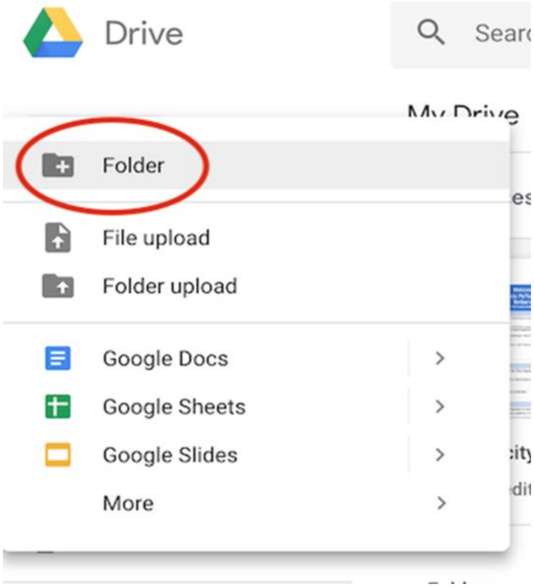


**Fig Ds.8 File upload**

While you're already in your Google Drive you can create a new Colab notebook. Just by clicking "New" and drop the menu down to "More" and then select "Colaboratory."

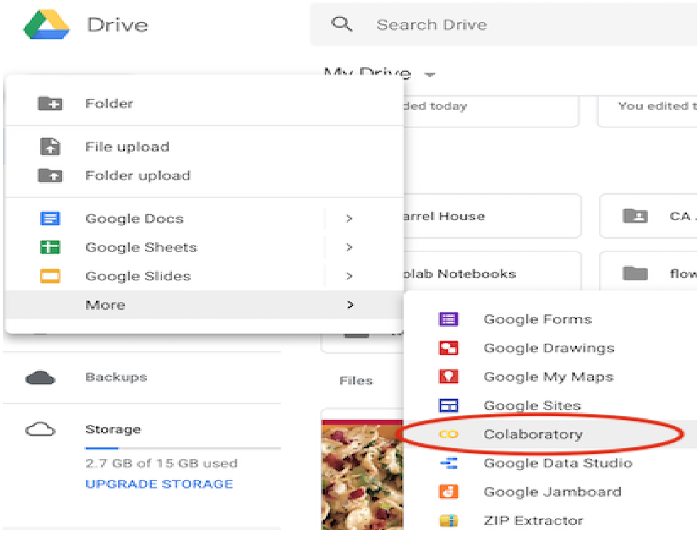
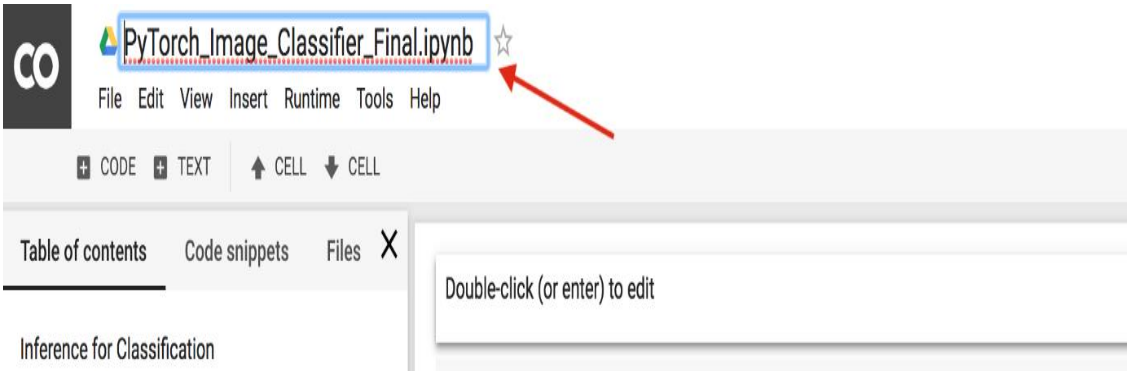


**Fig 3.9: New Book Creation**

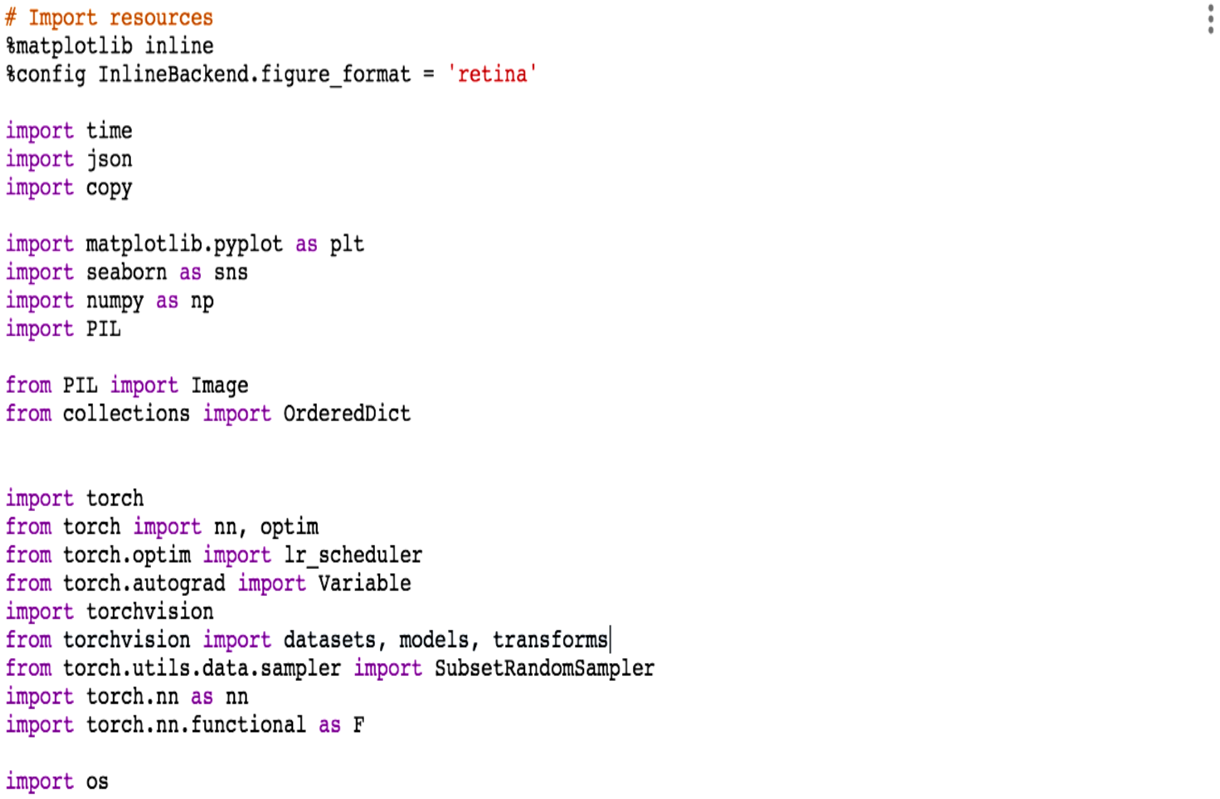
Otherwise, you can go directly to Google Colab. Now you're in Colab, you can rename your notebook by clicking on the name of the notebook and changing it or by dropping the "File" menu down to "Rename."


CO
PyTorch_Image_Classifier_Final.ipynb
File Edit View Insert Runtime Tools Help
CODE TEXT CELL CELL
Table of contents Code snippets Files X
Inference for Classification

Double-click (or enter) to edit

### 1. 2 Importing libraries

**For the most part, you can import your libraries by running import like you do in any other notebook.**



```
# Import resources
%matplotlib inline
%config InlineBackend.figure_format = 'retina'

import time
import json
import copy

import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import PIL

from PIL import Image
from collections import OrderedDict


import torch
from torch import nn, optim
from torch.optim import lr_scheduler
from torch.autograd import Variable
import torchvision
from torchvision import datasets, models, transforms
from torch.utils.data.sampler import SubsetRandomSampler
import torch.nn as nn
import torch.nn.functional as F

import os
```

Then you can continue with your imports.

**Fig Ds.10 Rename colab nootebook**

### 1. Data Science/Analytics

**Objective:** To extract valuable information for use in strategic decision making for real world problems, product development, trends analysis, and forecasting.

**Outcome:** Develop the ability to build and assess **data**-driven models using relevant programming skills.

**Content Overview:** Brief Introduction to Data Science/Data Analytics with its various applications.

**2. Basics Python Programming and Data Science**

**Objectives:**

1 To acquire programming skills in core Python.

2 To familiarize the Google Colab Environment

**Outcome:** Programming capability to solve various problems in the data science domain.

**Contents:**

1 Introduction to Python programming and Google colab environment.

2 Structure of a python program.

3 Important Libraries-Numpy, pandas, matplotlib.

4 Data preprocessing techniques- data exploration, correlational analysis, data wrangling, missing value treatment.

5 Functions

**3. Machine Learning**

**3.1 Regression analysis:**

**Objective:** To explain variability in dependent variables by means of one or more independent or control variables.

**Outcomes:** Learn how to apply linear regression models in practice: identify situations where linear regression is appropriate; build and fit linear regression models with software; interpret estimates and diagnostic statistics; produce exploratory graphs.

**Content Coverage:** Shall cover mainly linear regression, multiple linear regressions, logistic regression, polynomial regression and its practical implementations.

**3.2 Classification:**

**Objective:** To classify the different datasets depending on certain classification criterions.

**Outcome:** Learning how to predict the group memberships of individual's datasets.

**Content Coverage:** KNN Classifier, Decision-Tree Classifier, Logistic Regression, Naive-Bayes Classifier, Support Vector Machines, Finally, practical implementation of all.

**3.3 Clustering Analysis:**

**Objective:** The goal of cluster analysis is to assign observations to groups or clusters so that observations within each group are similar to one another with respect to variables or attributes of interest, and the groups them- selves stand apart from one another.

**Outcome:** Learning how to create clusters using various algorithms and their implementations.

**Content Coverage:** K-Means and Hierarchical

**3.4 Feature**

**Objectives:**

1 Preparing the proper input dataset, compatible with the machine learning algorithm requirements

2 Improving the performance of machine learning models

**Outcome:** Learning the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

**Content Coverage:** Imputation, Outliers, Binning, Log Transform, One-Hot Encoding, Grouping Operations, Scaling

**3.5 Model Validation:**

**Objective:** Evaluating the model performance using testing dataset.

**Outcome:** Learning the process of model evaluation to check its performance against a testing dataset.

**Content Coverage:** Hyperparameters, Cross Validation, Validation Curves

**3.6 Dimensionality Reduction:**

**Objective:** Reduce the dimensionality by projecting the data to a lower dimensional subspace that captures the data essence.

**Outcome:** Learning the process of reducing the number of input variables in the training data.

**Content Coverage:** PCA, LDA, ICA, Graph based Techniques

**4. Ensemble Learning:**

**Objective:** To introduce the concept of ensemble learning and understand its different methods such as stacking, blending, bagging and Boosting.

**Outcome:** Participants will be able to apply ensemble learning techniques to improve the results.

**Content: Ensemble Learning Techniques:** Stacking, Bagging, Boosting, Blending

**5. Recommender Systems**

**Objective:** The objective of recommender systems is to provide recommendations based on recorded information on the users' preferences.

**Outcome:** Learning information filtering techniques to process information and provide the user with potentially more relevant items.

**Content Coverage:** Basic concepts, Content-based filtering, collaborative filtering, Evaluation of recommender systems

**6. Time-Series Analysis and Forecasting**

**Objective:** Explain the dynamic properties of a physical system by analyzing the input time series and performing prediction.

**Outcome:** Learning the process of analyzing time-series data using different models and techniques.

**Content Coverage:**

1. Introductory Concepts
2. Autoregressive Integrated Moving Average Models
3. Seasonal ARIMA Models
4. Intervention Analysis and Outlier Detection
5. Multivariate Time Series Analysis and Forecasting

**7. Deep Leaning**

**7.1 Neural Networks and CNN**

**Objective:** The objective of an Artificial Neural Network is to compute output values from inputs using a computational model inspired by biological neurons for solving real world problems.

**Outcome:** Learning the key steps involved in Artificial Neural Network and Convolution Neural Networks and their application.

**Content Coverage:**

1. Perceptron Network, MLP and ANN
2. Gradient Descent & Backpropagation
3. Understanding Convolution Filters
4. Fully Connected layer & Classification
5. Other Popular Networks

**8. Computer Vision**

**Objective:** The purpose of computer vision is to program a computer to "understand" a scene or features in an image.

**Outcome:** Learning the process of detection, segmentation, and classification of objects in images (e.g., human faces). Learning key concepts of unsupervised learning applied to computer vision problems using Autoencoders and GAN.

**Content Coverage:**

1. Image Features
2. Image Segmentation
3. Image Classification
4. Autoencoders and GAN for Computer Vision

## 9. IOT Analytics

**Objective:** The objective of IoT analytics is to gain value from large volumes of data generated by devices connected via the Internet of Things (IoT).

**Outcome:** Learning the process of retrieving data through different sensors and then performing its analysis.

**Content Coverage:** Example Sensors, Various applications, Case Studies

## 10. Projects

**Objective:** to let the students apply the programming knowledge into the given problems.

**Outcome:** Participants will be able to apply machine learning and deep learning concepts to solve the given small problems.

**List of the Projects:**

1. BigMart Sales Prediction ML Project
2. Home Value Prediction Project
3. Stock Prices Predictor using Time Series Forecasting
4. Coupon Purchase Prediction
5. Retail Price Optimization using Machine Learning

## 11. ZoomIt: An Online Teaching Tool

ZoomIt is a tool for screen-zooming and annotations used mainly for presentations and instructional videos. It can have multiple applications as its simple nature makes it quite flexible. It runs in the background and it's activated by fast shortcuts that you can easily customize inside the app's options menu.

### ZoomIt Default Shortcuts

Remember to always open Zoom It at least once to "turn on" the functionality.

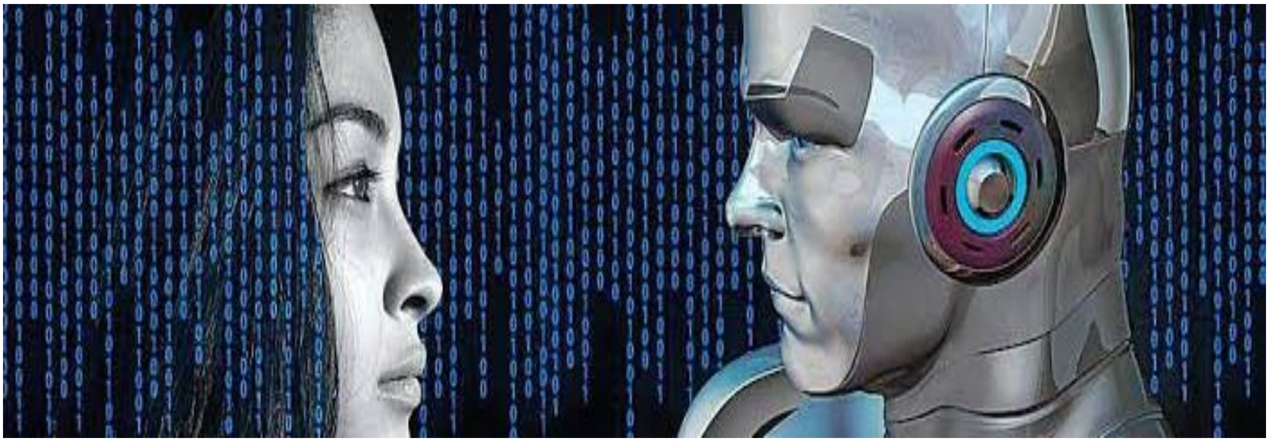
**Zoom:** Ctrl+1

**Exit Zoom:** Esc or Right Mouse Click

**Draw Mode:** Ctrl + D (Left mouse Click while in zoom mode)

- Ctrl + Z = Undo
- E = erase all
- Hold CTRL and press UP or DOWN arrow keys to change size of pen.
- Change pen color by pressing R(red), G(green), B(blue), O(orange), Y(yellow), and P(pink)
- Hold SHIFT to draw straight lines.
- Hold CTRL to draw boxes.

- Hold TAB to draw circles.
- Hold SHIFT+CTRL to draw arrows.
- W = white sketchpad
- K = Black sketchpad



- CTRL+C to copy screen
- CTRL+S to save screenshot RIGHT CLICK to exit drawing mode.

**Typing Mode**: T while in Zoom mode

ESC or LEFT CLICK to exit typing mode

**Reference**

1. https://www.google.com/search?q=introduction+to+data+science&oq=introduction+to+data+science&aqs=chrome..69i57j0i131i433i512l2j0i512j0i131i433i512.14839j0j4&client=ms-android-samsung-gj-rev1&sourceid=chrome-mobile&ie=UTF-8

# RPA – Robotic Process Automation

RPA- Robotic Process Automation

Ms. Priyanka Bhalere

Dr. Sunil Khilari

Mr. Shripad Kulkarni

Prof. Rahul Dwivedi

**4.1. Introduction**

**4.2. What is RPA**

**4.3. Myths About RPA**

- **a. Automation will replace humans in the workforce**
- **b. RPA software robots are precise and correct**
- **c. Any process that can be automated with RPA can also be automated using APIs and programming languages**
- **d. RPA will not work in my industry**
- **e. RPA is not value added or the cost of investment**

**4.4. Where RPA can be implemented**

- **a. Payroll Processing:**
- **b. Customer Complaint Processing:**
- **c. Client Information Updating:**

**4.5. Various RPA tools available**

**4.6. Why UiPath**

**4.7. UiPath Enterprise Platform (Lifecycle of automating RPA process)**

- **a. Requirement Gathering(discovery):**
- **b. Development (Build):**
- **c. Testing & Deployment (Manage):**
- **d. Human in Loop (Engage):**
- **e. Stabilize the process – Hyper care (Measure):**
- **f. Constant Improvement:**

**4.8. UiPath Platform**

- **a. UiPath Studio**
- **b. UiPath Orchestrator**
- **c. The Robot**

**4.9. Conclusion**

## 4.1. Introduction:

Automation has transformed the facets of the today's business. Furthermore, the occasion of applying robotic automation in corporate industry processes has been ahead more consideration as they challenge in a digital era, which requires perfect processes. With RPA Robotic Process Automation invention, businesses are automating knowledge-based, specialized service developments that don't request human interaction. Simultaneously, it is helping as a fundamental innovation to the conventional understanding of labor requirement.

Robotic Process Automation is the expertise that allows individuals to install, configure and computer software to match and integrate the movements of a human interaction within digital schemes to execute a corporate business process. It operates the user interface to capture data and manipulate claims just like people do. They can interpret, trigger, reactions and communicate with additional systems in order to perform on a huge variety of repetitive tasks. An RPA software robot brands very fewer mistakes and charges less than an employee.

RPA consists of three key technologies: screen scraping, workflow automation, and artificial intelligence.

a) **Screen scraping is** the process of gathering data from an inheritance application so that the data can be showed in a more contemporary user interface.

b) **The workflow automation** software eradicates the need for physical data entry and increases order fulfillment rates, including increased rapidity, efficiency, and accuracy. It decreases the chance of errors.

c) **Artificial Intelligence** includes the ability of computer systems to achieve tasks extra efficiently than a human can work

**Robotic Process Automation (RPA)** — the automation of multifaceted processes that substitutes humans through the execution of progressive software — is renovating the future of back office developments. Industries across the world are appreciating that RPA is the next momentous for digital transformation that will empower employees to stop working on tedious tasks. Robotic Process Automation permits employees to essence on more value adding creativities, which are commanding for the lowest floor of the business. To guarantee the accuracy and applicability of movements recognized in Robotic Process Automation.

In last decade various companies were seeking cost saving strategies in the form of labour requirement and as a result, several tasks were stimulated to low-cost countries throughout world. Organizations take benefits of Robotic Process Automation in their Business process. RPA allows businesses to radically progress cost effectiveness and quality enhancements in their transactional functions and processes. A key benefits of Robotic Process Automation is that dissimilar like previous IT transformations for example as Enterprise Resource Planning (ERP's), RPA does not require a massive upfront investment or a significant change to the current IT systems and processes. In fact, RPA can be implemented relatively quickly when compared to previous digital transformations, as it requires minimal capital or infrastructure. RPA can act as an additional employee that can work between the IT systems and with the back office processes in various functions. Similarly, to humans, RPA can acquire from people and copy their task and processes, eventually taking over the processes that humans once completed, at a much faster pace. Robotic Process Automation is going to continue to develop and work with increasingly complex processes and tasks.

Robotic Process Automation and Business Process Management have similarities, they are in fact fundamentally different. BPM works from the top down, standardizing all processes throughout its implementation. In comparison, Robotic Process Automation works from the bottom up, integrating itself with processes. While RPA automates processes, it does not standardize them, nor does it help to standardize processes. BPM standardizes processes, but does not automate them. Even though RPA does not standardize processes, having standardized processes is hugely beneficial for RPA, as can be seen by the fact that the majority of processes that will see an increase in RPA usage in the near future are also the processes that are currently the most standardized.

Robotic Process Automation is unquestionably the next wave in digital transformation—RPA is a software application that can reproduce processes humans would work to move information through and between different technology application platforms. Robotic automation uses software as a virtual platform to operate on existing application software (e.g. ERPs, e-CRMs, e-SCM) in the same way that an individual finishes a process. What is predominantly innovative about robotic automation software is that it ensures not necessarily require businesses to make variations to their strategic processes. Even if businesses are divided geographically or have numerous technological systems implemented, RPA is able to interconnect all systems and applications. Therefore, RPA may function as a quick win-win solution for progression of optimization. While knowledge-based automation techniques, tools and cognitive artificial intelligence systems are introducing the digital market, the majority of businesses are presently motivated on rules-based robotic automation solutions, which means that RPA can do well with multifaceted processes that have a precise set of repetitive instructions. Rules-based activity would comprise tasks in back-office work and processes, such as completing bill, gate pass, challan and invoices. As RPA becomes more commendable, businesses will start to implement knowledge-based automation, empowering robotic automation to do with many more exceptions. A classic example of knowledge-based automation would be in CRM,e-CRM,e-SCM functions, identified for information across systems and responding customer communication and request. Finally, while RPA has not developed into the all-cognitive border, experts positively realize potential for RPA to ultimately be able to contemplate for itself, doing alongside humans on value-adding creativities that are important for every enterprise foundation level

Robotic Process Automation is previously transforming back-office task and processes like Customer Service, support, Finance and Human Resource Management for rules-based activities. Robotic Process Automation is not only beneficial for employees and employer, but also for all type of businesses operations. Activities can be accomplished quickly, accurately, efficiently and at a lesser cost through the implementation of robotic automation solutions. Statistics show that robots work meaningfully faster than human beings.

## 4.2. What is RPA:

**RPA** (Robotic Process Automation) is the expertise that enables software 'robots' to carry out repetitive, rule-based digital tasks. Humans typically perform these tasks through the user interface, using the mouse and keyboard. RPA robots are capable of mimicking human actions, and they are typically more accurate, faster, and more consistent at it.

**Automation** is a term that describes more accurately these possibilities that exceed the sphere of basic RPA. Sometimes we use these two terms interchangeably, given that RPA is still at the core of automation.

But Nowadays RPA is not just about automating the rule-based processes, it's evolving by training the robot to make an intelligent decision and work on data. Combined with **Artificial Intelligence** (AI), RPA can target more sophisticated work. This opens up endless possibilities on the path towards a fully automated enterprise.

## 4.3. Myths about RPA:

a) Automation will replace humans in the workforce.

It’s true that many business tasks previously performed by human employees can now be automated with RPA.

Yet, even with the rise of artificial intelligence (AI), these expertise are not totally independent from humans, nor are they presently able to replicate the higher-level thinking of which humans are accomplished.

At the same time, the human workforce will certainly be augmented by RPA. Companies and their employees will benefit from it. For example:

1. RPA permits employees to upsurge their efficiency and productivity, and employees will be able to

focus on higher-level happenings, such as sales or marketing. These actions create business value and stand-in deeper assignation with customers.

2. Employee roles and responsibilities are redefined, and talent is moved to focus on customer-facing tasks in the front office. There will no lengthier be a need to focus on boring, back-office tasks.

b) RPA software robots are precise and correct

RPA software robots can make mistakes. The robots will follow the instructions given. If the process is flawed, since the RPA robot cannot improve any defects, the automation will be flawed.

It is up to human employees to spot early in the automation process any mistakes present in the instructions. Unless detected, these errors force that the work will need to be rebuilt, either physically or by re-automating tasks after the errors have been fixed.

c) Any process that can be automated with RPA can also be automated using APIs and programming languages

It is true, but automation done by APIs and programming languages consumes more resources. Typically, the processes need to be first redesigned and only then automated. RPA, on the other hand, is non-invasive, because it aims to automate the processes as they are carried out by human users. This makes it a scalable approach.

d) RPA will not work in my business/Industry

There's a common misapprehension that RPA can be use in certain industries, such as accounts and finance. However, automatable tasks exist in every business. RPA can be functional to almost any ordinary, rules-based, high-volume business action in any business. Here are a couple of scenarios in which RPA can be used:

- Order processing and execution in retail
- Claims processing and execution in the insurance industry
- Fraud detection and fixing in banks
- Communication and relationship with customers in the manufacturing business
- Patient appointments scheduling in the healthcare industry

e) RPA is not value added or the cost of investment.

RPA does have preliminary implementation costs, but the investment is not as important as a business process management software (BPM) or enterprise resource planning (ERP) operation.

At the same time, RPA delivers quick internal cost reduction and substantial increases in ROI (Return of Investment). Some of the paybacks are:

- Operating costs cut down
- Task efficiency increased
- Reduced compliance errors and risks
- Improved customer experience

- Increased employee satisfaction and engagement
- Accelerated digital transformation

## 4.4. Where RPA can be implemented:

One of the myths debunked above was "Automation will not work in my industry". In fact, automation can take on almost any repetitive task performed on a machine. Further you can find those in any industry.

Let's see some examples that are common in many businesses:

a) Payroll Processing:

Payroll processing refers to the actions that companies take to pay their employees—keeping track of their presence, of their salaries, bonuses, and taxes, etc.

Payroll processing needs physical interference month after month, every year that to be automated so that the bot can handle payroll processing as well as claim processing submitted by employees.

RPA service and system can be used to extract the facts that are required from saved timesheets by employees and estimate the pay from their specified contracts and pay them as well (by even ordering the essential bank dealings).

b) Customer Complaint Processing:

Irrespective of industry, customer complaints are always a part of any service for serving better customer service. Their feedback is an important indicator of the business's health and a good predictor of the future of the company.

Through RPA, customer complaints can be categorized based on keywords and other criteria, and practical solutions can be recommended to the customers right away. By automating this, the customer grievances can be replied 24 x 7 instead of 8 hours a day and only 7 days a week.

c) Client Information Updating:

Any organization that has implemented a CRM (Customer Relationship Management) faces all sorts of related issues: the client-base is varies across many topographies, there are frequent request calls to the back-end databases, updates, and save the changes are coming from all sources.

RPA solutions can process these requests in bulk instead of one after the other, reducing the load on the back-end schemes and ensuring improved performance and data quality across the whole application, also a failure of one request doesn’t stop the bot from processing other requests.

## 4.5. Various RPA tools:

There are lots of tools available in the market for RPA, each tool has its own feature and capabilities. Also, open-source RPA tools are there to automate the business processes. An enterprise should check the feasibility of the tool to complete the requirement and then should choose based on their requirement, the cost involved, etc.

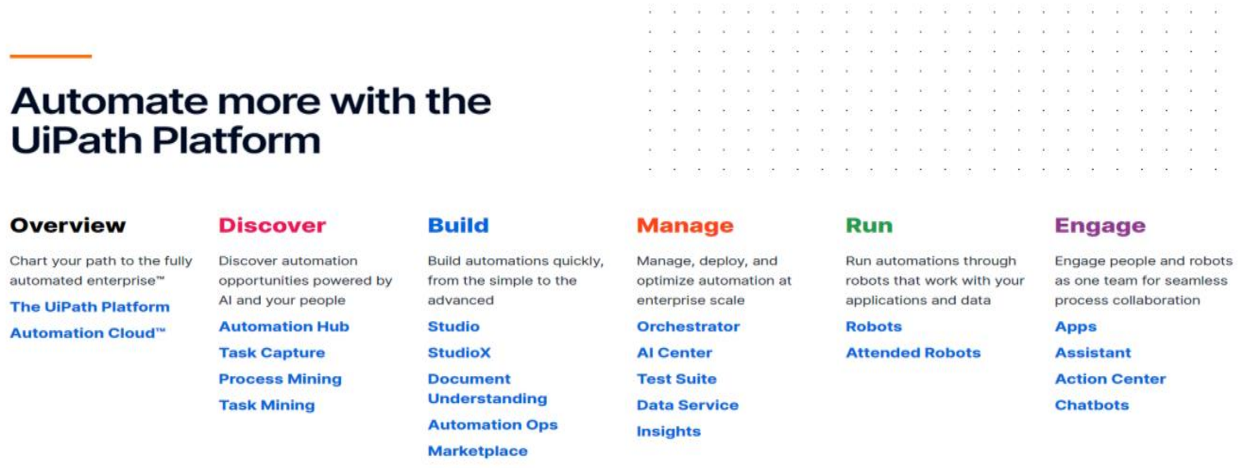
Below are few mostly used tools of RPA.

1. UiPath
2. Blue Prism
3. Automation Anywhere
4. Pega
5. WorkFusion

## 4.6. Why UiPath:

UiPath gives a wide range of products during the whole lifecycle of RPA. There are different products available for each phase of process automation, also it offers different integration with other tools. Also, UiPath provides the concept of an attended robot that enables the user to trigger the automation when they want and provide the input to complete the process run.

Below is the product suite of UiPath during each stage of automation.



Automate more with the UiPath Platform

**Overview**

Chart your path to the fully automated enterprise™

The UiPath Platform

Automation Cloud™

**Discover**

Discover automation opportunities powered by AI and your people

Automation Hub

Task Capture

Process Mining

Task Mining

**Build**

Build automations quickly, from the simple to the advanced

Studio

StudioX

Document Understanding

Automation Ops

Marketplace

**Manage**

Manage, deploy, and optimize automation at enterprise scale

Orchestrator

AI Center

Test Suite

Data Service

Insights

**Run**

Run automations through robots that work with your applications and data

Robots

Attended Robots

**Engage**

Engage people and robots as one team for seamless process collaboration

Apps

Assistant

Action Center

Chatbots

**Fig 4.1: Product Suite of UiPath**

Also, This Forrester Trend scrutinizes the rapidly maturing RPA market to appreciate and evaluate this developing space - highlighting what's significant to appearance for in RPA providers and providing an valuation of the vendors.

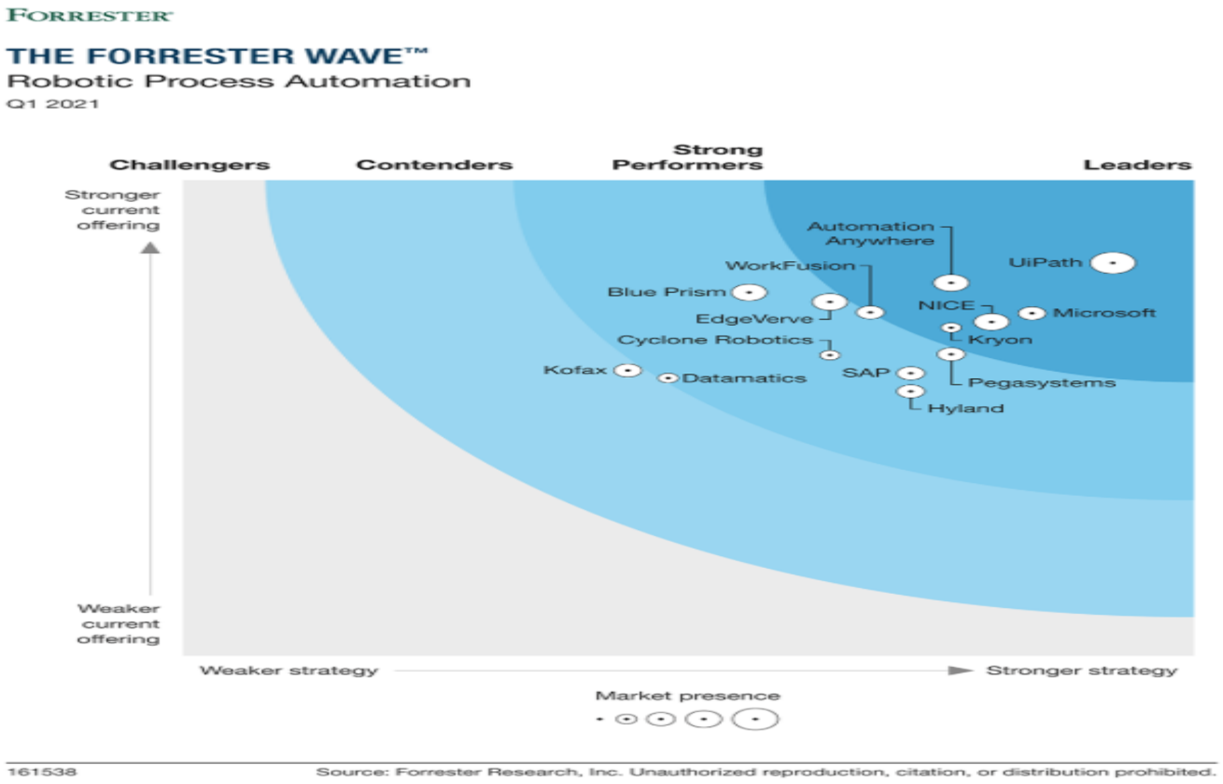


**Fig 4.2: Forrester Trend Scrutinizes**

By means of a thorough and clear evaluation methodology, The Forrester Wave™: Robotic Process Automation, Q1 2021 names UiPath a Leader with the maximum ranking in each of three classes: Current Offering, Strategy, and Market Presence.

## 4.7. UiPath Enterprise Platform (Lifecycle of automating RPA process)

Each project has its lifecycle that mapped to the model. RPA process also has a lifecycle to automate the processes It starts with requirement gathering, checking feasibility or doing POC (Proof of Concept), Documentation, building the automation, testing and deployment in the production environment, running the automations, notifying the unhandled cases for the process improvement.

Below is the lifecycle of process automation and how the UiPath product performs a role during this entire lifecycle of automation.

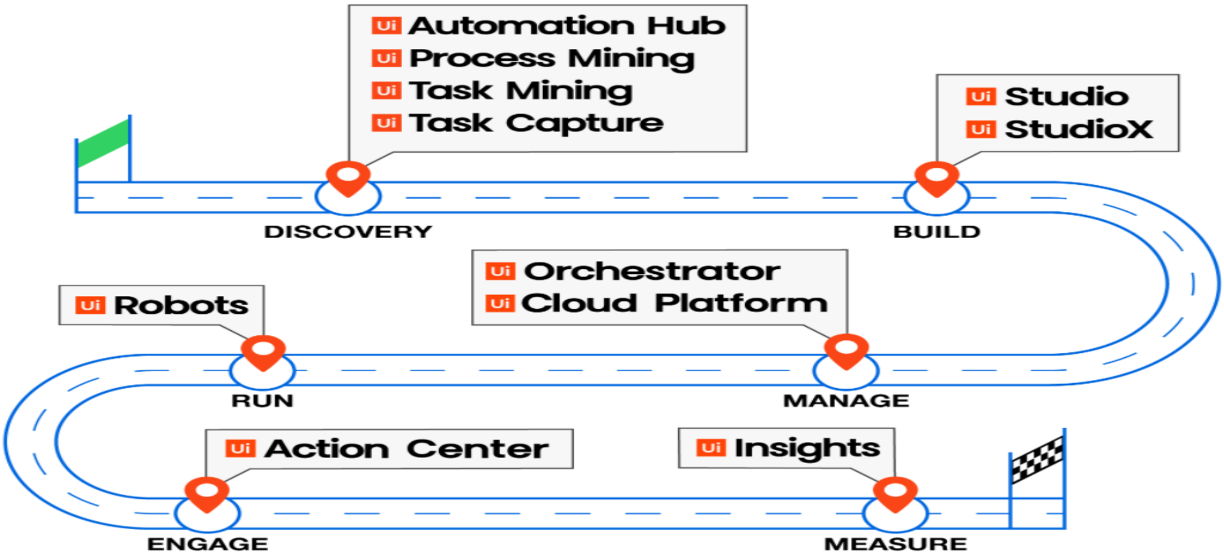


**Fig 4.3: Lifecycle of Automation**

a) Requirement Gathering (discovery):

The typical enterprise-scale RPA implementation starts with testing the capabilities of automation for one process. And the results are good, so naturally, you want to automate more. As the scaling starts, you need the best understanding of the "as is" processes in order to prioritize them for automation.

The challenges of process discovery

1. Process maps are built by business users or process experts and lack reliable data on types of exceptions and how often they occur.
2. Even when process data is available, it's difficult to bring all sources together to build objective and meaningful process maps.
3. Even when process maps are available, it's hard to assess their automation suitability and prioritize them.

**The Main Features of UiPath Task Mining**

**Centralized Tool with Easy Setup**

The tool is easy to setup and manage from UiPath Orchestrator, our platform component for managing RPA. It consists of a central console and agents installed on the machines of the users. The agents feed information into the central engine.

Once the process map is built, it can be easily exported to a Process Definition Document or UiPath Studio workflow file.

**Data-driven Process Map**

Based on the ways tasks are performed by users on their machines, UiPath Task Mining builds a centralized process map, containing everything from a number of steps, clicks, and execution time to actual screenshots.

**Process Insights & Automation Report**

Based on the scientific data analysis, UiPath Task Mining provides an RPA score for each task and a report with automation suggestions.

### The Main Features of UiPath Process Mining

#### Comprehensive View on Processes Based on Existing Data

UiPath Process Mining turns existing back-end data from your IT systems, databases, and flat files into an end-to-end process visualization. It connects easily with more than 40 ERP, CRM, and other applications and with any database without third-party tools.

#### Automation Opportunity Assessment

UiPath Process Mining offers data-driven insights to support the process of prioritizing the automation ideas in your organization.

#### Continuous Process Monitoring

The intuitive process visualization enables continuous monitoring of your processes for further improvement.

b) Development (Build):

UiPath Studio is at the core of our RPA capabilities. It's where RPA goes from idea to reality. and while it was created for professional developers, it shares with StudioX some capabilities that make it easy to use, such as the visual editor and the drag-and-drop activities.

How does it look like?

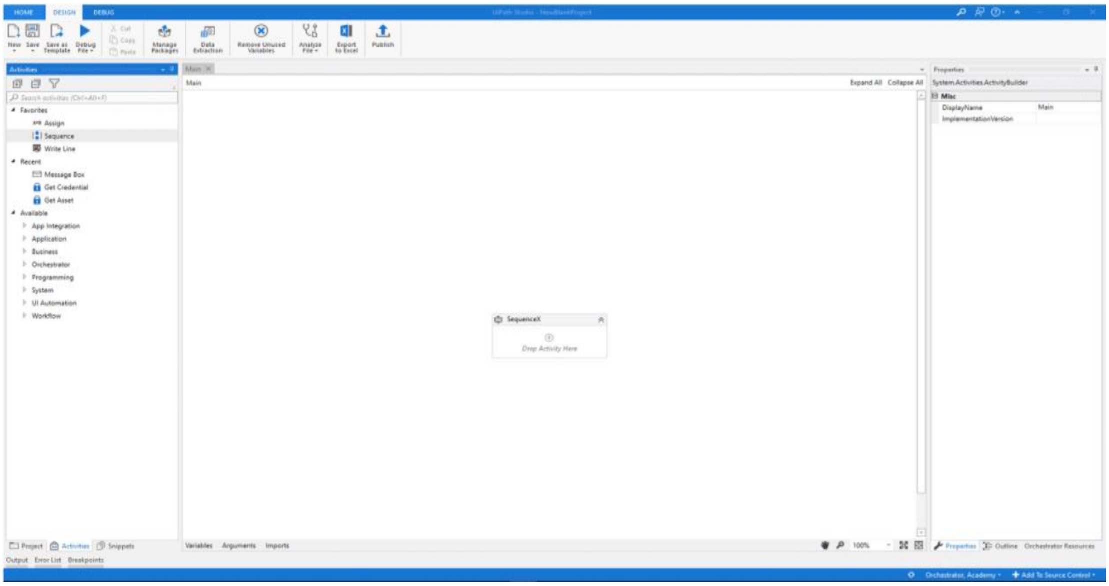


**Fig 4.4: UiPath StudioX**

c) Testing & Deployment (Manage):

On top of the main RPA management capabilities presented in the previous section, many large enterprise clients have certain needs, starting with having the solution installed on-premises or in a private cloud.

### What UiPath Orchestrator offers

UiPath Orchestrator has been historically our first solution for robot and license management. After the release of the UiPath Cloud Platform, it is mostly preferred by large enterprise customers with on-prem or private cloud deployments.

While sharing most of the features with UiPath Cloud Platform, the features below cater to these large enterprise customers:

### Enterprise-Scale Management Capabilities

Folders are available in UiPath Orchestrator to separate automation workflows and user rights. Together with the integration with Active Directory at the user and group level, it provides low-touch license management: as long as a user is part of an Active Directory group with access to licenses, a robot will be automatically provisioned and will have the group rights.

### Credential Store Integration

If the company has already a credential store set up, it can be integrated with Orchestrator. The users and robots will not only be able to access it, but they will also be able to edit.

### High-Availability

UiPath Orchestrator can be deployed as a multi-node, using a solution developed for UiPath. This solution ensures permanent availability and disaster recovery. At the same time, the multi-node infrastructure can be used to ensure a balanced distribution of robots.

### Queues with Predictive SLAs

As many large organizations tend to have clear and strict SLAs for processes, these should apply to automation. Queues in UiPath Orchestrator can be set up with SLAs and clear procedures when these are at risk or exceeded, such as provisioning extra robots.

d) Human in Loop (Engage):

RPA is about creating a true collaboration between robots and human users, with the activities split to match the strengths of each actor - speed and accuracy for the robot, and critical thinking and complex decision-making for humans.

Engage is about creating an ecosystem in which the way activities are carried out by humans and robots don't generate breaks in the processes, instead of ensuring a continuous and effective flow.

### What UiPath Action Center Offers?

### Tasks for Robot-Human Hand-off

- Whenever the robot reaches a human intervention point, a Task is being created and sent to Orchestrator. The robot is now freed to take the next job;
- The human user receives the task in a centralized inbox and provides input in different forms;
- A free robot (not necessarily the one that created the task) can resume the automation from where it was handed off to the human user.

**Flexibility**

- Human in the loop scenarios can be accommodated by the UiPath Enterprise RPA Platform through specific activities and processes in UiPath Studio and flexible integration with Orchestrator through Jobs, Tasks and Queues;
- Tasks can be accessed and processed by humans using the Orchestrator mobile application or desktop instance.

**End-to-End Process Monitoring**

Carrying out a long-running workflow as a single process is not enough, you need to be able to monitor it as a single process. And this is what we offer through the Process Monitoring capability in Orchestrator. You get:

- a process execution summary of robots, human, and triggers to identify and resolve bottlenecks;
- End-to-end visibility across the entire process in a single place, to be able to make decisions to optimize it.

e) Stabilize the process – Hyper care (Measure)

Measuring return on investment (ROI) and process indicators is something that most companies do for the growth in business. And RPA implementations should be measured both as regular business processes, and also using specific metrics.

Monitoring can start with the first RPA implementation and should become an important part of the cycle when companies scale RPA and look for continuous improvement.

**UiPath Insights is a powerful, embedded analytics tool that helps you to quantify, report, and bring into line RPA processes with tactical business results.**

f) Constant Improvement

The process automation performance is assessed, the benefits tracked for the constant delivery, and the changes managed for making the process automation stable, covering all scenarios and less prone to errors.

## 4.8. UiPath Platform

The UiPath Platform offers you the components you need to design and develop automation projects, execute the instructions automatically and manage your robot workforce. The components of the UiPath Platform are Studio (the workflow designer), Orchestrator (the robot management platform), and the Robot (the agent executing the instructions).

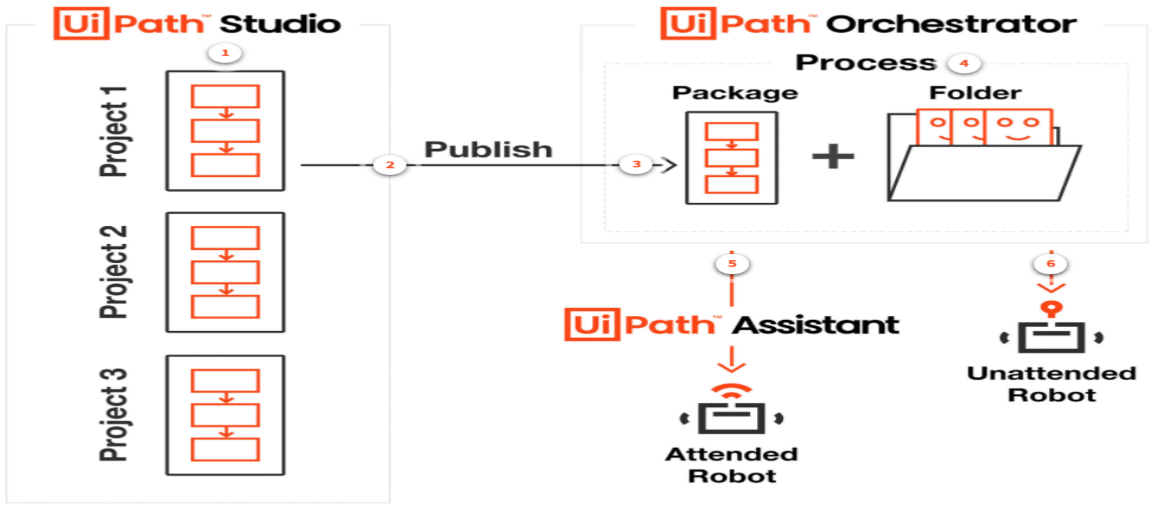


**Fig 4.5: UiPath Platform**

a) UiPath Studio

Helps you to design automation workflows visually, quickly, and with only basic or no programming knowledge. The Studio is where the automated processes are built in a visual way, using the built-in recorder, drag & drop activities, and best practice templates.

Installing UiPath Studio Community Edition

The best way to understand the UiPath Suite components is to start working with them. During this learning plan, you will have lots of opportunities to practice, so let's start by installing UiPath Studio. Note that there are three installation options:

**Community Edition**

Permanently free. Upgrade to Enterprise easily.

- 2 Studios for sketching and designing automation
- 3 Robots
- Cloud-hosted Orchestrator
- Forum and group support
- UiPath Conservatoire access

**Enterprise Server Edition**

On-premises initiative deployments for large industries.

- Limitless Studios for designing automation
- Unrestricted Robots
- On-premises Orchestrator
- Superior Support

- Balance as you grow
- Self-controlled updates
- UiPath certified training associates
-

### Enterprise Cloud Edition

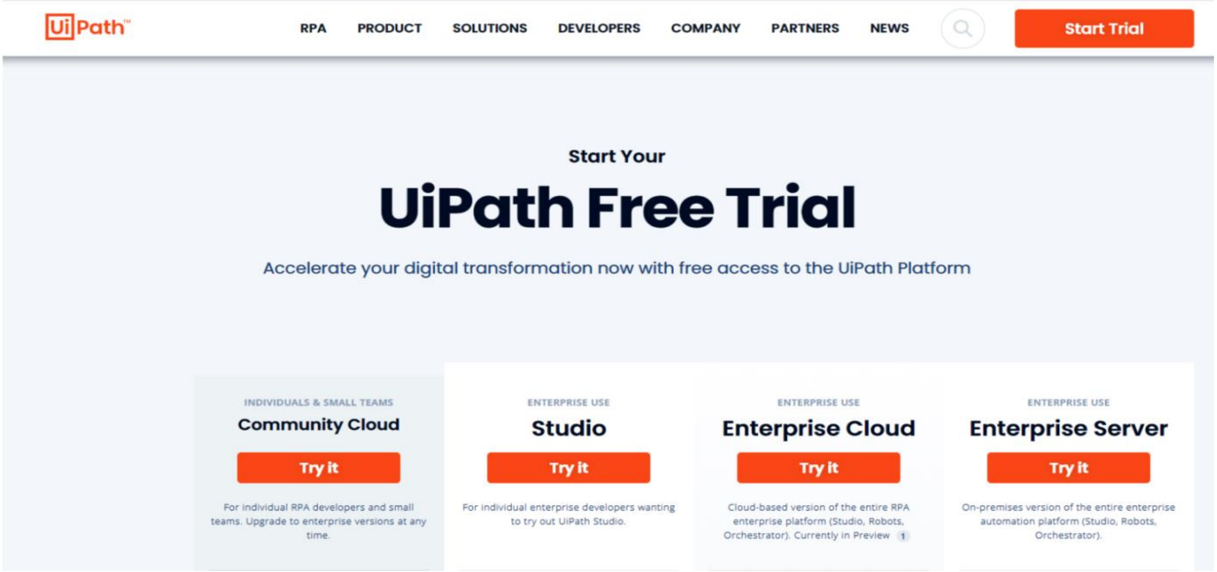
Cloud enterprise deployments for industries of any size. Presently in Showing.

- Limitless Studios for designing automation
- Unrestricted Robots
- Cloud-hosted Orchestrator
- Superior Support
- Balance as you grow
- Continually up to date
- Integrated user access management
- Safe and compliant
- UiPath authorized training buddies

### How to install UiPath Studio?

Follow these steps to install Studio using the UiPath Community Edition.

**Step 1 - Open the** UiPath Start Trial **webpage.**



Start Your

UiPath Free Trial

Accelerate your digital transformation now with free access to the UiPath Platform

INDIVIDUALS & SMALL TEAMS
Community Cloud
For individual RPA developers and small teams. Upgrade to enterprise versions at any time.

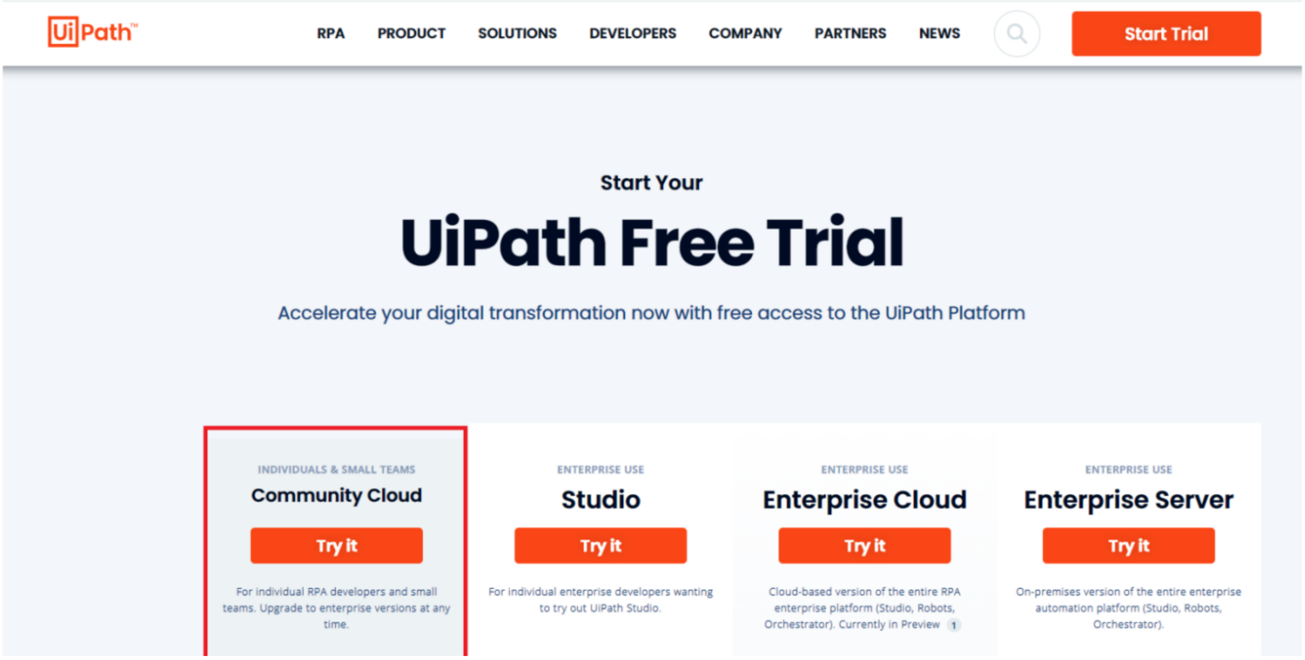
ENTERPRISE USE
Studio
For individual enterprise developers wanting to try out UiPath Studio.

ENTERPRISE USE
Enterprise Cloud
Cloud-based version of the entire RPA enterprise platform (Studio, Robots, Orchestrator). Currently in Preview 1

ENTERPRISE USE
Enterprise Server
On-premises version of the entire enterprise automation platform (Studio, Robots, Orchestrator).

**Fig 4.6: Step-1 Start of UiPath**

Step 2 - Select the **Community Cloud** option and log into your UiPath Platform account (you will need to create an account if you do not have one yet).



Start Your

UiPath Free Trial

Accelerate your digital transformation now with free access to the UiPath Platform

For individual RPA developers and small teams. Upgrade to enterprise versions at any time.

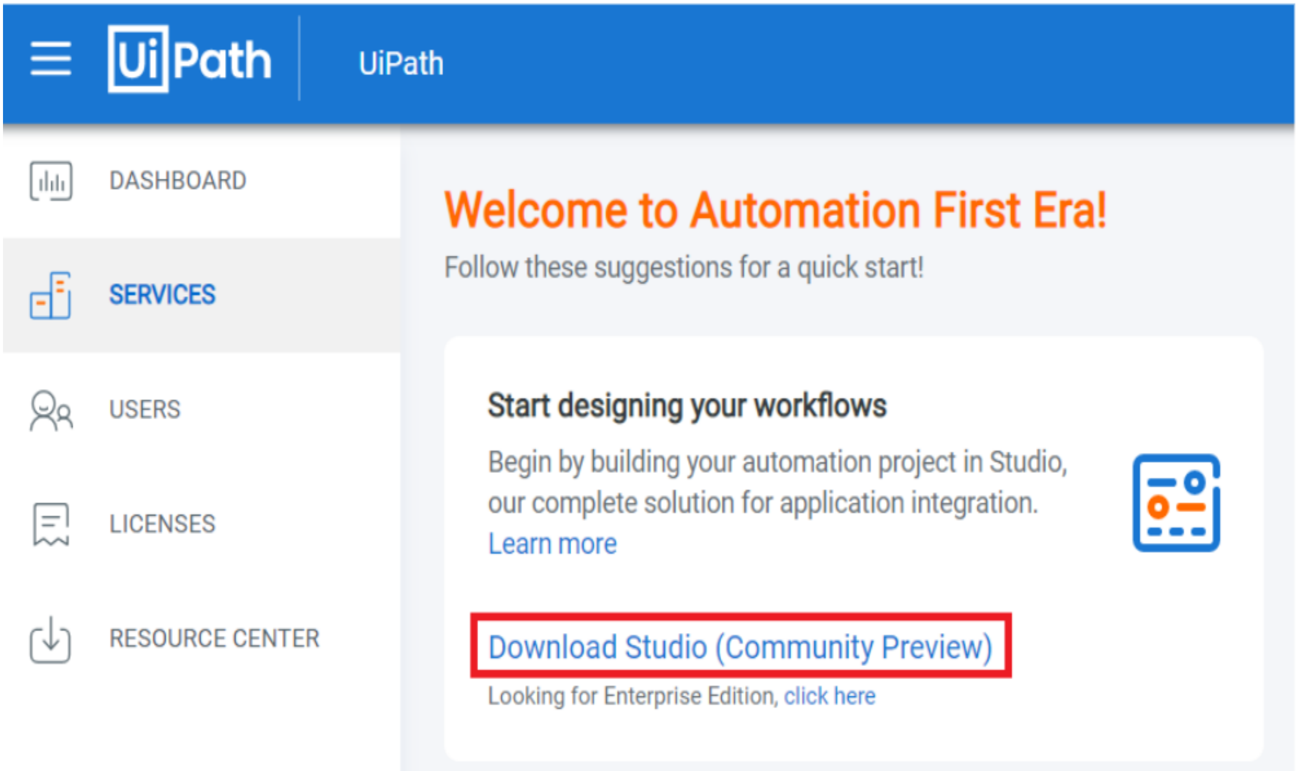
For individual enterprise developers wanting to try out UiPath Studio.

Cloud-based version of the entire RPA enterprise platform (Studio, Robots, Orchestrator). Currently in Preview 1

On-premises version of the entire enterprise automation platform (Studio, Robots, Orchestrator).

**Fig 4.7: Step-2 Start of UiPath**

Step 3 - Select the **Download Studio (Community Preview)** option from the Services tab and install the **UiPathStudioSetup** exec file.



Welcome to Automation First Era!

Follow these suggestions for a quick start!

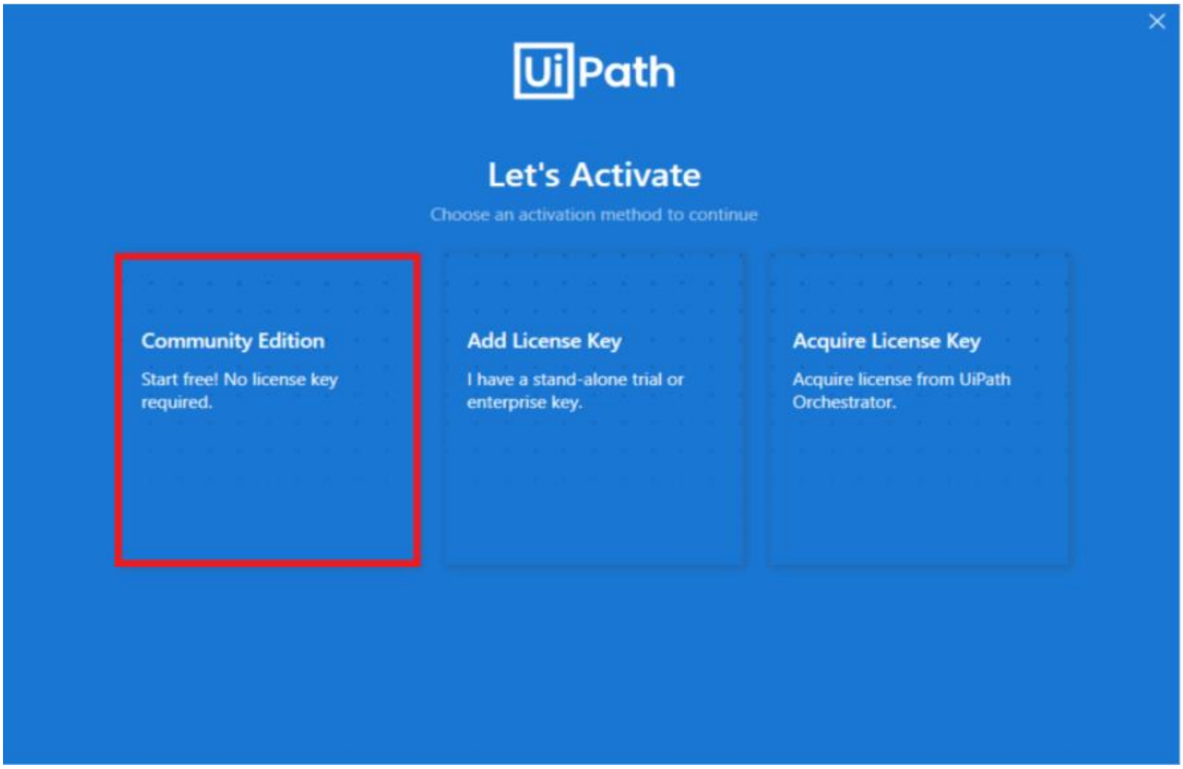
Start designing your workflows

Begin by building your automation project in Studio, our complete solution for application integration. Learn more

Looking for Enterprise Edition, click here

**Fig 4.8: Step-3 Automation First**

Step 4 - Select the **Community Edition** option from the Activation Method screen.



Let's Activate

Choose an activation method to continue

Community Edition

Start free! No license key required.

Add License Key

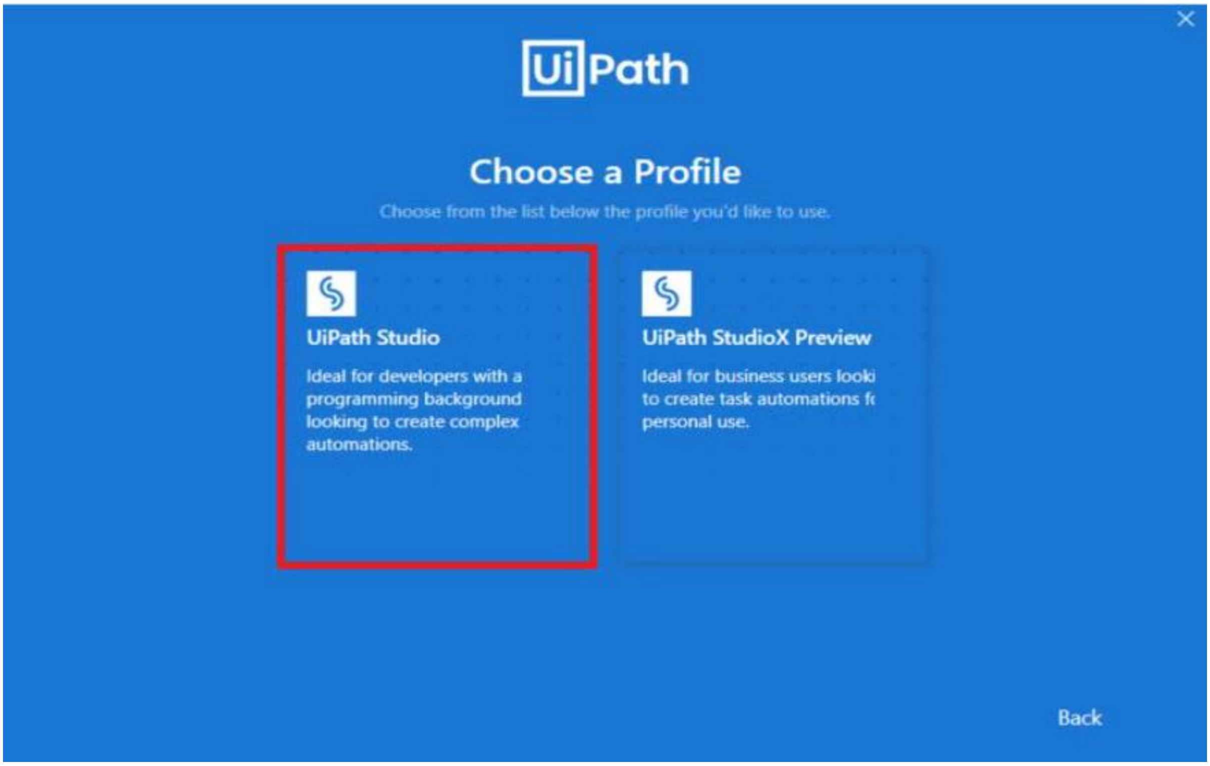
I have a stand-alone trial or enterprise key.

Acquire License Key

Acquire license from UiPath Orchestrator.

**Fig 4.9: Step-4 Activation**

Step 5 - Select the **UiPath Studio** option from the Profile screen.



Choose a Profile

Choose from the list below the profile you'd like to use.

UiPath Studio

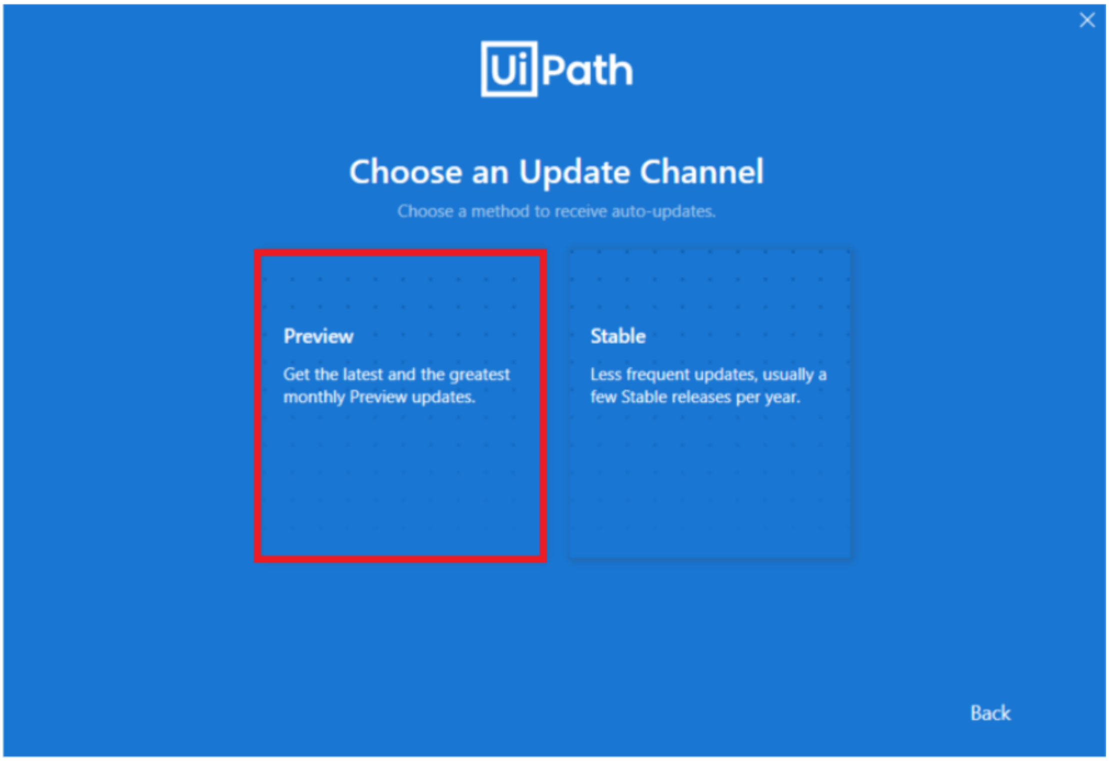
Ideal for developers with a programming background looking to create complex automations.

UiPath StudioX Preview

Ideal for business users looki to create task automations fc personal use.

**Fig 4.10: Profile Selection**

Step 6 – Select the **Preview** option from the Update Channel screen. You're now ready to start building your first automation.



Choose an Update Channel

Choose a method to receive auto-updates.

**Preview**

Get the latest and the greatest monthly Preview updates.

**Stable**

Less frequent updates, usually a few Stable releases per year.

**Fig.4.11: Selection of Channel**

b) UiPath Orchestrator

Orchestrator is the module of the UiPath Platform in control of the management of automation, users, and robots, as well as the management of the resources used in the development for execution or in running mechanization.

UiPath Orchestrator let you control, manage and monitor the robots. It is also the residence where libraries, recyclable components, assets and procedures used by the robots are deposited. Orchestrator is a server application retrieved via browser, through which the robotic workforce is administered and controlled, achieved and monitored:

- The interconnections with the robots are shaped and sustained, and the robots are assembled.
- The automated procedures are disseminated as activity to the robots
- The execution of tasks is registered and kept track of checking

**Orchestrator's main capabilities**

- **Provisioning**: creates and maintains the connection with robots and attended users
- **Control**: enables the creation, assignment, and maintenance of licenses, roles, permissions, groups, and folder hierarchies;
- **Storage and distribution**: allow the controlled storage and distribution of automation projects, assets, and credentials, as well as large files used in automation

• **Running automation jobs in unattended mode:** enables the creation and distribution of automation jobs in various ways, including through queues and triggers

• **Monitoring**: allows monitoring of jobs and robots and stores logs for auditing and analytics.

❖ **Deployment options**

There are 3 ways to deploy Orchestrator:

✓ **UiPath Cloud**

Orchestrator is available as a service inside Automation Cloud, our cloud platform. This is SaaS, by far the easiest to set up.

✓ **On-premises**

Orchestrator needs to be installed as a standalone product in the customer's infrastructure.

✓ **Private Cloud**

Similar to the on-premises solution, with the difference that it is installed in a private cloud managed by the customer's infrastructure team.

c) The Robot

Simply put, a software robot is an execution agent that runs automations built with the Studio family and then published as packages either locally, on the same machine as the robot, or via Orchestrator.

As introduced above, there are two types of UiPath robots and they differ both in the way they work and in the way they are licensed:

Executes the workflows and instructions sent locally or via Orchestrator. There are two types of robots:

• Attended – is triggered by user events, and operates alongside a human, on the same workstation.

• Unattended - run unattended in virtual environments and can automate any number of processes.

❖ **Attended Robots**

They are digital helpers for human users. They work on the same machines as the humans, during the same hours. They are triggered directly by humans (usually through UiPath Assistant) or by an event related to what the human user does. For example, opening an application or receiving an email.

There are two main categories of human users working with Attended Robots:

- **Automation Users**: they benefit from having Attended Robots by handing them over the repetitive and mundane tasks in their work. Some examples of automation users are contact center agents, financial analysts or support specialists;

- **Automation Developers**: their role involves (at least partially) developing automations. Thus, their license includes access to the Studio family on top of the attended robot.

❖ **Unattended Robots**

Although they run automations almost the same way as attended robots, these are meant to work non-stop, with as less input from human users as possible. They are deployed on separate machines and their jobs are triggered exclusively from Orchestrator.

Their interactions with the human users are typically handled with as less disruption as possible, by creating and sending requests for human input or validation as tasks. While these await to be processed, unattended robots can continue their work by picking up other jobs. When the human input is finally provided, unattended robots can resume their work on the process.

### 4.9. Conclusion

The RPA technology comprises huge potential in shifting the way industries function. Unquestionably, every corporate across the world will get advantage from the capability of an automation system in the forthcoming. As long as businesses are on the exploration for novel solutions, a sophisticated ROI and lesser overhead, RPA will remain becoming more widespread as well as sophisticated.

As discussed, RPA is a great tool to help you eliminate some mundane and repetitive processes while allowing your employees to focus on value-add work. The key is focusing on the processes that will provide you the most value by being automated. This value is both quantitative and qualitative. Quantitative in the sense of saving money, reallocating money elsewhere, or bringing in more money. While qualitatively, you may improve employee satisfaction if you can remove mundane tasks from your employee's plate.

Beginning your RPA journey by utilizing a consulting company with RPA experience, like UDig, can help get you off the ground running, bringing you expertise and best practices to make your implementation more efficient and worthwhile.

You only have one chance to make a good first impression with RPA at your company, so doing the upfront business case and process analysis to pick the right first use case is critical. If you are wondering where to start, check out some of the other RPA related blogs UDig has written. Check out what to look for when evaluating RPA tools and the pros and cons of the top RPA tools in the bazaar and market.

**References:**

1. https://www.ibm.com/cloud/architecture/architectures/roboticProcessAutomationDomain/reference-architecture/
2. https://www.uipath.com/rpa/robotic-process-automation
3. 3.https://www.i-scoop.eu/robotic-process-automation-rpa/
4. 4.https://en.m.wikipedia.org/wiki/Robotic_process_automation
5. 5.https://www.pega.com/rpa-survey?&utm_source=google&utm_medium=cpc&utm_campaign=&utm_term=%2Brpa%20%2Brobotics&gloc=1007788&utm_content=pcrid|482227154380|pkw|kwd-313344480603|pmt|b|pdv|m|&gclsrc=aw.ds&gclid=CjwKCAjwr56IBhAvEiwA1fuqGp0GWPCLoLCFnD6xV1PsnsuV3cflNT1i57OrInjTvtdCkCGq4c_RGRoC0rMQAvD_BwE

# 5. Internet of Things (IoT)

Mr. Anchal Koshta

Dr. Milind Godase

### 5.1. Introduction

The IoT is a emerging subject of technical, societal and financial importance. Consumer products, durable goods, cars and trucks, industrial and utility components, sensors and other everyday items are being integrated with Internet connectivity and powerful data analytics capabilities that promise the way you work live and play.

Estimates impact of IoT on Internet and economy are remarkable, with some expecting that by 2025 there will be more than 100 billion of connected IoT devices and a global financial effect of more than $ 11 trillion.

### 5.1.1. What is IoT?

Internet of Things is an idea of connecting any device to the Internet and other connected devices. IoT is a huge network of connected things and people - all of which gather and share data about how they use it and the environment around them.

Which comprises a fantastic number of items of all shapes and sizes - from smart microwaves to automatic cooking at just the right time, to self-driving cars, whose complex sensors find objects on their way, your heart bit rate for a wearable fitness device and the number of steps you took that day, then use that information to suggest exercise plans that suit you. There are also connected footballs that can track how far and fast they are thrown and record those figures through the app for forthcoming preparation purposes.

### 5.1.2. How does IoT work?

The devices and objects with built in sensors are connected to the IoT platform, which integrates data across different devices and applies analytics to share the most valuable information with applications intended to meet particular needs.

This powerful IoT platform can accurately tell which information is useful and which can be ignored. This information can be used to find patterns, make recommendations, and find them before potential problems arise.

Let’s say, if I have a car manufacturing business, I probably want to know which alternative components (leather seats or alloy wheels) are the most popular. With IoT, I can:

- Use sensors to identify which regions in a showroom are the most popular, and where consumers stay longest;
- Mining the available sales data to detect which components are selling fastest;
- Automatically line up sales data with supply, so that popular items don’t go out of stock.

The information picked up by the connected devices facilitates me to make smart decisions about which components to stock based on real-time information, which helps me save time and money.

Through the insight provided by advanced analytics gives the power to make processes more efficient. Smart objects and systems mean you can automate certain tasks, particularly when these are repetitive, ordinary, time-consuming or even risky. Let's look at one example to see what this looks like in real life.

### 5.1.3. IoT at your home

Imagine you wake up at 7 am every day to go to work. Your alarm clock does the job of waking you just fine. That is, until something goes wrong. Your train's cancelled and you have to drive to work instead. The only problem is that it takes longer to drive, and you would have needed to get up at 6.45 am to avoid being late. Oh, and it is driving in rain, so you'll need to drive slower than usual. IoT-enabled alarm clock would reset itself based on all these factors, to ensure you got to work on time. It could recognize that your usual train is cancelled, calculate the driving distance and travel time for your alternative route to work, check the weather and factor in slower travelling speed because of heavy rain, and calculate when it needs to wake you up so you are not late. If it is super-smart, if might even synchronize with your IoT-enabled coffee maker, to ensure your morning caffeine's ready to go when you get up.

## 5.2. Installation of Atmel Studio

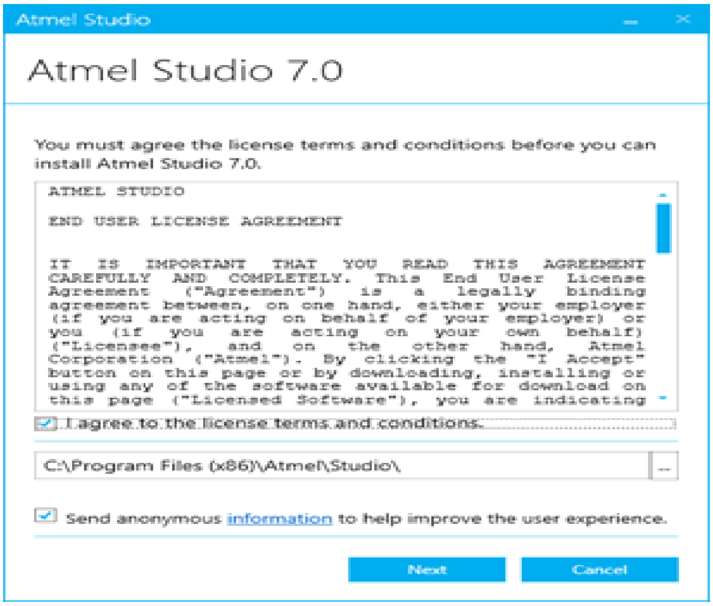
Following are the steps to install the Atmel Studio on your Laptop.

- **Run the Installer**

Open the location, wherever you downloaded the installer, and then run the installer:

```
as-installer-X.X.XXXX-full.exe
```

Dependent on your Windows security settings, you may be getting a message asking if you are sure you really want to run this program. Click on Yes button, if it prompts you.



Atmel Studio

Atmel Studio 7.0

You must agree the license terms and conditions before you can install Atmel Studio 7.0.

ATMEL STUDIO

END USER LICENSE AGREEMENT

IT IS IMPORTANT THAT YOU READ THIS AGREEMENT CAREFULLY AND COMPLETELY. This End User License Agreement ("Agreement") is a legally binding agreement between, on one hand, either your employer (if you are acting on behalf of your employer) or you (if you are acting on your own behalf) ("Licensee"), and on the other hand, Atmel Corporation ("Atmel"). By clicking the "I Accept" button on this page or by downloading, installing or using any of the software available for download on this page ("Licensed Software"), you are indicating

☑ I agree to the license terms and conditions.

C:\Program Files (x86)\Atmel\Studio\

☑ Send anonymous information to help improve the user experience.

Figure 5.1: Atmel Studio 7.0 installation – License Agreement

- **License Agreement and Choose Location for Installation**

Read the License Agreement, and then confirm to agree to the license terms and conditions.

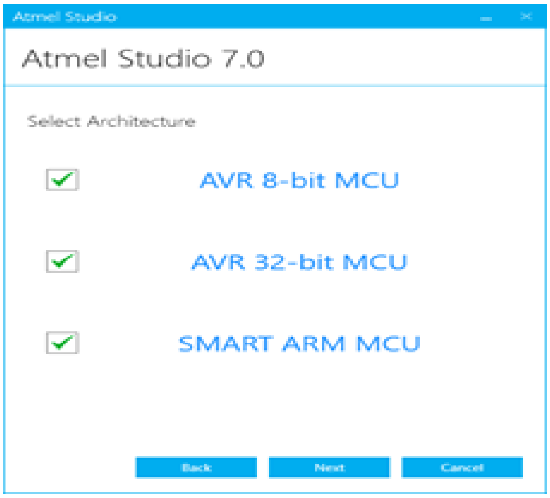
Pick the installation path. The default one is

```
C:\Program   Files(x86)\Atmel\Studio\
```

**Click Next**

- **Architecture you plan to work with – selection**
  - AVR 8-bit MCU

  **Click Next**



Atmel Studio 7.0

Select Architecture

☑ AVR 8-bit MCU

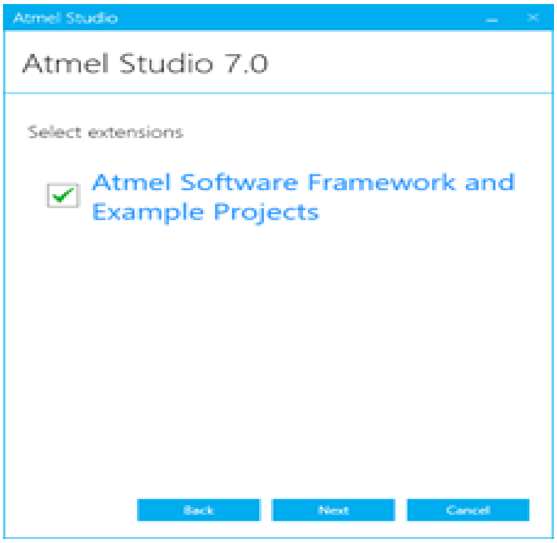
☑ AVR 32-bit MCU

☑ SMART ARM MCU

Figure 5.2: Atmel Studio 7.0 installation – Architecture Selection

- **Select - whether to install the Atmel Software Framework and Example Projects**

  Atmel Software Framework provides



Atmel Studio 7.0

Select extensions

☑ Atmel Software Framework and Example Projects

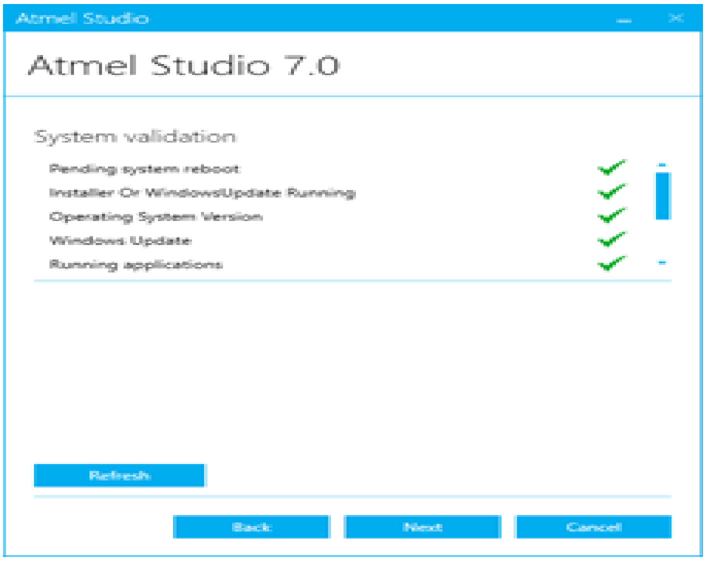
Figure 5.3: Atmel Studio 7.0 installation – Atmel Software Framework

8-bit AVR, 32-bit AVR and ARM Drivers for each MCU peripheral, Hardware components driver, Demo applications which uses all drivers RTOS-ready source code, Complete software framework in C code, optimization in assembly code Full projects compatible with GNU GCC.

It is designed to run on Atmel evaluation kits and reference design which can be easily portable to any other hardware platform

It is also designed to develop software applications for Atmel microcontrollers.

**Click Next**



Atmel Studio

Atmel Studio 7.0

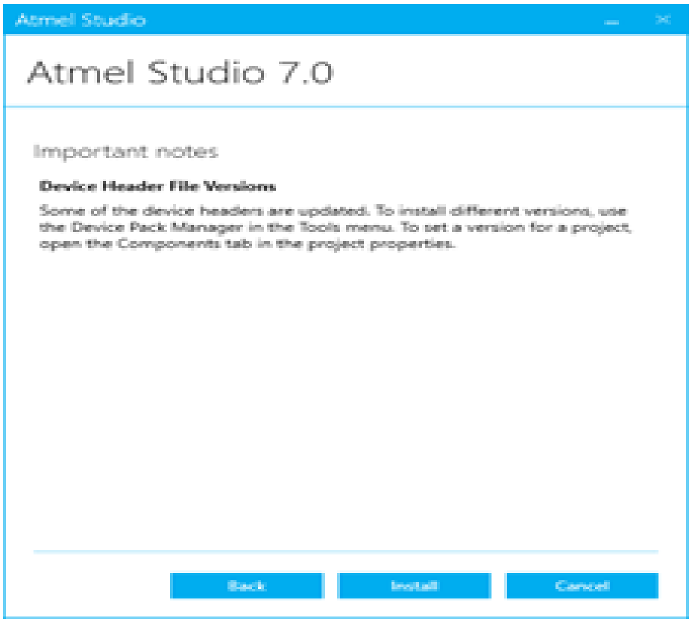
System validation

Pending system reboot ✓
Installer Or WindowsUpdate Running ✓
Operating System Version ✓
Windows Update ✓
Running applications ✓

Figure 5.4: Atmel Studio 7.0 installation – System Validation

- **Accept System Validation**

**Click Next**



Atmel Studio

Atmel Studio 7.0

Important notes

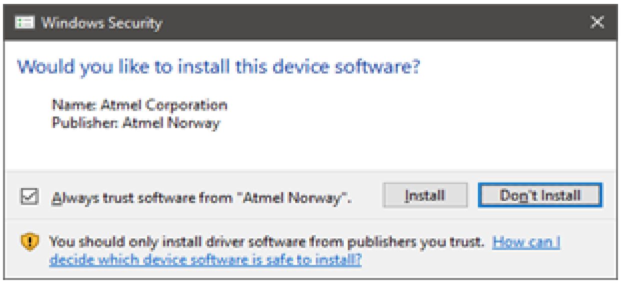
**Device Header File Versions**

Some of the device headers are updated. To install different versions, use the Device Pack Manager in the Tools menu. To set a version for a project, open the Components tab in the project properties.

Figure 5.5: Atmel Studio 7.0 installation – Accept System Validation

- **Begin Installation**

 **Click Next**



Would you like to install this device software?

Name: Atmel Corporation
Publisher: Atmel Norway

☑ Always trust software from "Atmel Norway".

You should only install driver software from publishers you trust. How can I decide which device software is safe to install?

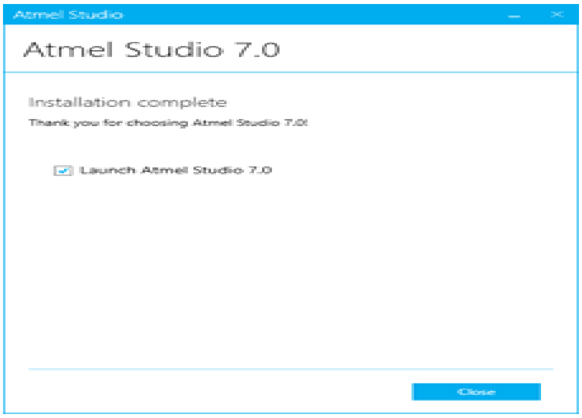
Figure 5.6: Atmel Studio 7.0 installation – Begin installation

- **Accepting the Secondary Installations**

This main installer may launch other secondary installers to install the components you selected in previous steps, allow these installers to run.

**Click on Install**

- **Installation Complete -** Click Close.



Installation complete

Thank you for choosing Atmel Studio 7.0!

☑ Launch Atmel Studio 7.0

Figure 5.7: Atmel Studio 7.0 installation – complete

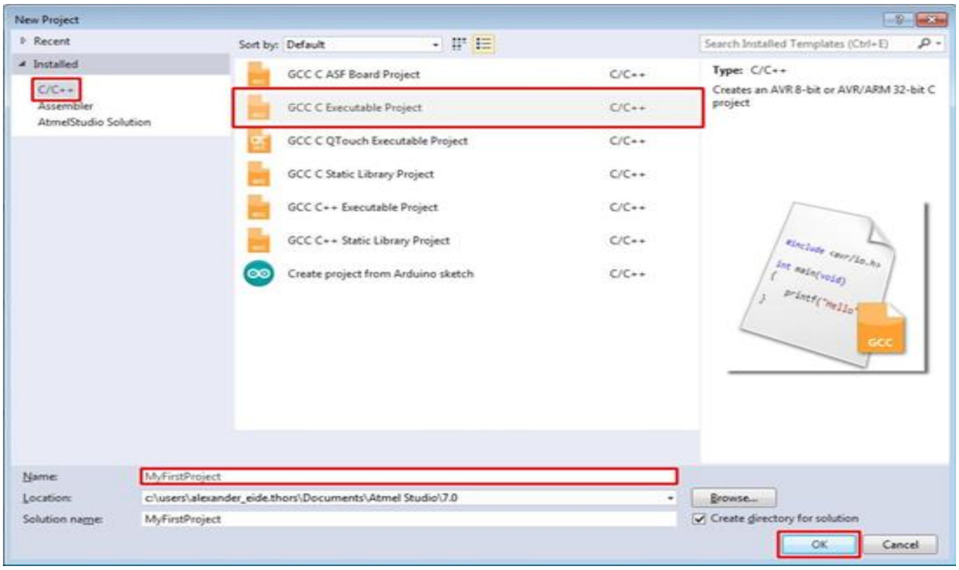
### 5.3. Creating a new Project in Atmel Studio 7

**Atmel Studio 7** has a New Project wizard which allows you to create a project. You can enter into this through the following options:

- File > New > Project from Main Menu
- Press Ctrl + Shift + N
- Click on the New Project icon

New project window will open, provides the option to specify the programming language and project template to be used.

Then choose 'GCC C Executable Project' option from the template list to generate a bare-bones executable project. Provide a project name as MyFirstProject and also provide the path where you want the project to be stored on your computer then Click on **OK.**



C/C++
Assembler
AtmelStudio Solution

| | |
|---|---|
| GCC C ASF Board Project | C/C++ |
| GCC C Executable Project | C/C++ |
| GCC C QTouch Executable Project | C/C++ |
| GCC C Static Library Project | C/C++ |
| GCC C++ Executable Project | C/C++ |
| GCC C++ Static Library Project | C/C++ |
| Create project from Arduino sketch | C/C++ |

Type: C/C++

Creates an AVR 8-bit or AVR/ARM 32-bit C project

```
#include <avr/io.h>
int main(void)
{
    printf("Hello"
}
```

Name: MyFirstProject

Location: c:\users\alexander_eide.thors\Documents\Atmel Studio\7.0

Solution name: MyFirstProject

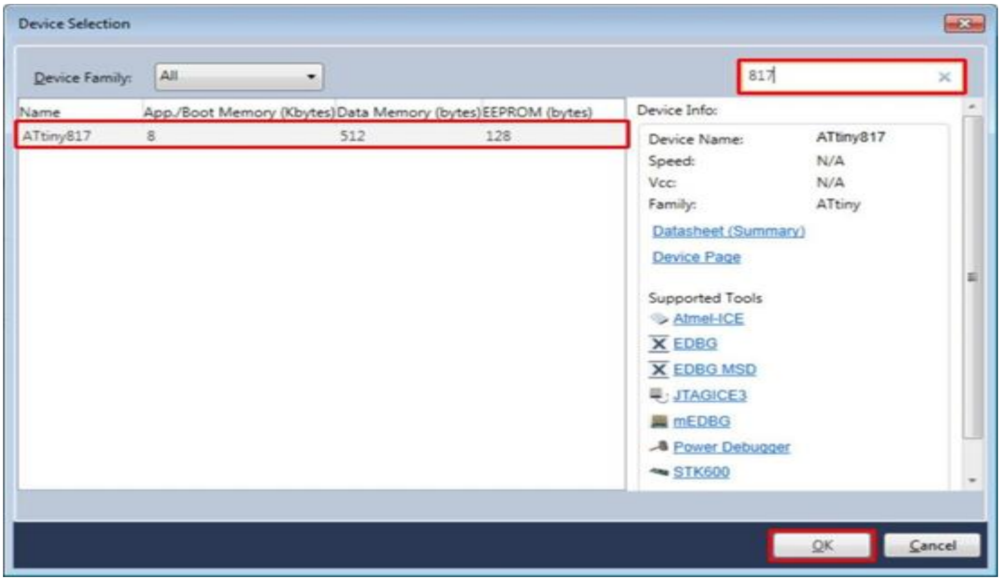
☑ Create directory for solution

Figure 5.8: Creating a new Project in Atmel Studio 7

All Atmel Studio projects belong to a solution. By default, Studio will use the same name for both the newly created solution and project. Solution name field can be used to manually specify the solution name.

The 'Create directory for solution' is selected by default. Then, Atmel Studio will generate a new folder with the specified solution name at the location specified by the 'Location' field.

Next, the *Device Selection* window will appear. It is necessary to specify which device the project will be developed for.



Device Family: All

817

| Name | App./Boot Memory (Kbytes) | Data Memory (bytes) | EEPROM (bytes) |
|---|---|---|---|
| ATtiny817 | 8 | 512 | 128 |

Device Info:

Device Name: ATtiny817
Speed: N/A
Vcc: N/A
Family: ATtiny

Datasheet (Summary)
Device Page

Supported Tools
Atmel-ICE
EDBG
EDBG MSD
JTAGICE3
mEDBG
Power Debugger
STK600

Figure 5.9: Device Selection in Atmel Studio 7

List of devices will be present in the Device Selection dialog, which can be scrolled through. It is possible to narrow the search by using the 'Device Family' drop- down menu or by using the search box.

In the search bar enter the key characters for the device you intend to use, then select the exact device from the list that appears. In this example, "817" was used to find and select the device **ATmega16**.

Then click **OK** to create the project.

A new GCC C Executable   project has now been created for the **ATmega16**   AVR device. The *Solution Explorer* on the right side of the window will list the contents of the newly generated solution.

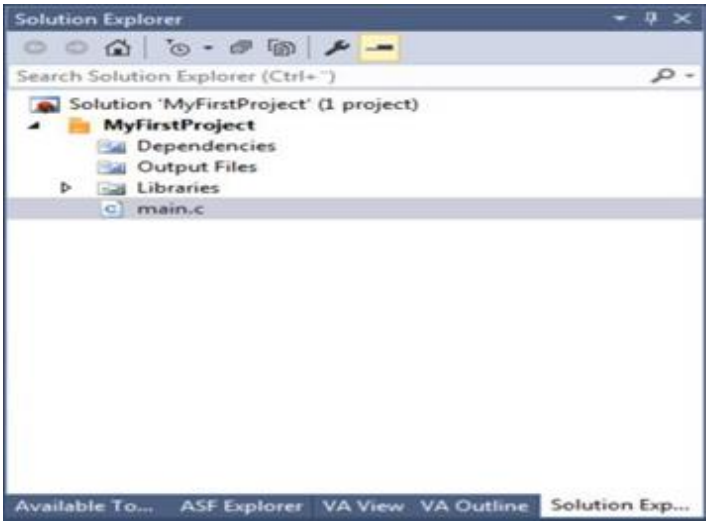
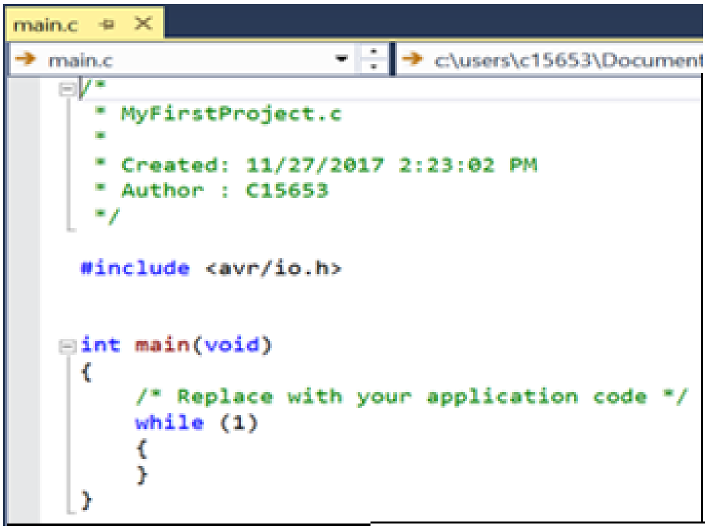


Figure 5.10: Solution Explorer in Atmel Studio 7

A *main.c*  file is automatically created with recommended #include file for the device selected.



```
/*
 * MyFirstProject.c
 *
 * Created: 11/27/2017 2:23:02 PM
 * Author : C15653
 */

#include <avr/io.h>


int main(void)
{
    /* Replace with your application code */
    while (1)
    {
    }
}
```

Figure 5.11: main.c Explorer in Atmel Studio 7

Your new project is now created and is ready for the application code to be developed!

### 5.4. Microprocessor

It is an integrated circuit which contains all the functions of CPU of a computer.

#### 5.4.1. Definition

Microprocessor, any of a type of miniature electronic device that contains the arithmetic, logic, and control circuitry necessary to perform the functions of a digital computer's central processing unit.

### 5.4.2. Block Diagram

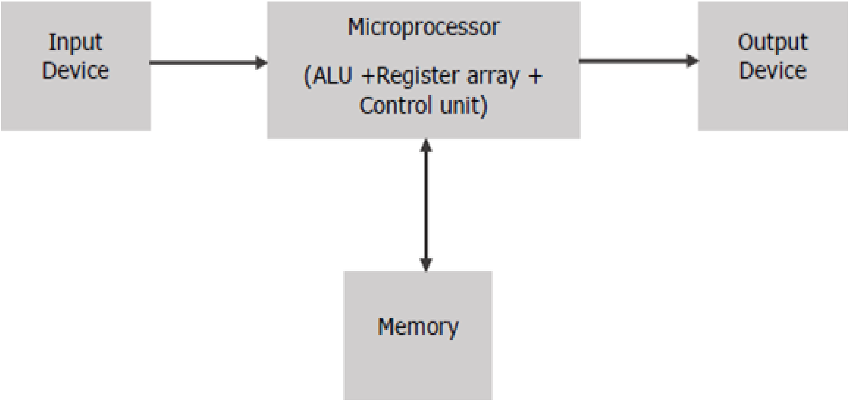


Figure 5.12: Block diagram - Microprocessor

## 5.5. Microcontroller

A microcontroller (MCU for microcontroller unit) is a small computer on a single metal-oxide-semiconductor (MOS) integrated circuit (IC) chip. A microcontroller contains one or more CPUs (processor cores) along with memory and programmable input/output peripherals. Program memory in the form of ferroelectric RAM, NOR flash or OTP ROM is also often included on chip, as well as a small amount of RAM. Microcontrollers are designed for embedded applications, in contrast to the microprocessors used in personal computers or other general purpose applications consisting of various discrete chips.

### 5.5.1. Introduction to Atmega16

- It is a 40-pin low power microcontroller, developed by using CMOS technology.
- CMOS is an advanced technology, mainly used for developing ICs, which is low power consumption and high noise immunity technology.
- It is an 8-bit controller based on AVR advanced RISC architecture. AVR is family of microcontrollers developed by Atmel in 1996.
- A single chip computer which provides CPU, ROM, RAM, EEPROM, Timers, Counters, ADC and four 8-bit ports called PORTA, PORTB, PORTC, PORTD where each port consists of 8 I/O pins.
- It has built-in registers which are used to make a connection between CPU and external peripherals components. It takes input by reading registers and gives output by writing registers.
- It has two 8-bit timers and one 16-bit timer, which are used as counters when they are optimized to count the external signal.
- All necessary peripherals required to run automatic functions are incorporated in this device like ADC, Analog comparator, USART, SPI, which make it economical as compared to a microprocessor that requires external peripheral to perform various functions.
- It comes with 1KB of static RAM which is volatile memory; stores information for short period of time and highly depends on the constant power supply. Whereas 16KB of flash memory,

also known as ROM, is also incorporated in the device; it is non-volatile in nature and can store information for long period of time and doesn't lose any information when the power supply is disconnected.

- It works on a maximum frequency of 16MHz where instructions are executed in one M/C cycle.
- It is always preferred over other microcontrollers like Atmel 8051 because it has much faster ability to execute instructions and consist of modified RISC processor.
- It has in-built flash that comes with features of a bootloader. It has built-in 10-bit ADC, SPI, PWM, and EEPROM.

Figure 5.13: Atmega16 Microcontroller chip

### 5.5.2. Atmega16 Pinout

Following figure shows the pin diagram of this AVR microcontroller Atmega16.

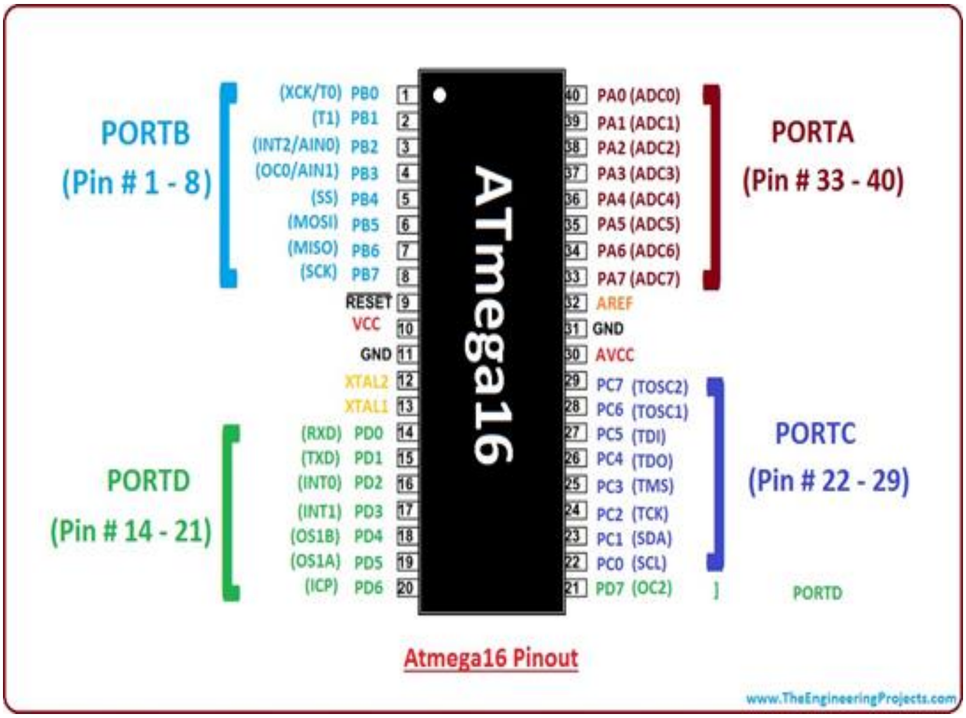


Figure 5.14: Pinout diagram - Atmega16

### 5.5.3. Pin Description of Atmega16

Atmega16 has 40 pins where each pin is used to perform a specific task. Out of these 40 pins, total 32 I/O pins and four ports and each port consist of 8 I/O pins.

- PORTA = 8 Pins ( Pin 33 - 40 )
- PORTB = 8 Pins ( Pin 1 - 8 )
- PORTC = 8 Pins ( Pin 22 - 29 )
- PORTD = 8 Pins ( Pin 14 - 21 )

**Reset:** Pin9 - an active low reset Pin. A low-level pulse for longer than minimum pulse length will produce a reset. Short pulses are unlikely to produce reset.

**VCC.** Pin10 - a power supply pin for this controller. 5 V is required to put this controller in a running condition.

**GND:** Pin11 - a ground pin.

**AREF:** Pin32 - an analog reference pin mainly used for A/D converter.

**AVCC:** Pin30 - an AVCC which is a supply voltage pin for PORTA and ADC. It is connected to VCC through a low pass filter in the presence of ADC. However, in the absence of ADC, AVCC is externally connected to VCC.

**Pin 12 & 13:** Connection for crystal oscillator. Atmega16 works at the internal frequency of 1MHZ; the oscillator is added to generate high clock pulses and frequency.

### 5.5.4. Applications

AVR controllers provide a wide range of applications where automation is required. Following are the main applications

- Medical equipment
- Home automation
- Embedded systems
- Arduino Projects
- Used in automobiles and industrial automation
- Home appliances and security systems
- Temperature and pressure control devices

### 5.6. LED (Light Emitting Diode)

Light Emitting Diodes (LEDs) are all around us. They are in our homes, our cars, even our phones. LEDs come in a variety of shapes and sizes, this gives designers the ability to tailor them to their product. Any time something electronic lights up, there’s a good chance that an LED is behind it. Their low power and small sizes make them a great choice for many different products as they can be worked into the design more seamlessly to make it an overall better device.

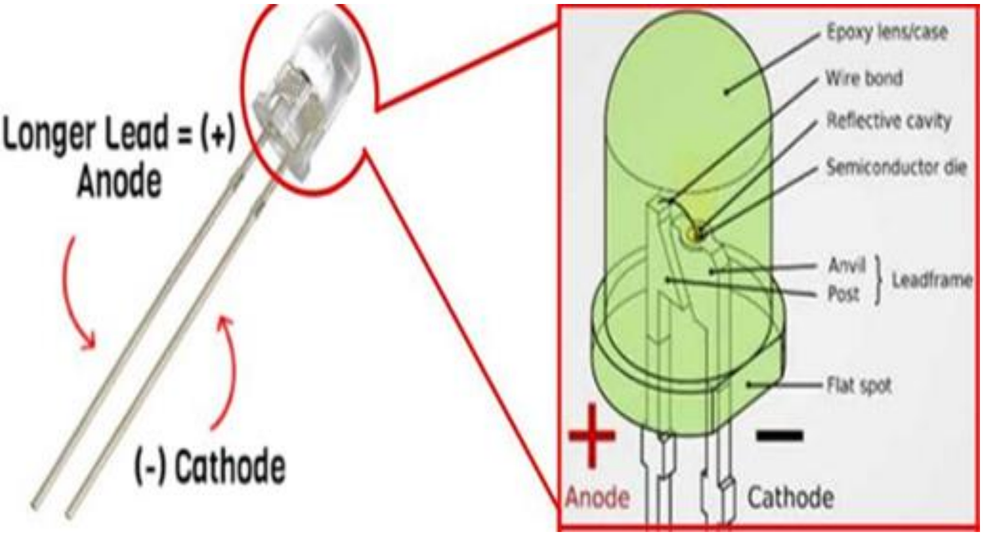


Figure 5.15: LED (Light Emitting Diode)

### 5.6.1. How to interface

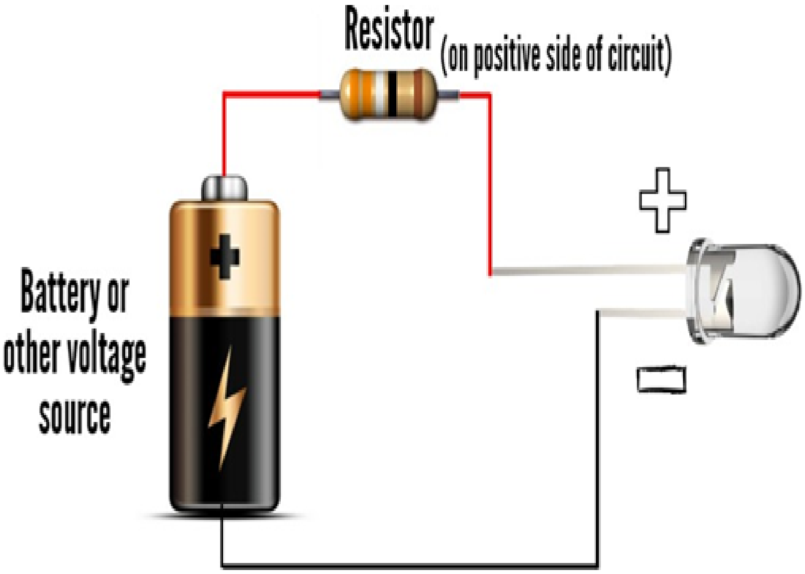


Figure 5.16: LED interfacing

### 5.7. Dual-Channel Relay Module

It is similar like a single - channel relay module, but with some extra features like optical isolation. It can be used to switch mains powered loads from the pins of a microcontroller.

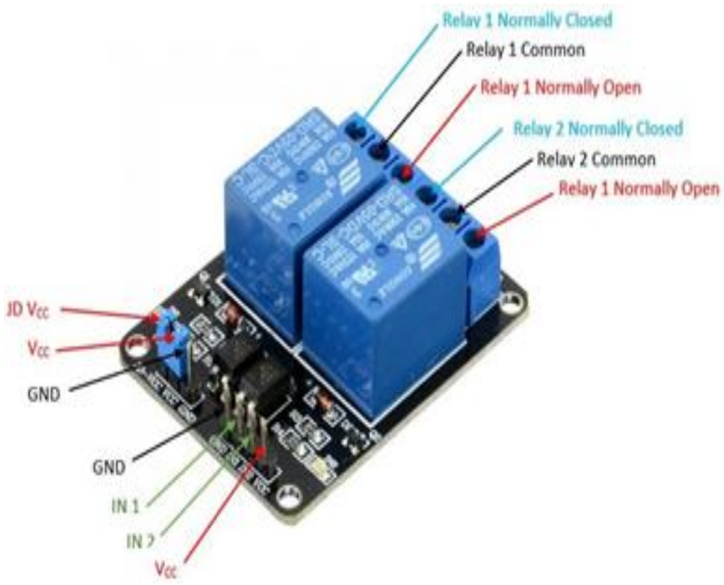


Figure 5.17: Dual-Channel Relay Module

### 5.7.1. Dual-Channel Relay Module Pinout

| Pin Number | Pin Name | Description |
|---|---|---|
| 1 | JD-VCC | Input for isolated power supply for relay coils |
| 2 | VCC | Input for directly powering the relay coils |
| 3 | GND | Input ground reference |
| 4 | GND | Input ground reference |
| 5 | IN1 | Input to activate the first relay |
| 6 | IN2 | Input to activate the second relay |
| 7 | VCC | VCC to power the optocouplers, coil drivers, and associated circuitry |

Table 5.1: Dual-Channel Relay Module Pinout

### 5.7.2. Specifications of Dual-Channel Relay Module

- 3.75V to 6V – Power supply voltage
- 5mA - Trigger current
- ~70mA (single), ~140mA (both) - Current when relay is active
- 250VAC, 30VDC - Relay maximum contact voltage
- 10A - Relay maximum current

### 5.7.3. Interfacing

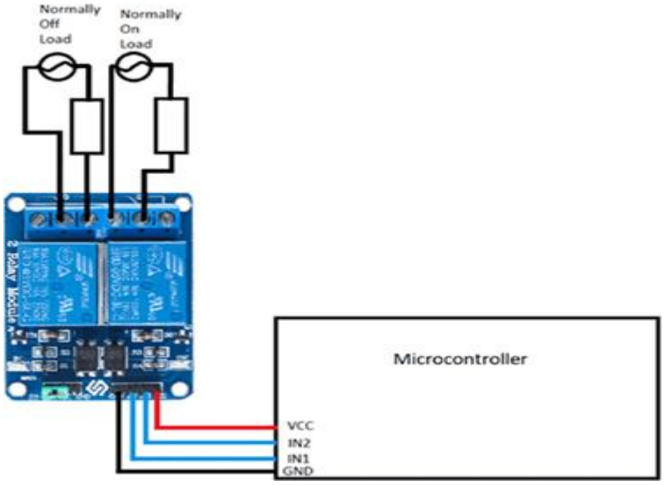


Figure 5.18: Dual-Channel Relay Module interfacing

### 5.7.4. Applications of Dual-Channel Relay Module

- Switching mains loads
- Home automation
- Battery backup
- High current load switching

## 5.8. LDR (Light Dependent Resistor)

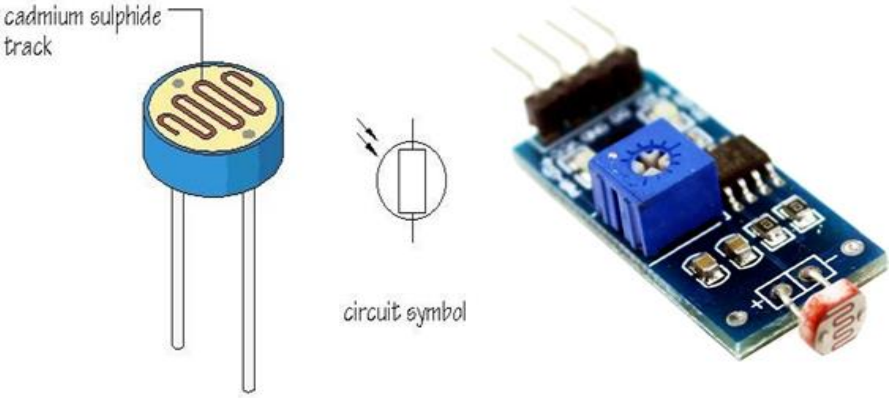


Figure 5.19: LDR (Light Dependent Resistor)

An LDR or light dependent resistor is also known as photo resistor, photocell, photo-conductor. It is a one type of resistor whose resistance varies depending on the amount of light falling on its surface. When the light falls on the resistor, then the resistance changes. These resistors are often used in many circuits where it is required to sense the presence of light. These resistors have a variety of functions and resistance. For instance, when the LDR is in darkness, then it can be used to turn ON a light or to turn OFF a light when it is in the light. A typical light dependent resistor has a resistance in the darkness of 1MOhm, and in the brightness a resistance of a couple of K Ohm.

### 5.8.1. Working Principle of LDR

These devices depend on the light, when light falls on the LDR then the resistance decreases, and increases in the dark. When a LDR is kept in the dark place, its resistance is high and, when the LDR is kept in the light its resistance will decrease.

If a constant "V' is applied to the LDR, the intensity of the light increased and current increases.

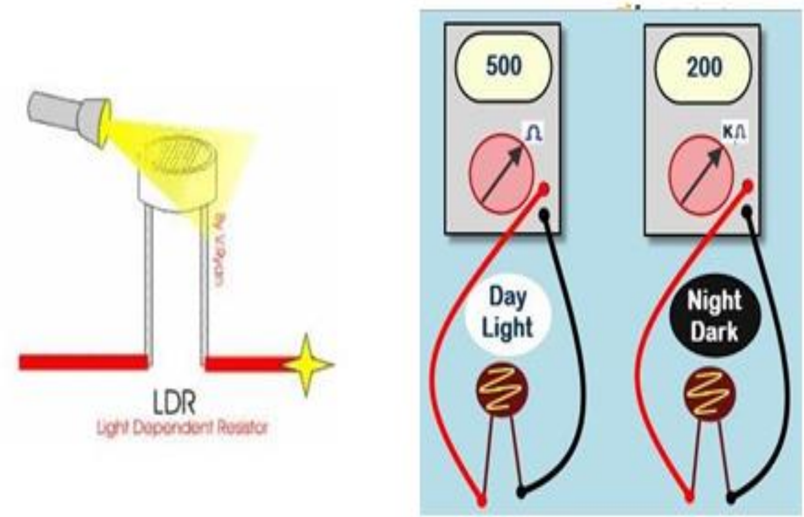


Figure 5.20: Working of LDR

### 5.8.2. Circuit Diagram of LDR Module

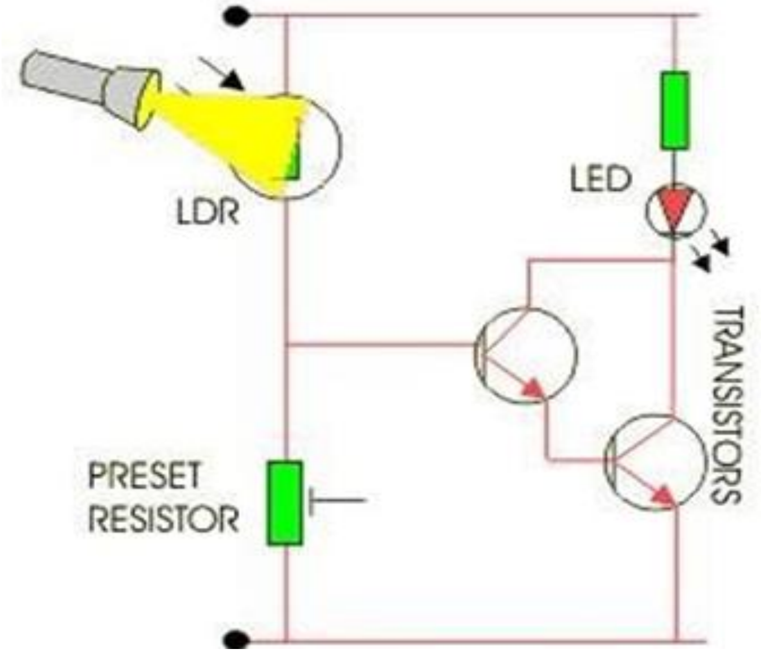


Figure 5.21: Circuit diagram – LDR

### 5.9. UART (Universal Asynchronous receiver – transmitter)

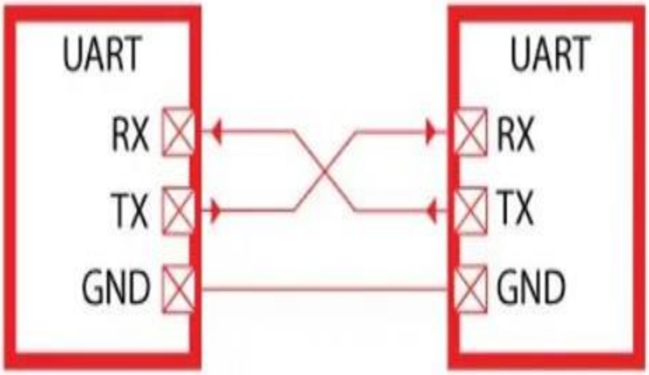


Figure 5.22: UART

UART Communication stands for Universal asynchronous receiver- transmitter. It is a dedicated hardware device that performs asynchronous serial communication. It provides features for the configuration of data format and transmission speeds at different baud rates. A driver circuit handles electric signaling levels between two circuits. A Universal asynchronous receiver-transmitter (UART) Communication is usually an individual component or part of an integrated circuit. We can use it for communications over a computer or its peripheral devices such as a mouse, monitor or printer. In microcontroller chips, there are usually a number of dedicated UART hardware peripherals available.

UART or Serial communication is one of the most simple communication protocols between two devices. It transfers data between devices by connecting two wires between the devices, one is the transmission line while the other is the receiving line. The data transfers bit by bit digitally in form of bits from one device to another. The main advantage of this communication protocol is that its not necessary for both the devices to have the same operating frequency. For example, two microcontrollers operating at different clock frequencies can communicate with each other easily via serial communication. However, a predefined bit rate that is referred to as baud rate usually set in the flash memory of both microcontrollers for the instruction to be understood by both the devices.

#### 5.9.1. How UART communication works

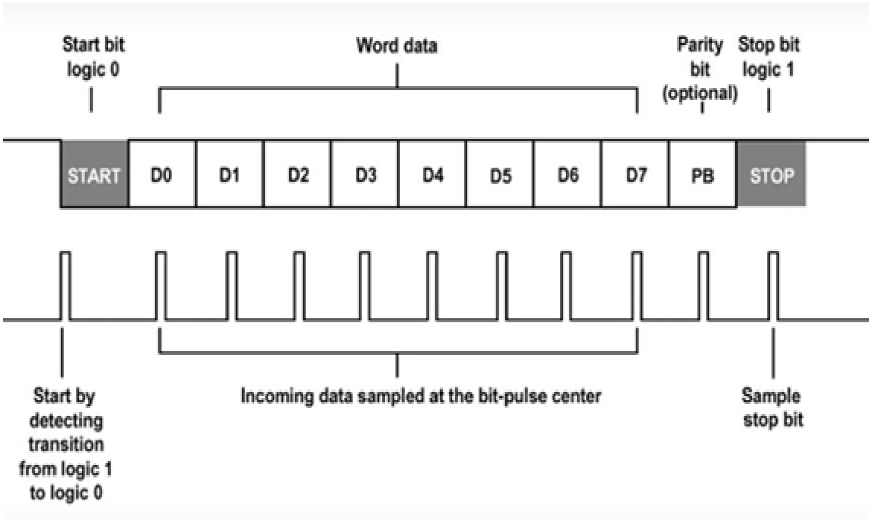


Figure 5.23: UART communication working

### 5.9.2. Baud Rate

The communication between two devices via UART Protocol occurs by transmission of bits. A total of 8 bits are sent one right after the other to transmit a byte. A bit is either a logical low or high. The time interval between two bits is called the baud rate or bit rate. it must be defined in both devices so the sending device can encode the data into bits with this specific time interval and the receiver expects the successive bits at the right time. The most commonly used baud rates is 9600 bits per second. Although other baud rates are also used, but the higher the bit rate, the more chances there are of data corruption. Lower bit rates are used when there is greater physical distance between two devices because the length of the wire increases resistance and thus deteriorates the signal.

## 5.10. ADC (Analog to Digital Converter)

In electronics, an analog-to-digital converter (ADC, A/D, or A-to-D) is a system that converts an analog signal, such as a sound picked up by a microphone or light entering a digital camera, into a digital signal. An ADC may also provide an isolated measurement such as an electronic device that converts an input analog voltage or current to a digital number representing the magnitude of the voltage or current. Typically the digital output is a two's complement binary number that is proportional to the input, but there are other possibilities.

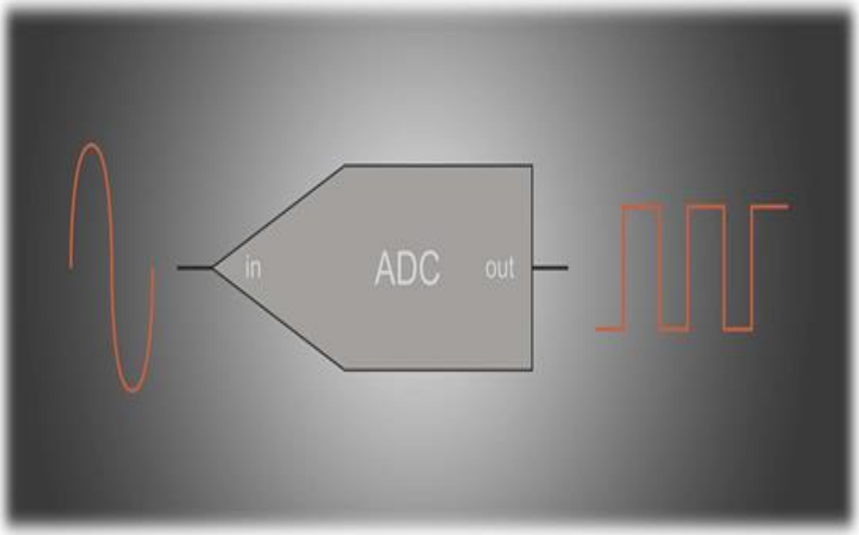
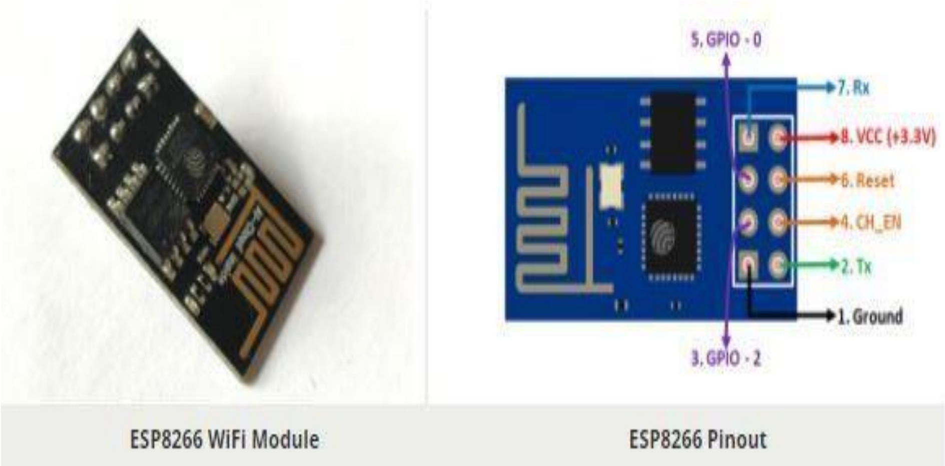


Figure 5.24: ADC

There are several ADC architectures. Due to the complexity and the need for precisely matched components, all but the most specialized ADCs are implemented as integrated circuits (ICs). These typically take the form of metal–oxide–semiconductor (MOS) mixed-signal integrated circuit chips that integrate both analog and digital circuits.

### 5.11. Wi-Fi Module (ESP8266)

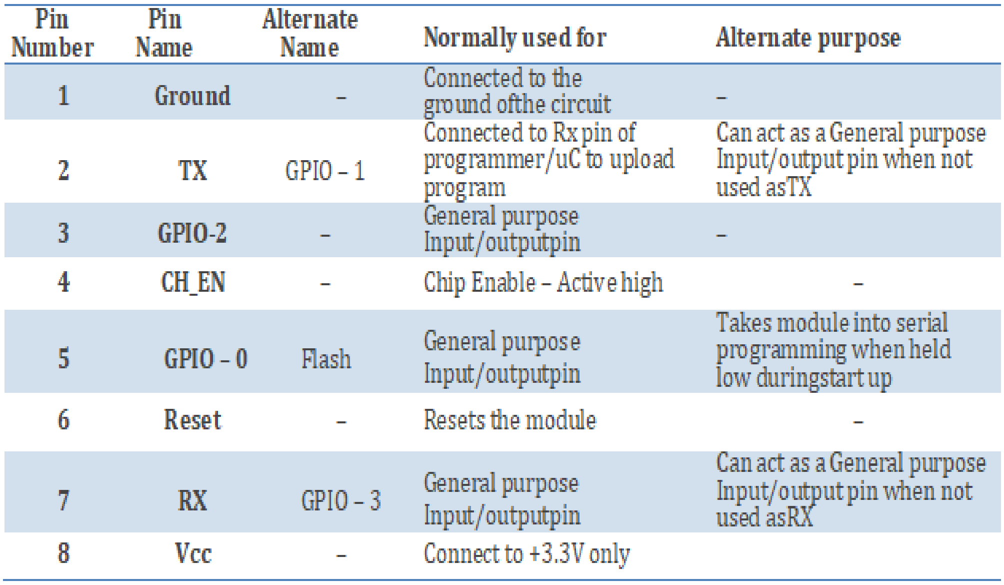


ESP8266 WiFi Module

ESP8266 Pinout

Figure 5.25: Wi-Fi Module (ESP8266)

#### 5.11.1. ESP8266 Pin Configuration



| Pin Number | Pin Name | Alternate Name | Normally used for | Alternate purpose |
|---|---|---|---|---|
| 1 | Ground | - | Connected to the ground ofthe circuit | - |
| 2 | TX | GPIO – 1 | Connected to Rx pin of programmer/uC to upload program | Can act as a General purpose Input/output pin when not used asTX |
| 3 | GPIO-2 | - | General purpose Input/outputpin | - |
| 4 | CH_EN | - | Chip Enable – Active high | - |
| 5 | GPIO – 0 | Flash | General purpose Input/outputpin | Takes module into serial programming when held low duringstart up |
| 6 | Reset | - | Resets the module | - |
| 7 | RX | GPIO – 3 | General purpose Input/outputpin | Can act as a General purpose Input/output pin when not used asRX |
| 8 | Vcc | - | Connect to +3.3V only | |

Table 5.2: ESP8266 Pins

#### 5.11.2. ESP8266 Features

- Low cost, compact and powerful Wi-Fi Module
- Power Supply: +3.3V only
- Current Consumption: 100mA
- I/O Voltage: 3.6V (max)

- I/O source current: 12mA (max)
- Built-in low power 32-bit MCU @ 80MHz
- 512kB Flash Memory
- Can be used as Station or Access Point or both combined
- Supports Deep sleep (<10uA)
- Supports serial communication hence compatible with many development platform like Arduino
- Can be programmed using Arduino IDE or AT-commands or Lua Script

### 5.12. How to Collect Data with New Channel

Following example shows how to create a new channel to collect analyzed data. You can read data from the public ThingSpeak channel 12397 - Weather Station, and can write it into your new channel. To learn how to post data to a channel from devices, see Write Data to Channel and  the API Reference.

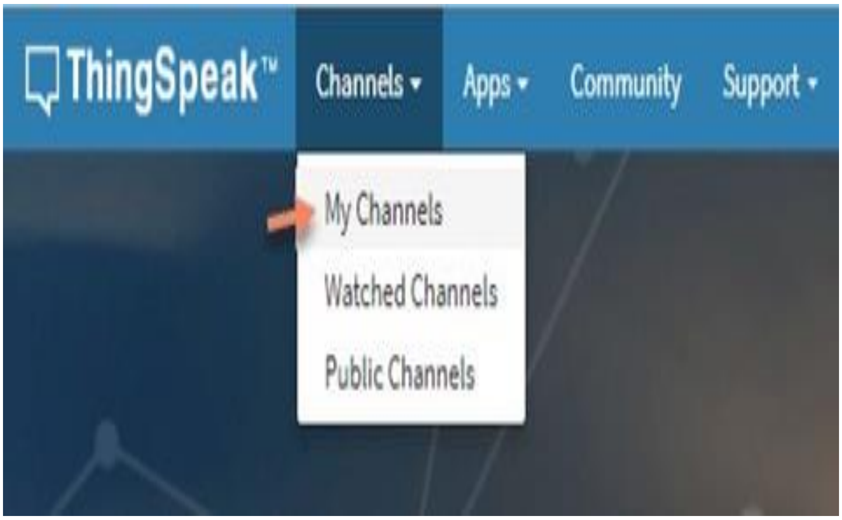


Figure 5.26: ThingSpeak channel

#### 5.12.1. Steps to create a Channel

Sign in first to ThingSpeak by using your MathWorks Account credentials, or create a new account.

- Click **Channels** > **MyChannels**.
- On the Channels page, click **New Channel**.
- Check the boxes next to Fields 1–3. Enter these channel setting values:

  **Name**: Dew Point Measurement

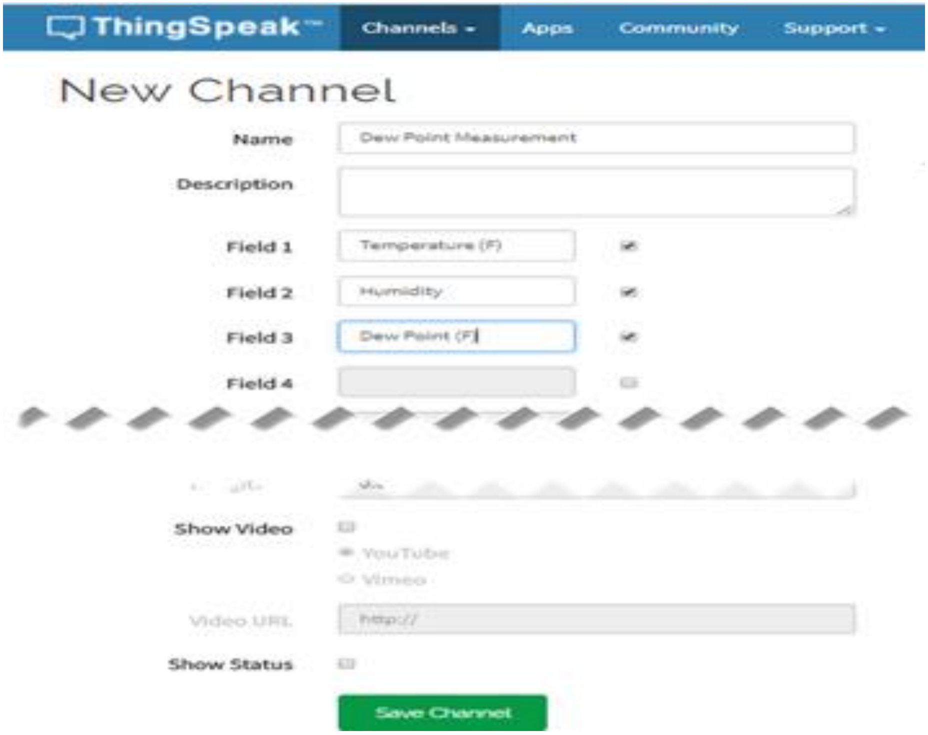
  **Field 1:** Temperature (F)

  **Field 2:** Humidity

  **Field 3:** Dew Point

- Click **Save Channel** at the bottom of the settings.

You now see these tabs:



New Channel

Name: Dew Point Measurement

Description:

Field 1: Temperature (F) ☑

Field 2: Humidity ☑

Field 3: Dew Point (F) ☑

Field 4: □

Show Video □
● YouTube
○ Vimeo

Video URL: http://

Show Status □

Figure 5.27: ThingSpeak New channel creation

1. **Private View**: This tab displays information about your channel that only you can see.
2. **Public View**: If you choose to make your channel publicly available, use this tab to display selected fields and channel visualizations.
3. **Channel Settings**: This tab shows all the channel options you set at creation. You can edit, clear, or delete the channel from this tab.
4. **Sharing**: This tab shows channel sharing options. You can set a channel as private, shared with everyone (public), or shared with specific users.
5. **API Keys**: This tab displays your channel API keys. Use the keys to read from and write to your channel.
6. **Data Import/Export**: This tab enables you to import and export channel data.

Next Steps

Your channel is available for future use by clicking **Channels → My Channels.**

## 5.13. Conclusion

**Atmel Studio 7** is an integrated development platform (IDP) for developing and debugging all AVR and SAM microcontroller applications. The Atmel Studio 7 IDP gives you a seamless and easy-to-use environment to write, build, and debug your applications written in C/C++ or assembly code. It also connects seamlessly to the debuggers, programmers, and development kits that support AVR and SAM devices. The development experience between Atmel START and Studio 7 has been optimized. Iterative developments of START-based projects in Studio 7 are supported through re-configure and merge functionality.

Atmel Studio 7 has a New Project wizard that steps you through the process of creating a project. In the search bar enter the key characters for the device you intend to use, and then select the exact device from the list that appears. In this example, "817" was used to find and select the device **ATmega16**. Then click **OK** to create the project.

A new GCC C Executable project has now been created for the **ATmega16** AVR device. The *Solution Explorer* on the right side of the window will list the contents of the newly generated solution.

**References:**

1. https://www.internetsociety.org/resources/doc/2015/iot-overview/
2. https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/
3. https://wspublishing.net/avr-c/installing-atmel-studio-7/
4. https://www.elprocus.com/atmega16-next-generation-micro-controller/
5. https://www.ledsupply.com/blog/how-does-a-5mm-led-work/
6. https://components101.com/switches/5v-dual-channel-relay-module-pinout-features-applications-working-datasheet
7. https://www.watelectronics.com/light-dependent-resistor-ldr-with-applications/
8. https://microcontrollerslab.com/uart-communication-working-applications/
9. https://components101.com/wireless/esp8266-pinout-configuration-features-datasheet
10. https://www.mathworks.com/help/thingspeak/collect-data-in-a-new-channel.html

# A survey of blended learning methodologies and a framework to assure quality

Dr Chandrani Singh[1]

Prof. Archana Nair [2]

Dr. Manisha Kumbhar [3]

E-learning market in the past decade has continued to Shift, Develop and Advance with the growing prevalence across geographies, emerging trends and technologies, social/collaborative learning trends and a complicated Learning and Development Industry.E-Learning Market size was estimated at over USD 150 billion in 2016 and is predicted to grow at over 7% CAGR from 2017 to 2024 and 331 billion by 2025[1].While E-learning has taken center stage, Blended learning a pedagogical initiative has come to fore across the global education industry. Blended learning model, a combination of traditional (classroom) training with digital on-line content, promising higher degree of interaction between stakeholders, was developed with a notion such that teachers could spend more time to create preferred learning pathways for students with respect to their needs using learner centered approach.

Technology being an enabler, a documented technology plan should be prepared to ensure the assurance of quality. Institutional policies should have an appropriate grievance registration and addressable system for all stakeholders and should adopt good ethical practices. The Quality Assurance agencies or third parties involved in giving accreditations should make assessment supportive for blended learning, by being receptive to new pedagogical methods, devise a maturity model for blended learning. For the government to assure quality and accreditation in higher education it should cultivate drivers for novelty and appraise governing regulations and practices, encourage and accelerate innovation with an expression for change through national strategies.

This short paper tries to assure the quality of blended framework by instilling a top-down approach with due regard to national legal and statutory guidelines and continuous governmental interventions/assistance with regard to legal and policy updates. It also highlights that innovations, scalability of the virtual learning infrastructure etc. are essential constituents to improvise on the blended adoptions and encourages stringent guidelines to internal cyclical monitoring and review by external agency to assure quality.

**Keywords**: Pedagogy, Personalized, Scalability, Sustainability, Accessibility

## 6.1. Introduction to Blended learning

While E-learning has taken center stage, blended learning a pedagogical initiative has come to fore across the global education industry. Blended learning model, a combination of traditional (classroom) training with digital on-line content, promising higher degree of interaction between stakeholders, was developed with a notion such that teachers could spend more time to create preferred learning pathways for students with respect to their needs using learner centered approach. The Sloan Consortium currently known as Online learning consortium expresses blended courses as having 30 -79 pc of their content delivered online, and the online courses being delivered 80 pc of the time through virtual mode [6]. The increased internet access and availability of virtual platforms in current era have encouraged, blended learning to have paved the way to augment quality, parity, and entrée to learning opportunities for the lifetime of an individual. With aggressive growth in digital learning technologies, blended formats have shown their prominence in curriculums which are fully online with selected days required for presence in classrooms or computer labs or fully online courses imparted in classroom or lab on a daily basis. Blended formats also impart learning beyond classrooms and schooldays where instructions incorporate components or integration of online resources. The analysis of the data would enable the policy makers to design a framework for the pedagogical use of ICT for learning, linking ICT usage to assessable cognitive outcomes. Asia being the most populous and disaster-prone region, the propagation and adoption of blended initiatives would have been highly impactful, for learners. But with internet penetration rate at 50 pc of the total population and China with 642 million users as per 2014 reports (greater than US, India and Japan combined), the usage of online modes seems to be more prominent and promising in advanced economies. While in the least developing countries (Afghanistan, Timor-Leste) with penetration rate ranging between 2 to 20 pc, online modes of teaching and learning is a future project in itself altogether [3].

Some of the noble initiatives taken in the Asia Pacific region with regard to online and blended modes range from making available massive open online courses by IndonesiaX, to using a Korean Developed System named Math Cloud by Bhutan and

supported by Asian Development Bank and finally to University of South Pacific using blended modes to make University training accessible to the students of remote outer islands. Several other initiatives as implementation of blended formats across class, course and program level in East China Normal University to University of Western Australia implementing, blended modes in teaching and learning as an institutional level strategy, to formulation of Malaysian Blueprint for Globalized Online Learning are evidences of positive infiltration. Few other known facts and successful implementations of blended approaches are boasted by Sunway University in Malaysia, implementing blended learning in a range of forms across university degrees and the advanced technological initiatives of Chiang Mai University (CMU) in Thailand which helped them to incorporate the approach with success. NTU Singapore's integrated campus wide approach, wherein diverse systems and tools impeccably supplemented and reinforced one another, as well as the 'professor friendly' attitude, ere crucial features impelling the high acceptance and usage scenarios of blended learning. Similarly the Education University of Hong Kong has implemented capacity building strategies that target to augment learning rendezvous and results, and gauge blended learning practices in the faculty. Seoul National University in Republic of Korea with the support of internal and external institutions have developed an infrastructure for online and blended learning, providing instructional and technical assistance to the faculty fraternity and conduct research and development to build a smart campus. On the other hand few Chinese Universities are yet to transit from exploration stage to adoption and growth stage through strong institutional support. In India National Mission on Education through Information and Communication Technology (NMEICT) has been instrumental in leveraging the prospect of ICT in qualitative personalization of education for the learners in Higher Education Institution, in anytime, anywhere mode. Blended formats or innovative pedagogical approaches are being used on an experimental basis to impart e-learning and conducting sessions through virtual laboratories, by bringing together learners and teachers on the same platform. Several other initiatives as virtual/online university for disseminating courses developed by BITS Pilani, IGNOU's virtual campus initiative, Internet Based Online Interactive Courseware and eb based intelligent tutoring by IIT, Delhi, spoken tutorial courses by IIT Bombay, Digital Library Inflibnet, MOOC's and NIC's e-learning portal, launch of integrated National Knowledge Network for collaborative research have been some intrepid steps taken to promote online learning using innovative pedagogical approaches. While attempts were made to encourage promoting inclusion and access to learning across the entire sub-continent there arose issues with regard to quality which institutions and regulatory bodies had to perceive in due course. Scalability and sustainability of blended learning having been a major challenge to assure quality for blended initiatives. A framework designed for Blended learning can yield a structure as follows in Figure 1:

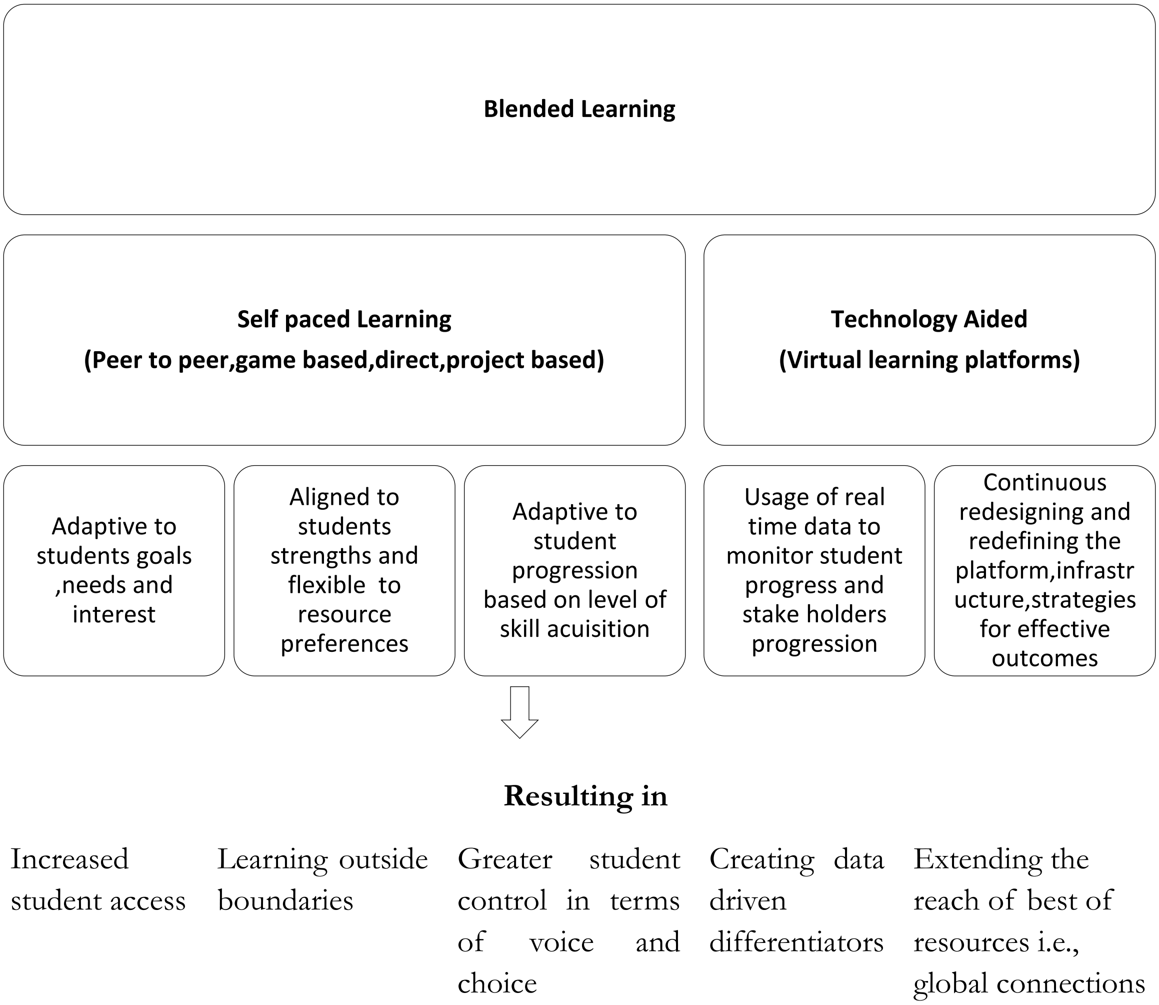

Blended Learning
Self paced Learning
(Peer to peer,game based,direct,project based)
Technology Aided
(Virtual learning platforms)
Adaptive to students goals ,needs and interest
Aligned to students strengths and flexible to resource preferences
Adaptive to student progression based on level of skill acuisition
Usage of real time data to monitor student progress and stake holders progression
Continuous redesigning and redefining the platform,infrastructure,strategies for effective outcomes
Resulting in
Increased student access
Learning outside boundaries
Greater student control in terms of voice and choice
Creating data driven differentiators
Extending the reach of best of resources i.e., global connections

**Fig 6.1: Blended learning Framework**

## 6.2. Global scenario in adoption of blended learning by countries

The countries leading the way in adoption of Online/ Blended learning on a year-on-year basis are as follows as shown in table 1: It is seen that few of the European and Asian countries like Japan, South Korea, France and United Kingdom have taken substantive effort with respect to implementation of blended learning.

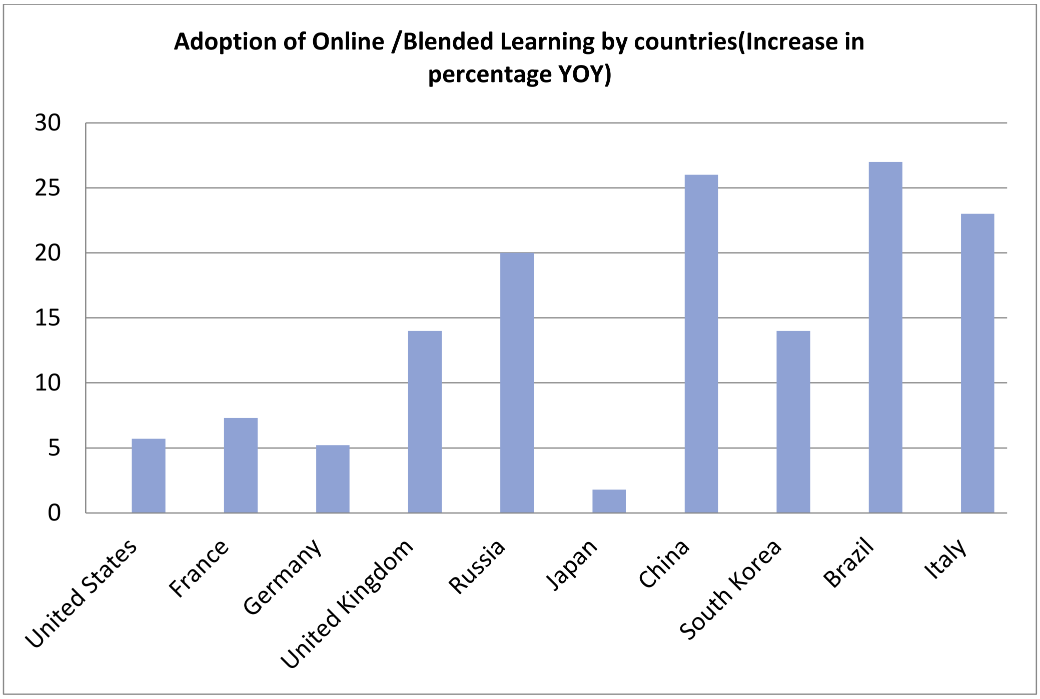


**Table 6.1: Adoption of online/blended learning by countries**

### a. Implementation of Blended Learning in the US

In the US, recent trend at the level of primary education is to increase equity and access for all; hence a large number of high-quality online courses and instructors are made available for the needy and the marginalized. These courses which have different structures and outcomes are managed by state, district, charter schools, universities, profit and non-profit organizations. Till date as per the report, over 80 percent of the schools in the US have adopted online or integrated modes of imparting education in the K -12 segment and more than 6 million students are taking at least one distance course in the higher education segment, having increased the enrolment by 3.9% over the previous year's[6] While more than 14 pc of the students in the above Higher Education segment have opted for exclusively distance courses in the recent past as per statistics, around 16 pc have taken a combination of distance and non-distance courses. As of 2016, the most common digital learning resources used for K-12 education in the United States were online educational videos, educational apps or software and websites and key players were the public institutions responsible for hosting two third of the student population

In the United States, about 33% of all post-secondary students take at least one online course necessarily. The Online Learning Consortium (formerly Sloan-C) defines a blended learning course as one where 30%-79% of the content is delivered online, while an online course could have anything over 79% of its content offered online.

### b. The European Commission's initiatives on implementation of Blended Learning

In Europe, as per the eurostat data generated three to four years before, there was a persistent digital divide between north and south Europe vis a vis the west and the east in terms of outcomes, capability and access which impacted learning in one way or the other. Hence the European Commission under the Digital Education action plan has taken rigorous efforts to seize the opportunities of digital revolution and has designed an extensive execution strategy for the establishment of European Hub of Digitally Innovative Education institutions, showcasing and steering pioneering ICT-based instructional and legislative practices, accompanied by improving education through better data analysis and foresight. Several other initiatives as SELFIE and concept of connected classrooms are also in pipeline with the PISA questionnaire forming an essential tool to collect data from 31 countries (EU 28, Iceland, Norway and Turkey) [2]

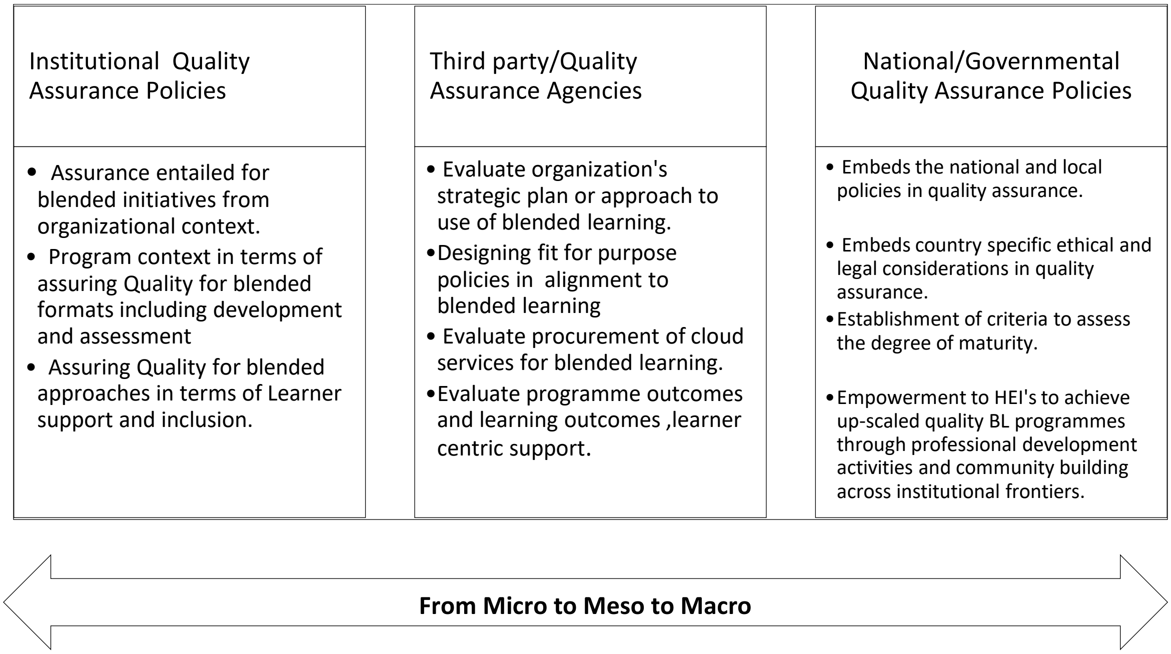
The European Union's initiative is in creating an online platform for Higher Education institutions (HEIs) in Europe using digital technologies to:

- improve the quality and make learning and teaching relevant
- support internationalization and greater cooperation across HEI's, research institutions and recruiting
- organizations within Europe.
- Making Higher Education accessible to all categories of students in alignment to the principal of inclusion.

The online platform will be central to the existing national and regional platforms exchanging and disseminating best practices and dealing with all forms of online teaching and learning institutions and campuses across the globe thus promoting collaboration and co-creation of knowledge and content. A big leap by the European Commission towards the effective and strategic implementation of blended learning was to initiate a funded project on devising maturity models for blended learning under innovation in Higher Education wherein strategies, policies, guidelines are devised and a reference model is being created in association with the partner institutions/universities embracing all levels, which will acts as an instrument to measure stakeholder focused outcomes [2]. In addition, a monitor is being designed and validated to map blended learning practices.

### 6.3. Blended learning Framework - Assuring the quality

The quality assurance in blended learning can be viewed from dual perspectives such as focussing on the quality of the designed environment on one hand and teaching learning outcomes on the other hand, based on the extensive interactions between the teacher and the learner. Hence an important requisite is not to limit the scope to design and construction of the environment only rather to extend it to the incessant learning outcomes. It is further to be noted that a technology mishap might ruin the quality of the teaching learning process in the blended approach. Hence organizations imbibing blended approaches should set a mandate. Quality Assurance in blended learning should be an amalgamation of best practices adopted and followed by Institutions, Governments and Quality Assurance Agencies as shown in Fig 2.



Institutional Quality Assurance Policies

- Assurance entailed for blended initiatives from organizational context.
- Program context in terms of assuring Quality for blended formats including development and assessment
- Assuring Quality for blended approaches in terms of Learner support and inclusion.

Third party/Quality Assurance Agencies

- Evaluate organization's strategic plan or approach to use of blended learning.
- Designing fit for purpose policies in alignment to blended learning
- Evaluate procurement of cloud services for blended learning.
- Evaluate programme outcomes and learning outcomes ,learner centric support.

National/Governmental Quality Assurance Policies

- Embeds the national and local policies in quality assurance.
- Embeds country specific ethical and legal considerations in quality assurance.
- Establishment of criteria to assess the degree of maturity.
- Empowerment to HEI's to achieve up-scaled quality BL programmes through professional development activities and community building across institutional frontiers.

From Micro to Meso to Macro

**Fig 6.2: Institution of best practices adopted in Blended Learning by stakeholders**

### a. Quality Assurance at the Institutional Level

Institutional quality policies have to align themselves to the following with respect to blended initiatives:

–Vision and Philosophy

–Curriculum and Academic Delivery

–Professional Development and Learning Support

–Infrastructure and Support Facilities

–Institutional Structure and Policies, Partnerships

–Research and Evaluation

with respect to the blended initiatives to assure the qualitative outcomes envisaged with respect to the stakeholders involved. Transitioning from macro to micro, quality policy for any blended learning initiatives should not only include assessment at the institutional level but also at the course and degree level. As per a case study taken into consideration East China Normal University sites that to device an appropriate quality assurance mechanism it is mandatory that gaps/issues, effectiveness, and strategies to be identified and devised at the course, class and program level for blended learning to become a success [4]. To assure quality, the instructional design and delivery mechanism from content creation to lecture and guided practice and from interactive discussion sessions to project and case-based simulations, from classroom to computer mediated, from synchronous to asynchronous and from instructor-led to social learning, each aspects outcome have to be qualitatively determined and monitored. Target group specific blended learning initiatives for a course should aptly monitor

–course aim and prerequisites

–content delivery and tutor skills

–learning target and outcomes

–knowledge transfer and didactic rules

–organizational frameworks, and media platforms adaptability

–a channel for course information exposition and

–rules for split up of the course content

The curriculum development should implement the 7 R's as discussed by Ron Ritchart. The 7 R's being real, rigorous, requires independence, rich in thinking, reflective, rewarding, revealing and reflective [5]. People involved in curriculum design should be a hybrid bunch of instructional designers and content and technical experts with roles and responsibilities very well defined. It is important to have an exchange of agreements with other educational institutions, for students' virtual mobility, providing e-learning programs.

Technology being an enabler, a documented technology plan should be prepared to ensure the assurance of quality. Appropriate learning management tools used in individual and collaborative learning should be in agreement with the IT infrastructure available, with desired connectivity, learning adaptivity, tutoring skills etc. Operating, security and recovery procedures should be in place with a performance management application for troubleshooting purpose. Support for building and maintaining the b-learning infrastructure should be addressed by a centralized system.

Tutors should have access to documented resources to render support to student-centric issues with regard to electronically-accessed data. Under the Institutional policy, there should exist an e-portfolio service and an e-repository for students to present their dynamics to concerned entities. Appropriate indexing and archiving of e-learning materials in the repository should also exist.

External training service providers should be introduced and they should adhere to all national and international legal requirements and policies. Learning outcomes should match to a national level of qualification and be placed in a context with wider dimensions.

The institution should have a clear policy with regard to reviews and updates on learning outcomes and the acquirement and valuation of transferable skills, including e-skills.

Institutional policies should have an appropriate grievance registration and addressable system for all stakeholders and should adopt good ethical practices. Periodic reviews with formative evaluation should be carried out to ensure the quality and performance of b-learning systems.

**b. External Agencies** - On the other hand the Quality Assurance agencies or third parties involved in giving accreditation's have been informed to make assessment supportive for blended learning by being receptive to new pedagogical methods, devise a maturity model for blended learning ,evaluate tutor skills, document standards and collaborate for adoption of good practices and have a repository of blended learning experts or on-board them into the team.

**c. Government/Non-Government Entities** - For the government to assure quality and accreditation in higher education it should cultivate drivers for novelty and quality and appraise governing regulations and practices, encourage and accelerate innovation with an expression for change through national strategies.

Summarizing the above, a top-down approach with due regard/mandates to national legal and statutory guidelines and scalability of the virtual learning infrastructure etc., with a detailed alignment to guidelines supported by cyclical monitoring and review by external agency and also the internal quality assurance committee would be an apt framework to assure quality.

### 6.4. Conclusion

With a global health crisis going on from 2019, it is the responsibility of the central and state governments of every country to put within their national education policy the blended formats . This would be possible through extensive use of technology where student's continuity can be granted and would entail equity, excellence and expansion. The issue of lack of bandwidth for online learning needs us to think of ways to improve access for educational platforms and providing teachers and students the support by making devices and networks and infrastructure cost effective. There is a need for a paradigm shift requires to create a technology aided environment enabler, conducive for barrier free learning.

### 6.5. References

1. Preeti Wadhwani, Saloni Gankar, "E-Learning Market Size by Technology (Online E-Learning, Learning Management System (LMS), Mobile E-Learning, Rapid E-Learning, Virtual Classroom), By Provider (Service, Content), By Application (Academic [K-12, Higher Education, Vocational Training], Corporate [SMBs, Large Enterprises], Government), Industry Analysis Report, Regional Outlook, Growth Potential, Competitive Market Share & Forecast, 2020 – 2026", Published Date: May 2020, Report ID: GMI215.

2. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, The European Commission's contribution to the Leaders' meeting in Gothenburg, 17 November 2017.

3. Bridging the Contenents, International cooperation of ASEM Higher Education, Published by Erasmus+ National Agency for EU Higher Education Cooperation DAAD – Deutscher Akademischer Austauschdienst German Academic Exchange Service, Edition 150/February 2019

4. Lim, Cher Ping, Wang Libing. Blended learning for quality higher education: selected case studies on implementation from Asia-Pacific, Author ISBN: 978-92-9223-565-9 (electronic), Year of publication: 2016

5. Ron Ritchhart, The Seven R's of a Quality Curriculum, Project Zero, Harvard Graduate School of Education, Published by Education Quarterly Australia, 2007.

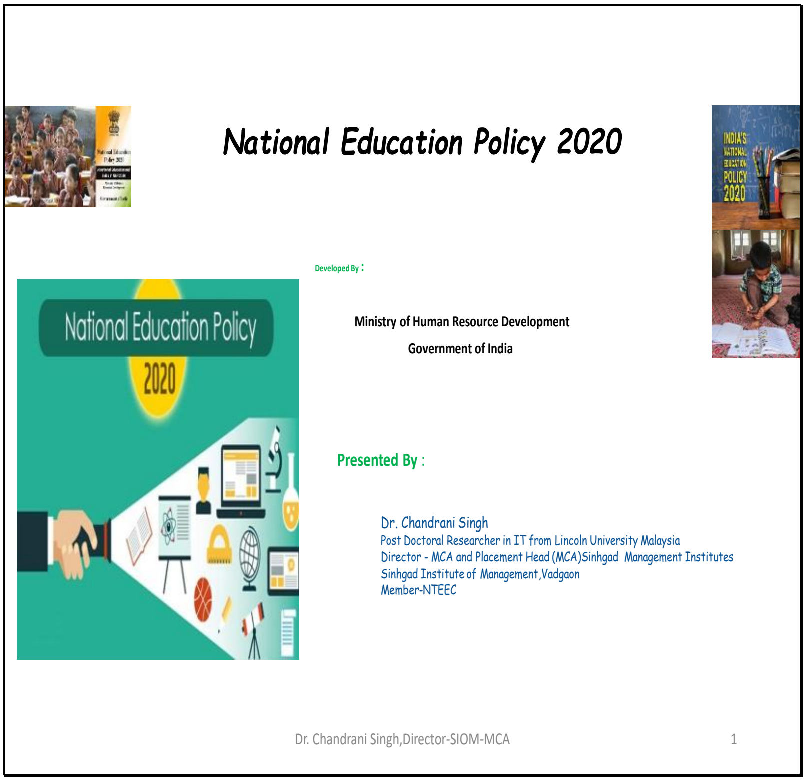
6. Kaufmann, Renee; Buckner, Marjorie MRevisiting "Power in the Classroom": Exploring Online Learning and Motivation to Study Course Content. Interactive Learning Environments, v27 n3 p402-409 2019.

7. Garneli, Varvara & Giannakos, Michail & Chorianopoulos, Konstantinos. (2015). Computing Education in K-12 Schools: A Review of the Literature. 2015. 10.1109/EDUCON.2015.7096023.

# National Education Policy (NEP-2020)

Dr. Shailesh Kasande

Dr. Chandrani Singh



## National Education Policy 2020

National Education Policy
2020
INDIA'S NATIONAL EDUCATION POLICY 2020

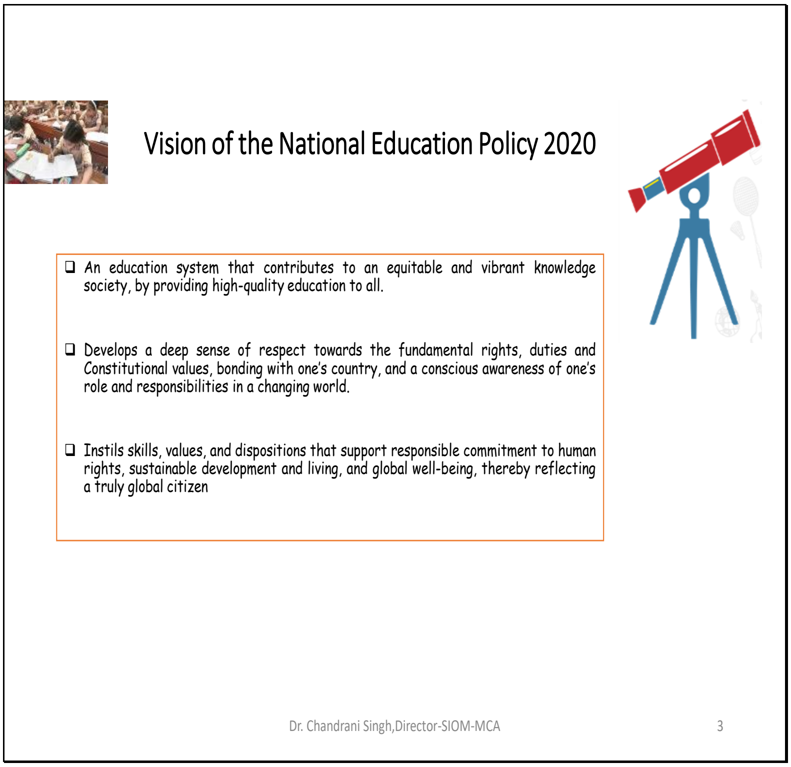
Developed By :

Ministry of Human Resource Development
Government of India

Presented By :

Dr. Chandrani Singh
Post Doctoral Researcher in IT from Lincoln University Malaysia
Director - MCA and Placement Head (MCA)Sinhgad Management Institutes
Sinhgad Institute of Management,Vadgaon
Member-NTEEC

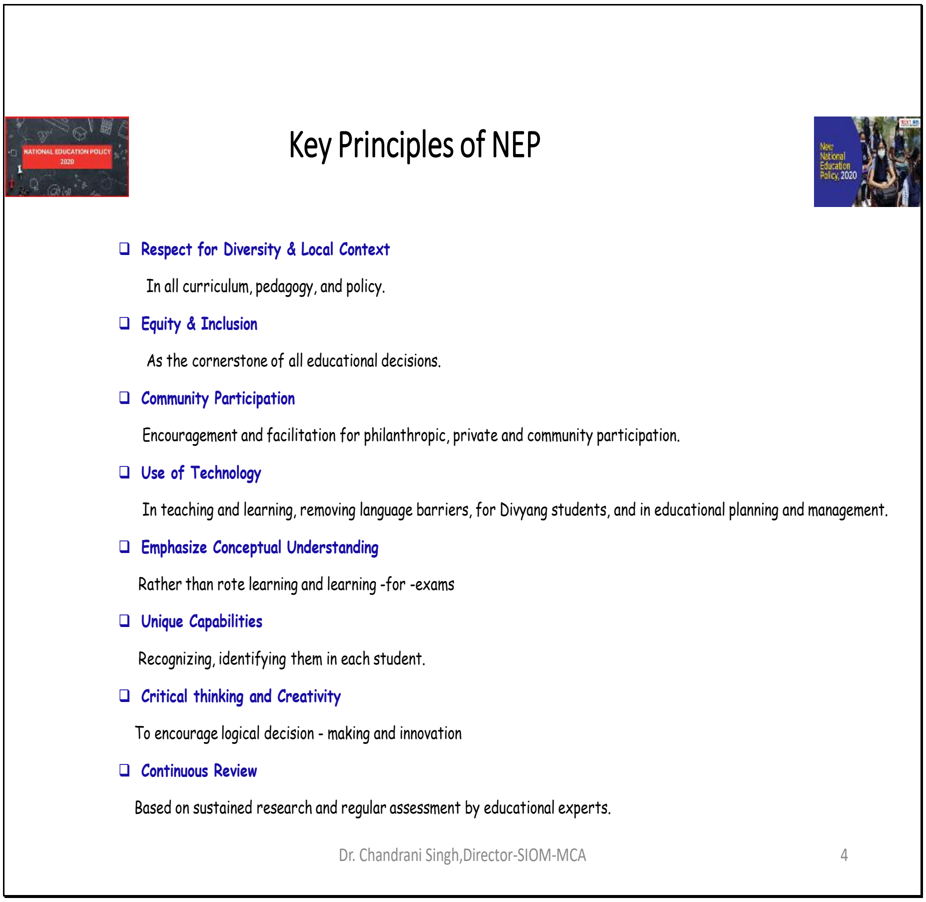
Dr. Chandrani Singh,Director-SIOM-MCA 1



## Vision of the National Education Policy 2020

- ❑ An education system that contributes to an equitable and vibrant knowledge society, by providing high-quality education to all.
- ❑ Develops a deep sense of respect towards the fundamental rights, duties and Constitutional values, bonding with one's country, and a conscious awareness of one's role and responsibilities in a changing world.
- ❑ Instils skills, values, and dispositions that support responsible commitment to human rights, sustainable development and living, and global well-being, thereby reflecting a truly global citizen

Dr. Chandrani Singh,Director-SIOM-MCA 3


NATIONAL EDUCATION POLICY 2020
New National Education Policy 2020

# Key Principles of NEP

- **Respect for Diversity & Local Context**

  In all curriculum, pedagogy, and policy.

- **Equity & Inclusion**

  As the cornerstone of all educational decisions.

- **Community Participation**

  Encouragement and facilitation for philanthropic, private and community participation.

- **Use of Technology**

  In teaching and learning, removing language barriers, for Divyang students, and in educational planning and management.

- **Emphasize Conceptual Understanding**

  Rather than rote learning and learning -for -exams

- **Unique Capabilities**

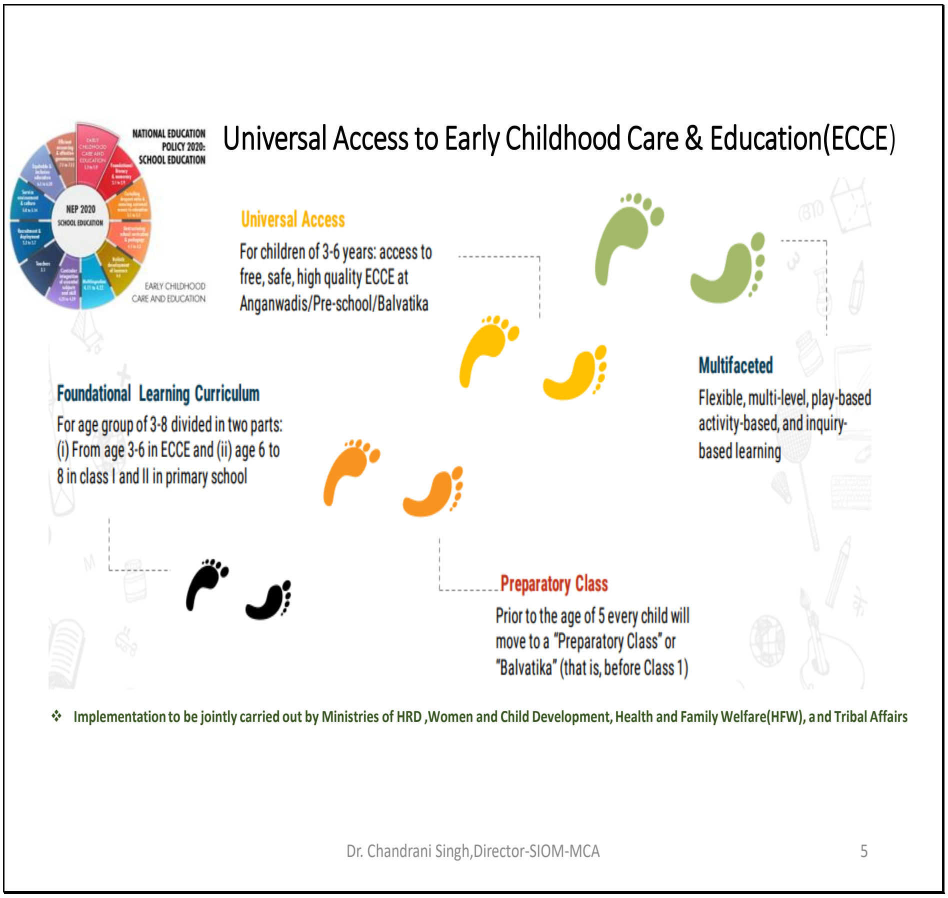
  Recognizing, identifying them in each student.

- **Critical thinking and Creativity**

  To encourage logical decision - making and innovation

- **Continuous Review**

  Based on sustained research and regular assessment by educational experts.


NATIONAL EDUCATION POLICY 2020: SCHOOL EDUCATION
NEP 2020
SCHOOL EDUCATION
EARLY CHILDHOOD CARE AND EDUCATION

# Universal Access to Early Childhood Care & Education(ECCE)

**Universal Access**

For children of 3-6 years: access to free, safe, high quality ECCE at Anganwadis/Pre-school/Balvatika

**Multifaceted**

Flexible, multi-level, play-based activity-based, and inquiry-based learning

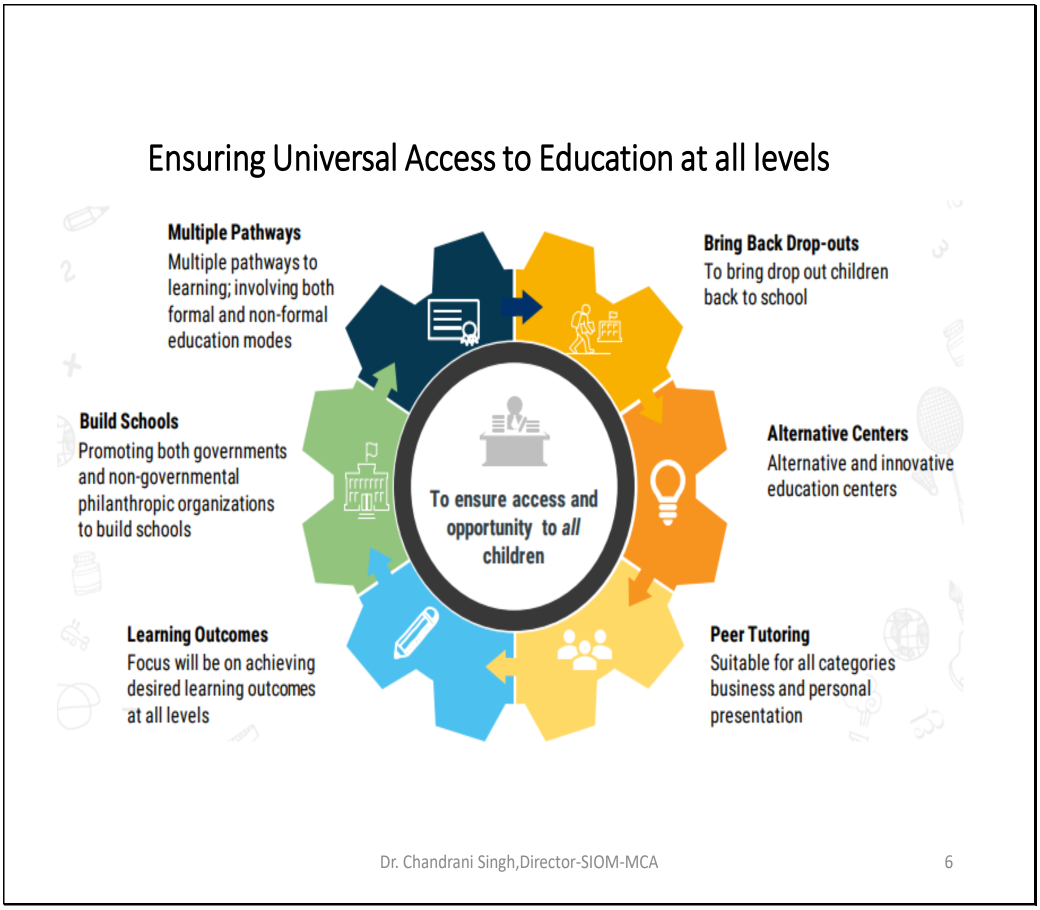
**Foundational Learning Curriculum**

For age group of 3-8 divided in two parts: (i) From age 3-6 in ECCE and (ii) age 6 to 8 in class I and II in primary school

**Preparatory Class**

Prior to the age of 5 every child will move to a "Preparatory Class" or "Balvatika" (that is, before Class 1)

- Implementation to be jointly carried out by Ministries of HRD ,Women and Child Development, Health and Family Welfare(HFW), and Tribal Affairs



## Ensuring Universal Access to Education at all levels

**Multiple Pathways**
Multiple pathways to learning; involving both formal and non-formal education modes

**Bring Back Drop-outs**
To bring drop out children back to school

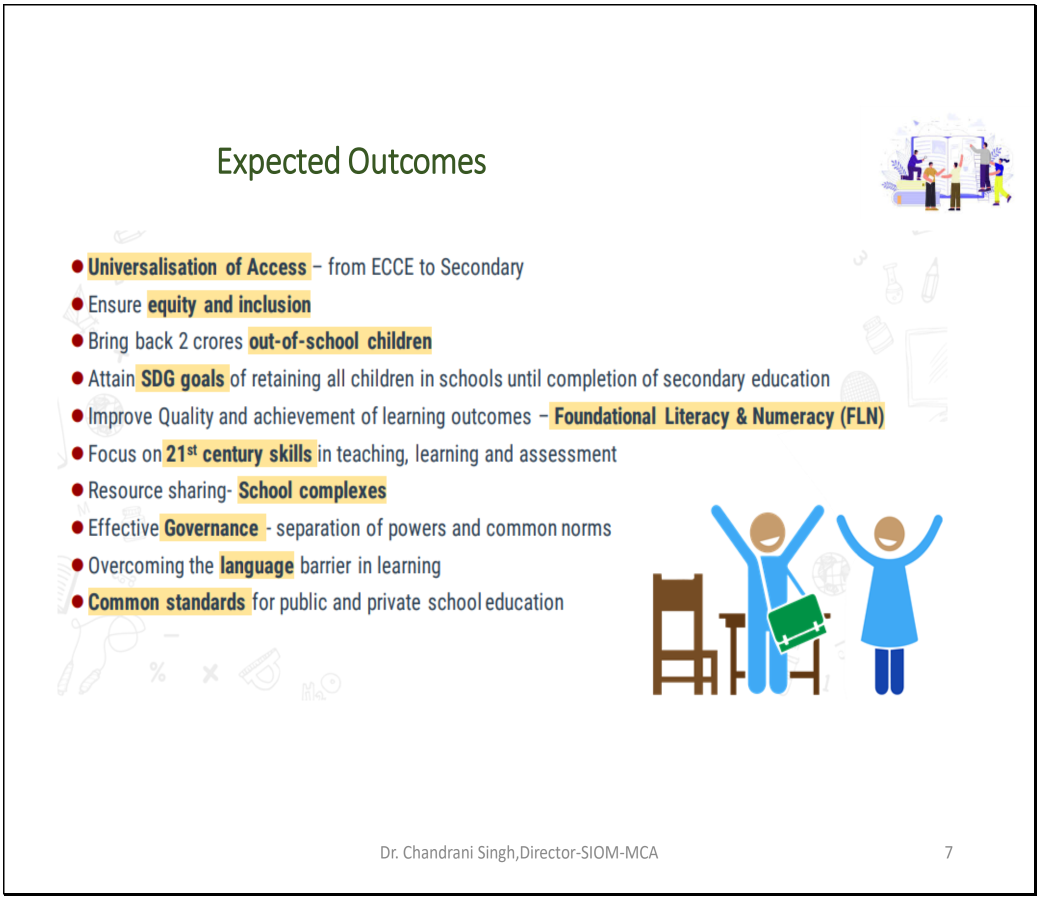
**Build Schools**
Promoting both governments and non-governmental philanthropic organizations to build schools

**To ensure access and opportunity to *all* children**

**Alternative Centers**
Alternative and innovative education centers

**Learning Outcomes**
Focus will be on achieving desired learning outcomes at all levels

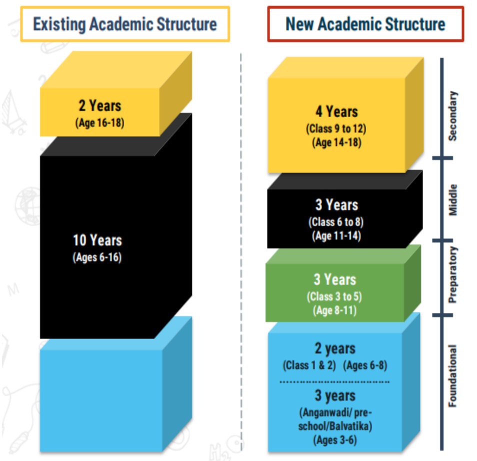
**Peer Tutoring**
Suitable for all categories business and personal presentation



## Expected Outcomes

- **Universalisation of Access** – from ECCE to Secondary
- Ensure **equity and inclusion**
- Bring back 2 crores **out-of-school children**
- Attain **SDG goals** of retaining all children in schools until completion of secondary education
- Improve Quality and achievement of learning outcomes – **Foundational Literacy & Numeracy (FLN)**
- Focus on **21st century skills** in teaching, learning and assessment
- Resource sharing- **School complexes**
- Effective **Governance** - separation of powers and common norms
- Overcoming the **language** barrier in learning
- **Common standards** for public and private school education

# Transforming Curricular & Pedagogical Structure



Existing Academic Structure

2 Years (Age 16-18)

10 Years (Ages 6-16)

New Academic Structure

4 Years (Class 9 to 12) (Age 14-18) — Secondary

3 Years (Class 6 to 8) (Age 11-14) — Middle

3 Years (Class 3 to 5) (Age 8-11) — Preparatory

2 years (Class 1 & 2) (Ages 6-8)

3 years (Anganwadi/ pre-school/Balvatika) (Ages 3-6) — Foundational

**New pedagogical and curricular structure of school education (5+3+3+4): 3 years in Anganwadi/pre-school and 12 years in school**

- Secondary Stage(4) multidisciplinary study, greater critical thinking, flexibility and student choice of subjects
- **Middle Stage (3)** experiential learning in the sciences, mathematics, arts, social sciences, and humanities
- Preparatory Stage (3) play, discovery, and activity-based and interactive classroom learning
- Foundational stage (5) multilevel, play/activity-based learning

# ECCE Framework

## NCPFECE

National Curricular and Pedagogical Framework for Early Childhood Education (NCPFECE) will be drafted by NCERT

## Research and Best Practices

NCPFECE will be aligned with the latest research on ECCE, and national and international best practices

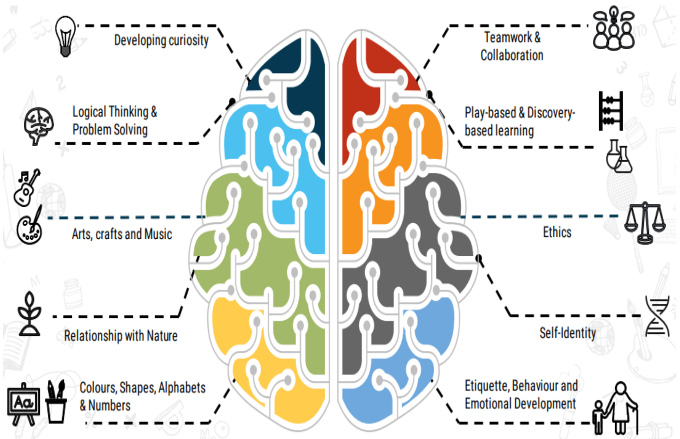
## Multi-faceted Framework

Comprising of alphabets, languages, numbers, counting, colours, shapes, indoor and outdoor play, puzzles and logical thinking, problem-solving, drawing, painting and other visual art, craft, drama and puppetry, music and movement

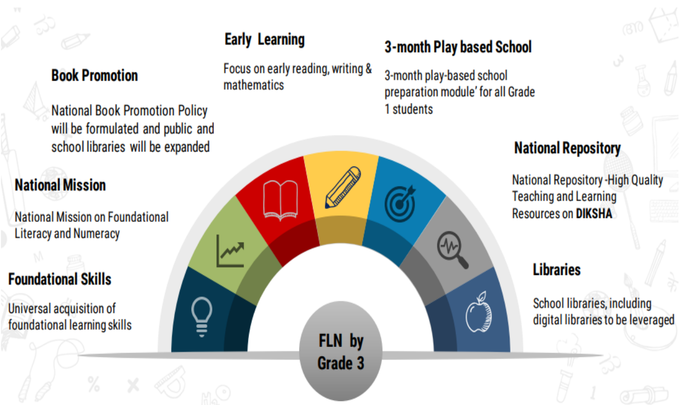
## School Preparation Module

A 3-month play-based 'school preparation module' for all Grade 1 students to be developed by NCERT

## Early Childhood Education: Learning in the Formative Years


Developing curiosity
Teamwork & Collaboration
Logical Thinking & Problem Solving
Play-based & Discovery-based learning
Arts, crafts and Music
Ethics
Relationship with Nature
Self-Identity
Colours, Shapes, Alphabets & Numbers
Etiquette, Behaviour and Emotional Development

## Attainment of Foundational Literacy & Numeracy by Grade 3 in Mission mode


**Early Learning**

Focus on early reading, writing & mathematics

**3-month Play based School**

3-month play-based school preparation module' for all Grade 1 students

**Book Promotion**

National Book Promotion Policy will be formulated and public and school libraries will be expanded

**National Repository**

National Repository -High Quality Teaching and Learning Resources on **DIKSHA**

**National Mission**

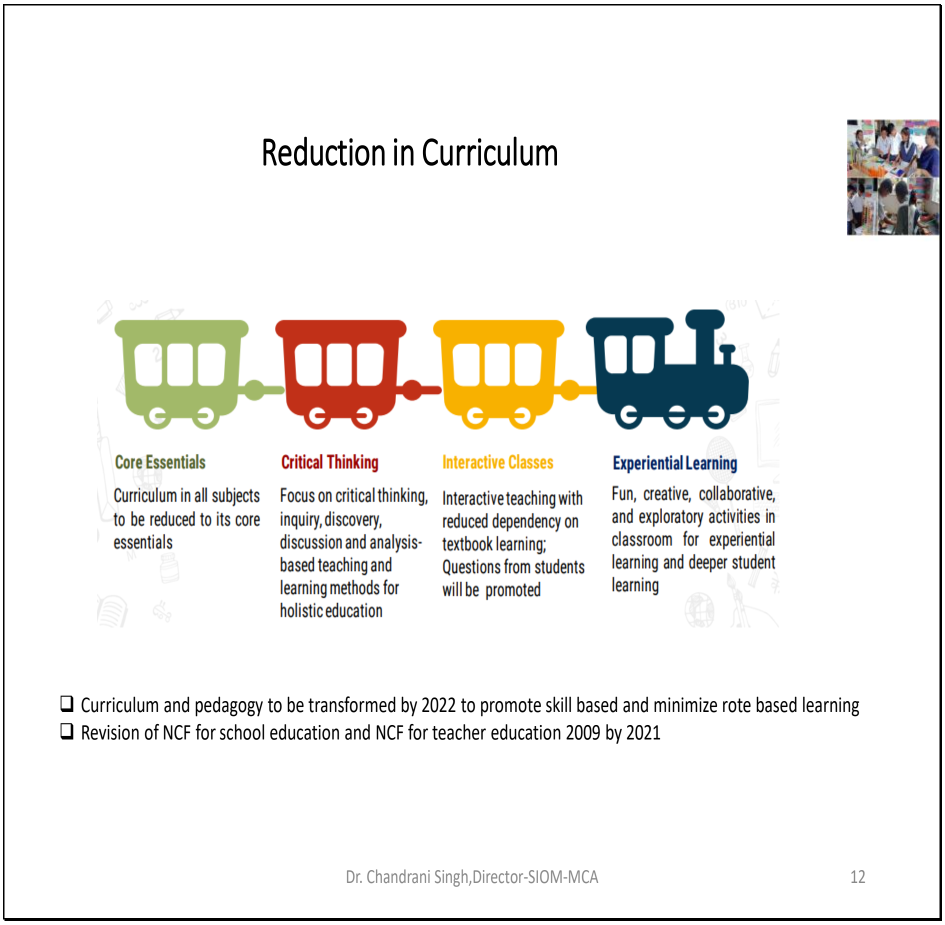
National Mission on Foundational Literacy and Numeracy

**Libraries**

School libraries, including digital libraries to be leveraged

**Foundational Skills**

Universal acquisition of foundational learning skills



# Reduction in Curriculum

**Core Essentials**

Curriculum in all subjects to be reduced to its core essentials

**Critical Thinking**

Focus on critical thinking, inquiry, discovery, discussion and analysis-based teaching and learning methods for holistic education

**Interactive Classes**

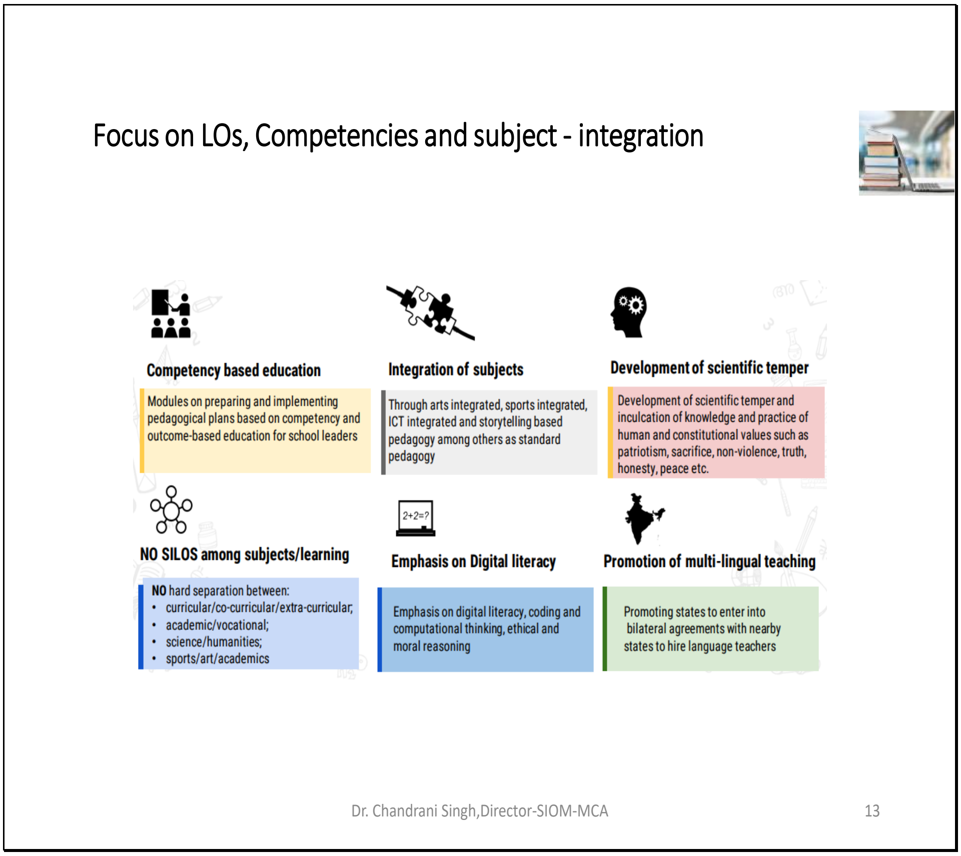
Interactive teaching with reduced dependency on textbook learning; Questions from students will be promoted

**Experiential Learning**

Fun, creative, collaborative, and exploratory activities in classroom for experiential learning and deeper student learning

❑ Curriculum and pedagogy to be transformed by 2022 to promote skill based and minimize rote based learning

❑ Revision of NCF for school education and NCF for teacher education 2009 by 2021



# Focus on LOs, Competencies and subject - integration

**Competency based education**

Modules on preparing and implementing pedagogical plans based on competency and outcome-based education for school leaders

**Integration of subjects**

Through arts integrated, sports integrated, ICT integrated and storytelling based pedagogy among others as standard pedagogy

**Development of scientific temper**

Development of scientific temper and inculcation of knowledge and practice of human and constitutional values such as patriotism, sacrifice, non-violence, truth, honesty, peace etc.

**NO SILOS among subjects/learning**

NO hard separation between:

- curricular/co-curricular/extra-curricular;
- academic/vocational;
- science/humanities;
- sports/art/academics

**Emphasis on Digital literacy**

Emphasis on digital literacy, coding and computational thinking, ethical and moral reasoning

**Promotion of multi-lingual teaching**

Promoting states to enter into bilateral agreements with nearby states to hire language teachers

# Mental and physical health and well-being

**Health check ups**
Annual health check up for all students

**Reduce weight of school bags**
Reduced weight of school bags and textbooks through suitable changes in curriculum and pedagogy

**Mandatory skills : Health and Wellness**
Mandatory skills to be imbibed by all students - health, nutrition, physical education, fitness, wellness, sports. In addition- Basic training in preventive health care, mental health, first aid, personal and public hygiene will be included in the curriculum

**Hiring counsellors in school complexes**
State governments will be encouraged to hire adequate number of counsellors and teachers (to be shared across school complex)

**Focus on children with disability**
Differentiated interventions and suitable infrastructure development at schools to make access easier for children with disability

**Inclusive and caring culture at school**
The role and expectations of principal and teachers will explicitly include developing a caring and inclusive culture at school

- Mandatory for students to acquire skills in: health and nutrition; physical education, fitness, wellness, and sports

# Innovative Pedagogy: Transforming teaching learning process

**Experiential Learning**
- Focus on experiential, inquiry and discovery based teaching learning methods

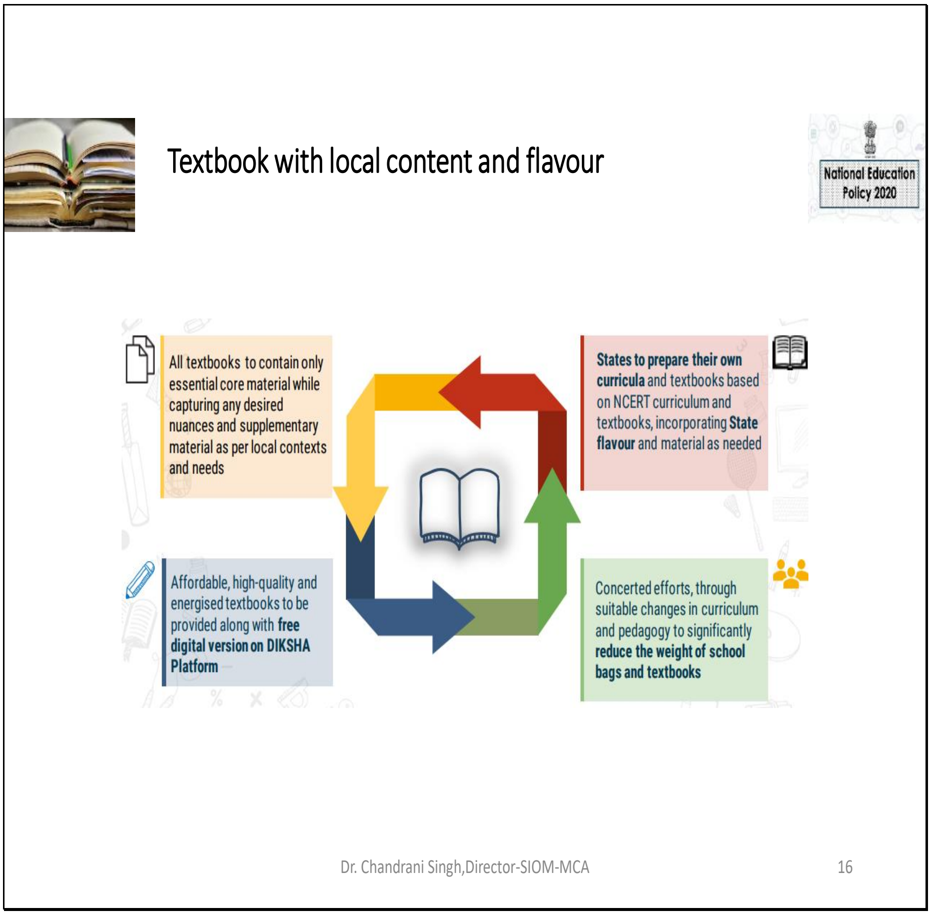
**Integrated Pedagogy**
- Arts, sports, and story-telling and ICT-integrated pedagogy

**Promotion of peer tutoring**
- Promoting peer tutoring as voluntary and joyful activity under the supervision of teachers

**Equal Weightage**
- No hard separation between curricular, co-curricular and extra curricular area.
- Freedom of choosing a variety of subject combination to be provided

**Bagless Days**
- Bagless days to be scheduled in academic calendar

**Use and integration of technology**
- Integration of technology enabled pedagogy in classes 6-12



National Education Policy 2020

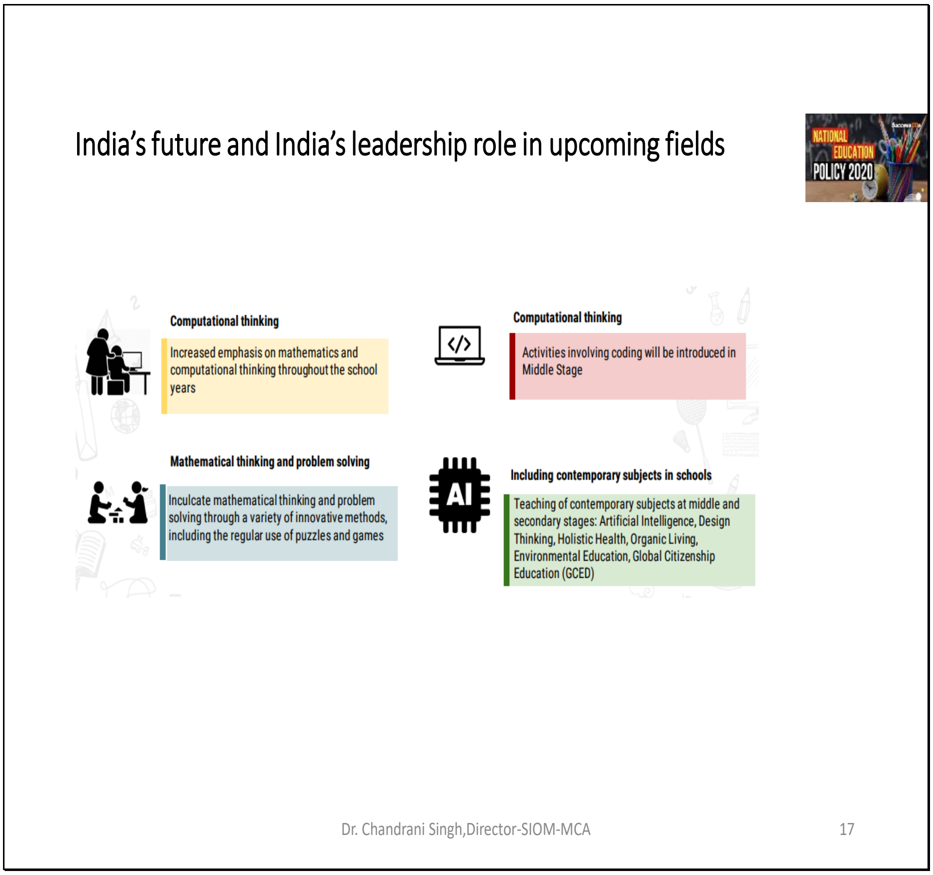
## Textbook with local content and flavour

All textbooks to contain only essential core material while capturing any desired nuances and supplementary material as per local contexts and needs

**States to prepare their own curricula** and textbooks based on NCERT curriculum and textbooks, incorporating **State flavour** and material as needed

Affordable, high-quality and energised textbooks to be provided along with **free digital version on DIKSHA Platform**

Concerted efforts, through suitable changes in curriculum and pedagogy to significantly **reduce the weight of school bags and textbooks**



NATIONAL EDUCATION POLICY 2020

## India's future and India's leadership role in upcoming fields

**Computational thinking**

Increased emphasis on mathematics and computational thinking throughout the school years

**Computational thinking**

Activities involving coding will be introduced in Middle Stage

**Mathematical thinking and problem solving**

Inculcate mathematical thinking and problem solving through a variety of innovative methods, including the regular use of puzzles and games

**Including contemporary subjects in schools**

Teaching of contemporary subjects at middle and secondary stages: Artificial Intelligence, Design Thinking, Holistic Health, Organic Living, Environmental Education, Global Citizenship Education (GCED)

# Knowledge of India

Video documentaries on inspirational luminaries of India, in science and beyond

Students will be given a logical framework for making ethical decisions at a young age.

In later years, expanded along themes of cheating, violence, plagiarism, littering, tolerance, equality, empathy.

Traditional Indian values and all basic human and Constitutional values will be developed in all students.

Excerpts from the Indian Constitution will also be considered essential reading for all students.

Basic training in health, mental health, good nutrition, personal and public hygiene, disaster response and first-aid will also be included.

Will be incorporated in an **accurate and scientific manner** wherever relevant.

Indian Knowledge Systems, including **tribal knowledge** and **indigenous and traditional ways of learning,** will be covered.

Specific courses in tribal ethno-medicinal practices, forest management, traditional (organic) crop cultivation, natural farming, etc. will also be made available.

Curriculum to include knowledge from ancient India to modern India as well as future aspirations.

Scientific explanations of the detrimental and damaging effects of alcohol, tobacco, and other drugs will be part of curriculum.

# Examination in grade 1 to 8

### Key stage assessments

Census assessments at key stage in classes 3, 5 and 8 to track achievement

### Achievement of critical LOs

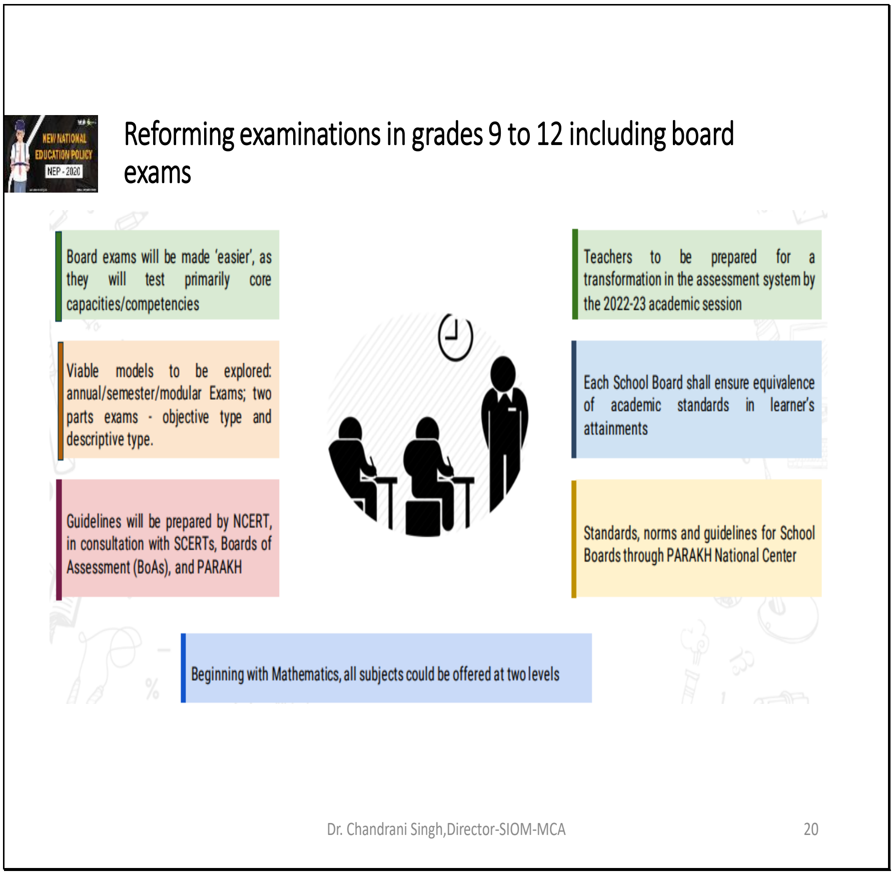
Testing to focus on achievement of essential learning outcomes

### Moving away from rote learning

Assessment of core concepts and knowledge, higher-order skills and its application in real-life situations. Moving away from rote learning.

### Results of school examinations

The results of school examinations will be used only for developmental purposes and for continuous monitoring and improvement of the schooling system



## Reforming examinations in grades 9 to 12 including board exams

Board exams will be made 'easier', as they will test primarily core capacities/competencies

Viable models to be explored: annual/semester/modular Exams; two parts exams - objective type and descriptive type.

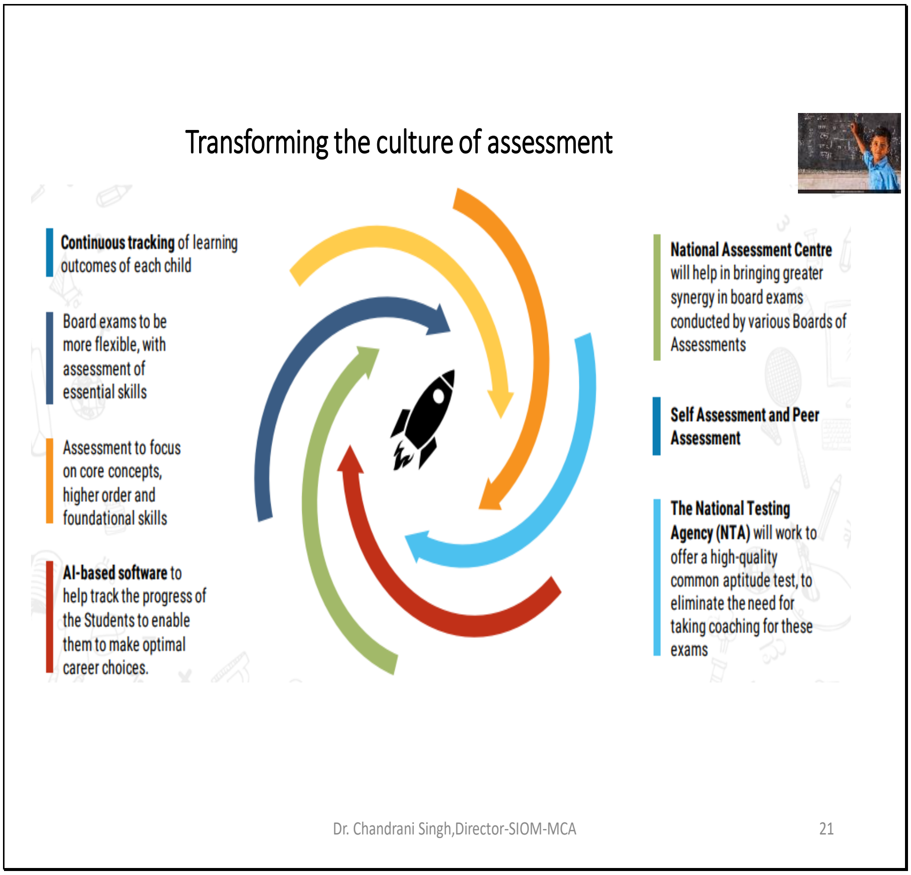
Guidelines will be prepared by NCERT, in consultation with SCERTs, Boards of Assessment (BoAs), and PARAKH

Teachers to be prepared for a transformation in the assessment system by the 2022-23 academic session

Each School Board shall ensure equivalence of academic standards in learner's attainments

Standards, norms and guidelines for School Boards through PARAKH National Center

Beginning with Mathematics, all subjects could be offered at two levels



## Transforming the culture of assessment

**Continuous tracking** of learning outcomes of each child

Board exams to be more flexible, with assessment of essential skills

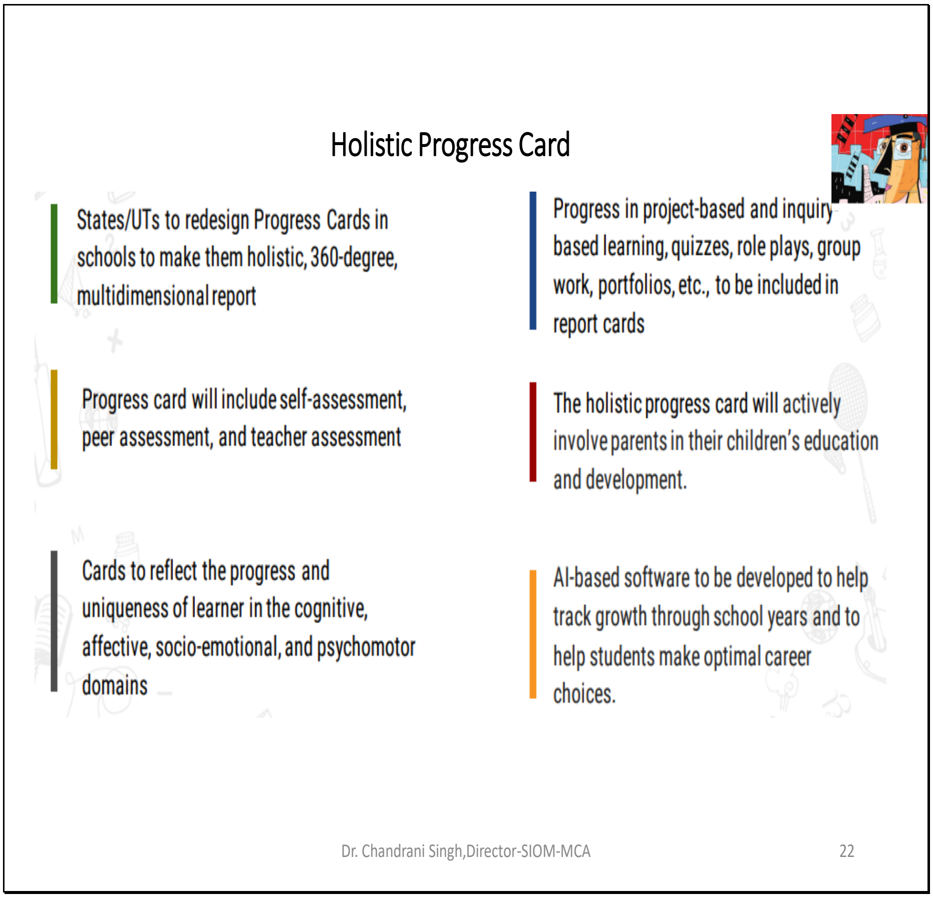
Assessment to focus on core concepts, higher order and foundational skills

**AI-based software** to help track the progress of the Students to enable them to make optimal career choices.

**National Assessment Centre** will help in bringing greater synergy in board exams conducted by various Boards of Assessments

**Self Assessment and Peer Assessment**

**The National Testing Agency (NTA)** will work to offer a high-quality common aptitude test, to eliminate the need for taking coaching for these exams



## Holistic Progress Card

States/UTs to redesign Progress Cards in schools to make them holistic, 360-degree, multidimensional report

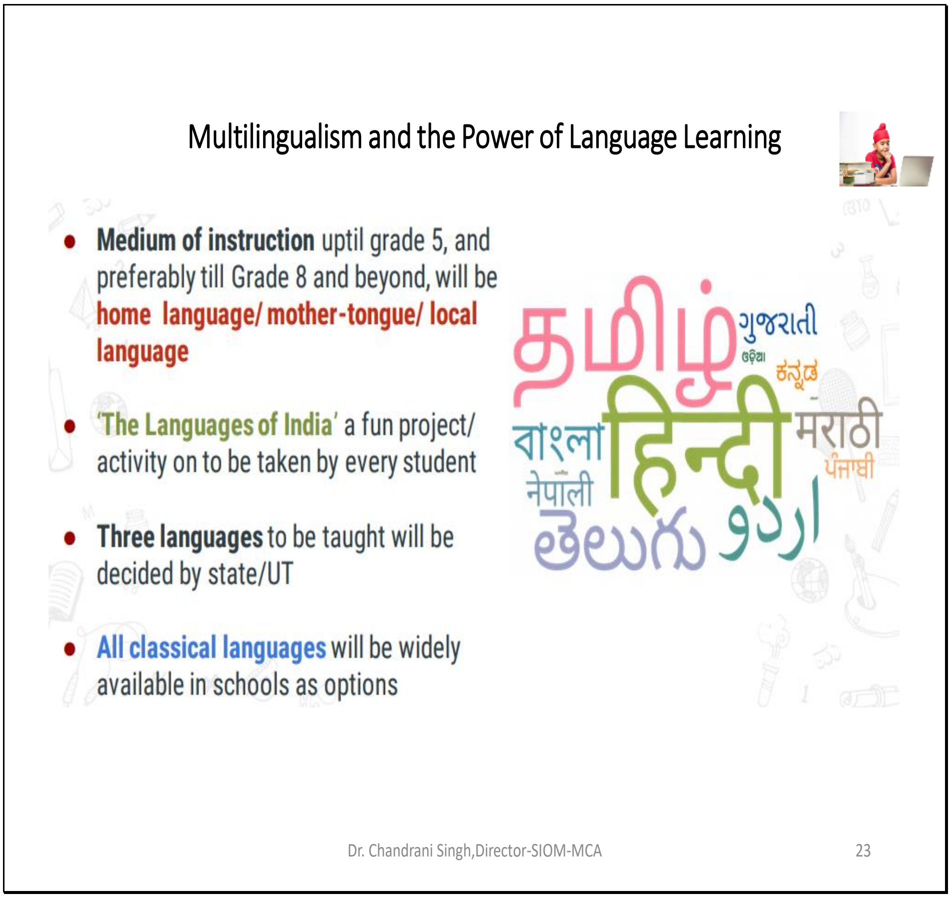
Progress card will include self-assessment, peer assessment, and teacher assessment

Cards to reflect the progress and uniqueness of learner in the cognitive, affective, socio-emotional, and psychomotor domains

Progress in project-based and inquiry-based learning, quizzes, role plays, group work, portfolios, etc., to be included in report cards

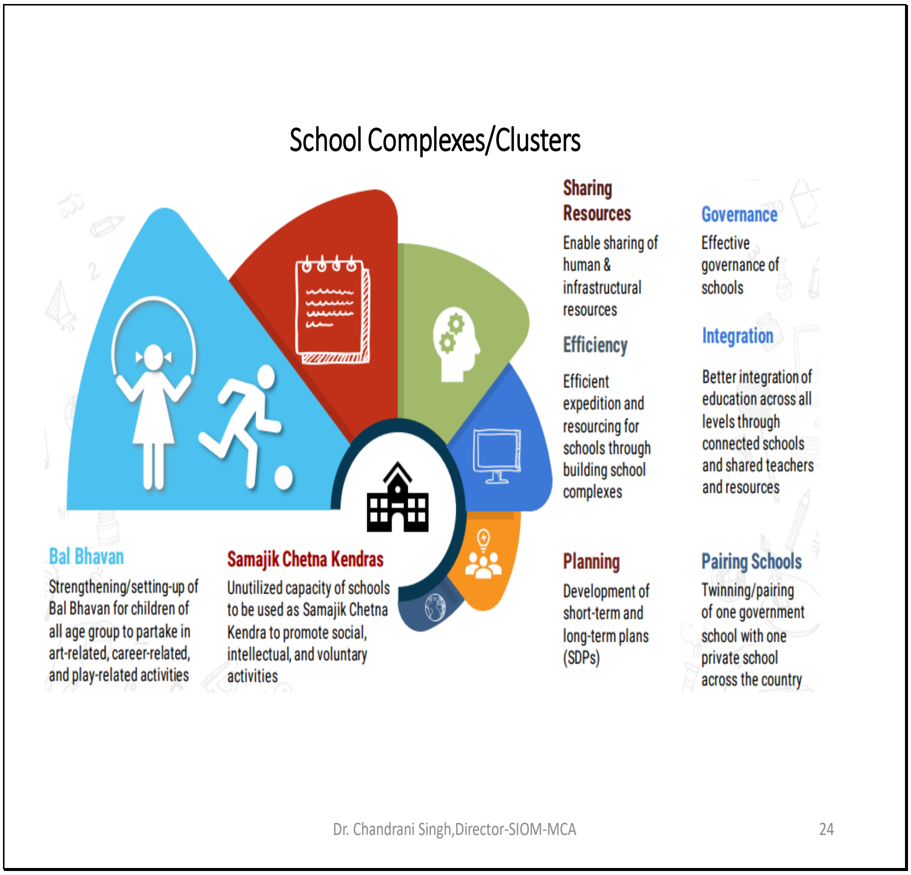
The holistic progress card will actively involve parents in their children's education and development.

AI-based software to be developed to help track growth through school years and to help students make optimal career choices.



## Multilingualism and the Power of Language Learning

- **Medium of instruction** uptil grade 5, and preferably till Grade 8 and beyond, will be **home language/ mother-tongue/ local language**
- **'The Languages of India'** a fun project/ activity on to be taken by every student
- **Three languages** to be taught will be decided by state/UT
- **All classical languages** will be widely available in schools as options



## School Complexes/Clusters

**Sharing Resources**
Enable sharing of human & infrastructural resources

**Governance**
Effective governance of schools

**Efficiency**
Efficient expedition and resourcing for schools through building school complexes

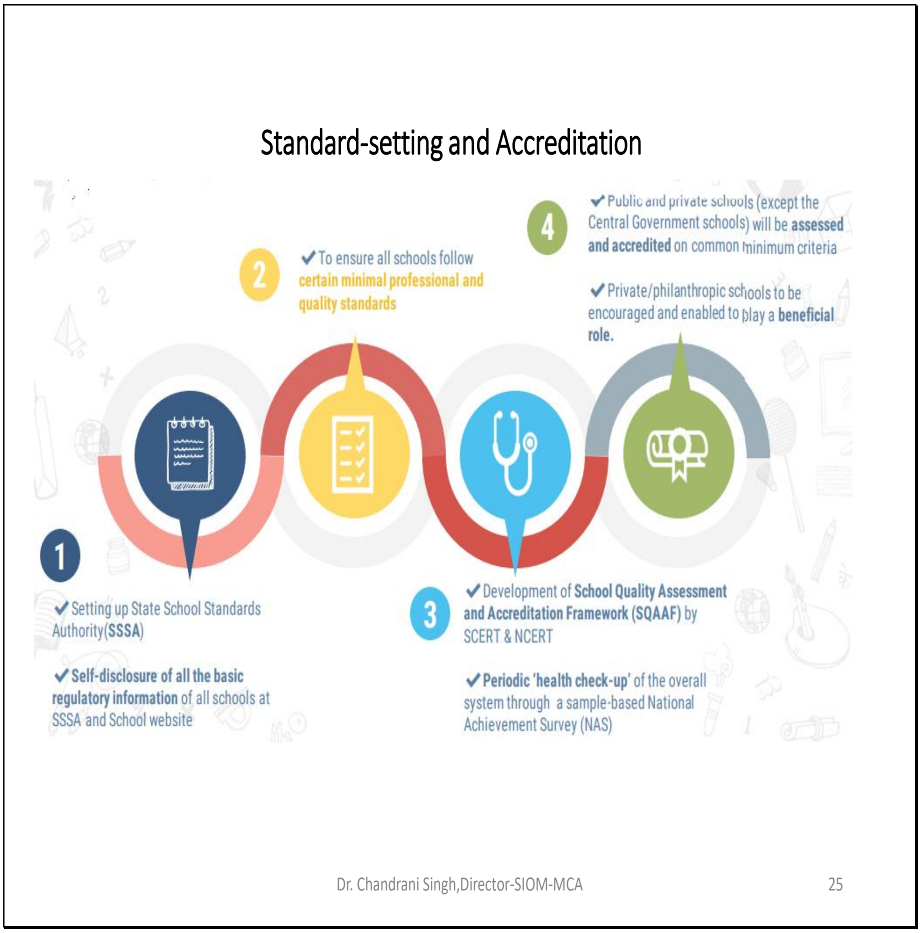
**Integration**
Better integration of education across all levels through connected schools and shared teachers and resources

**Bal Bhavan**
Strengthening/setting-up of Bal Bhavan for children of all age group to partake in art-related, career-related, and play-related activities

**Samajik Chetna Kendras**
Unutilized capacity of schools to be used as Samajik Chetna Kendra to promote social, intellectual, and voluntary activities

**Planning**
Development of short-term and long-term plans (SDPs)

**Pairing Schools**
Twinning/pairing of one government school with one private school across the country



## Standard-setting and Accreditation

1
✔ Setting up State School Standards Authority(**SSSA**)

✔ **Self-disclosure of all the basic regulatory information** of all schools at SSSA and School website

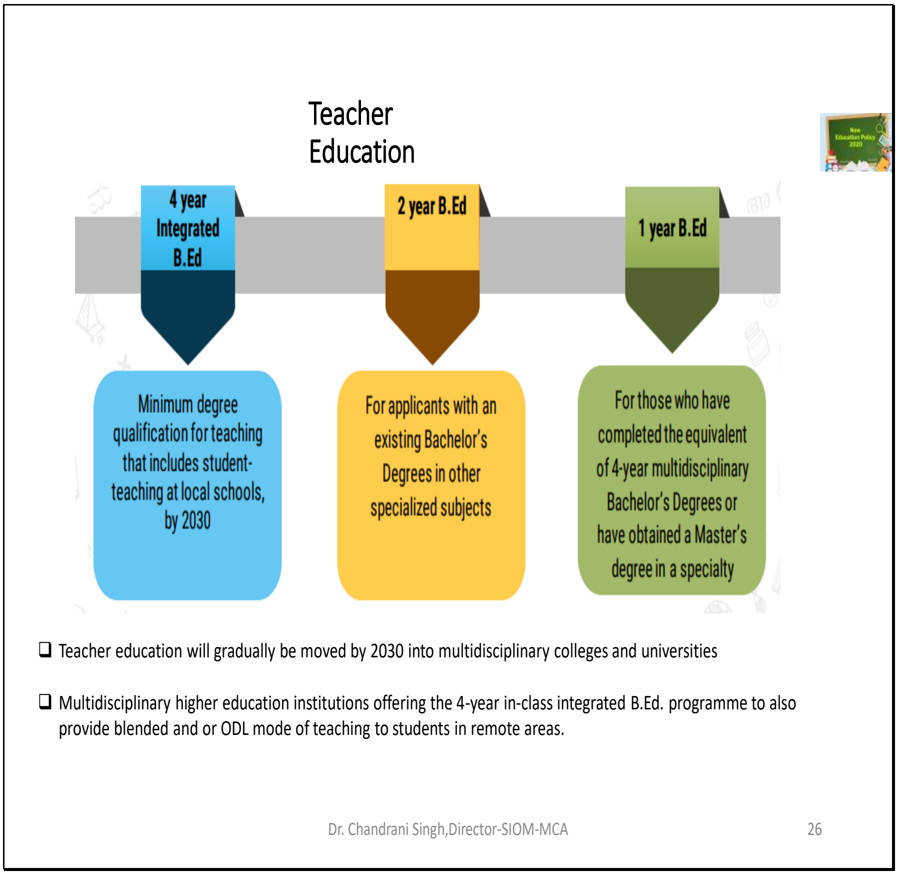
2
✔ To ensure all schools follow **certain minimal professional and quality standards**

3
✔ Development of **School Quality Assessment and Accreditation Framework (SQAAF)** by SCERT & NCERT

✔ **Periodic 'health check-up'** of the overall system through a sample-based National Achievement Survey (NAS)

4
✔ Public and private schools (except the Central Government schools) will be **assessed and accredited** on common minimum criteria

✔ Private/philanthropic schools to be encouraged and enabled to play a **beneficial role.**



## Teacher Education

4 year Integrated B.Ed

Minimum degree qualification for teaching that includes student-teaching at local schools, by 2030

2 year B.Ed

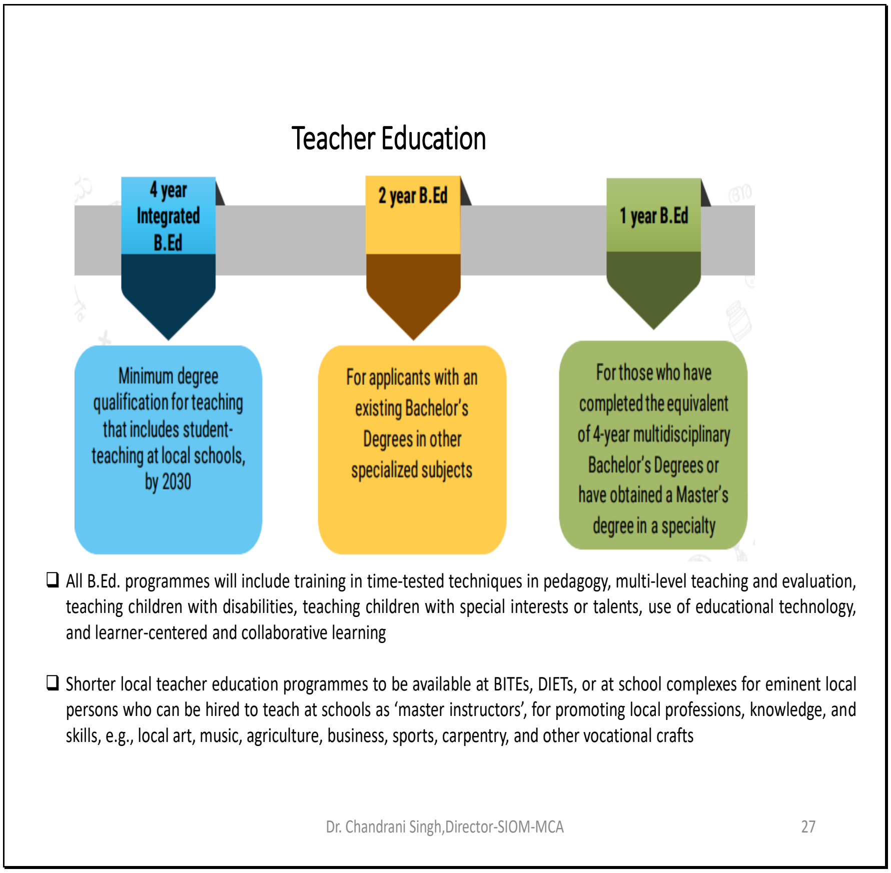
For applicants with an existing Bachelor's Degrees in other specialized subjects

1 year B.Ed

For those who have completed the equivalent of 4-year multidisciplinary Bachelor's Degrees or have obtained a Master's degree in a specialty

❑ Teacher education will gradually be moved by 2030 into multidisciplinary colleges and universities

❑ Multidisciplinary higher education institutions offering the 4-year in-class integrated B.Ed. programme to also provide blended and or ODL mode of teaching to students in remote areas.



## Teacher Education

4 year Integrated B.Ed

Minimum degree qualification for teaching that includes student-teaching at local schools, by 2030

2 year B.Ed

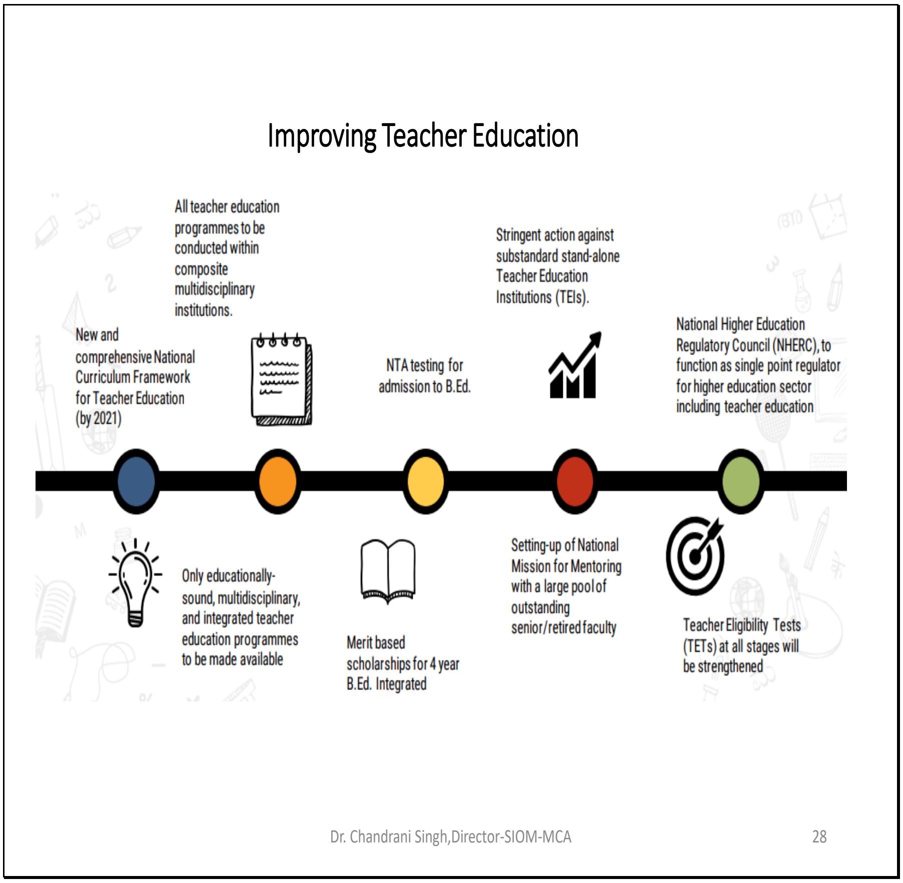
For applicants with an existing Bachelor's Degrees in other specialized subjects

1 year B.Ed

For those who have completed the equivalent of 4-year multidisciplinary Bachelor's Degrees or have obtained a Master's degree in a specialty

❑ All B.Ed. programmes will include training in time-tested techniques in pedagogy, multi-level teaching and evaluation, teaching children with disabilities, teaching children with special interests or talents, use of educational technology, and learner-centered and collaborative learning

❑ Shorter local teacher education programmes to be available at BITEs, DIETs, or at school complexes for eminent local persons who can be hired to teach at schools as 'master instructors', for promoting local professions, knowledge, and skills, e.g., local art, music, agriculture, business, sports, carpentry, and other vocational crafts



# Improving Teacher Education

New and comprehensive National Curriculum Framework for Teacher Education (by 2021)

Only educationally-sound, multidisciplinary, and integrated teacher education programmes to be made available

All teacher education programmes to be conducted within composite multidisciplinary institutions.

Merit based scholarships for 4 year B.Ed. Integrated

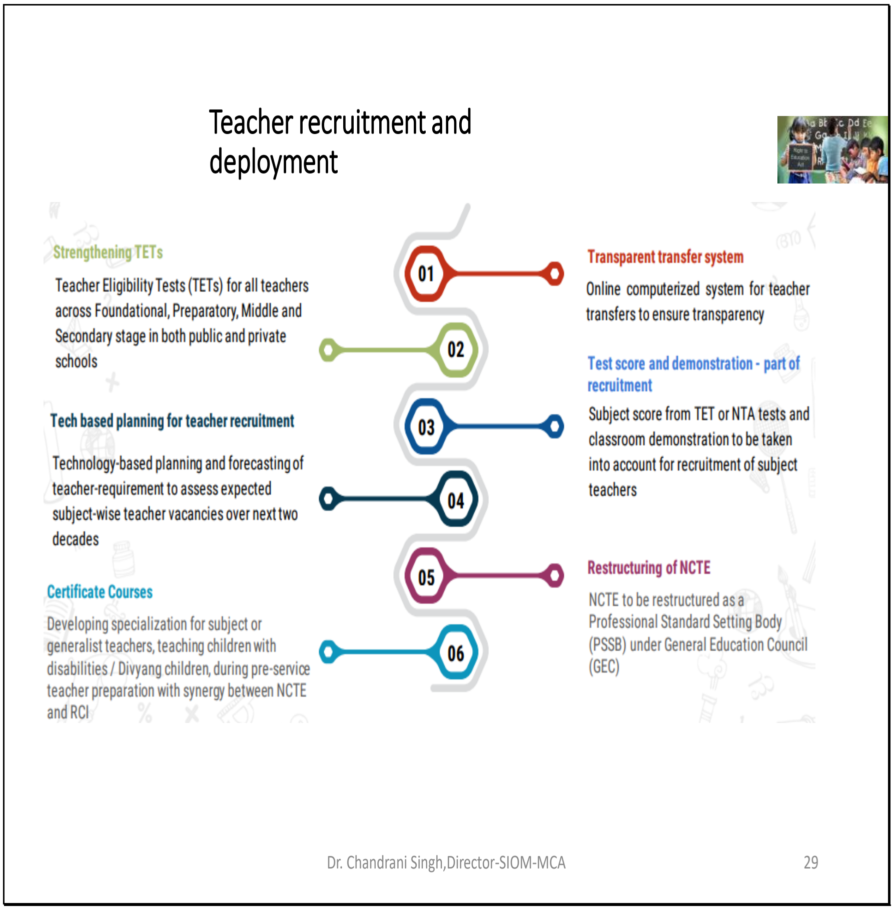
NTA testing for admission to B.Ed.

Setting-up of National Mission for Mentoring with a large pool of outstanding senior/retired faculty

Stringent action against substandard stand-alone Teacher Education Institutions (TEIs).

Teacher Eligibility Tests (TETs) at all stages will be strengthened

National Higher Education Regulatory Council (NHERC), to function as single point regulator for higher education sector including teacher education



# Teacher recruitment and deployment

**Transparent transfer system**

Online computerized system for teacher transfers to ensure transparency

**Strengthening TETs**

Teacher Eligibility Tests (TETs) for all teachers across Foundational, Preparatory, Middle and Secondary stage in both public and private schools

**Test score and demonstration - part of recruitment**

Subject score from TET or NTA tests and classroom demonstration to be taken into account for recruitment of subject teachers

**Tech based planning for teacher recruitment**

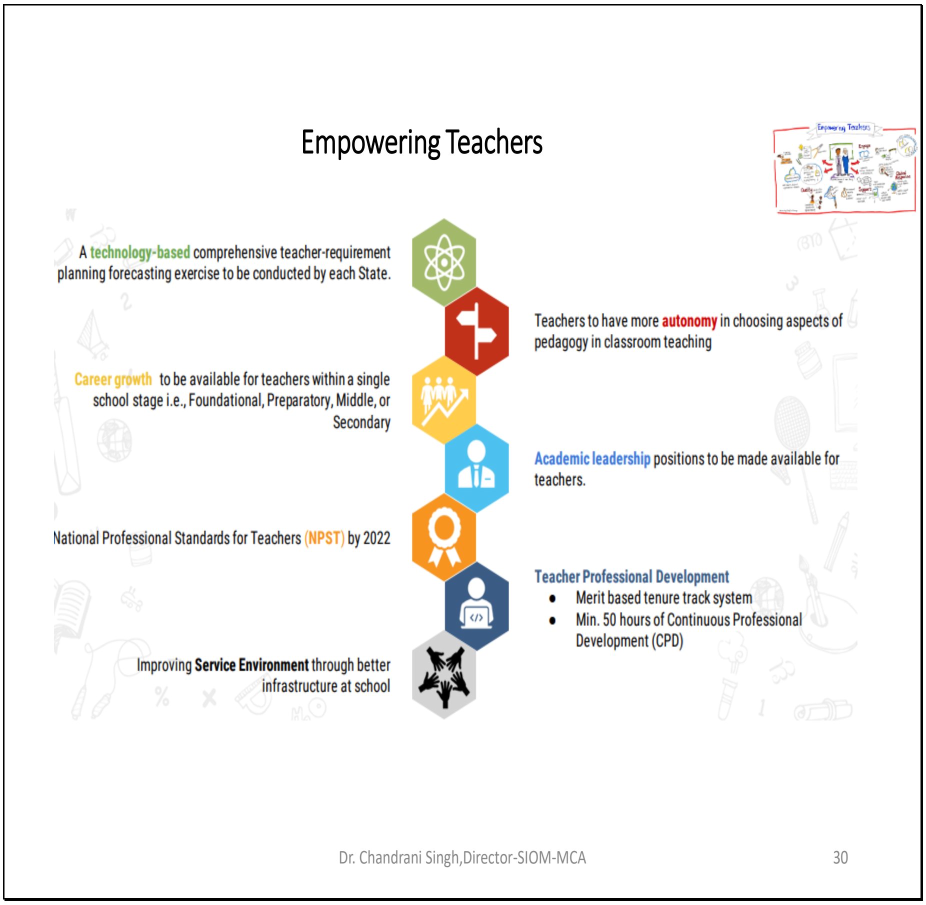
Technology-based planning and forecasting of teacher-requirement to assess expected subject-wise teacher vacancies over next two decades

**Restructuring of NCTE**

NCTE to be restructured as a Professional Standard Setting Body (PSSB) under General Education Council (GEC)

**Certificate Courses**

Developing specialization for subject or generalist teachers, teaching children with disabilities / Divyang children, during pre-service teacher preparation with synergy between NCTE and RCI



## Empowering Teachers

A technology-based comprehensive teacher-requirement planning forecasting exercise to be conducted by each State.

Teachers to have more **autonomy** in choosing aspects of pedagogy in classroom teaching

Career growth to be available for teachers within a single school stage i.e., Foundational, Preparatory, Middle, or Secondary

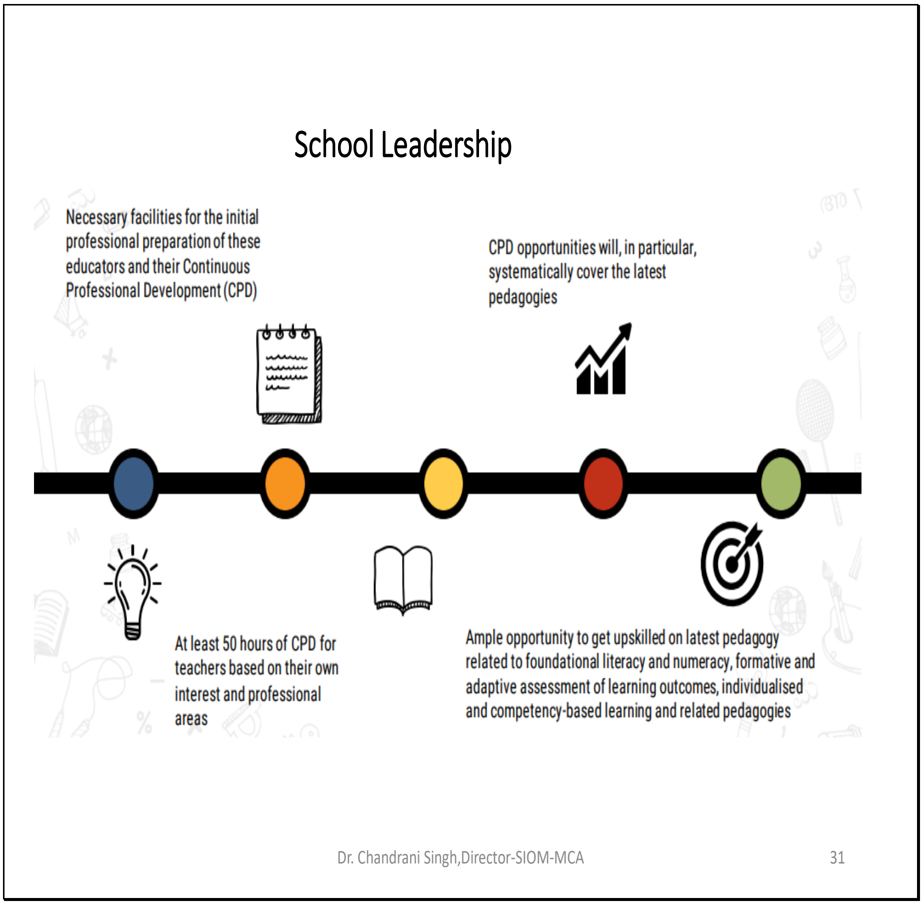
Academic leadership positions to be made available for teachers.

National Professional Standards for Teachers (NPST) by 2022

**Teacher Professional Development**

- Merit based tenure track system
- Min. 50 hours of Continuous Professional Development (CPD)

Improving **Service Environment** through better infrastructure at school



## School Leadership

Necessary facilities for the initial professional preparation of these educators and their Continuous Professional Development (CPD)

CPD opportunities will, in particular, systematically cover the latest pedagogies

At least 50 hours of CPD for teachers based on their own interest and professional areas

Ample opportunity to get upskilled on latest pedagogy related to foundational literacy and numeracy, formative and adaptive assessment of learning outcomes, individualised and competency-based learning and related pedagogies

# Focus on Socio-Economically Disadvantaged Groups (SEDGs)

**SEDGs** can be broadly categorized based on:

- **Gender identities** (particularly female and transgender individuals),
- **Socio-cultural identities** (such as Scheduled Castes, Scheduled Tribes, OBCs, and minorities),
- **Geographical identities** (such as students from villages, small towns, and aspirational districts),
- **Disabilities** (including learning disabilities), and
- **Socio-economic conditions** (such as migrant communities, low income households, children in vulnerable situations, victims of or children of victims of trafficking, orphans including child beggars in urban areas, and the urban poor).

❑ **Separate strategies will be formulated for focused attention for reducing each of the category-wise gaps in school education.**

# Ensuring Equity

**Interventions**

The critical problems and recommendations regarding ECCE, foundational literacy and numeracy, access, enrolment and attendance will be targeted in a concerted way for Socio=Economically Disadvantaged groups - SEDGs.

**SEZs**

Large populations from SEDGs to be declared Special Education Zones (SEZs)

**Fee Waivers**

Fee waivers and scholarships will be offered to meritorious students from all SEDGs on a larger scale

**Special Mechanisms**

Special mechanisms for children belonging to tribal groups to receive quality education

**Counsellors**

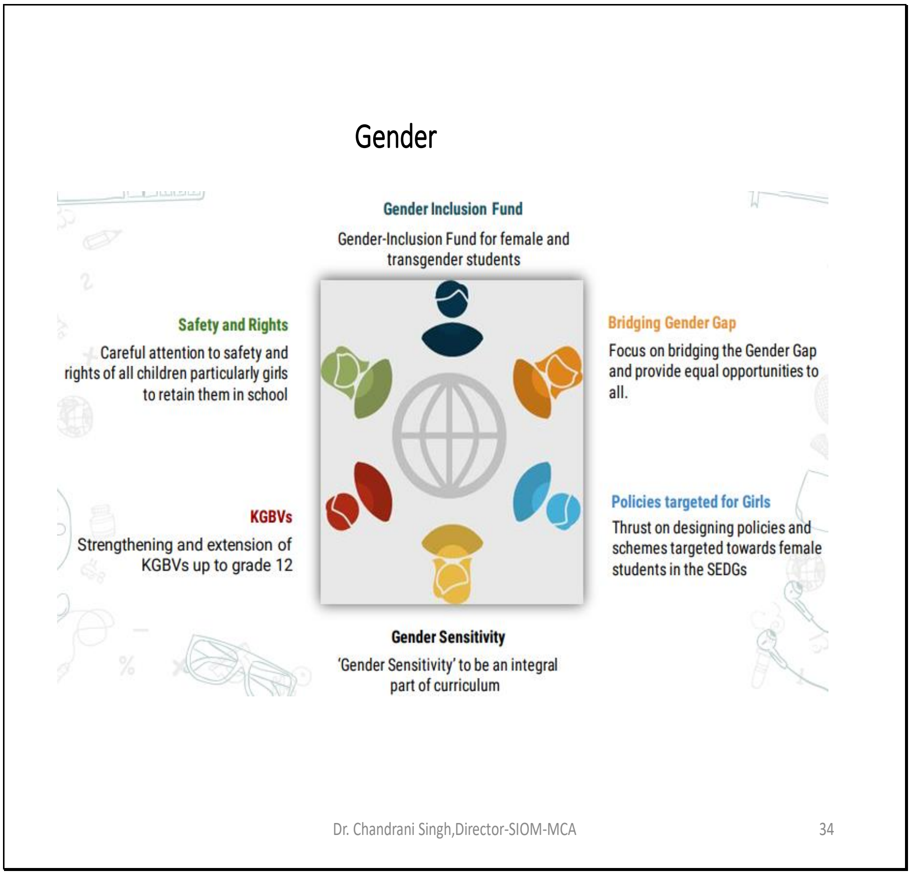
Recruitment of counsellors in schools

**Learning Outcomes**

Focus on attainment of learning outcomes of children belonging to SC/ST/OBC

**Additional Schools**

Setting-up of additional JNVs and KVS in aspirational districts/SEZs



## Gender

**Gender Inclusion Fund**

Gender-Inclusion Fund for female and transgender students

**Bridging Gender Gap**

Focus on bridging the Gender Gap and provide equal opportunities to all.

**Policies targeted for Girls**

Thrust on designing policies and schemes targeted towards female students in the SEDGs

**Gender Sensitivity**

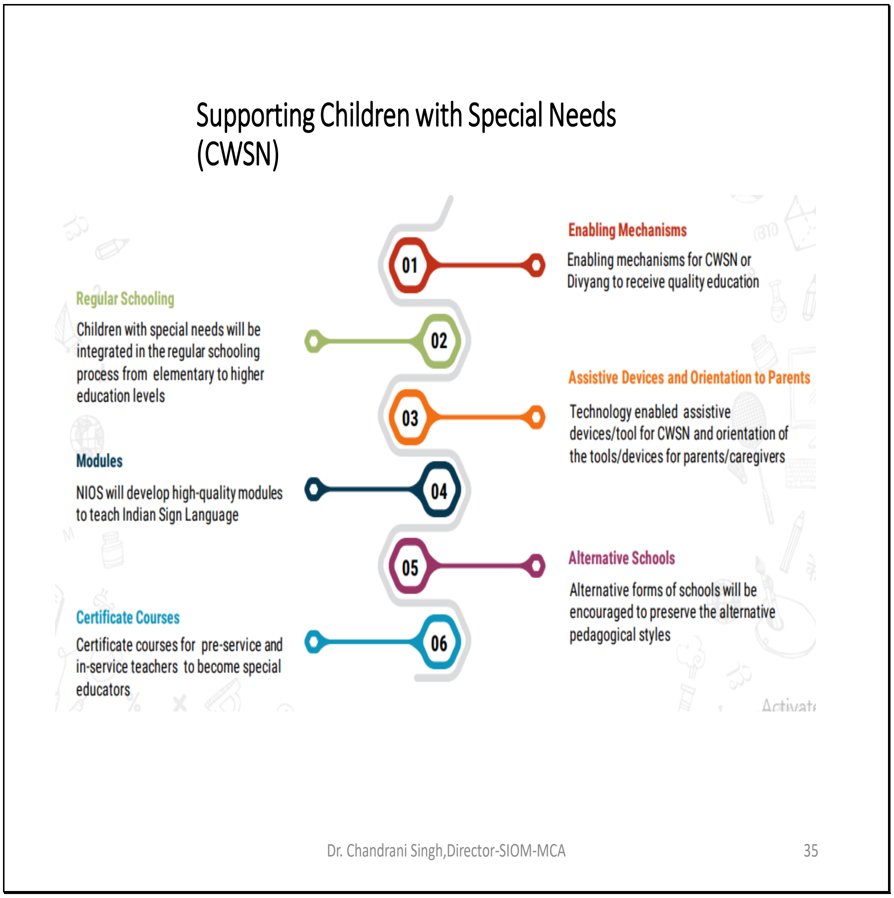
'Gender Sensitivity' to be an integral part of curriculum

**KGBVs**

Strengthening and extension of KGBVs up to grade 12

**Safety and Rights**

Careful attention to safety and rights of all children particularly girls to retain them in school



## Supporting Children with Special Needs (CWSN)

**Enabling Mechanisms**

Enabling mechanisms for CWSN or Divyang to receive quality education

**Regular Schooling**

Children with special needs will be integrated in the regular schooling process from elementary to higher education levels

**Assistive Devices and Orientation to Parents**

Technology enabled assistive devices/tool for CWSN and orientation of the tools/devices for parents/caregivers

**Modules**

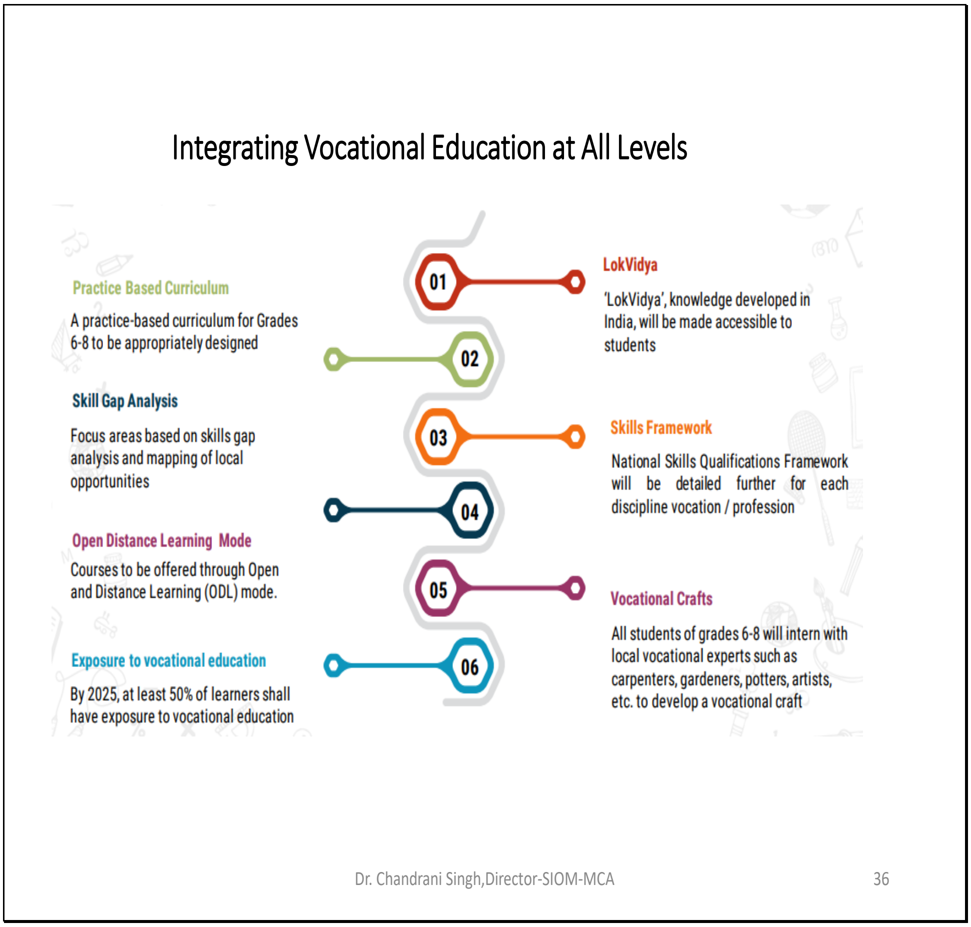
NIOS will develop high-quality modules to teach Indian Sign Language

**Alternative Schools**

Alternative forms of schools will be encouraged to preserve the alternative pedagogical styles

**Certificate Courses**

Certificate courses for pre-service and in-service teachers to become special educators



## Integrating Vocational Education at All Levels

01 **LokVidya**
'LokVidya', knowledge developed in India, will be made accessible to students

02 **Practice Based Curriculum**
A practice-based curriculum for Grades 6-8 to be appropriately designed

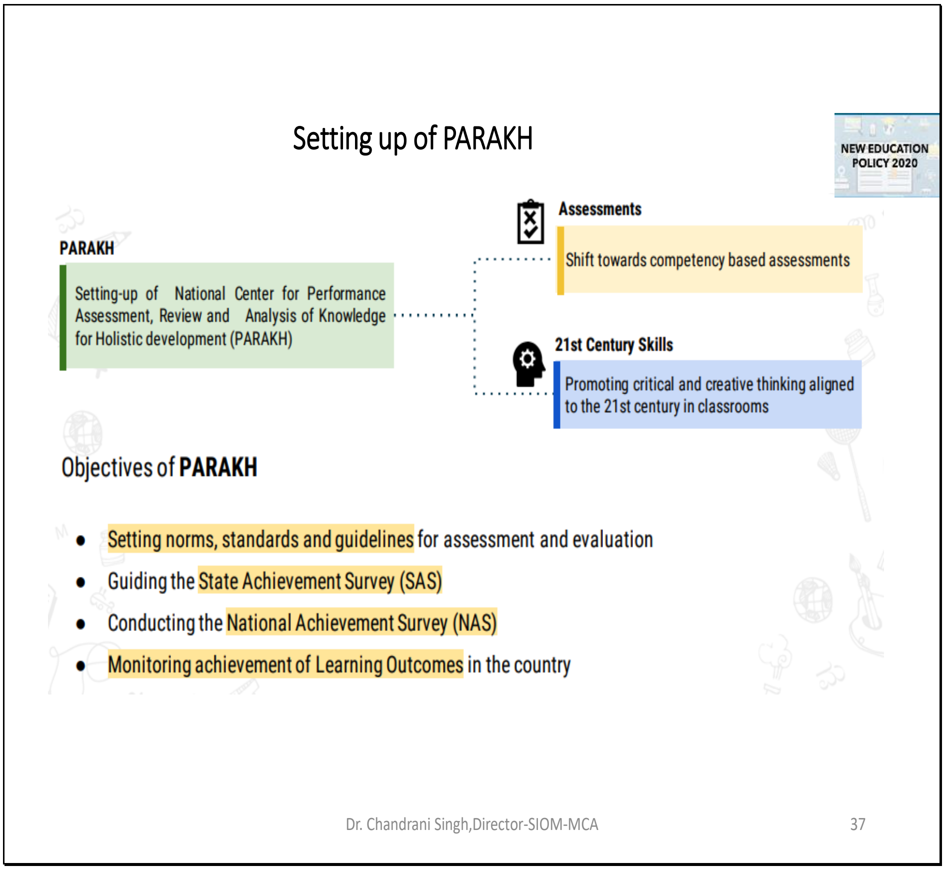
03 **Skills Framework**
National Skills Qualifications Framework will be detailed further for each discipline vocation / profession

04 **Skill Gap Analysis**
Focus areas based on skills gap analysis and mapping of local opportunities

05 **Vocational Crafts**
All students of grades 6-8 will intern with local vocational experts such as carpenters, gardeners, potters, artists, etc. to develop a vocational craft

05 **Open Distance Learning Mode**
Courses to be offered through Open and Distance Learning (ODL) mode.

06 **Exposure to vocational education**
By 2025, at least 50% of learners shall have exposure to vocational education



## Setting up of PARAKH

NEW EDUCATION POLICY 2020

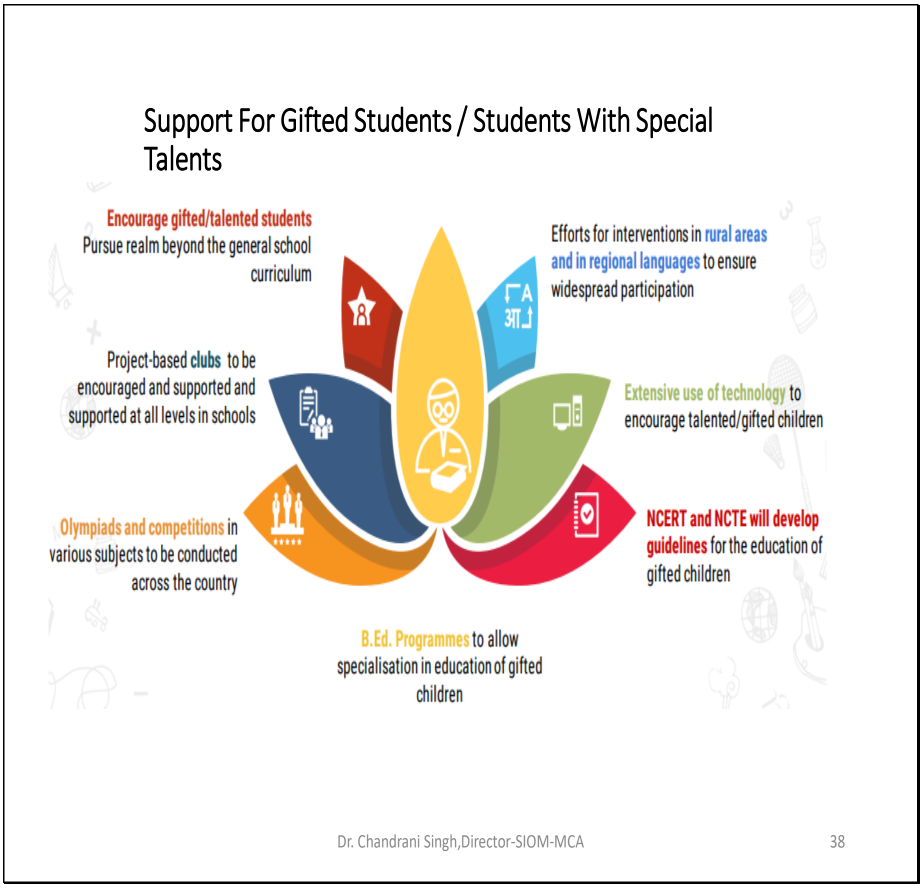
**PARAKH**
Setting-up of National Center for Performance Assessment, Review and Analysis of Knowledge for Holistic development (PARAKH)

**Assessments**
Shift towards competency based assessments

**21st Century Skills**
Promoting critical and creative thinking aligned to the 21st century in classrooms

### Objectives of **PARAKH**

- Setting norms, standards and guidelines for assessment and evaluation
- Guiding the State Achievement Survey (SAS)
- Conducting the National Achievement Survey (NAS)
- Monitoring achievement of Learning Outcomes in the country



## Support For Gifted Students / Students With Special Talents

**Encourage gifted/talented students** Pursue realm beyond the general school curriculum

Project-based **clubs** to be encouraged and supported and supported at all levels in schools

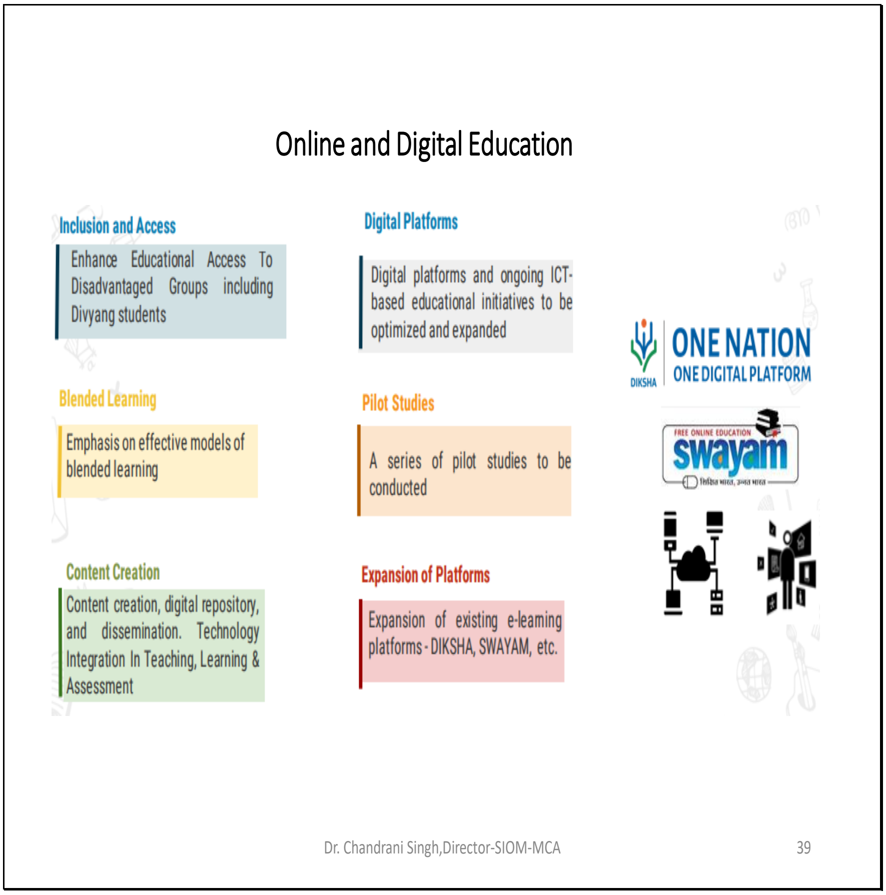
**Olympiads and competitions** in various subjects to be conducted across the country

Efforts for interventions in **rural areas and in regional languages** to ensure widespread participation

**Extensive use of technology** to encourage talented/gifted children

**NCERT and NCTE will develop guidelines** for the education of gifted children

**B.Ed. Programmes** to allow specialisation in education of gifted children



## Online and Digital Education

**Inclusion and Access**

Enhance Educational Access To Disadvantaged Groups including Divyang students

**Blended Learning**

Emphasis on effective models of blended learning

**Content Creation**

Content creation, digital repository, and dissemination. Technology Integration In Teaching, Learning & Assessment

**Digital Platforms**

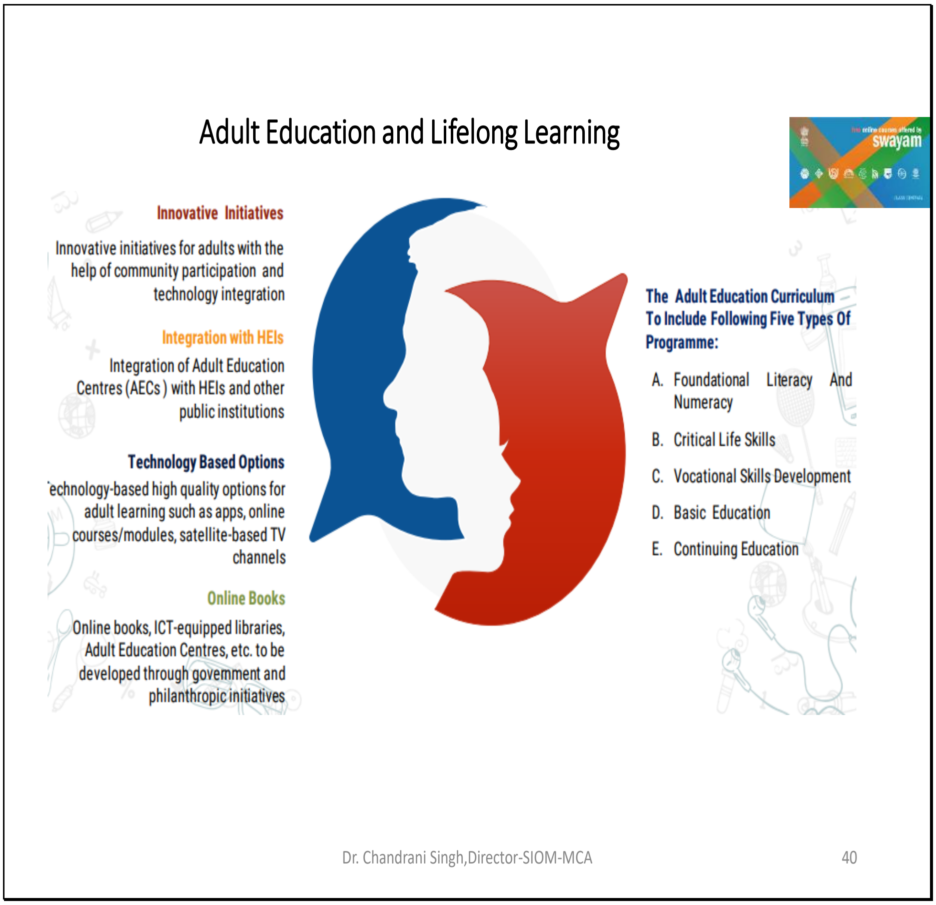
Digital platforms and ongoing ICT-based educational initiatives to be optimized and expanded

**Pilot Studies**

A series of pilot studies to be conducted

**Expansion of Platforms**

Expansion of existing e-learning platforms - DIKSHA, SWAYAM, etc.



## Adult Education and Lifelong Learning

**Innovative Initiatives**

Innovative initiatives for adults with the help of community participation and technology integration

**Integration with HEIs**

Integration of Adult Education Centres (AECs ) with HEIs and other public institutions

**Technology Based Options**

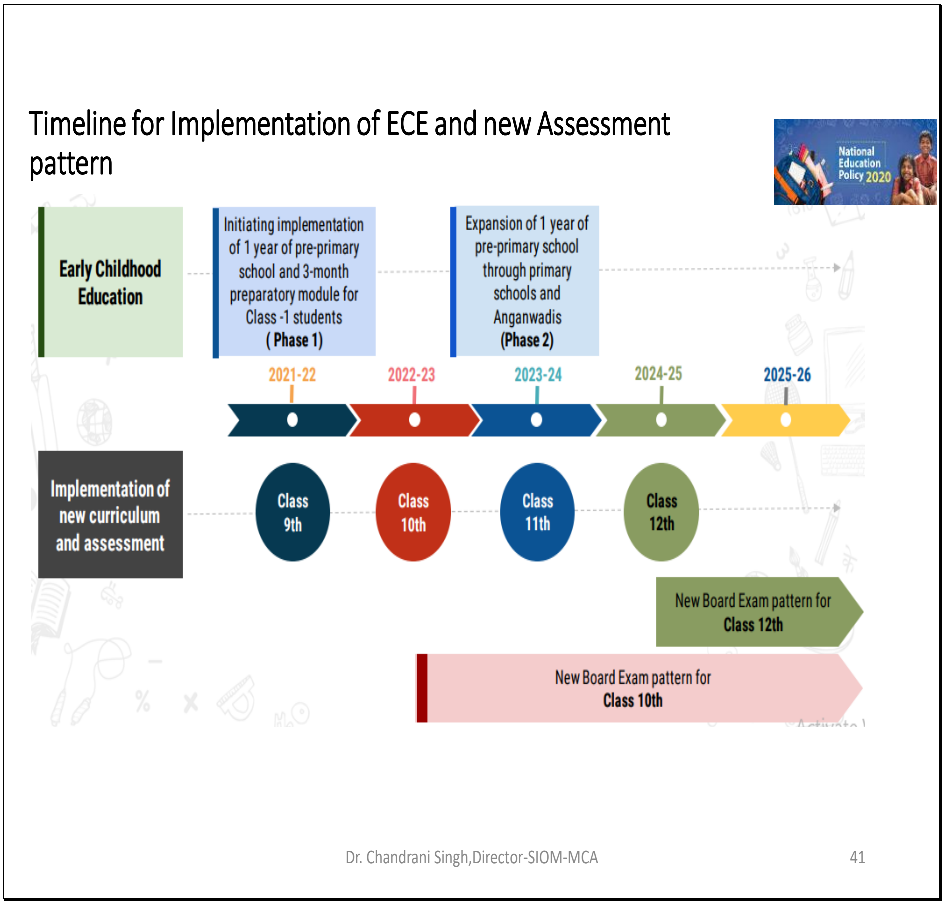
'echnology-based high quality options for adult learning such as apps, online courses/modules, satellite-based TV channels

**Online Books**

Online books, ICT-equipped libraries, Adult Education Centres, etc. to be developed through government and philanthropic initiatives

**The Adult Education Curriculum To Include Following Five Types Of Programme:**

A. Foundational Literacy And Numeracy
B. Critical Life Skills
C. Vocational Skills Development
D. Basic Education
E. Continuing Education



## Timeline for Implementation of ECE and new Assessment pattern

**Early Childhood Education**

Initiating implementation of 1 year of pre-primary school and 3-month preparatory module for Class -1 students **( Phase 1)**

Expansion of 1 year of pre-primary school through primary schools and Anganwadis **(Phase 2)**

2021-22 2022-23 2023-24 2024-25 2025-26

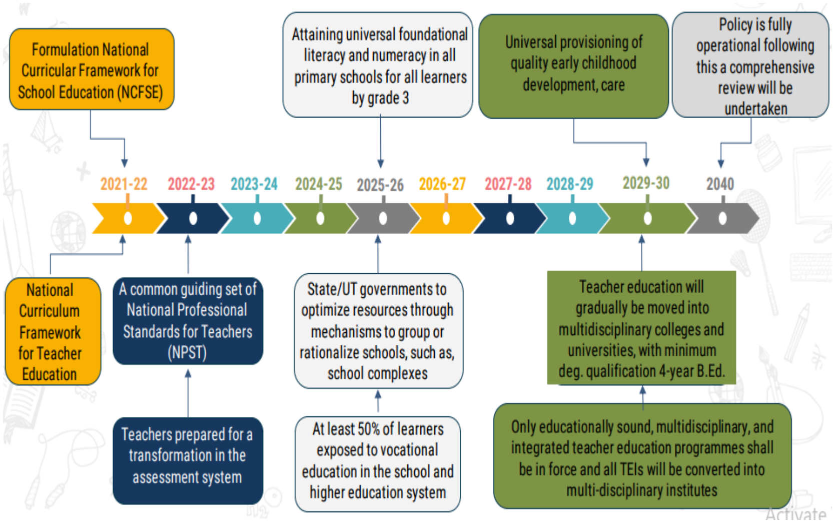
**Implementation of new curriculum and assessment**

Class 9th
Class 10th
Class 11th
Class 12th

New Board Exam pattern for **Class 12th**

New Board Exam pattern for **Class 10th**

# Timelines for Implementation in NEP 2020



Formulation National Curricular Framework for School Education (NCFSE)

Attaining universal foundational literacy and numeracy in all primary schools for all learners by grade 3

Universal provisioning of quality early childhood development, care

Policy is fully operational following this a comprehensive review will be undertaken

National Curriculum Framework for Teacher Education

A common guiding set of National Professional Standards for Teachers (NPST)

Teachers prepared for a transformation in the assessment system

State/UT governments to optimize resources through mechanisms to group or rationalize schools, such as, school complexes

At least 50% of learners exposed to vocational education in the school and higher education system

Teacher education will gradually be moved into multidisciplinary colleges and universities, with minimum deg. qualification 4-year B.Ed.

Only educationally sound, multidisciplinary, and integrated teacher education programmes shall be in force and all TEIs will be converted into multi-disciplinary institutes

# New Features of the Policy

## 1 Preparation for Schooling and Elementary Schooling Level

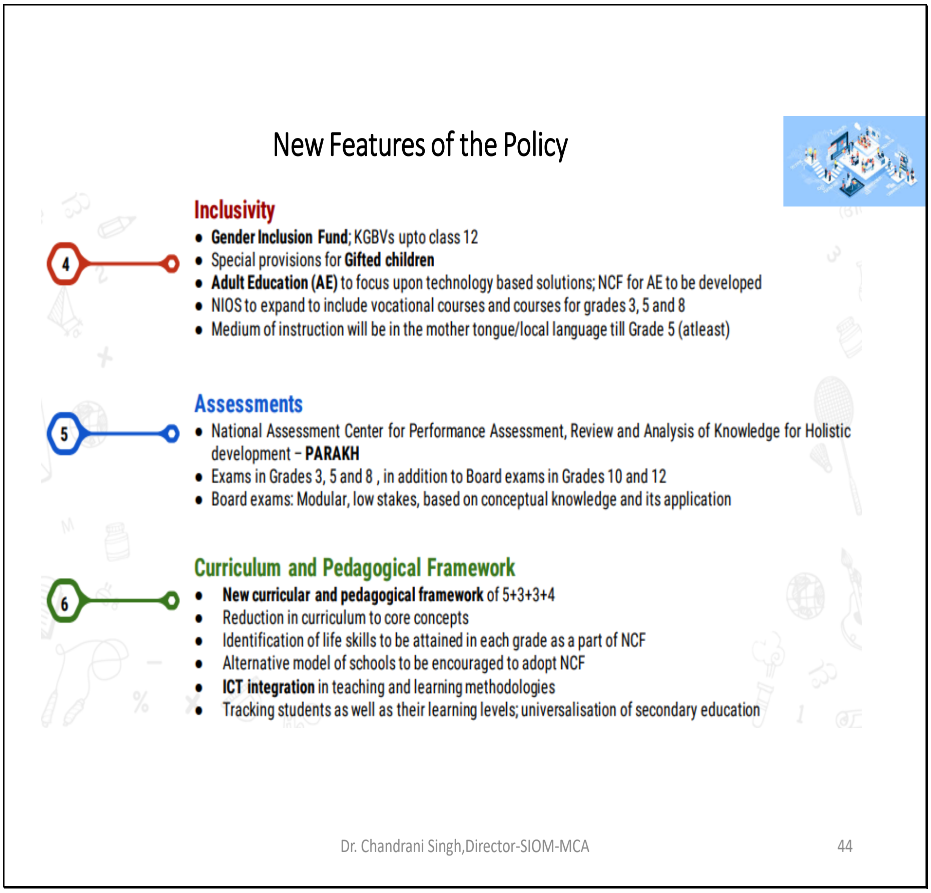
- ECCE for all by 2030: National Curriculum Framework for ECCE
- Achieve 100% Gross Enrolment Ratio in school education by 2030
- Preparatory class/**Balvatika** for 5-6 year old children in Anganwadis/pre-schools
- School Preparation module for all class 1 entrants
- **National Foundational Literacy and Numeracy Mission**
- Setup of Bal Bhavans

## 2 School Infrastructure and Resources

- Special Education Zones (SEZ)
- Utilize unused capacity of schools as Samajik Chetna Kendras
- School complex/clusters for resource sharing

## 3 Holistic Development of the Student

- No hard separation of curricular, extra and co-curricular, arts and science, sports and vocational crafts. Curriculum to integrate Indian culture and ethos
- **Innovative pedagogies** to be explored such as experiential teaching/learning methods
- Book promotion policy and digital libraries
- **Holistic Report card** – use AI for identifying specific aptitude of child
- **Vocational education** integration from primary grades and a ten days (no bag days) internship with local trades/craftsperson for Grades 6-8
- *Lok Vidya – local artists as master instructors in schools*



## New Features of the Policy

4

### Inclusivity

- **Gender Inclusion Fund**; KGBVs upto class 12
- Special provisions for **Gifted children**
- **Adult Education (AE)** to focus upon technology based solutions; NCF for AE to be developed
- NIOS to expand to include vocational courses and courses for grades 3, 5 and 8
- Medium of instruction will be in the mother tongue/local language till Grade 5 (atleast)

5

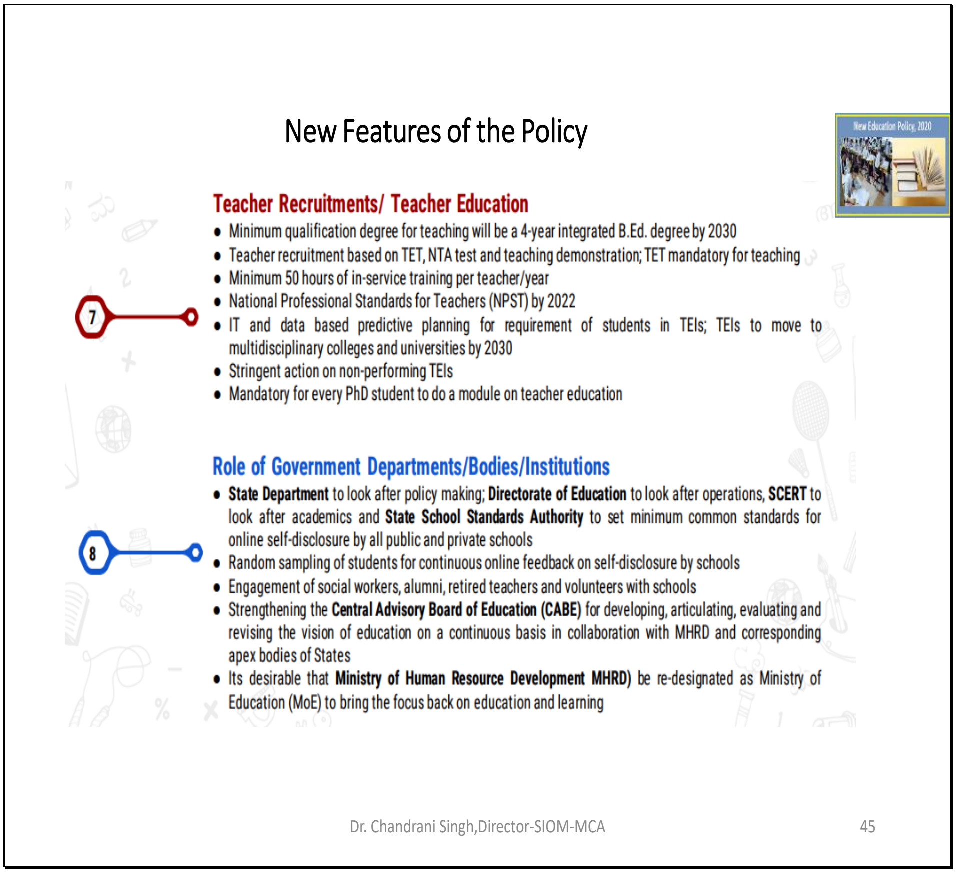
### Assessments

- National Assessment Center for Performance Assessment, Review and Analysis of Knowledge for Holistic development – **PARAKH**
- Exams in Grades 3, 5 and 8 , in addition to Board exams in Grades 10 and 12
- Board exams: Modular, low stakes, based on conceptual knowledge and its application

6

### Curriculum and Pedagogical Framework

- **New curricular and pedagogical framework** of 5+3+3+4
- Reduction in curriculum to core concepts
- Identification of life skills to be attained in each grade as a part of NCF
- Alternative model of schools to be encouraged to adopt NCF
- **ICT integration** in teaching and learning methodologies
- Tracking students as well as their learning levels; universalisation of secondary education



New Education Policy, 2020

## New Features of the Policy

7

### Teacher Recruitments/ Teacher Education

- Minimum qualification degree for teaching will be a 4-year integrated B.Ed. degree by 2030
- Teacher recruitment based on TET, NTA test and teaching demonstration; TET mandatory for teaching
- Minimum 50 hours of in-service training per teacher/year
- National Professional Standards for Teachers (NPST) by 2022
- IT and data based predictive planning for requirement of students in TEIs; TEIs to move to multidisciplinary colleges and universities by 2030
- Stringent action on non-performing TEIs
- Mandatory for every PhD student to do a module on teacher education

8

### Role of Government Departments/Bodies/Institutions

- **State Department** to look after policy making; **Directorate of Education** to look after operations, **SCERT** to look after academics and **State School Standards Authority** to set minimum common standards for online self-disclosure by all public and private schools
- Random sampling of students for continuous online feedback on self-disclosure by schools
- Engagement of social workers, alumni, retired teachers and volunteers with schools
- Strengthening the **Central Advisory Board of Education (CABE)** for developing, articulating, evaluating and revising the vision of education on a continuous basis in collaboration with MHRD and corresponding apex bodies of States
- Its desirable that **Ministry of Human Resource Development MHRD)** be re-designated as Ministry of Education (MoE) to bring the focus back on education and learning

# Thank You

**Python Library**

**Library:** The library is having a **collection of related functionality of codes**that allows you to perform many tasks without writing your code. It is a **reusable chunk of code** that we can use by importing it in our program, we can just use it by importing that **library and calling the method of that library with period(.)**.

**NumPY Library** NumPy is a Python library.

NumPy is used for working with arrays.

NumPy is short for "Numerical Python".

Installation of NumPy-pip install numpy

```
import numpy

arr = numpy.array([1, 2, 3, 4, 5])

print(arr)
```

```
    [1 2 3 4 5]
```

NumPy is used to work with arrays. The array object in NumPy is called ndarray.

create a NumPy ndarray object by using the array() function. To create an ndarray, we can pass a list, tuple or any array-like object into the array() method, and it will be converted into an ndarray

```
import numpy as np

arr = np.array((1, 2, 3, 4, 5))# create an array using tuple

print(arr)
```

```
    [1 2 3 4 5]
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5])#1-D array

print(arr)
import numpy as np

arr = np.array([[[1, 2, 3], [4, 5, 6]], [[1, 2, 3], [4, 5, 6]]])#3-D a
```

```
print(arr)
print(arr.ndim)#check no of dimensions
```

```
[1 2 3 4 5]
[[[1 2 3]
  [4 5 6]]

 [[1 2 3]
  [4 5 6]]]
3
```

```
import numpy as np

arr = np.array([1, 2, 3, 4])

print(arr[2] + arr[3])#accessing two elements and add them
import numpy as np

arr = np.array([[1,2,3,4,5], [6,7,8,9,10]])

print('5th element on 2nd dim: ', arr[1, 4])
```

```
7
5th element on 2nd dim:  10
```

```
import numpy as np

arr = np.array([[1,2,3,4,5], [6,7,8,9,10]])

print('Last element from 2nd dim: ', arr[1, -1])#last element from sec
```

```
Last element from 2nd dim:  10
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7])

print(arr[-3:-1])#slicing array
```

```
[5 6]
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7])

print(arr[1:5:2])#step slicing print every other element from the rang
```

```
[2 4]
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7])

print(arr[::2])#step slicing print every other element from the entire
```

```
[1 3 5 7]
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7])

print(arr[-3:-1])#negative slicing
```

```
[5 6]
```

```
import numpy as np

arr = np.array([[1, 2, 3, 4], [5, 6, 7, 8]])

print(arr.shape)#Print the shape of a 2-D array:
```

```
(2, 4)
```

Reshaping means changing the shape of an array.

The shape of an array is the number of elements in each dimension.

By reshaping we can add or remove dimensions or change number of elements in each dimension.

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12])

newarr = arr.reshape(4, 3)#1-D to 2-D

print(newarr)
```

```
[[ 1  2  3]
 [ 4  5  6]
 [ 7  8  9]
 [10 11 12]]
```

```
import numpy as np
```

```
arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12])

newarr = arr.reshape(2, 3, 2)#1-d to 3-d

print(newarr)
```

```
[[[ 1  2]
  [ 3  4]
  [ 5  6]]

 [[ 7  8]
  [ 9 10]
  [11 12]]]
```

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8])

newarr = arr.reshape(2, 2, -1)#1-d to 3-d

print(newarr)
```

```
[[[1 2]
  [3 4]]

 [[5 6]
  [7 8]]]
```

```
#Random module for random numbers
#Random number does NOT mean a different number every time. Random mea
#Random integer numbers
from numpy import random
x = random.randint(100)#Generate a random integer from 0 to 100
print(x)

from numpy import random
x=random.randint(100, size=(5))#Generate a 1-D array containing 5 ranc
print(x)

from numpy import random
x = random.randint(100, size=(3, 5))#Generate a 2-D array with 3 rows,
print(x)
```

```
88
[51 33 44  6 72]
[[86 47 77 53 41]
 [ 4 56 14 69 10]
 [35 54 54 69 45]]
```

```
#Random float numbers
#Generate a random float from 0 to 1
from numpy import random
x = random.rand()
print(x)

from numpy import random
x = random.rand(5)
print(x)

from numpy import random
x = random.rand(3, 5)
print(x)

from numpy import random
x = random.choice([3, 5, 7, 9])#The choice() method takes an array as
print(x)

from numpy import random
x = random.choice([3, 5, 7, 9], size=(3, 5))#2-D array
print(x)
```

```
0.8681724259898607
[0.16960403 0.49190983 0.7250351  0.44974501 0.13680861]
[[0.57528065 0.57083986 0.06665918 0.46740613 0.43777742]
 [0.78017591 0.32299414 0.13111266 0.58300421 0.47867802]
 [0.95904096 0.38082992 0.74753883 0.04835569 0.58241216]]
3
[[7 3 9 3 9]
 [7 9 9 5 3]
 [5 9 7 7 3]]
```

```
#The random number generator needs a number to start with (a seed valu
import random
random.seed(10)
print(random.random())
```

```
0.5714025946899135
```

**SCiPy Library**

SciPy is a scientific computation library that uses NumPy underneath.

SciPy stands for Scientific Python.

It provides more utility functions for optimization, stats and signal processing.

Like NumPy, SciPy is open source so we can use it freely.

SciPy has optimized and added functions that are frequently used in NumPy and Data Science.

pip install scipy

```
#SCIPY library
#contants module from scipy library
from scipy import constants

print(constants.pi)
```

```
3.141592653589793
```

```
#Optimizers are a set of procedures defined in SciPy that either find
#NumPy is capable of finding roots for polynomials and linear equatior
#it can not find roots for non linear equations like x + cos(x)
#For that you can use SciPy's optimze.root function.
#This function takes two required arguments
#fun - a function representing an equation.
#x0 - an initial guess for the root.
#The function returns an object with information regarding the solutic

from scipy.optimize import root
from math import cos

def eqn(x):
  return x + cos(x)

myroot = root(eqn, 0)

print(myroot.x)
```

```
[-0.73908513]
```

**Pandas Library**

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis"

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

Relevant data is very important in data science.

install-pip install pandas

```
#PANDAS Library
```

```
#A Pandas Series is like a column in a table.It is a one-dimensional a
import pandas as pd

a = [1, 7, 2] #labels by default starts With 0.

myvar = pd.Series(a)

print(myvar)
# create own label using index argument
import pandas as pd

a = [1, 7, 2]

myvar = pd.Series(a, index = ["x", "y", "z"])

print(myvar)
```

```
0    1
1    7
2    2
dtype: int64
x    1
y    7
z    2
dtype: int64
```

```
# A Pandas DataFrame is a 2 dimensional data structure, like a 2 dimer
#refer to the row index
import pandas as pd

data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}

#load data into a DataFrame object:
df = pd.DataFrame(data)

print(df)
print(df.loc[0])
#use a list of indexes
print(df.loc[[0, 1]])
#Add a list of names to give each row a name:
import pandas as pd

data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
  duration : [50, 40, 45]
}

df = pd.DataFrame(data, index = ["day1", "day2", "day3"])

print(df)
```

```
   calories  duration
0       420        50
1       380        40
2       390        45
calories    420
duration     50
Name: 0, dtype: int64
   calories  duration
0       420        50
1       380        40
      calories  duration
day1       420        50
day2       380        40
day3       390        45
```

```
#Load the CSV into a DataFrame:
import pandas as pd

df = pd.read_csv('data.csv')

print(df.to_string()) #use to_string() to print the entire DataFrame c
```

```
110        60   102       124     325.2
111        45   107       124     275.0
112        15   124       139     124.2
113        45   100       120     225.3
114        60   108       131     367.6
115        60   108       151     351.7
116        60   116       141     443.0
117        60    97       122     277.4
118        60   105       125       NaN
119        60   103       124     332.7
120        30   112       137     193.9
121        45   100       120     100.7
122        60   119       169     336.7
123        60   107       127     344.9
124        60   111       151     368.5
125        60    98       122     271.0
126        60    97       124     275.3
127        60   109       127     382.0
128        90    99       125     466.4
129        60   114       151     384.0
130        60   104       134     342.5
131        60   107       138     357.5
132        60   103       133     335.0
133        60   106       132     327.5
134        60   103       136     339.0
135        20   136       156     189.0
136        45   117       143     317.7
```

```
137        45    115       137     318.0
138        45    113       138     308.0
139        20    141       162     222.4
140        60    108       135     390.0
141        60     97       127       NaN
142        45    100       120     250.4
143        45    122       149     335.4
144        60    136       170     470.2
145        45    106       126     270.8
146        60    107       136     400.0
147        60    112       146     361.9
148        30    103       127     185.0
149        60    110       150     409.4
150        60    106       134     343.0
151        60    109       129     353.2
152        60    109       138     374.0
153        30    150       167     275.8
154        60    105       128     328.0
155        60    111       151     368.5
156        60     97       131     270.4
157        60    100       120     270.4
158        60    114       150     382.8
159        30     80       120     240.9
160        30     85       120     250.4
161        45     90       130     260.4
162        45     95       130     270.0
163        45    100       140     280.9
164        60    105       140     290.8
165        60    110       145     300.0
166        60    115       145     310.2
167        75    120       150     320.4
168        75    125       150     330.4
```

```
import pandas as pd

df = pd.read_csv('data.csv')

print(df.head())#Print the first 5 rows of the DataFrame
print(df.tail())#Print the last 5 rows of the DataFrame
print(df.info())
```

```
   Duration  Pulse  Maxpulse  Calories
0        60    110       130     409.1
1        60    117       145     479.0
2        60    103       135     340.0
3        45    109       175     282.4
4        45    117       148     406.0
     Duration  Pulse  Maxpulse  Calories
164        60    105       140     290.8
165        60    110       145     300.0
166        60    115       145     310.2
167        75    120       150     320.4
168        75    125       150     330.4
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 169 entries, 0 to 168
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
```

```
---  ------    --------------  -----
 0   Duration  169 non-null    int64
 1   Pulse     169 non-null    int64
 2   Maxpulse  169 non-null    int64
 3   Calories  164 non-null    float64
dtypes: float64(1), int64(3)
memory usage: 5.4 KB
None
```

```
#removing duplicates
import pandas as pd

df = pd.read_csv('data.csv')
print(df.info)
print(df.duplicated())
print(df.drop_duplicates(inplace = True))
```

```
<bound method DataFrame.info of      Duration  Pulse  Maxpulse  Calories
0          60    110       130     409.1
1          60    117       145     479.0
2          60    103       135     340.0
3          45    109       175     282.4
4          45    117       148     406.0
..        ...    ...       ...       ...
164        60    105       140     290.8
165        60    110       145     300.0
166        60    115       145     310.2
167        75    120       150     320.4
168        75    125       150     330.4

[169 rows x 4 columns]>
0      False
1      False
2      False
3      False
4      False
       ...
164    False
165    False
166    False
167    False
168    False
Length: 169, dtype: bool
None
```

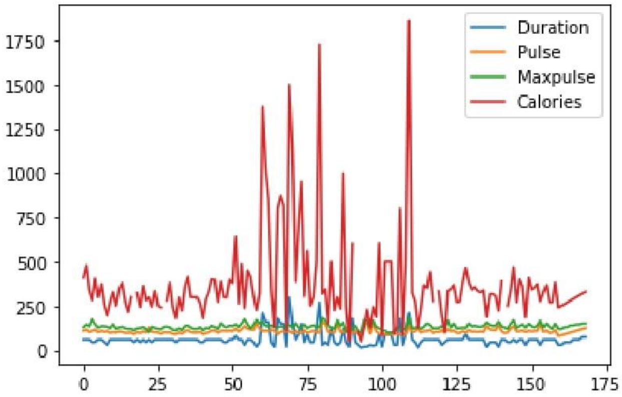
```
#correlation
#the corr() method calculates the relationship between each column in
#The corr() method ignores "not numeric" columns.
import pandas as pd

df = pd.read_csv('data.csv')
print(df.info)
df.corr()#good(0.922717), bad(.009403) and perfect coorealtion(1.00000
#three methods pearson, kendall and spearman method
```

```
#default method pearson
```

```
<bound method DataFrame.info of      Duration  Pulse  Maxpulse  Calories
0          60    110       130     409.1
1          60    117       145     479.0
2          60    103       135     340.0
3          45    109       175     282.4
4          45    117       148     406.0
..        ...    ...       ...       ...
164        60    105       140     290.8
165        60    110       145     300.0
166        60    115       145     310.2
167        75    120       150     320.4
168        75    125       150     330.4

[169 rows x 4 columns]>
```

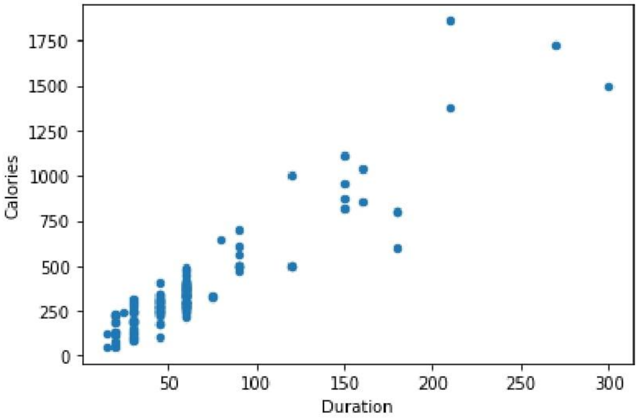
| | Duration | Pulse | Maxpulse | Calories |
|---|---|---|---|---|
| **Duration** | 1.000000 | -0.155408 | 0.009403 | 0.922717 |
| **Pulse** | -0.155408 | 1.000000 | 0.786535 | 0.025121 |
| **Maxpulse** | 0.009403 | 0.786535 | 1.000000 | 0.203813 |
| **Calories** | 0.922717 | 0.025121 | 0.203813 | 1.000000 |

```
#pandas plotting
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('data.csv')

df.plot()

plt.show()
```

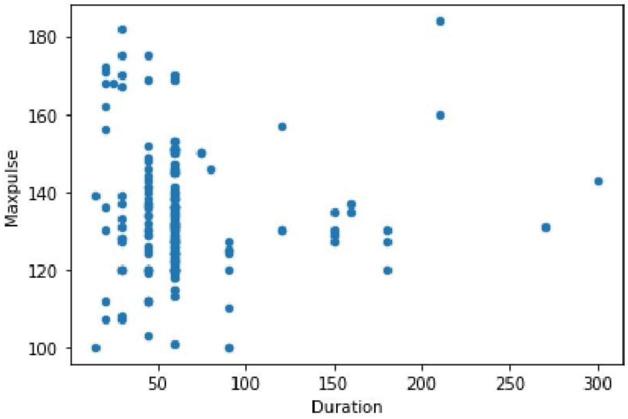


```
#scatter plot
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('data.csv')

df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')

plt.show()
```



```
#bad relationship can be seen using scatter plt
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('data.csv')

df.plot(kind = 'scatter', x = 'Duration', y = 'Maxpulse')

plt.show()
```

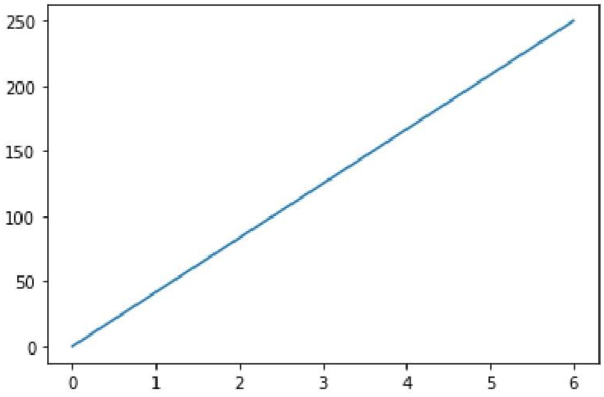


Matplotlib is a low level graph plotting library in python that serves as a visualization utility.

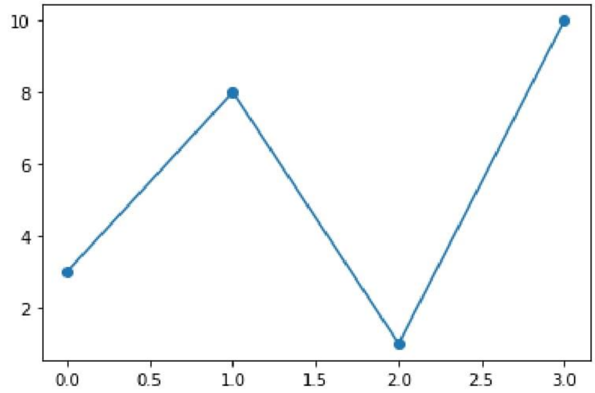
Matplotlib is open source and we can use it freely.

Installation-pip install matplotlib

```
#MATPLOTLIB library
import matplotlib.pyplot as plt
import numpy as np

xpoints = np.array([0, 6])
ypoints = np.array([0, 250])

plt.plot(xpoints, ypoints)#plot a line from position (0,0) to positior
plt.show()
```



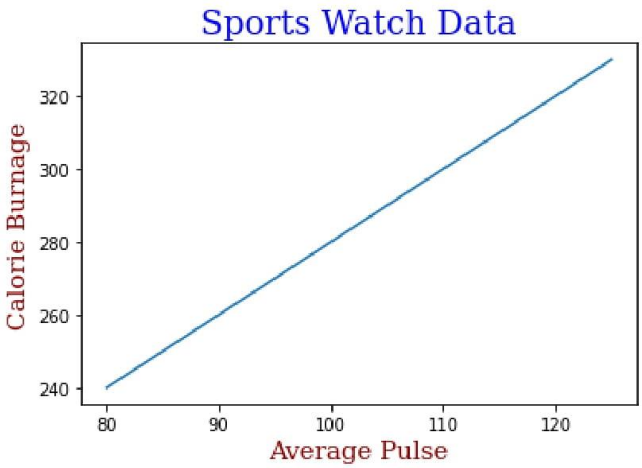
```
#plot with marker
import matplotlib.pyplot as plt
import numpy as np

ypoints = np.array([3, 8, 1, 10])

plt.plot(ypoints, marker = 'o')
plt.show()
plt.plot(ypoints, marker = '*')
plt.show()
```



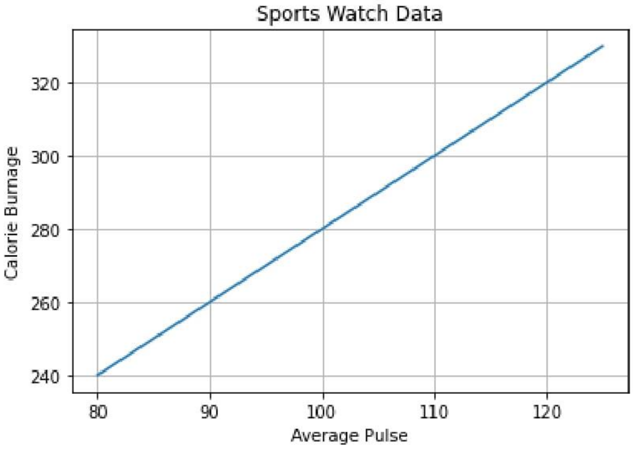
```
#linestyles
import matplotlib.pyplot as plt
import numpy as np

ypoints = np.array([3, 8, 1, 10])

plt.plot(ypoints, linestyle = 'dotted')
plt.show()
plt.plot(ypoints, linestyle = 'dashed')
plt.show()
```

```
#labels and title
import numpy as np
import matplotlib.pyplot as plt

x = np.array([80, 85, 90, 95, 100, 105, 110, 115, 120, 125])
y = np.array([240, 250, 260, 270, 280, 290, 300, 310, 320, 330])

plt.plot(x, y)
plt.title("Sports Watch Data") #Title
plt.xlabel("Average Pulse")     #x axis label
plt.ylabel("Calorie Burnage")   #y axis label

plt.show()

plt.title("Sports Watch Data", loc = 'left')#position of the title
plt.plot(x, y)
plt.xlabel("Average Pulse")     #x axis label
plt.ylabel("Calorie Burnage")   #y axis label

plt.show()
```

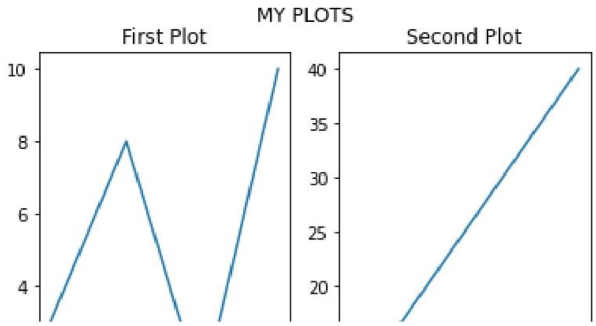
```
#set font property for label and title
import numpy as np
import matplotlib.pyplot as plt

x = np.array([80, 85, 90, 95, 100, 105, 110, 115, 120, 125])
y = np.array([240, 250, 260, 270, 280, 290, 300, 310, 320, 330])

font1 = {'family':'serif','color':'blue','size':20}
font2 = {'family':'serif','color':'darkred','size':15}

plt.title("Sports Watch Data", fontdict = font1)
plt.xlabel("Average Pulse", fontdict = font2)
plt.ylabel("Calorie Burnage", fontdict = font2)

plt.plot(x, y)
plt.show()
```



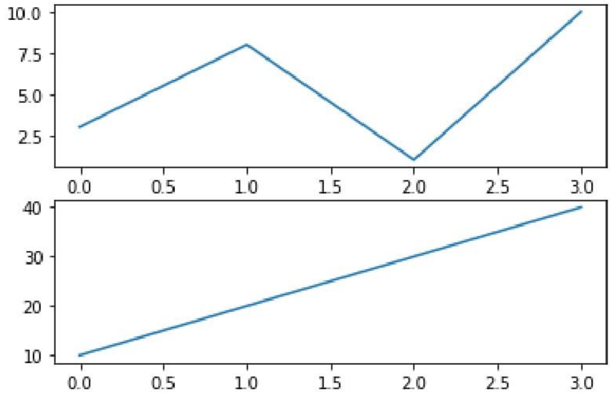
```
#add grid to plot
import numpy as np
import matplotlib.pyplot as plt

x = np.array([80, 85, 90, 95, 100, 105, 110, 115, 120, 125])
y = np.array([240, 250, 260, 270, 280, 290, 300, 310, 320, 330])

plt.title("Sports Watch Data")
plt.xlabel("Average Pulse")
plt.ylabel("Calorie Burnage")

plt.plot(x, y)
plt.plot(x, y)

plt.grid()

plt.show()
```



```
#subplots- multiple plots in one figure
import matplotlib.pyplot as plt
import numpy as np

#plot 1:
x = np.array([0, 1, 2, 3])
y = np.array([3, 8, 1, 10])

plt.subplot(1, 2, 1)#the figure has 1 row, 2 columns, and this plot is
plt.plot(x,y)
plt.title("First Plot")

#plot 2:
x = np.array([0, 1, 2, 3])
y = np.array([10, 20, 30, 40])

plt.subplot(1, 2, 2)#the figure has 1 row, 2 columns, and this plot is
plt.plot(x,y)
plt.title("Second Plot")

plt.suptitle("MY PLOTS")
plt.show()
```

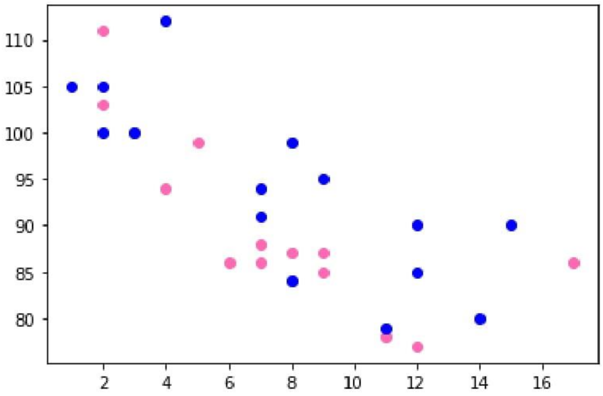


```
#subplots- multiple plots in one figure
import matplotlib.pyplot as plt
import numpy as np

#plot 1:
x = np.array([0, 1, 2, 3])
y = np.array([3, 8, 1, 10])

plt.subplot(2, 1, 1)#the figure has 2 rows, 1 column, and this plot is
plt.plot(x,y)

#plot 2:
x = np.array([0, 1, 2, 3])
y = np.array([10, 20, 30, 40])

plt.subplot(2, 1, 2)#the figure has 2 rows, 1 column, and this plot is
plt.plot(x,y)

plt.show()
```

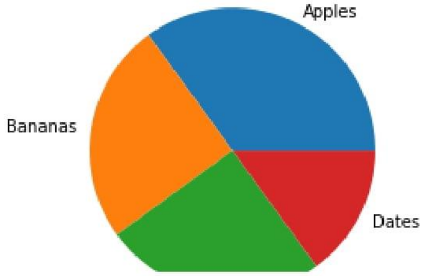


```
#scatter plots
import matplotlib.pyplot as plt
import numpy as np
```

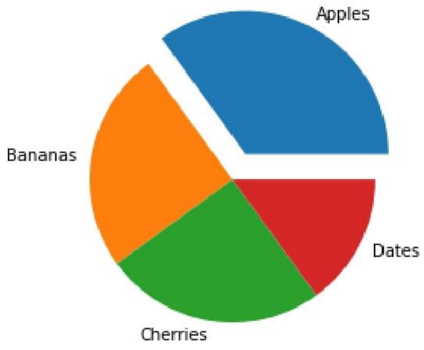
```
x = np.array([5,7,8,7,2,17,2,9,4,11,12,9,6])
y = np.array([99,86,87,88,111,86,103,87,94,78,77,85,86])
plt.scatter(x, y, color = 'hotpink')

x = np.array([2,2,8,1,15,8,12,9,7,3,11,4,7,14,12])
y = np.array([100,105,84,105,90,99,90,95,94,100,79,112,91,80,85])
plt.scatter(x, y, color = 'Blue')

plt.show()
```



```
#bar graphs
import matplotlib.pyplot as plt
import numpy as np

x = np.array(["A", "B", "C", "D"])
y = np.array([3, 8, 1, 10])

#plt.bar(x, y, color = "red")
#plt.bar(x, y, width = 0.1)
#plt.barh(x, y, height = 0.1)
plt.show()


#pie charts
import matplotlib.pyplot as plt
import numpy as np

y = np.array([35, 25, 25, 15])
mylabels = ["Apples", "Bananas", "Cherries", "Dates"]
plt.pie(y, labels = mylabels)
plt.show()
```



```
#exploded pie chart
import matplotlib.pyplot as plt
import numpy as np

y = np.array([35, 25, 25, 15])
mylabels = ["Apples", "Bananas", "Cherries", "Dates"]
myexplode = [0.2, 0, 0, 0]

plt.pie(y, labels = mylabels, explode = myexplode)
plt.show()
```



data.csv ✕

151 to 160 of 169 entries Filter

| Duration | Pulse | Maxpulse | Calories |
|---|---|---|---|
| 60 | 106 | 134 | 343 |
| 60 | 109 | 129 | 353.2 |
| 60 | 109 | 138 | 374 |
| 30 | 150 | 167 | 275.8 |
| 60 | 105 | 128 | 328 |
| 60 | 111 | 151 | 368.5 |
| 60 | 97 | 131 | 270.4 |
| 60 | 100 | 120 | 270.4 |
| 60 | 114 | 150 | 382.8 |
| 30 | 80 | 120 | 240.9 |

Show 10 per page 1 10 15 16 17

```
import numpy as np
import matplotlib.pyplot as mpt
import pandas as pd
```

```
data_set= pd.read_csv('data.csv
```

```
print(data_set)
```

```
     Duration  Pulse  Maxpulse
0          60    110       130
1          60    117       145
2          60    103       135
3          45    109       175
4          45    117       148
..        ...    ...       ...
164        60    105       140
165        60    110       145
166        60    115       145
167        75    120       150
168        75    125       150

[169 rows x 4 columns]
```

```
x= data_set.iloc[:,0:3].values
```

```
print(x)
```

```
 [ 60 102 124]
 [ 45 107 124]
 [ 15 124 139]
 [ 45 100 120]
 [ 60 108 131]
 [ 60 108 151]
 [ 60 116 141]
 [ 60  97 122]
 [ 60 105 125]
 [ 60 103 124]
 [ 30 112 137]
 [ 45 100 120]
 [ 60 119 169]
 [ 60 107 127]
 [ 60 111 151]
 [ 60  98 122]
 [ 60  97 124]
 [ 60 109 127]
 [ 90  99 125]
 [ 60 114 151]
 [ 60 104 134]
 [ 60 107 138]
 [ 60 103 133]
 [ 60 106 132]
 [ 60 103 136]
 [ 20 136 156]
 [ 45 117 143]
 [ 45 115 137]
 [ 45 113 138]
 [ 20 141 162]
 [ 60 108 135]
 [ 60  97 127]
 [ 45 100 120]
 [ 45 122 149]
 [ 60 136 170]
 [ 45 106 126]
 [ 60 107 136]
 [ 60 112 146]
 [ 30 103 127]
 [ 60 110 150]
 [ 60 106 134]
 [ 60 109 129]
 [ 60 109 138]
 [ 30 150 167]
 [ 60 105 128]
 [ 60 111 151]
 [ 60  97 131]
 [ 60 100 120]
 [ 60 114 150]
 [ 30  80 120]
 [ 30  85 120]
 [ 45  90 130]
 [ 45  95 130]
 [ 45 100 140]
 [ 60 105 140]
 [ 60 110 145]
 [ 60 115 145]
 [ 75 120 150]
 [ 75 125 150]]
```

```
y= data_set.iloc[:,2].values
```

```
print(y)
```

```
[130 145 135 175 148 127 136 134
 123 125 131 119 101 132 126 126
 127 120 120 129 112 126 122 138
 175 146 121 144 172 152 160 137
 127 127 146 125 134 141 130 131
 127 137 107 100 171 168 128 168
 120 184 124 124 139 120 131 151
 124 127 125 151 134 138 133 132
 170 126 136 146 127 150 134 129
 130 140 140 145 145 150 150]
```

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python.

It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

pip install scikit-learn

```
from sklearn.impute import SimpleImputer  #Imputation transform
imputer= SimpleImputer(missing_values =np.nan, strategy='mean')
#Fitting imputer object to the independent variables x.
imputerimputer1= imputer.fit(x[:, 1:3])
#Replacing missing data with the calculated mean value
x[:, 1:3]= imputer.transform(x[:, 1:3])
```

Double-click (or enter) to edit

```
print (x)
```

```
[ 60 102 124]
[ 45 107 124]
[ 15 124 139]
[ 45 100 120]
[ 60 108 131]
[ 60 108 151]
[ 60 116 141]
[ 60  97 122]
[ 60 105 125]
[ 60 103 124]
[ 30 112 137]
[ 45 100 120]
[ 60 119 169]
[ 60 107 127]
[ 60 111 151]
[ 60  98 122]
[ 60  97 124]
[ 60 109 127]
[ 90  99 125]
[ 60 114 151]
[ 60 104 134]
[ 60 107 138]
[ 60 103 133]
[ 60 106 132]
[ 60 103 136]
[ 20 136 156]
[ 45 117 143]
[ 45 115 137]
[ 45 113 138]
[ 20 141 162]
```

```
 [ 20 141 162]
 [ 60 108 135]
 [ 60  97 127]
 [ 45 100 120]
 [ 45 122 149]
 [ 60 136 170]
 [ 45 106 126]
 [ 60 107 136]
 [ 60 112 146]
 [ 30 103 127]
 [ 60 110 150]
 [ 60 106 134]
 [ 60 109 129]
 [ 60 109 138]
 [ 30 150 167]
 [ 60 105 128]
 [ 60 111 151]
 [ 60  97 131]
 [ 60 100 120]
 [ 60 114 150]
 [ 30  80 120]
 [ 30  85 120]
 [ 45  90 130]
 [ 45  95 130]
 [ 45 100 140]
 [ 60 105 140]
 [ 60 110 145]
 [ 60 115 145]
 [ 75 120 150]
 [ 75 125 150]]
```

```
#Catgorical data
#for Country Variable
data_set= pd.read_csv('data1.csv')
x= data_set.iloc[:,0:3].values
from sklearn.impute import SimpleImputer  #Imputation transform
imputer= SimpleImputer(missing_values =np.nan, strategy='mean')
#Fitting imputer object to the independent variables x.
imputerimputer1= imputer.fit(x[:, 1:3])
#Replacing missing data with the calculated mean value
x[:, 1:3]= imputer.transform(x[:, 1:3])
from sklearn.preprocessing import LabelEncoder
label_encoder_x= LabelEncoder()
x[:, 0]= label_encoder_x.fit_transform(x[:, 0])  #Fit label enc
```

```
print(x)
```

```
 [1 109.0 133.0]
 [2 98.0 124.0]
 [2 103.0 147.0]
 [2 100.0 120.0]
 [0 106.0 128.0]
 [1 104.0 132.0]
 [2 98.0 123.0]
 [2 98.0 120.0]
 [2 100.0 120.0]
 [2 90.0 112.0]
 [2 103.0 123.0]
 [2 97.0 125.0]
 [2 108.0 131.0]
 [0 100.0 119.0]
 [1 130.0 101.0]
 [2 105.0 132.0]
 [2 102.0 126.0]
 [2 100.0 120.0]
 [2 92.0 118.0]
 [2 103.0 132.0]
 [2 100.0 132.0]
 [2 102.0 129.0]
 [2 92.0 115.0]
 [2 90.0 112.0]
 [2 101.0 124.0]
 [2 93.0 113.0]
 [2 107.0 136.0]
 [2 114.0 140.0]
 [2 102.0 127.0]
 [2 100.0 120.0]
 [2 100.0 120.0]
 [2 104.0 129.0]
 [2 90.0 112.0]
 [2 98.0 126.0]
 [2 100.0 122.0]
 [2 111.0 138.0]
 [2 111.0 131.0]
 [2 99.0 119.0]
 [2 109.0 153.0]
 [2 111.0 136.0]
 [2 108.0 129.0]
 [2 111.0 139.0]
 [2 107.0 136.0]
 [4 123.0 146.0]
 [2 106.0 130.0]
 [2 118.0 151.0]
 [5 136.0 175.0]
 [2 121.0 146.0]
 [2 118.0 121.0]
 [2 115.0 144.0]
 [5 153.0 172.0]
 [2 123.0 152.0]
 [4 108.0 145.0]
 [3 110.0 137.0]
 [3 109.0 135.0]
 [2 118.0 141.0]
 [0 110.0 130.0]
 [0 90.0 130.0]
 [0 105.0 135.0]]
```

## Assignment # Regression Dt.02/07/2021

1. Apply linear regression on the following dataset.

   X = (1, 2, 3, 4, 5, 6, 7, 8, 8, 9)

   Y = (2, 3, 4, 5, 5, 6, 6, 7, 8, 9)

Calculate the following.

(a) Coefficient of Determination (b) Intercept (c) Slope

2. Please check whether linear regression is suitable here or not?

   x = [89,43,36,36,95,10,66,34,38,20,26,29,48,64,6,5,36,66,72,40]
   y = [21,46,3,35,67,95,53,72,58,10,26,34,90,33,38,20,56,2,47,15]

3. Apply the regressor.

   x = [1,2,3,5,6,7,8,9,10,12,13,14,15,16,18,19,21,22]
   y = [100,90,80,60,60,55,60,65,70,70,75,76,78,79,90,99,99,100]

==============================================================

**Q.1 Apply linear regression on the following dataset.**

**X = (1, 2, 3, 4, 5, 6, 7, 8, 8, 9)**

**Y = (2, 3, 4, 5, 5, 6, 6, 7, 8, 9)**

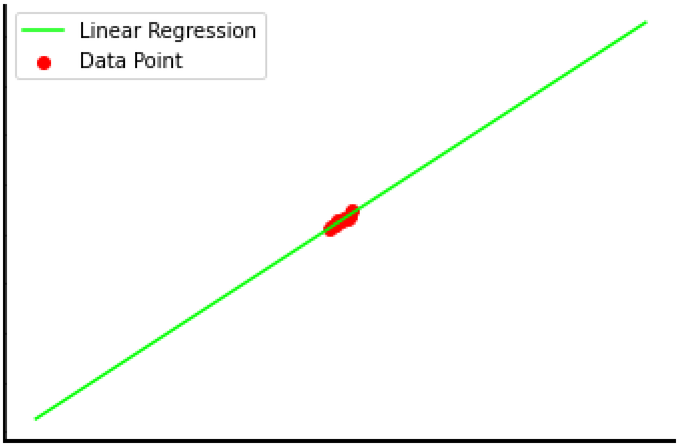
**Calculate the following.**

**a) Coefficient of Determination (b) Intercept (c) Slope**

Code & Output :

```
#import libraries

%matplotlib inline

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

X = (1, 2, 3, 4, 5, 6, 7, 8, 8, 9)

Y = (2, 3, 4, 5, 5, 6, 6, 7, 8, 9)

x_mean = np.mean(X)
```

```
y_mean = np.mean(Y)
n = len(X)
numerator = 0
denominator = 0
for i in range(n):
    numerator += (X[i] - x_mean) * (Y[i] - y_mean)
    denominator += (X[i] - x_mean) ** 2

b1 = numerator / denominator
b0 = y_mean - (b1 * x_mean)
#printing the coefficient
print(b1, b0)
#plotting values
x_max = np.max(X) + 100
x_min = np.min(X) - 100
#calculating line values of x and y
x = np.linspace(x_min, x_max, 1000)
y = b0 + b1 * x
#plotting line
plt.plot(x, y, color='#00ff00', label='Linear Regression')
#plot the data point
plt.scatter(X, Y, color='#ff0000', label='Data Point')
# x-axis label
plt.xlabel('Head Size (cm^3)')
#y-axis label
plt.ylabel('Brain Weight (grams)')
plt.legend()
plt.show()
```

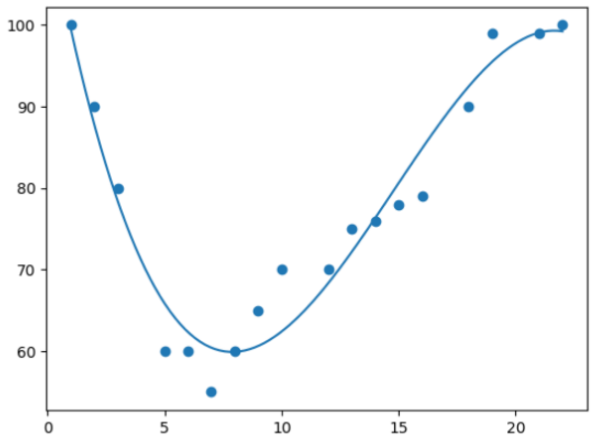


**Q.3 Apply the repressor.**

x = [1,2,3,5,6,7,8,9,10,12,13,14,15,16,18,19,21,22]
y = [100,90,80,60,60,55,60,65,70,70,75,76,78,79,90,99,99,100]

**Code and Output**

```
import numpy

import matplotlib.pyplot as plt

x = [1,2,3,5,6,7,8,9,10,12,13,14,15,16,18,19,21,22]

y = [100,90,80,60,60,55,60,65,70,70,75,76,78,79,90,99,99,100]


mymodel = numpy.poly1d(numpy.polyfit(x, y, 3))


myline = numpy.linspace(1, 22, 100)




plt.scatter(x, y)

plt.plot(myline, mymodel(myline))

plt.show()
```



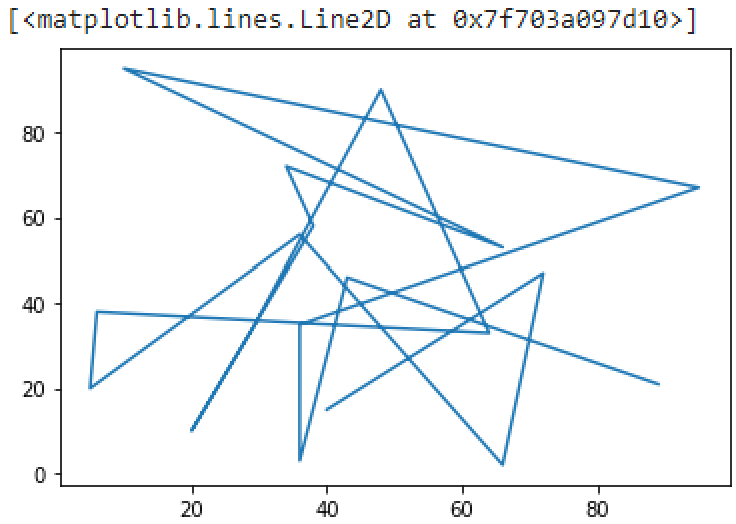
**Q-2 Please check whether linear regression is suitable here or not?**

**x = [89,43,36,36,95,10,66,34,38,20,26,29,48,64,6,5,36,66,72,40]**

**y = [21,46,3,35,67,95,53,72,58,10,26,34,90,33,38,20,56,2,47,15]**

Code & Output

```
# Step 1 import library

import numpy as np

from sklearn.linear_model import LinearRegression

from matplotlib import pyplot as plt


#Step 2: Provide data

x = np.array([89,43,36,36,95,10,66,34,38,20,26,29,48,64,6,5,36,66,72,40]).reshape((-1, 1))

y = np.array([21,46,3,35,67,95,53,72,58,10,26,34,90,33,38,20,56,2,47,15])


# Step 4 plot the graph

plt.plot(x,y)
```



```
[<matplotlib.lines.Line2D at 0x7f703a097d10>]
```

Out[4]:

```
[<matplotlib.lines.Line2D at 0x23da6f7b520>]
```

from the analysis of graph it is clear that linear regression is not suitable here.

In [ ]:

**Assignment # 2**

**Q1. Create a decision tree that can be used to decide if any new shows are worth attending to.**

**Dataset : Shows**

**We can use the Decision Tree to predict new values.**

- **(a)** **Example: Should I go see a show starring a 40 years old American comedian, with 10 years of experience, and a comedy ranking of 7?**
- **(b)** **What would the answer be if the comedy rank was 6?**

Q2. Calculate the number of clusters and plot the clusters.

Dataset: SUV_Data

**Q1. Create a decision tree that can be used to decide if any new shows are worth attending to.**

**Dataset: Shows**

**We can use the Decision Tree to predict new values.**

(a) Example: Should I go see a show starring a 40 years old American comedian, with 10 years of experience, and a comedy ranking of 7?

(b) What would the answer be if the comedy rank was 6?

➔

```
import pandas as pd

df = pd.read_csv('shows (2).csv')

X = df[['Age','Experience','Rank']]

y = df['Go']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.30,random_state=91)

from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()

model.fit(X_train,y_train)

model.score(X_test,y_test), model.score(X_train,y_train)

y_pre = model.predict([[40,10,7]])

y_pre

y_pre = model.predict([[40,10,6]])

y_pre
```

Q2. Calculate the number of clusters and plot the clusters.

Dataset: SUV_Data

→

```
import pandas as pd

import matplotlib.pyplot as pl

df = pd.read_csv('suv_data.csv')

X = df[['Age','EstimatedSalary']]

plt.scatter(X['Age'], X['EstimatedSalary'])
```

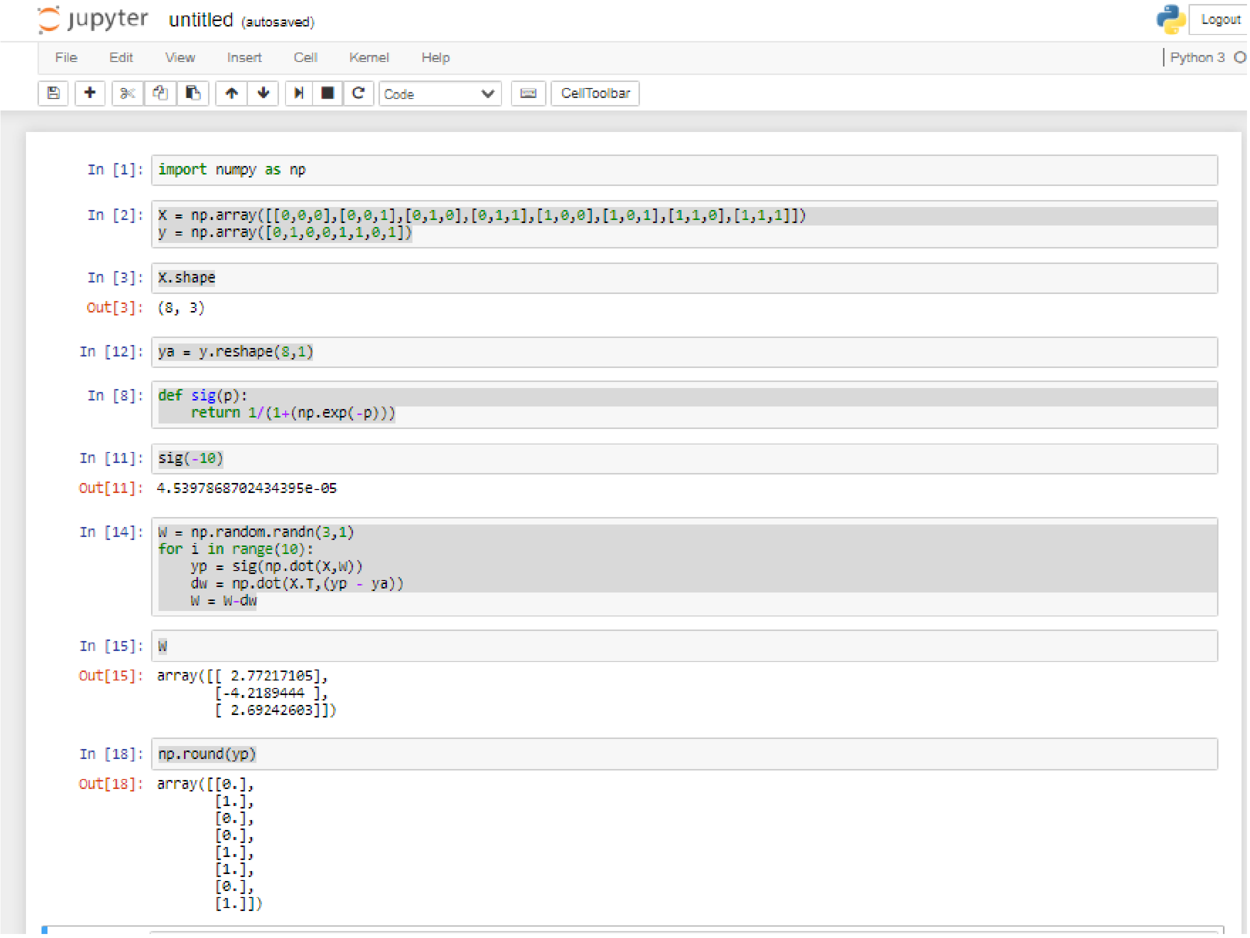
```
from sklearn.cluster import KMean
model = KMeans(n_clusters=5)
model.fit(X)
model.cluster_centers_
cluster_number = model.predict(X)
len(cluster_number)
c0 = X[cluster_number==0]
c1 = X[cluster_number==1]
c2 = X[cluster_number==2]
c3 = X[cluster_number==3]
c4 = X[cluster_number==4]
plt.scatter(c0['Age'], c0['EstimatedSalary'],c='red')
plt.scatter(c1['Age'], c1['EstimatedSalary'],c='blue')
plt.scatter(c2['Age'], c2['EstimatedSalary'],c='yellow')
plt.scatter(c3['Age'], c3['EstimatedSalary'],c='cyan')
plt.scatter(c4['Age'], c4['EstimatedSalary'],c='green')
```

## Artificial Neural Network (ANN)

### A) Gradient Decent Colab Code

```
Import the dataset
from tensorflow.keras.datasets import mnist

X  = np.array([[0,0,0],[0,0,1],[0,1,0],[0,1,1],[1,0,0],[1,0,1],[1,1,0],[1,1,1]])
y = np.array([0,1,0,0,1,1,0,1])

X.shape
ya = y.reshape(8,1)
def sig(p):
   return 1/(1+(np.exp(-p)))
sig(-10)
```

```
W = np.random.randn(3,1)
for i in range(10):
    yp = sig(np.dot(X,W))
    dw = np.dot(X.T,(yp - ya))
    W = W-dw
np.round(yp)
```



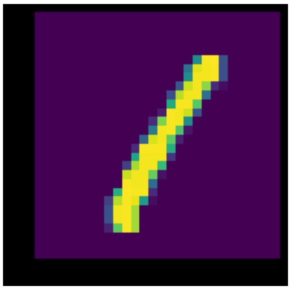
jupyter untitled (autosaved)
Logout
File Edit View Insert Cell Kernel Help
Python 3
Code
CellToolbar

```
In [1]: import numpy as np

In [2]: X = np.array([[0,0,0],[0,0,1],[0,1,0],[0,1,1],[1,0,0],[1,0,1],[1,1,0],[1,1,1]])
        y = np.array([0,1,0,0,1,1,0,1])

In [3]: X.shape

Out[3]: (8, 3)

In [12]: ya = y.reshape(8,1)

In [8]: def sig(p):
            return 1/(1+(np.exp(-p)))

In [11]: sig(-10)

Out[11]: 4.5397868702434395e-05

In [14]: W = np.random.randn(3,1)
         for i in range(10):
             yp = sig(np.dot(X,W))
             dw = np.dot(X.T,(yp - ya))
             W = W-dw

In [15]: W

Out[15]: array([[ 2.77217105],
                [-4.2189444 ],
                [ 2.69242603]])

In [18]: np.round(yp)

Out[18]: array([[0.],
                [1.],
                [0.],
                [0.],
                [1.],
                [1.],
                [0.],
                [1.]])
```

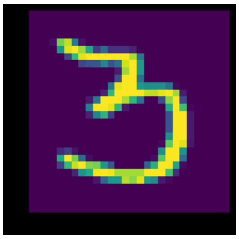
## B) Code: Data Training, Testing and Prediction – Model

```
# Import the dataset
from tensorflow.keras.datasets import mnist
(x_train, y_train), (x_test, y_test) = mnist.load_data()

x_train.shape
```

```
x_test.shape

y_train.shape

import matplotlib.pyplot as plt

img0 = x_train[3]

plt.imshow(img0)
```



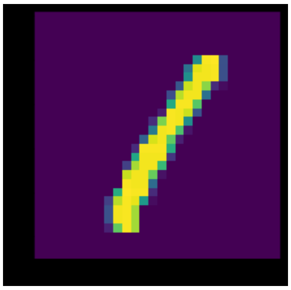
```
y_train[:4]

# Preprocessing -> Scaling /255

X_train = x_train/255

X_test  = x_test/255

# Preprocessing -> reshape images as flattened input

X_train = X_train.reshape(X_train.shape[0],X_train.shape[1]*X_train.shape[2]) # X_train.shape[1]*X_train.shape[2]

X_test = X_test.reshape(10000,784)

X_train.shape

y_train

# preprocessing -> y_train/y_test to categorical data

from tensorflow.keras.utils import to_categorical

y_train = to_categorical(y_train, num_classes=10)

y_test = to_categorical(y_test, num_classes=10)

y_train.shape

# Design the model  50-50-10

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

model = Sequential()

model.add(Dense(50, activation='relu', input_shape=(784,)))

model.add(Dense(50,activation='relu'))

model.add(Dense(10, activation='softmax'))
model.compile(loss='categorical_crossentropy', metrics=['accuracy'])
model.fit(X_train, y_train, batch_size=64, epochs=10, validation_data=(X_test,y_test))
# its a prediction time?????
img0 = x_test[780]
plt.imshow(img0)
```



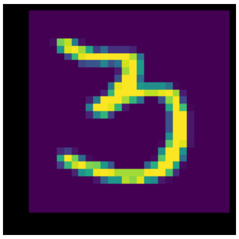
```
# preprocess the input
img = img0/255
img = img.reshape(1,784)
model.predict(img)

def get_number(img):
    img = img/255
    img = img.reshape(1,784)
    return model.predict(img).argmax()
get_number(x_test[78])
import cv2

A = cv2.imread('/content/8.jpg')
A.shape
```

**Assignments:**

1. Gray scale
2. resize 28 X 28
3. get_number



```
y_train[:4]

# Preprocessing -> Scaling /255

X_train = x_train/255

X_test  = x_test/255

# Preprocessing -> reshape images as flattened input

X_train = X_train.reshape(X_train.shape[0],X_train.shape[1]*X_train.shape[2]) # X_train.shape[1]*X_train.shape[2]

X_test = X_test.reshape(10000,784)

X_train.shape

y_train

# preprocessing -> y_train/y_test to categorical data

from tensorflow.keras.utils import to_categorical

y_train = to_categorical(y_train, num_classes=10)

y_test = to_categorical(y_test, num_classes=10)

y_train.shape

# Design the model  50-50-10

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

model = Sequential()

model.add(Dense(50, activation='relu', input_shape=(784,)))

model.add(Dense(50,activation='relu'))

model.add(Dense(10, activation='softmax'))

model.compile(loss='categorical_crossentropy', metrics=['accuracy'])

model.fit(X_train, y_train, batch_size=64, epochs=10, validation_data=(X_test,y_test))

# its a prediction time?????

img0 = x_test[780]

plt.imshow(img0)
```



```
# preprocess the input
img = img0/255
img = img.reshape(1,784)
model.predict(img)

def get_number(img):
    img = img/255
    img = img.reshape(1,784)
    return model.predict(img).argmax()
get_number(x_test[78])
import cv2

A = cv2.imread('/content/8.jpg')
A.shape
```

**Assignments:**

1. Gray scale
2. resize 28 X 28
3. get_number