

STUDENT BEHAVIOUR ANALYSIS USING UNSUPERVISED APPROACH

A MAJOR PROJECT REPORT

Submitted by

G.SAI PRAMIDHA

G.SVETHALINA

K.NAGA SOWMYA

18RH1A0476

18RH1A0479

18RH1A04B6

Under the Esteemed Guidance of

Mr. S.SRIKANTH
Assistant Professor of ECE

in partial fulfillment of the Academic Requirements for the Degree of

BACHELOR OF TECHNOLOGY

Electronics & Communication Engineering



MALLA REDDY ENGINEERING COLLEGE FOR WOMEN
(Autonomous Institution-UGC, Govt. of India)

Accredited by NBA & NAAC with 'A' Grade

NIRF Indian Ranking, Accepted by MHRD, Govt. of India | Rank Band (6th-25th) by ARIIA, Accepted by MHRD, Govt. of India
Approved by AICTE, Affiliated to JNTUH, ISO 9001:2015 Certified Institution

Platinum Rated by AICTE-CII Survey, AAAA+ Rated by Digital Learning Magazine, AAA+ Rated by Careers 360, National Ranking-Top 100 Rank band by Outlook Magazine,
3rd Rank by CSR, National Ranking-Top 100 Rank band by Times News Magazine, 141 Rank by India Today-Best Engineering Colleges of India Rankings-2020.
Maisammaguda, Dhulapally, Secunderabad, Kompally-500100.

2021-22



MALLA REDDY ENGINEERING COLLEGE FOR WOMEN

(Autonomous Institution-UGC, Govt. of India)

Accredited by NBA & NAAC with 'A' Grade

NIRF Indian Ranking, Accepted by MHRD, Govt. of India / Rank Band (6th-25th) by ARIIA, Accepted by MHRD, Govt. of India
Approved by AICTE, Affiliated to JNTUH, ISO 9001:2015 Certified Institution

Platinum Rated by AICTE-CII Survey, AAAA+ Rated by Digital Learning Magazine, AAA+ Rated by Careers 360, 3rd Rank by CSR,
National Ranking-Top 100 Rank band by Outlook Magazine, , National Ranking-Top 100 Rank band by Times News Magazine,
141 Rank by India Today-Best Engineering Colleges of India Rankings-2020.
Maisammaguda, Dhulapally, Secunderabad, Kompally-500100.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that the Mini Project work entitled "**STUDENT BEHAVIOUR ANALYSIS USING UNSUPERVISED APPROACH**" is carried out by **G.Sai Pramidha (18RH1A0476), G.Svethalina (18RH1A0479), K.Naga Sowmya (18RH1A04B6)** in partial fulfillment for the award of degree of **BACHELOR OF TECHNOLOGY** in Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Hyderabad during the academic year 2021-2022.

SUPERVISOR'S SIGNATURE

Mr. S. SRIKANTH
Assistant Professor of ECE

HEAD OF THE DEPARTMENT

Dr. N.SREEKANTH
Professor of ECE

EXTERNAL EXAMINER

PANHYD29/MAJOR PROJECT/2021-2022

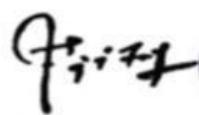
COMPLETION CERTIFICATE
TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Mr./Ms.G.saipramidha**, Roll Number – **(18RH1A0476)**, who is pursuing **Electronics and Communication Engineering**, Department at **Malla Reddy College of Engineering for Women** has successfully completed his/her Major Project at, **Pantech E Learning Pvt Limited** on (“**Student behaviour analysis using unsupervised approach**”) and has submitted the report.

During the Major Project period, the candidate has shown keen interest and commitment towards learning and his/her performance was good.

Yours truly,

Pantech E Learning Pvt.Ltd.



Srinivasan.N

(Branch Manager)

PANHYD27/MAJOR PROJECT/2021-2022

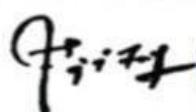
COMPLETION CERTIFICATE
TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Mr./Ms.G. Svethalina**, Roll Number – **(18RH1A0479)**, who is pursuing **Electronics and Communication Engineering**, Department at **Malla Reddy College of Engineering for Women** has successfully completed his/her Major Project at, **Pantech E Learning Pvt Limited** on (“**Student behaviour analysis using unsupervised approach**”) and has submitted the report.

During the Major Project period, the candidate has shown keen interest and commitment towards learning and his/her performance was good.

Yours truly,

Pantech E Learning Pvt.Ltd.



Srinivasan.N

(Branch Manager)

PANHYD28/MAJOR PROJECT/2021-2022

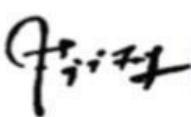
COMPLETION CERTIFICATE**TO WHOMSOEVER IT MAY CONCERN**

This is to certify that **Mr./Ms.K.Naga Sowmya**, Roll Number – **(18RH1A04B6)**, who is pursuing **Electronics and Communication Engineering**, Department at **Malla Reddy College of Engineering for Women** has successfully completed his/her Major Project at, **Pantech E Learning Pvt Limited** on (“**Student behaviour analysis using unsupervised approach**”) and has submitted the report.

During the Major Project period, the candidate has shown keen interest and commitment towards learning and his/her performance was good.

Yours truly,

Pantech E Learning Pvt.Ltd.,



Srinivasan.N

(Branch Manager)

Pantech E Learning Pvt Ltd

4th Floor, Delta Chambers,

Behind Chennai Shopping Mall, Ameerpet, Hyderabad, Telangana – 500 016

Phone:91 040-40077960. | hr@pantechmail.com

ACKNOWLEDGMENT

We feel ourselves honored and privileged to place our warm salutation to our college **Malla Reddy Engineering College for Women** and Department of **Electronics and Communication Engineering** which gave us the opportunity to have expertise in engineering and profound technical knowledge.

We would like to deeply thank our Honorable Minister of Telangana State **Sri.Ch. Malla Reddy Garu**, founder chairman MRGI, the largest cluster of institutions in the state of Telangana for providing us with all the resources in the college to make our project success.

We wish to convey gratitude to our Principal **Dr. Y. Madhavee Latha** for providing us with the environment and mean to enrich our skills and motivating us in our endeavor and helping us to realize our full potential.

We express our sincere gratitude to **Dr. N.Sreekanth**, Head of the Department of Electronics and Communication Engineering for inspiring us to take up a project on this subject and successfully guiding us towards its completion.

We would also like to thank our Project coordinator **Mr. A. Bhanu Chandar**, for his kind encouragement and overall guidance in viewing this program a good asset with profound gratitude.

We would like to thank our internal guide **Mr. S. Srikanth**, and all the Faculty members for their valuable guidance and encouragement towards the completion of our project work.

With Regards and Gratitude

G.Pramidha(18RH1A0476)

G.Svethalina(18RH1A0479)

K.Sowmya(18RH1A04B6)

DECLARATION

We hereby declare that our mini project entitled "**STUDENT BEHAVIOUR ANALYSIS USING UNSUPERVISED APPROACH**" submitted to **Malla Reddy Engineering College for Women**, affiliated to **Jawaharlal Nehru Technological University, Hyderabad** for the award of the degree of Bachelor of Technology in **Electronics and Communication Engineering** is a result of original research work done by us.

It is declared that the project report or any part thereof has not been previously submitted to any University or Institute for the award of degree.

G.Sai Pramidha(18RH1A0476)

G.Svethalina(18RH1A0479)

K.Sowmya(18RH1A04B6)

ABSTRACT

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. This paper proposes a complete EDM framework in a form of a rule based recommender system that is not developed to analyze and predict the student's performance only, but also to exhibit the reasons behind it. The proposed framework analyzes the students' demographic data, study related and psychological characteristics to extract all possible knowledge from students, teachers and parents. Seeking the highest possible accuracy in academic performance prediction using a set of powerful data mining techniques. The framework succeeds to highlight the student's weak points and provide appropriate recommendations. The realistic case study that has been conducted on 200 students proves the outstanding performance of the proposed framework in comparison with the existing ones.

TABLE OF CONTENTS

TITLE	PAGE.NO
CHAPTER 1: INTRODUCTION	1-7
1.1 Introduction to Domain	1
1.2 Introduction to Project	2
1.3 Literature Review	3
1.4 Existing System	4
1.5 Proposed System	4
1.5.1 Block Diagram	5
1.5.2 Requirements	6
1.5.3 Flowchart	6
1.5.4 Principle of Operation	7
1.5.5 Advantages	7
1.5.6 Applications	7
CHAPTER 2: MACHINE LEARNING	8-20
2.1 Machine Learning	8
2.2 Software Development Life Cycle	19
2.3 Applications of Machine learning	20
CHAPTER 3: SOFTWARE	24-31
3.1 Python	24
3.2 Anaconda Navigator	27
3.3 NumPy	28
3.4 VNC Viewer	29

CHAPTER 4: HARDWARE	32-36
4.1 Raspberry pi	32
CHAPTER 5: RESULT ANALYSIS	37-45
5.1 Working Principle	37
5.2 Result	39
CONCLUSION	46
FUTURE SCOPE	47
REFERENCES	

TABLE OF FIGURES

TITLE	PAGE.NO
1.5.1 Block Diagram	5
1.5.3 Flow Chart	6
2.1.1 Types of Machine Learning	8
2.1.2 Pictorial representation of ML	9
2.1.3 Building Model of ML	11
2.1.4 Algorithms of ML	15
2.1.5 SVM	16
2.1.6 KNN	17
2.1.7 Naive Bayes	18
2.2 Software Development Life Cycle	19
2.3 Applications of ML	21
3.1 Features of Python	26
3.3 NumPy	28
3.4 VNC Viewer	31
4.1.1 Raspberry pi	32
4.1.2 Power supply of raspberry pi	34
4.1.3 Circuit connections	35
4.1.4 Pin diagram with connections	35
4.1.5 GPIO port connections	36

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO DOMAIN

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

- **Supervised learning:** In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
- **Unsupervised learning:** This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.
- **Semi-supervised learning:** This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.
- **Reinforcement learning:** Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

1.2 INTRODUCTION TO PROJECT

Data mining in higher education is a new emerging field. Educational Data Mining (EDM) is concerned with developing the methods of exploring the unique types of data that come from educational settings. Through EDM the education has been benefited from a real E-Revolution as Virtual Learning Environment (VLE). The Internet also helped e-learning through the resources available for both teachers and students to share information. The goals of a teacher have changed from teaching facts into helping students to learn and how to find relevant information, how to access it, how to organize in different ways. Meanwhile, learning has moved towards more student-centre, problem-based, challenge-based or cooperative learning. In traditional learning, the teacher is instructed to teach or facilitates all the sessions in classroom. In this the teacher usually talks more than the student. The teacher teaches the lesson according to the study program with the prescribed curriculum. In traditional learning the teachers take more effort than students, which had become unsuitable to today's curriculums. In Virtual Learning Environment, the relationship between the teacher and the learner is more effective through the beloved categories. Learning achievement Students in the VLE achieve better learning performance than in the traditional environment. Self-efficacy Students in the VLE report higher levels of computer self-efficacy than there in the traditional environment .Satisfaction students in the VLE report higher levels of satisfaction than students in the traditional environment. Learning methodology students in the VLE report higher levels of learning climate than there in the traditional environment.

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. This paper proposes a complete EDM framework in a form of a rule based recommender system that is not developed to analyze and predict the student's performance only, but also to exhibit the reasons behind it. The proposed framework analyzes the students' demographic data, study related and psychological characteristics to extract all possible knowledge from students, teachers and parents. Seeking the highest possible accuracy in academic performance prediction using a set of powerful data mining techniques. The framework succeeds to highlight the student's weak points and provide appropriate recommendations. The realistic case study that has been conducted on 200 students proves the outstanding performance of the proposed framework in comparison with the existing ones.

1.3 LITERATURE REVIEW

Trilok Chand Sharma, Manoj Jain , the article titled as “WEKA Approach for Comparative Study of Classification Algorithm” discussed about classification of different decision tree algorithm as data mining techniques to process a dataset and identify the relevance of classification test data. Lukman, R., &Krajnc, M. (2012) , the article titled as “Exploring Non-traditional Learning Methods in Virtual and Real-world Environments”. It represents an important step for the university and its further performance using non-traditional learning methods, due to the fact that most of the lectures carried out at the university are still done in a traditional way (lecturer-centred). In the virtual-class, more efforts should be directed towards decreasing frustrations by improving motivation and interactivity.

Motivation could be strengthening by creating a sense of community and by building trust between students. Peter W. Stonebraker and James E. Hazeltine North-eastern Illinois University, Chicago, Illinois, USA titled as “Virtual learning effectiveness an examination of the process “This study defines, examines, and measures the effectiveness of a corporate virtual learning program. Initially, distinctions between traditional and virtual learning and university and corporate programs are defined. Ángel del Blanco¹, Javier Torrente¹, Pablo Moreno-Ger¹, BaltasarFernández-Manjón¹ titled as “Enhancing Adaptive Learning and Assessment in Virtual Learning Environments with Educational Games” The rising acceptance of Virtual Learning Environments (VLE) in the e-Learning field poses new challenges such as producing studentcentered courses which can be automatically tailored to each student's needs. Morten FlatePaulsen titled as “Online Education Systems: Discussion and Definition of Terms”This paper is written in order to establish a common framework of terms for the Web Education Systems Project (Web-edu), which is supported by the European Leonardo da vinci program.

1.4 EXISTING SYSTEM

The previous predictive models only focused on using the student's demographic data like gender, age, family status, family income and qualifications. In addition to the study related attributes including the homework and study hours as well as the previous achievements and grades. These previous work were only limited to provide the prediction of the academic success or failure, without illustrating the reasons of this prediction. Most of the previous researches have focused to gather more than 40 attributes in their data set to predict the student's academic performance. These attributes were from the same type of data category whether demographic, study related attributes or both, that lead to lack of diversity of predicting rules.

Disadvantages:

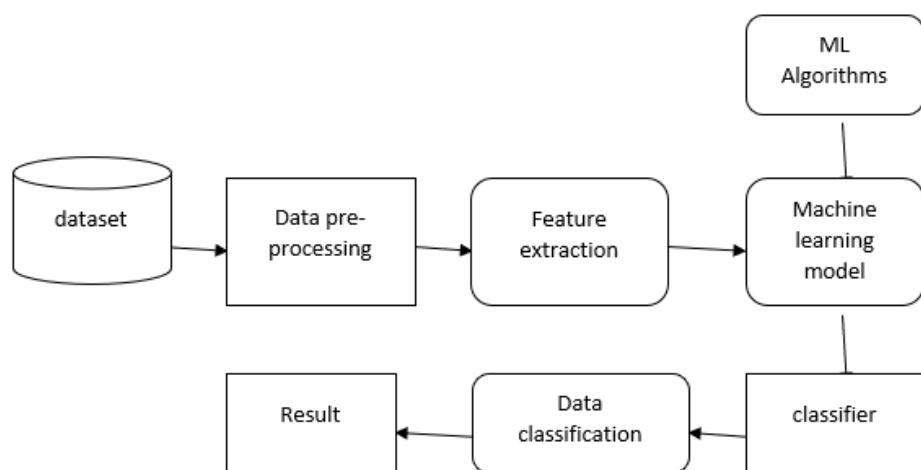
- As a result, these generated rules did not fully extract the knowledge for the reasons behind the student's dropout.
- Apart from the previously mentioned work, there were previous statistical analysis models from the perspective of educational psychology that conducted a couple of studies to examine the correlation between the mental health and the academic performance.
- The type of the recommendations was too brief, they missed illustrating the methodologies to apply them.

1.5 PROPOSED SYSTEM.

The proposed framework firstly focuses on merging the demographic and study related attributes with the educational psychology fields, by adding the student's psychological characteristics to the previously used data set (i.e., the students' demographic data and study related ones). After surveying the previously used factors for predicting the student's academic performance, we picked the most relevant attributes based on their rationale and correlation with the academic performance. An Unsupervised Ensemble Clustering used in this research.

Advantages:

- The proposal aims to analyze student's demographic data, study related details and psychological characteristics in terms of final state to figure whether the student is on the right track or struggling or even failing.
- In addition to extensive comparison of our proposed model with the other previous related models.

1.5.1 BLOCK DIAGRAM

1.5.2 REQUIREMENTS

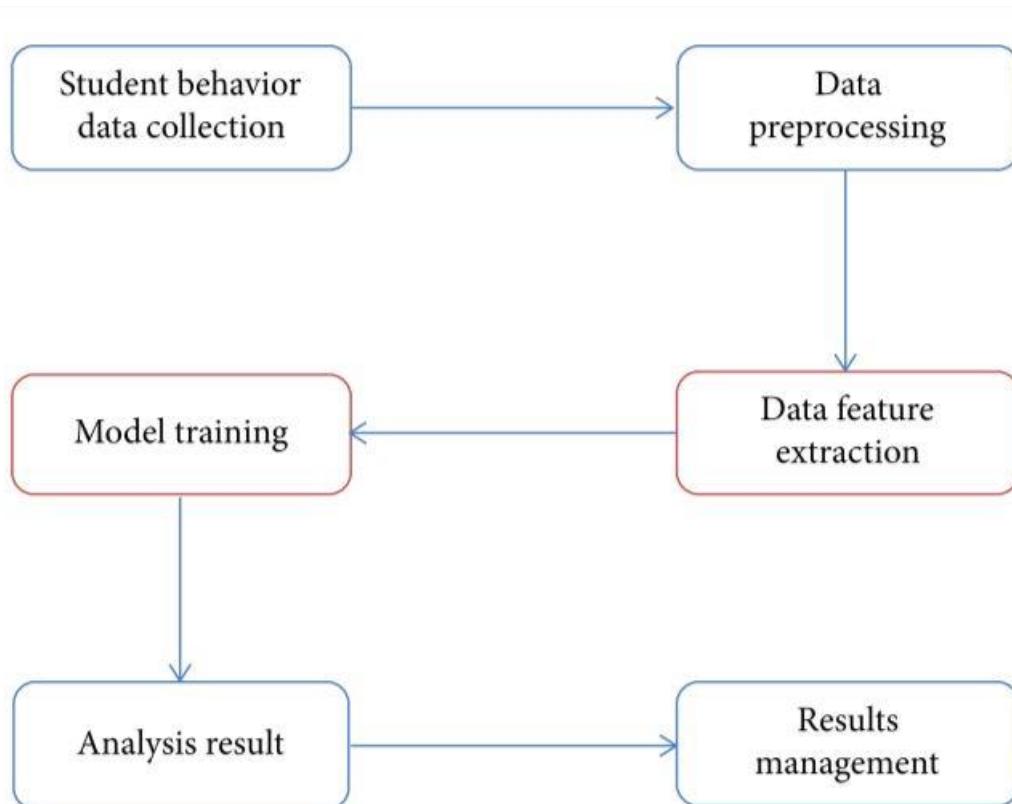
1.Hardware Requirements

- Windows 7,8,9,10,11.
- RAM 4GB.

2.Software Requirements

- Anaconda Navigator
- Python Language
- Jupyter notebook

1.5.3 FLOW CHART



1.5.4 PRINCIPLE OF OPERATION

The complexity of individual college students makes it impossible for school administrators to understand students' dynamics in real time. For some students with abnormal behaviours, the students around them may be inconvenient or embarrassed to inform the management staff of the specific situation, which causes a certain lag in the work of the student management staff. In order to understand the behaviour and habits of students in real time, the management of students will be transformed from passive to active. The daily behaviour data of students needs to be displayed intuitively on the student behaviour analysis system. On the premise of ensuring the privacy and safety of students, the data of the system mainly comes from the data collected by the school's digital systems, and the machine learning technology is used to analyze the data.

First, the historical data stored in the school's digital management systems need to be integrated and stored. Second, establish a data analysis model based on the dimensions of student behaviour data analysis. Third, establish a student behaviour data analysis system to achieve student management goals based on data services. Finally, a real-time monitoring system for student behaviour is established to monitor abnormal students in real time.

1.5.5 ADVANTAGES

- To assist the school to improve the management level of students and provide decision making support for the management
- To analyze the behaviour characteristic of students.

1.5.6 APPLICATIONS

- Traditional classroom education
- Corporate Training
- Distance education

CHAPTER – 2

MACHINE LEARNING

2.1 Machine Learning

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases. Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data. Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

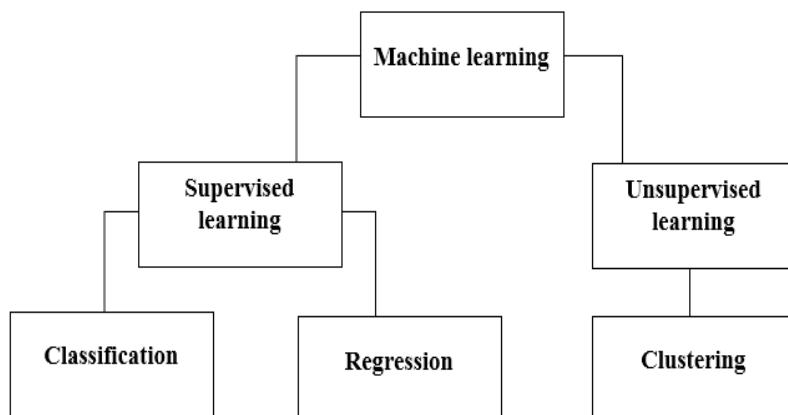


Fig.2.1.1 Types of Machine Learning

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics. The term machine learning was coined in 1959 by Arthur Samuel, an American and pioneer in the field of computer gaming and artificial intelligence. Also the synonym self-teaching computers was used in this time period. A representative book of the machine learning research during the 1960s was the Nilsson's book on Learning Machines, dealing mostly with machine learning for pattern classification. Interest related to pattern recognition continued into the 1970s, as described by Duda and Hart in 1973. In 1981 a report was given on using teaching strategies so that a neural network learns to recognize 40 characters (26 letters, 10 digits, and 4 special symbols) from a computer terminal.

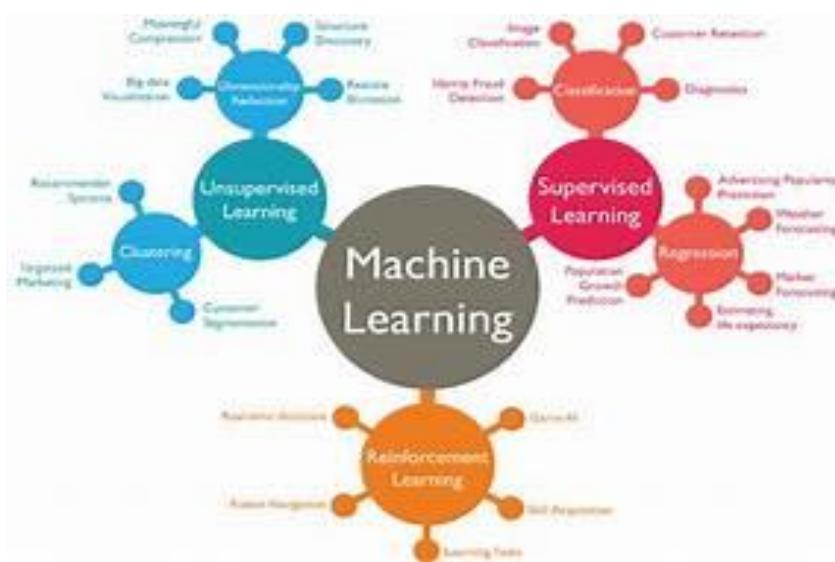


Fig.2.1.2 Pictorial Representation of Machine Learning

Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."^[18] This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?". Modern day machine learning has two objectives, one is to classify data based on models which have been developed, the other purpose is to make predictions for future outcomes based on these models. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with supervised learning in order to train it to classify the cancerous moles. Where as, a machine learning algorithm for stock trading may inform the trader of future potential predictions.

7 Steps to be followed in Machine Learning:

1. Gathering Data.
2. Preparing that Data.
3. Choosing a model.
4. Training.
5. Evaluation.
6. Hyper parameter Tuning
7. Prediction.

Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions. Machine learning contains a set of algorithms that work on a huge amount of data. Data is fed to these algorithms to train them, and on the basis of training, they build the model and perform a specific task.

These ML algorithms help to solve different business problems like Regression, Classification, Forecasting, Clustering, and Associations, etc.

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Reinforcement Learning

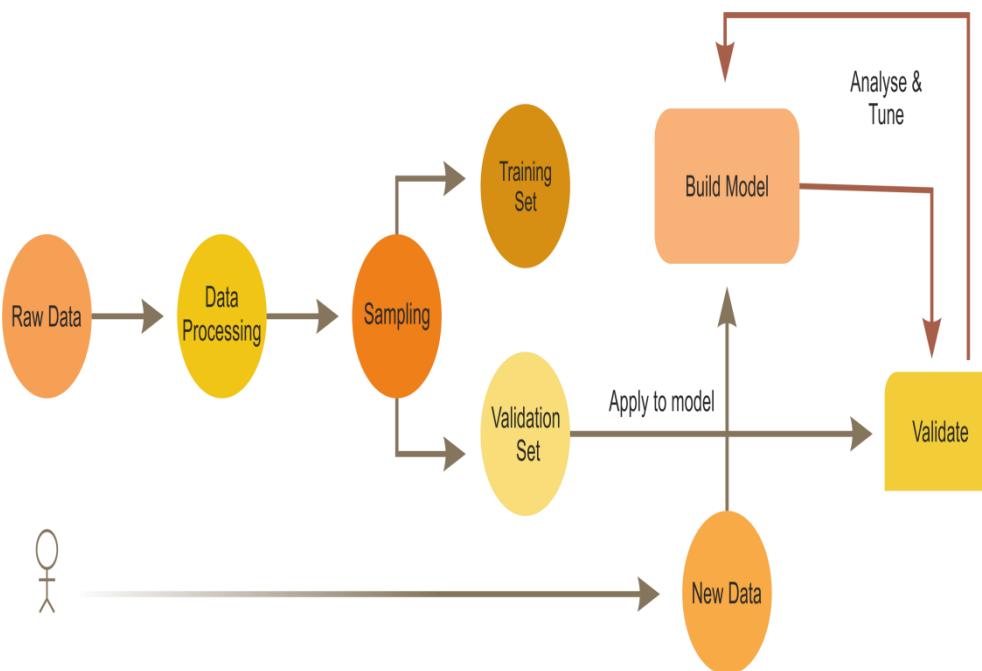


Fig.2.1.3 Building Model of ML

Supervised Learning

As its name suggests, Supervised machine learning

is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output. More precisely, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.

Supervised machine learning can be classified into two types of problems, which are given below:

- Classification
- Regression

a) Classification

Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "Yes" or No, Male or Female, Red or Blue, etc. The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are Spam Detection, Email filtering, etc.

Some popular classification algorithms are given below:

- Random Forest Algorithm
- Decision Tree Algorithm
- Logistic Regression Algorithm
- Support Vector Machine Algorithm

b) Regression

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

Some popular Regression algorithms are given below:

- Simple Linear Regression Algorithm
- Multivariate Regression Algorithm
- Decision Tree Algorithm
- Lasso Regression

Unsupervised Learning

Unsupervised Learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision. It means, in unsupervised machine learning, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.

In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision. The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences.

Let's take an example to understand it more preciously; suppose...

Some of the popular clustering algorithms are given below:

- K-Means Clustering algorithm
- Mean-shift algorithm
- DBSCAN Algorithm
- Principal Component Analysis
- Independent Component Analysis

2) Association

Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset. The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit. This algorithm is mainly applied in Market Basket analysis, Web usage mining, continuous production, etc.

1. Apriori
2. Eclat
3. F-P Growth Algorithm

Reinforcement Learning

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent gets rewarded for each good action and get punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards.

In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only. The reinforcement learning process is similar to a human being; for example, a child learns various things by experiences in his day-to-day life. An example of reinforcement learning is to play a game, where the Game is the environment.

Categories of Reinforcement Learning

Reinforcement learning is categorized mainly into two types of methods/algorithms:

- Positive Reinforcement Learning: Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.
- Negative Reinforcement Learning: Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

There are four Machine Learning algorithms which are used in the proposed system, namely,

- Random Forest(RF)
- Support Vector Machine(SVM)
- K-Nearest Neighbour(KNN)
- Naïve Bayes Algorithm(NB)

ALGORITHMS

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

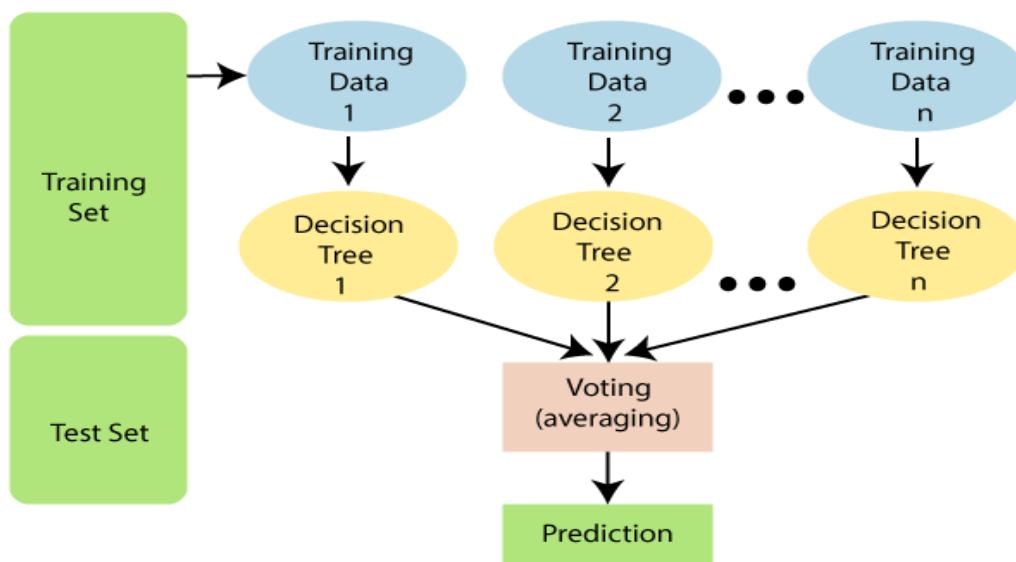


Fig.2.1.4 Algorithms of Machine Learning

Features of Random Forest:

- 1. Diversity-** Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- 2. Immune to the curse of dimensionality-** Since each tree does not consider all the features, the feature space is reduced.
- 3. Parallelization-** Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build random forests.
- 4. Train-Test split-** In a random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- 5. Stability-** Stability arises because the result is based on majority voting/ averaging.

Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine

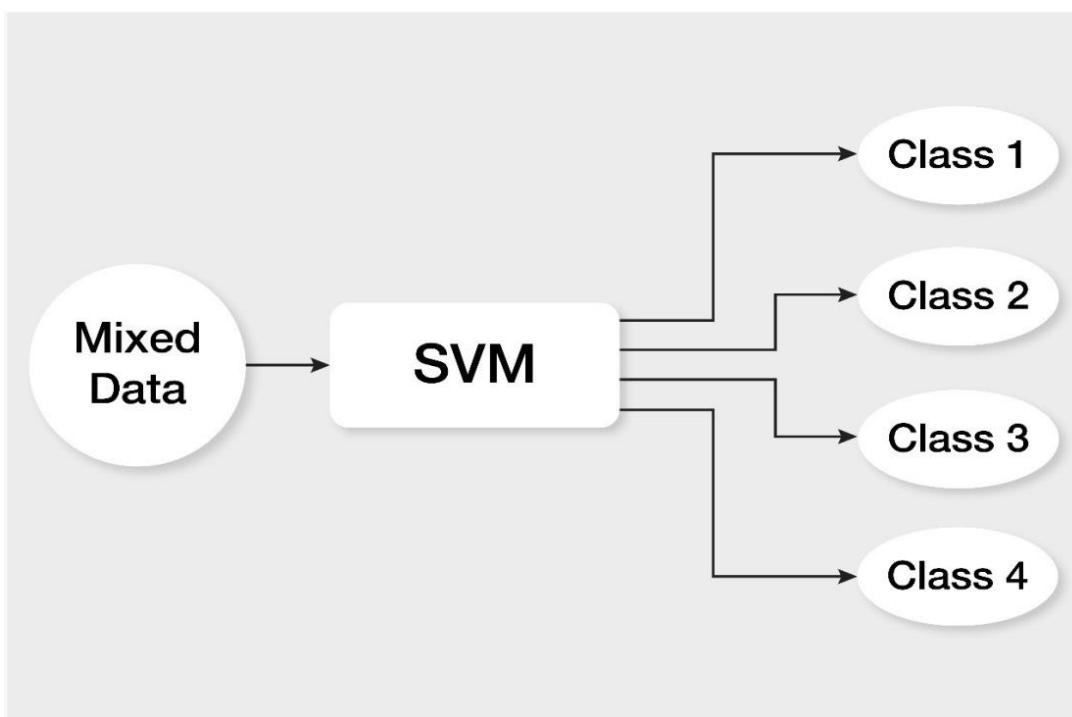


Fig.2.1.5 Algorithms of Machine Learning

K-Nearest Neighbor (KNN)

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. This algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using K- NN algorithm

K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the newdata.

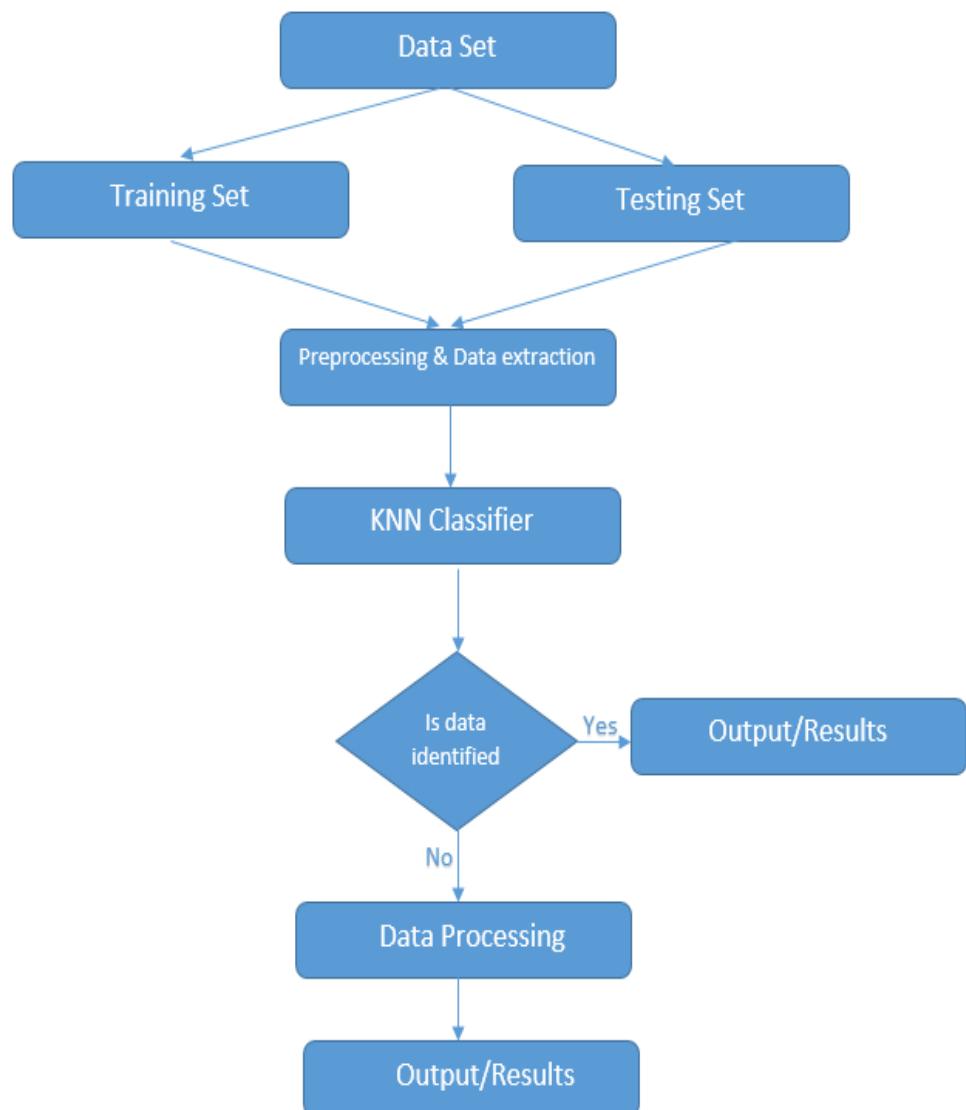


Fig.2.1.6 KNN

NAIVE BAYES

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

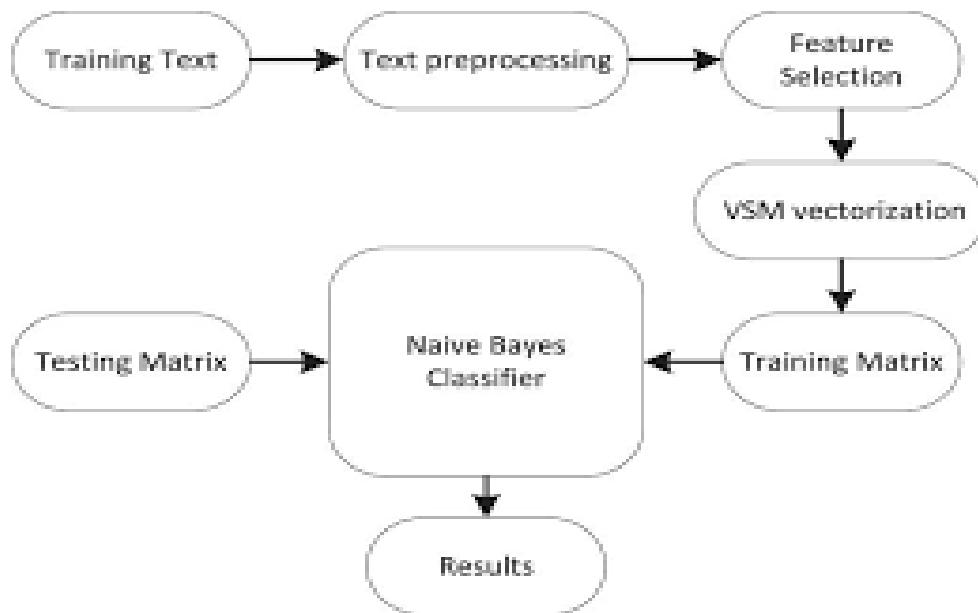


Fig.2.1.7 Naïve Bayes

2.2 SOFTWARE DEVELOPMENT LIFE CYCLE:-

There are various software development approaches defined and designed which are used/employed during development process of software, these approaches are also referred as "Software Development Process Models". Each process model follows a particular life cycle in order to ensure success in process of software development.

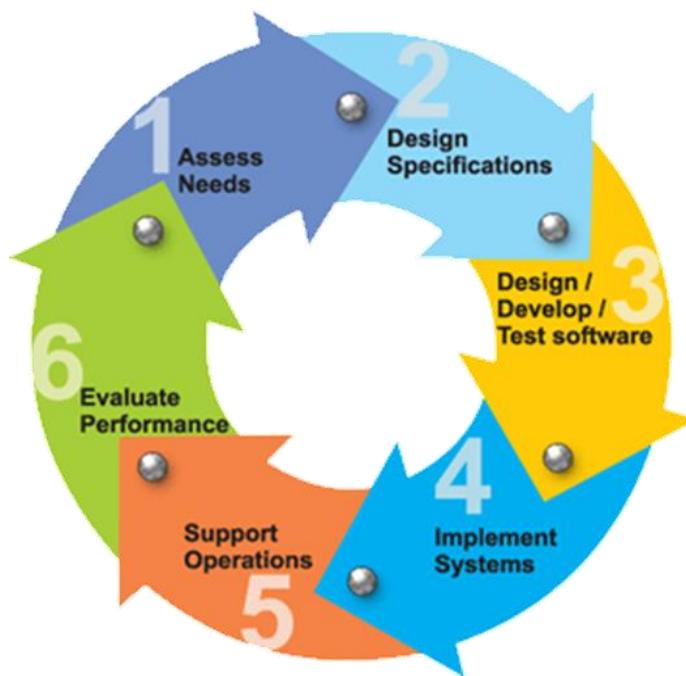


Fig.2.1.4 Algorithms of Machine Learning

Requirements

Business requirements are gathered in this phase. This phase is the main focus of the project managers and stakeholders. Meetings with managers, stakeholders and users are held in order to determine the requirements. Who is going to use the system? How will they use the system? What data should be input into the system? What data should be output by the system? These are general questions that get answered during a requirements gathering phase.

This produces a nice big list of functionality that the system should provide, which describes functions the system should perform, business logic that processes data, what data is stored and used by the system, and how the user interface should work. The overall result is the system as a whole and how it performs, not how it is actually going to do it.

Design

The software system design is produced from the results of the requirements phase. Architects have the ball in their court during this phase and this is the phase in which their focus lies. This is where the details on how the system will work is produced. Architecture, including hardware and software, communication, software design (UML is produced here) are all part of the deliverables of a design phase.

Implementation

Code is produced from the deliverables of the design phase during implementation, and this is the longest phase of the software development life cycle. For a developer, this is the main focus of the life cycle because this is where the code is produced. Implementation may overlap with both the design and testing phases. Many tools exist (CASE tools) to actually automate the production of code using information gathered and produced during the design phase.

2.3: 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.

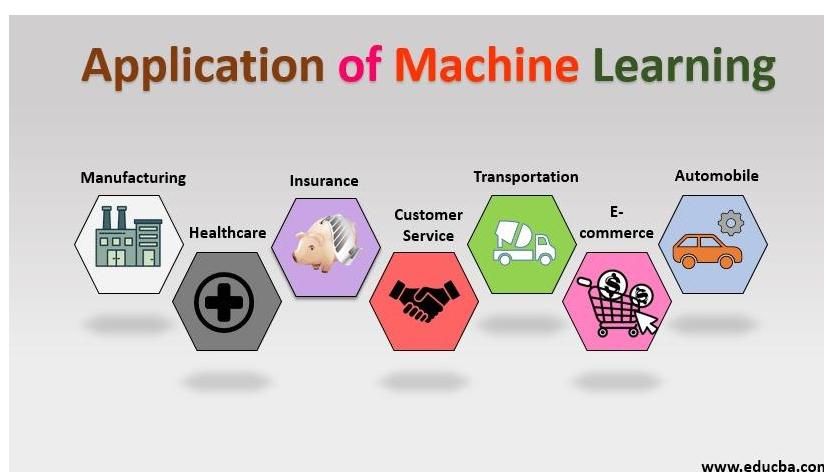


Fig.2.2 Applications of Machine Learning

1. Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, Automatic friend tagging suggestion

Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's face detection and recognition algorithm. It is based on the Facebook project named "Deep Face," which is responsible for face recognition and person identification in the picture.

2. Speech Recognition:

While using Google, we get an option of "Search by voice," it comes under speech recognition, and it's a popular application of machine learning. Speech recognition is a process of converting voice instructions into text, and it is also known as "Speech to text", or "Computer speech recognition." At present, machine learning algorithms are widely used by various applications of speech recognition. Google assistant, Siri, Cortana, and Alexa are using speech recognition technology to follow the voice instructions.

3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions. It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

- Real Time location of the vehicle from Google Map app and sensor
- Average time has taken on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations:

Machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning. Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest. As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. Self-driving cars:

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:

- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

Some machine learning algorithms such as Multi-Layer Perceptron, Decision tree, and Naïve Bayes classifier are used for email spam filtering and malware detection.

7. Virtual Personal Assistant:

We have various virtual personal assistants such as Google assistant, Alexa, Cortana, Siri. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc. These virtual assistants use machine learning algorithms as an important part.

8. Online Fraud Detection:

Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as fake accounts, fake ids, and steal money in the middle of a transaction. So to detect this, Feed Forward Neural network helps us by checking whether it is a genuine transaction or a fraud transaction.

For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

9. Stock Market trading:

Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's longshort term memory neural network is used for the prediction of stock market trends.

10. Medical Diagnosis

In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain. It helps in finding brain tumors and other brain-related diseases easily.

CHAPTER 3

SOFTWARE

3.1 PYTHON

- Anaconda Navigator
- Python
- Python built-in modules
 - Numpy
 - Pandas
 - Matplotlib
 - Sklearn
 - Seaborn

It is an interpreted, high-level and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python interpreters are available for many operating systems. Python is managed by the non-profit Python Software Foundation. Python features a dynamic types system and automatic memory management. It supports multiple programming paradigms, including object oriented, functional and procedural and has a large and comprehensive standard library. Python is easy to learn yet powerful and versatile scripting language, which makes it attractive for Application Development.

Python is a multi-paradigm programming language and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by meta programming and meta objects (magic methods)). Many other paradigms are supported via extensions, including design by contract and logic programming. It uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution. Python's design offers some support for functional programming in the Lisp tradition. It has filter, map and reduce functions; list comprehensions, dictionaries, sets and generator expressions. The standard library has two modules (itertools and functools) that implement functional tools borrowed from Haskell and Standard ML.

Python is used extensively to implement complex Internet services like search engines, cloud storage and tools, social media and so on. Whenever you use any of these services, you are actually very close to Python, although you wouldn't know it. Many developing tools are implemented in Python. More and more everyday use applications are being written in Python. Lots of scientists have abandoned expensive proprietary tools and switched to Python. Lots of IT project testers have started using Python to carry out repeatable test procedures. The list is long. Python is a general-purpose, versatile and popular programming language. It's great as a first language because it is concise and easy to read, and it is also a good language to have in any programmer's stack as it can be used for everything from web development to software development.

Features of Python

A simple language which is easier to learn, Python has a very simple and elegant syntax. It's much easier to read and write Python programs compared to other languages like: C++, Java, C#. Python makes programming fun and allows you to focus on the solution rather than syntax. If you are a newbie, it's a great choice to start your journey with Python.

1. Free and open source

You can freely use and distribute Python, even for commercial use. Not only can you use and distribute software's written in it, you can even make changes to the Python's source code. Python has a large community constantly improving it in each iteration.

2. Portability

You can move Python programs from one platform to another, and run it without any changes. It runs seamlessly on almost all platforms including Windows, Mac OS X and Linux.

3. Extensible and Embeddable

Suppose an application requires high performance. You can easily combine pieces of C/C++ or other languages with Python code. This will give your application high performance as well as scripting capabilities which other languages may not provide out of the box.

4.A high-level, interpreted language

Unlike C/C++, you don't have to worry about daunting tasks like memory management, garbage collection and so on. Likewise, when you run Python code, it automatically converts your code to the language your computer understands. You don't need to worry about any lower level operations.

5.Large standard libraries to solve common tasks

Python has a number of standard libraries which makes life of a programmer much easier since you don't have to write all the code yourself. For example: Need to connect MySQL database on a Web server You can use MySQLdb library using import MySQL db Standard libraries in Python are well tested and used by hundreds of people. So you can be sure that it won't break your application.

6.Object-oriented

Everything in Python is an object. Object oriented programming (OOP) helps you solve a complex problem intuitively. With OOP, you are able to divide these complex problems into smaller sets by creating objects.

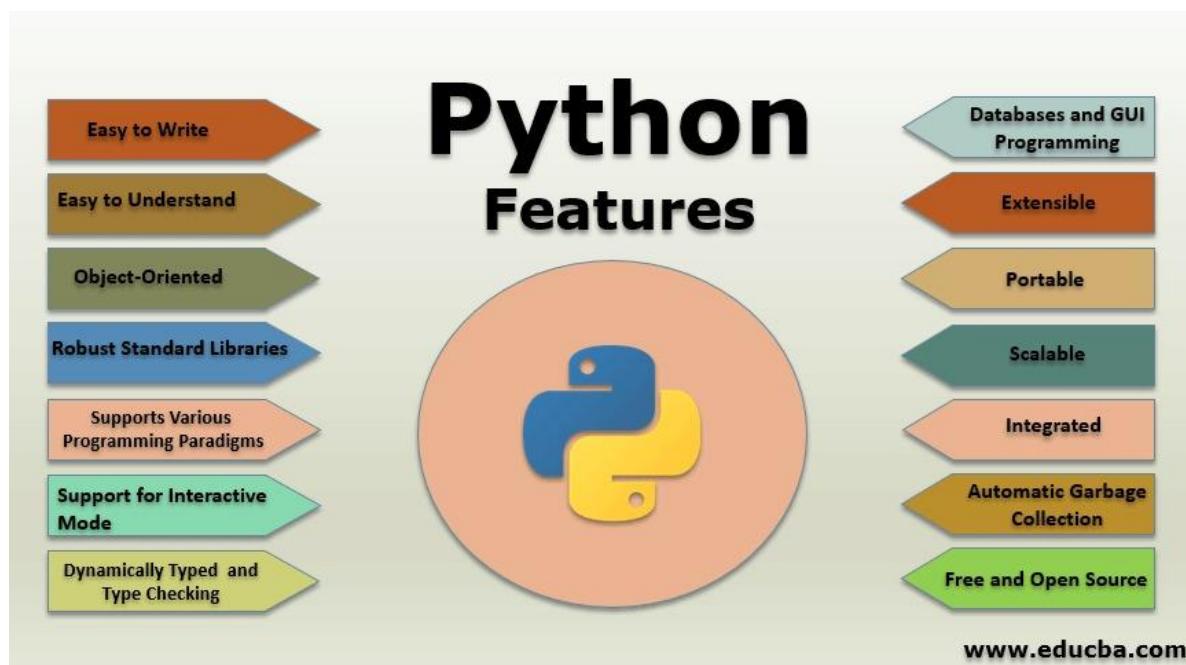


Fig.3 Features of Python

3.2 ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage anaconda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, mac OS and Linux.

Why use Navigator?

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages, and use multiple environments to separate these different versions.

The command line program anaconda is both a package manager and an environment manager, to help data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

WHAT APPLICATIONS CAN I ACCESS USING NAVIGATOR?

The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- QTConsole
- Spyder
- VSCode
- Glueviz
- Orange 3 App
- Rodeo
- RStudio

3.3 NUMPY

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more. At the core of the NumPy package, is the `ndarray` object.

The Numeric Python extensions (NumPy henceforth) is a set of extensions to the Python programming language which allows Python programmers to efficiently manipulate large sets of objects organized in grid-like fashion. These sets of objects are called arrays, and they can have any number of dimensions: one dimensional arrays are similar to standard Python sequences, two-dimensional arrays are similar to matrices from linear algebra. Note that one-dimensional arrays are also different from any other Python sequence, and that two-dimensional matrices are also different from the matrices of linear algebra, in ways which we will mention later in this text. All users of NumPy, whether interested in image processing or not, are encouraged to follow the tutorial with a working NumPy installation at their side, testing the examples, and, more importantly, transferring the understanding gained by working on images to their specific domain. The best way to learn is by doing – the aim of this tutorial is to guide you along this “doing.”



Fig.3.3 Numpy

In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy. Arrays are very frequently used in data science, where speed and resources are very important. NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently. This behavior is called locality of reference in computer science. This is the main reason why NumPy is faster than lists. Also it is optimized to work with latest CPU architectures.

3.4 VNC VIEWER

VNC stands for Virtual Network Computing. It is a cross-platform screen sharing system that was created to remotely control another computer. This means that a computer's screen, keyboard, and mouse can be used from a distance by a remote user from a secondary device as though they were sitting right in front of it. VNC works on a client/server model. A server component is installed on the remote computer (the one you want to control), and a VNC viewer, or client, is installed on the device you want to control from. This can include another computer, a tablet, or a mobile phone. When the server and viewer are connected, the server transmits a copy of the remote computer's screen to the viewer.

Not only can the remote user see everything on the remote computer's screen, but the program also allows for keyboard and mouse commands to work on the remote computer from afar, so the connected user has full control (after being granted permission from the remote computer). VNC was created in Cambridge in the late 1990s by the founders of RealVNC, and was commercialized in 2002 when the company was established.

What's a VNC Server?

A server is a piece of computer hardware or software that provides capabilities for other programs called "clients." This is called the client-server model, whereas a server can provide services such as data or resource sharing to one or multiple clients. One server can serve multiple clients in this way, and one single client can use multiple servers.

A client will send a request to a server, which then sends a response back to the client. A computer with VNC Server software installed can be accessed and controlled from a different device in a different location. The software allows a broadcast of the device desktop to a secondary device with VNC Viewer installed. Connected VNC Viewer users send a request, and then (with permission) can see the same thing as the person sitting in front of the remote computer.

What's a VNC Viewer?

A viewer, on the other hand, is a program that renders the contents of a digital file on screen. VNC Viewer is used for local computers and mobile devices you want to control from. A device such as a computer, tablet, or smart phone with VNC Viewer software installed can access and take control of a computer in another location. It is a graphical desktop sharing system that allows a user to remotely control the desktop of a remote computer (running VNC Server) from your device, and it transmits the keyboard and mouse or touch events to VNC Server, so that once you are connected, you have control over the computer you've accessed. If you're using your mobile phone, for example, you would be able to use the computer you've remotely accessed as though you were sitting right in front of it. Remote Framebuffer, or RFB, is the protocol that governs the format of the data that passes between the client and server within the VNC system. This is what allows a client to view and control another computer remotely. It is applicable to all windowing applications and systems, which means that it works across platforms such as Windows, macOS, Linux, and other popular operating systems.

The place where the user sits, with the display, mouse, and keyboard capabilities, is called the RFB client or viewer. The place where the framebuffer changes originate (as in the windowing system) is called the RFB server. Remote Framebuffer is designed so that clients can run on the widest range of hardware and so that implementing a client is as simple as possible, with very few requirements needed from the client. RFB started as a very simple protocol but has been enhanced to include features such as file transfer, more refined compression, and stronger security measures as it has developed. Seamless cross-compatibility between VNC clients and servers is made possible because they are able to negotiate a connection which uses the best RFB version, as well as security and compression options that are supported by both. RFB was developed as a remote display

Technology in Cambridge, UK, by some of the original developers of VNC and the current RFB protocol specifications for version 6 are published on the RealVNC website.



Fig.3.4 VNC Viewer

- Access all your Windows remote devices and servers with VNC® Connect – easily and without any fuss
- Transfer files, print and chat from anywhere
- Provide instant remote support on your computers whenever and wherever needed
- Deliver exceptional IT support through remote control to Windows devices
- Quickly and instantly connect to attended computers without having to pre-install any software
- Our patent-protected technology automatically optimizes to the speed of your network
- Low-bandwidth and high-latency environments? No problem – we've got you covered
- Deploy in bulk, offline, or remotely using standard enterprise tooling with VNC Connect Enterprise

CHAPTER 4

HARDWARE

4.1 RASPBERRY PI

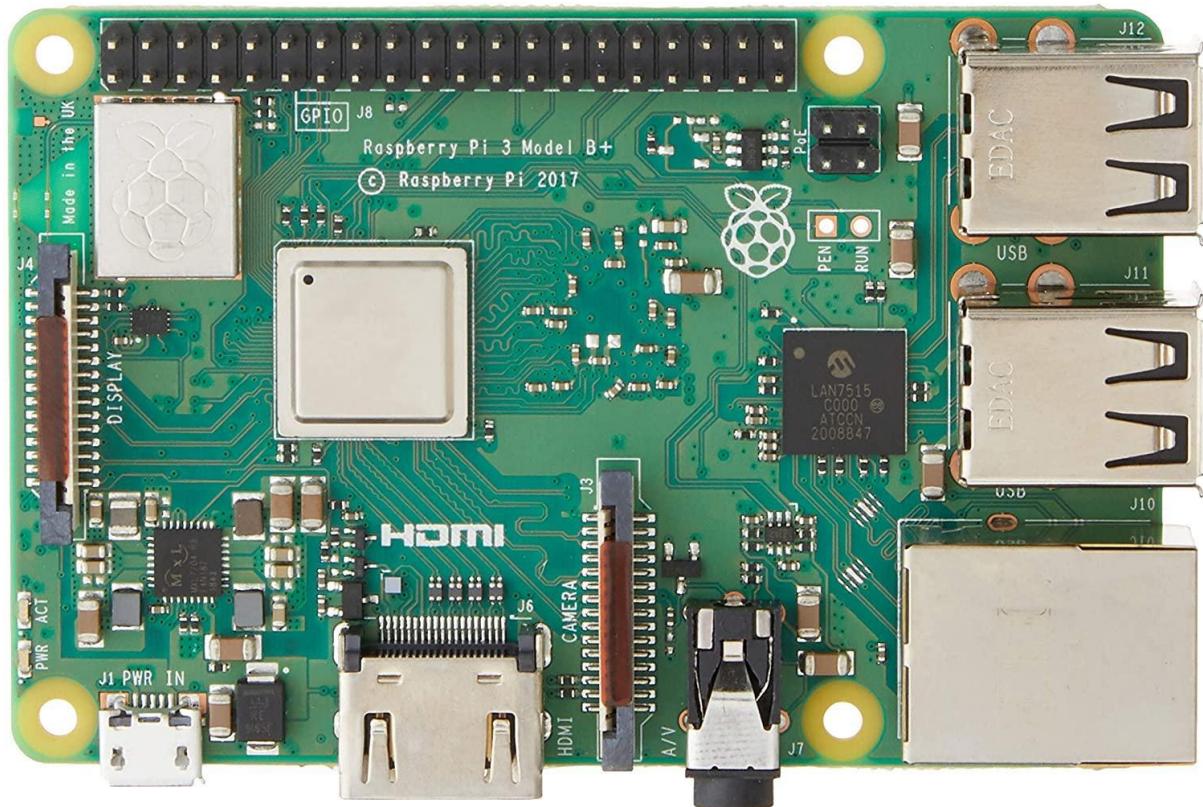


Fig 4.1.1 Raspberry pi

Raspberry Pi is a small single-board Computer developed in UK by Raspberry Pi foundation to promote the teaching of computer science in schools and in developing countries. Original model become far more popular than anticipated sealing outside of its target market, for uses such as robots.

History

Raspberry Pi has mainly three generations Raspberry Pi 1, Raspberry Pi 2, Raspberry Pi 3 and also a reduce simple inexpensive Raspberry Pi zero.

- The first model of Raspberry Pi was launched in February 2012 i.e. Raspberry Pi 1 Model B followed by a simple inexpensive Model A.
- In April 2014 “Compute Model” for embedded application Raspberry Pi 1 model B+ improved versions of A and B was launched.
- In November 2015 with reduced I/O and GPIO Raspberry Pi zero came into market.
- In February 2015 advance model with 40 GPIO pins, Ethernet, 4 USB slots Raspberry Pi 2 was launched.
- In February 2016 an upgraded model with inbuilt Bluetooth and Wi-Fi Raspberry Pi 3 Model B was launched.
- Recently in February 2017, “Raspberry Pi – Zero W” with in built Wi-Fi and Bluetooth come into the market.

Features

The heart of the Raspberry Pi is a Broadcom System on Chip (SOC) which includes ARM compatible CPU and on-chip graphic processing unit and Vediocore IV.

The key feature from First generation to the Third generation includes:

- CPU speed ranges from 700 MHz to 1.2 GHz.
- On board Memory (RAM) ranges from 256 MB to 1 GB.
- USB slot differs from 1 slot to USB slots.
- HDMI, composite video output and 3.5mm phone jack.
- Low level output is provided by GPIO pins which support common. protocols like I2C (inter-integrated circuit).
- Ethernet 8 Position 8 Contact (8P8C).

Processor:

The processor at the heart of the Raspberry Pi is a Broadcom BCM28XX. This is the Broadcom System on Chip (SOC) chip used in the Raspberry Pi. The processor from first to third generations include:

- Raspberry Pi 1: Broadcom BCM2835 SOC with 700MHz CPU speed, L2 cache of 128kb with ARM compatibility AR1176JZF-S (ARMv6) 32-bit RISC ARM.
 - Raspberry Pi 2: Broadcom BCM 2836 SOC with 900MHz CPU speed, L2 cache of 256kb with 32-bit quad-core ARM cortex-A7 (ARMv7).
 - Raspberry Pi 3: Broadcom BCM2837 SOC with 1.2GHz 64-bit quad-core –A53 with 512 kb shared L2 cache (64-bit instruction set ARMv8).

Raspberry pi power supply

Model B+ Power Supply to make the B+ more reliable and actually reduce the current draw, the power supply is completely redesigned.

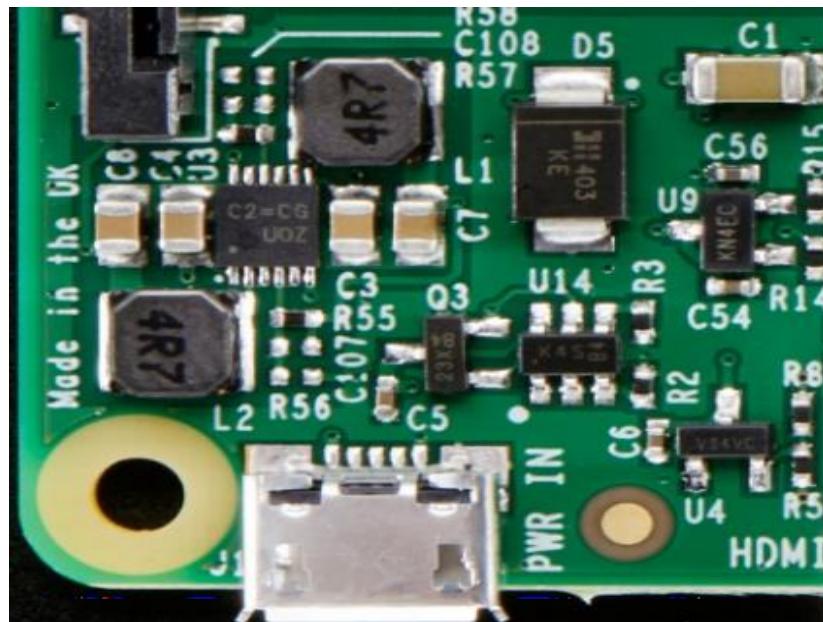


Fig 4.1.2 Power Supply of Raspberry pi

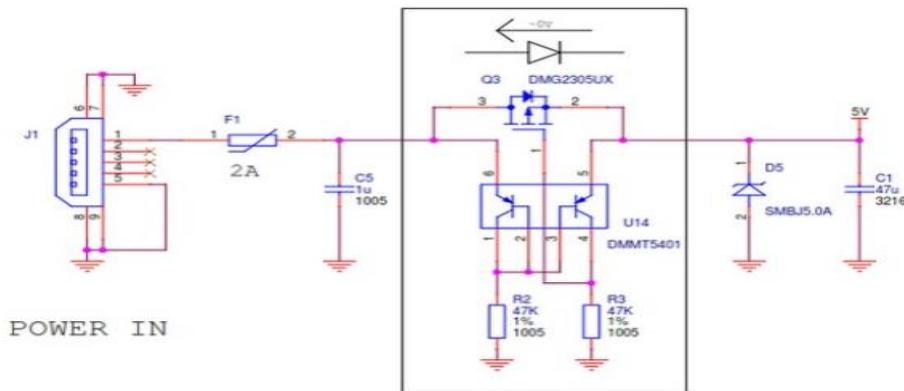


Fig 4.1.3 Circuit Connection

There's still the microUSB jack on the left, and the 1A fuse has been upgraded to a 2A fuse. There's also a DMG2305UX (<http://adafru.it/dGU>) P-Channel MOSFET. This acts as a polarity protection switch but is much lower 'drop-out' than a diode. It has only 52mW resistance so @ 2A its about 0.1V voltage drop. Most diodes would be at least 0.5V. Watch this great video about this technique here: To the right is a protection TVS diode (D5 part #SMBJ5) which protects from overvoltages. So not a lot has changed here (other than putting in a protection FET) There is a PNP-matched-pair action going on around the polarity FET, but its 3AM and I'm not 100% sure what it's for so I'll wait till I get some rest before doing any analysis. Let's look at the 3.3V & 1.8V supplies:

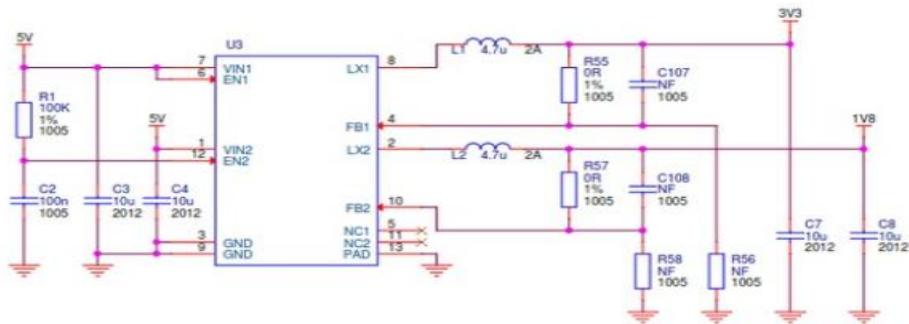


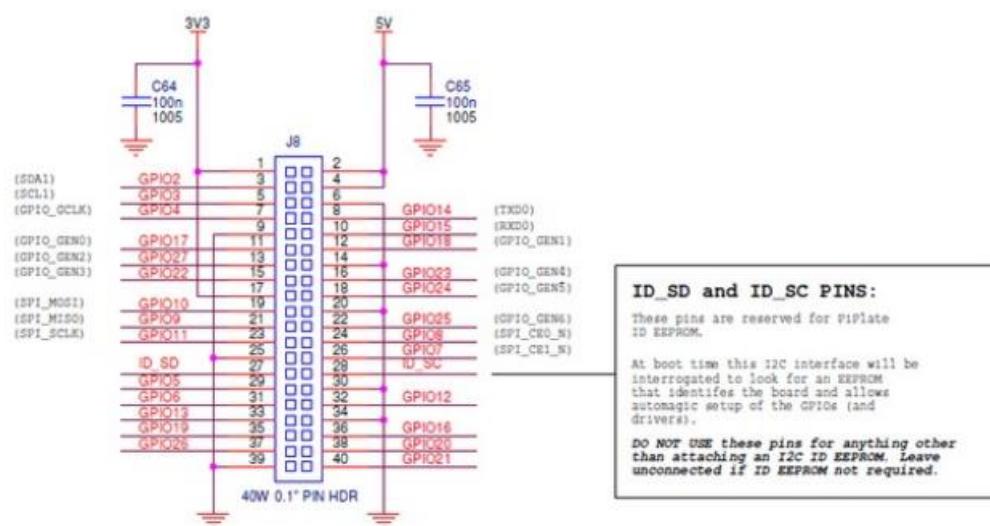
Fig 4.1.4 Pin diagram with connections

Raspberry Pi Model B+ GPIO Port:

First thing to notice, the top 26 pins of the 40-pin connector are the same as the original. That means that most/many Pi Plates that plug into the Model B will plug into the B+ just fine. They won't sit in the same location - they'll be slid down just a bit but electrically-wise it's the same.



Fig 4.1.5 GPIO Port Connections



CHAPTER 5

RESULT ANALYSIS

5.1 WORKING PRINCIPLE

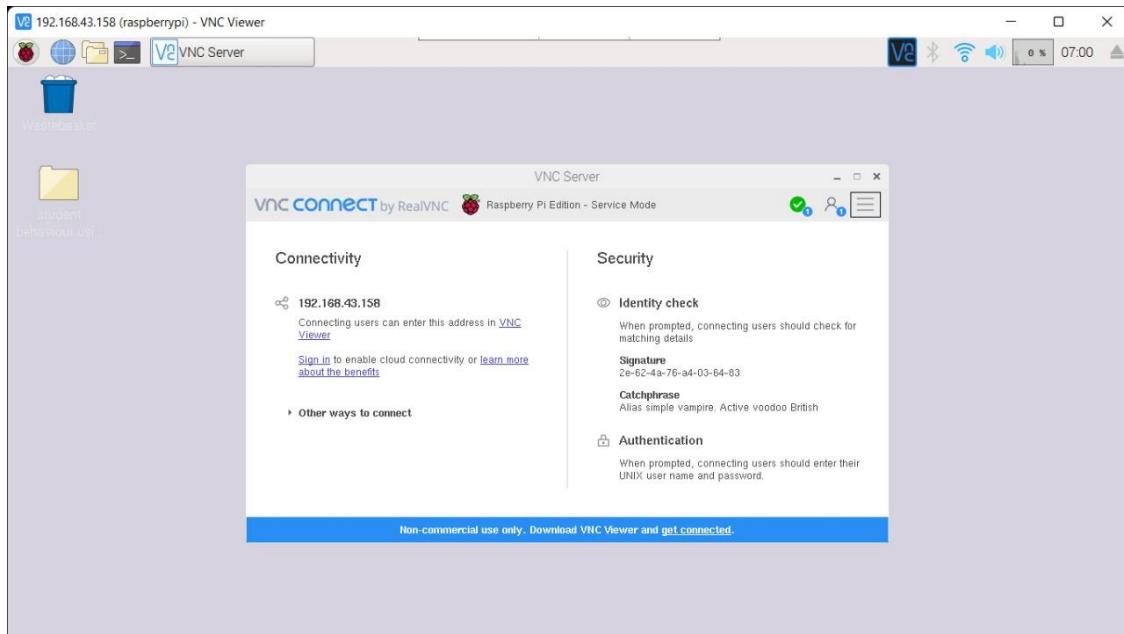
To find groups of learners with similar online engagement in an unsupervised manner, we follow the procedure summarised. We first create a similarity matrix between learners using a dynamic time warping kernel. This matrix is transformed into a similarity graph using a sparsification based on the Relaxed Minimum Spanning Treea procedure that retains global network connectivity while discarding weak similarities that can be explained through longer chains of strong similarities. Through this process, we create a graph where the nodes are learners linked by edges weighted according to their time-course similarity. Hence, two learners that complete the tasks of the course in a similar manner will be linked by a strong edge. The constructed similarity graph is then analysed using Markov Stability (MS), a multiscale graph partitioning algorithm that uses a Markov process to scan the graph across Markov time in order to find optimised and robust partitions of the graph at any level of resolution. The partitions are found by maximising a resolution-dependent cost function (the Markov Stability) at all levels of resolution, as given by the Markov time, t . We then select robust partitions in the following sense: (i) they are persistent across scales (i.e., optimal over an extended Markov time t , as given by a plateau with a low value of $VI(t, t')$), and (ii) robust to the small changes in the optimisation (i.e., consistently found as a good partition over those scales, as given by a relative dip in $VI(t)$). Such robust partitions identify clusters of learners that exhibit similar online temporal patterns. The definitions of the different measures and some details of the Markov Stability framework are given in Methods. summarises the results of our multiscale clustering method applied to the time-series of task completion of six online courses by 81 learners pursuing a post-graduate part-time Management degree at Imperial College Business School over one year. See Methods for further details about the data. As the Markov time is increased, the level of resolution is decreased and the method reveals robust partitions of decreasing granularity. we illustrate the partitions found from ten clusters down to two clusters, with a notably robust partition into six clusters. Note the quasi-hierarchical aggregation of the finer clusters into coarser ones, a feature that is intrinsic to the data and not imposed by our clustering algorithm.

We have applied the methodology to analyse task completion time-series data from a second cohort of 46 learners taking the online management course at Imperial College Business School. The results we obtain are similar, as shown in the multiscale clustering presented in Supplementary and the detailed analysis of the 6-cluster partition. In this case, we identified a robust 9-cluster partition (with four major clusters and five single learner clusters) and a robust 6-cluster partition (with three major clusters and three single outliers). The major clusters in the 6-way partition showed similar behaviours to those observed in the first cohort we analysed. In particular, the green cluster corresponds to the ‘Early Birds’ and ‘On time’ groups whereas the blue cluster is similar to the group of task-skipping ‘Low Engagers’ group and the purple cluster exhibits similar traits to the ‘Crammers’ cluster. Within this 6-cluster partition, we found that of the 8 low-performance learners, 4/8 were located in the massed learning cluster, 2/8 were sporadic outliers, and 1/8 was in the low engagement cluster. Only 1/8 was located in the distributed learning cluster. Moreover, using standard classification procedures in Supplementary we found that our methodology was superior at grouping learners with similar performance. These findings highlight the consistency of the methodology across the cohorts, yet attuned to particularities of the data.

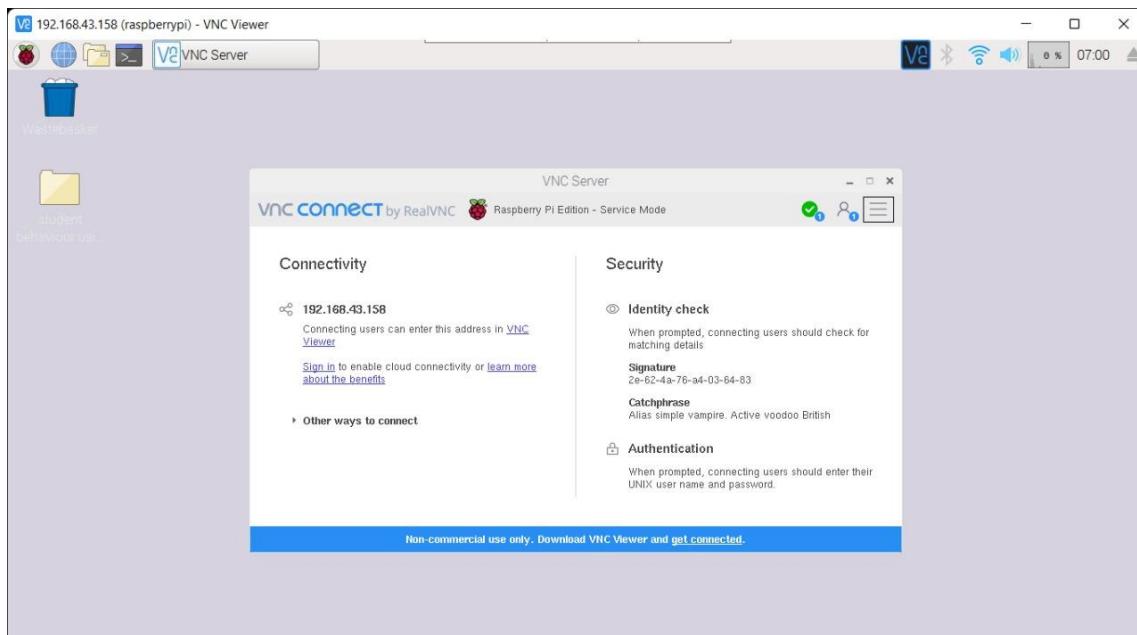
The types of temporal engagement data collected from learners will differ across educators or institutions depending on the particularities of the Learning Management System. To test the methodology on a different kind of data, we have studied a set of 100 learners undertaking an anonymised course within the Open University. The OULAD data set differs from our data set in several ways: (i) the time-stamp data in OULAD corresponds to page clicks and not necessarily to task completion; (ii) the time stamps were coarse-grained to days; (iii) pages could be revisited. The results of applying our methodology to the OULAD data set show that the multiscale clustering is robust to the sparsification implicit in the graph creation step. A robust 3-way partition is consistently found in our analysis, with two major clusters and a minor cluster of outliers. The two major clusters corresponded to a separation of learners who exhibited higher massed learning and lower task engagement versus learners with a distributed learning. We found that 6/7 of the low-performance learners (<60%) were located in the cluster associated with massed learning, while one low-performance learner was located in the minor outlier cluster and none were in the distributed learning group.

5.2 RESULT

1. Connect same hotspot connection to both monitor and to your PC. Open VNC Server in the monitor and it will display as shown below. Copy the IP address displayed and paste it in the VNC Server app in your PC.



2. Open Command Prompt and run Jupyter Notebook



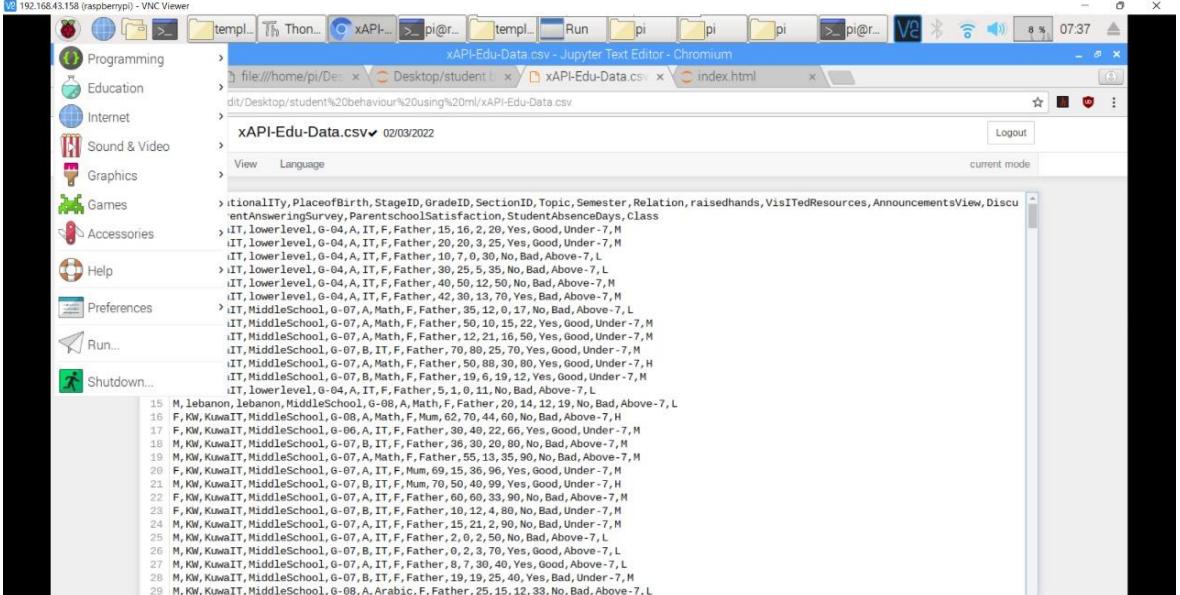
3. After Jupyter Notebbok, code is displayed. And then run it.

```
1 import numpy as np
2 import pandas as pd
3 from flask import Flask, request, jsonify, render_template
4 import pickle
5
6 app = Flask(__name__)
7 model = pickle.load(open('model.pkl', 'rb'))
8
9
10
11 @app.route('/')
12 def home():
13     return render_template('index.html')
14
15 @app.route('/predict', methods=['POST'])
16 def predict():
17     ...
18     For rendering results on HTML GUI
19     ...
20     float_features = [float(x) for x in request.form.values()]
21
22     final_features = [np.array(float_features)]
23     prediction = model.predict( final_features )
24
25     if prediction==1:
26         pred="Congratulations! Student Behaviour Satisfactory"
27     else:
28         pred="Student Behaviour UNSatisfactory"
29     output = pred
```

4. After running it, you'll get an IP address. Copy the IP address.

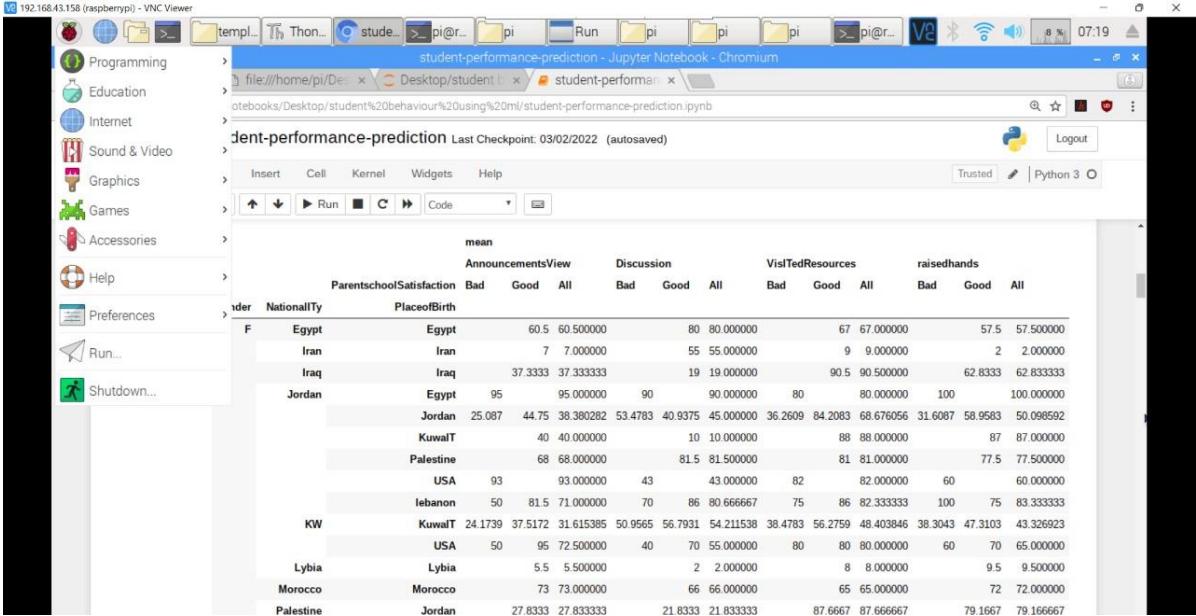
A screenshot of a Raspberry Pi desktop environment. On the left, there's a vertical dock with icons for Programming, Education, Internet, Sound & Video, Graphics, Games, Accessories, Help, Preferences, Run..., and Shutdown... Below this dock is a terminal window showing Python code for a Flask application. The code includes imports for Flask, request, jsonify, and render_template, and defines a route for 'index.html' with a POST method. A 'Stop' button is visible in the Thonny interface. The desktop background shows icons for Programming, Education, Internet, Sound & Video, Graphics, Games, Accessories, Help, Preferences, Run..., and Shutdown... At the top right, there's a system tray with icons for battery level (10%), signal strength, and the current time (07:05). The desktop also features a 'regular mode' switch.

5. StudentData dumped into the Raspberry Pi



The screenshot shows a VNC session on a Raspberry Pi. The terminal window displays a large amount of data from a CSV file named "xAPI-Edu-Data.csv". The data includes columns such as Nationality, PlaceofBirth, StageID, GradeID, SectionID, Topic, Semester, Relation, raisedhands, VisitedResources, AnnouncementsView, Discussion, and Class. The data is presented in a tabular format with many rows.

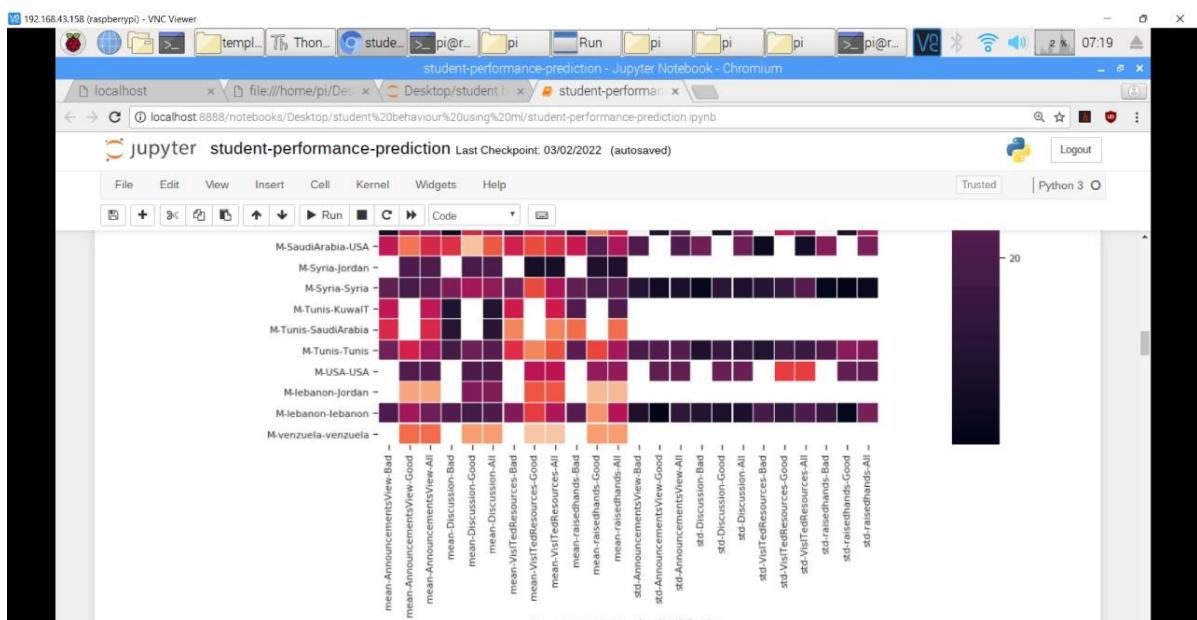
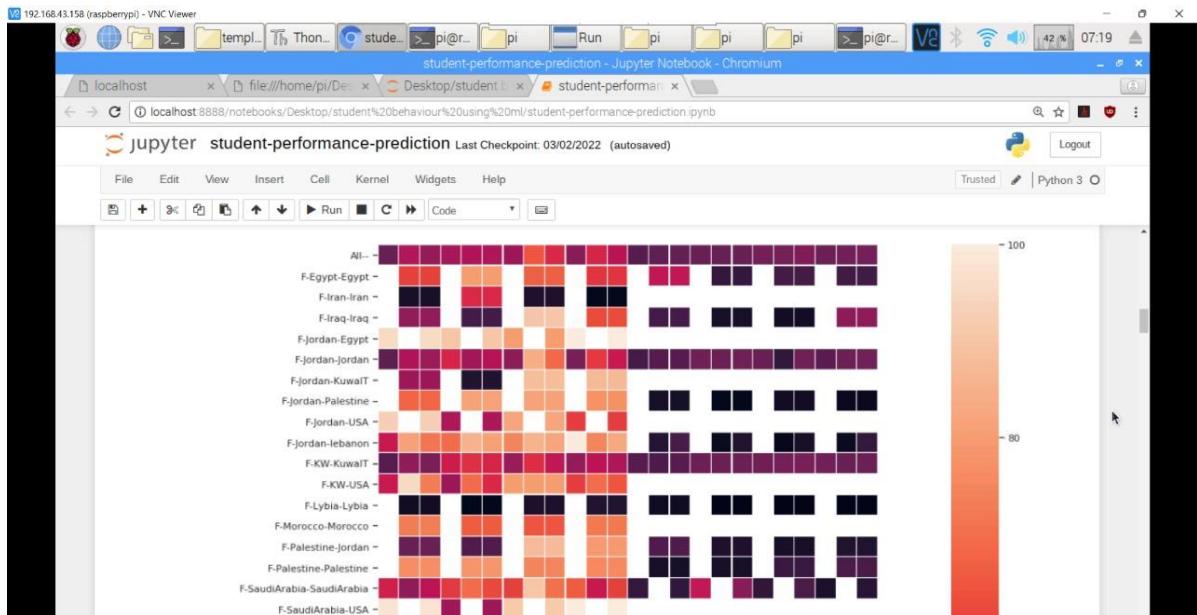
6. Mean values of considering gender, nationality,place of birth and various attributes.



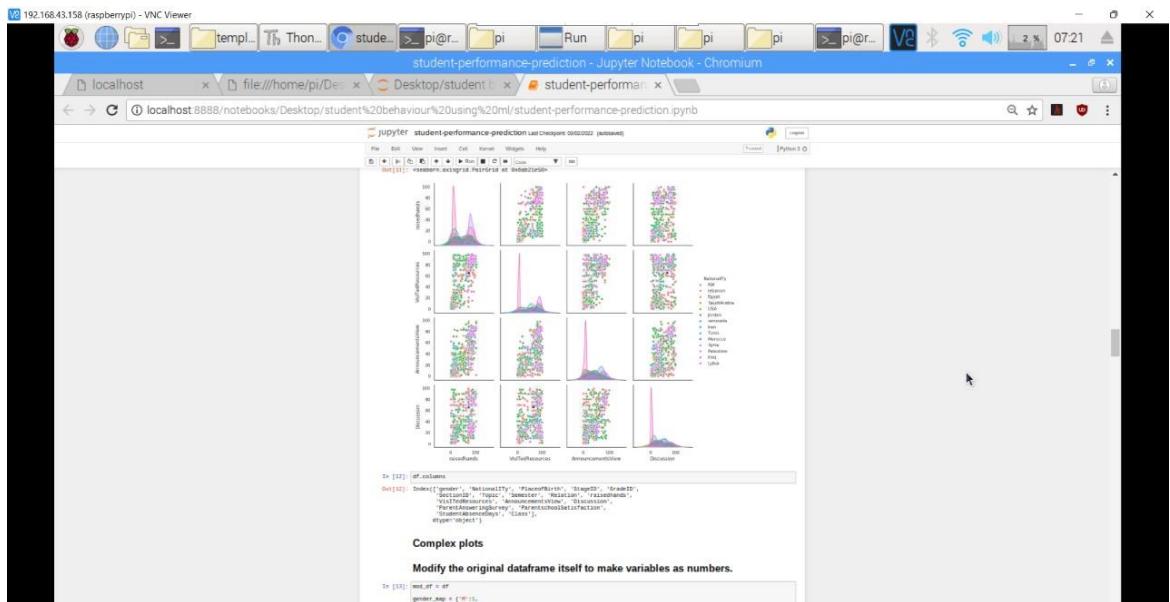
The screenshot shows a Jupyter Notebook interface on a Raspberry Pi. The notebook cell displays a table titled "mean" showing mean values for various attributes across different categories. The categories include Nationality (F, Jordan, KW) and PlaceofBirth (Egypt, Iran, Iraq, Lebanon, Palestine, USA, Kuwait, Libya, Morocco, Palestine). The attributes include AnnouncementsView, Discussion, VisitedResources, and raisedhands, broken down by gender (Bad, Good, All) and class (Bad, Good, All).

	Gender	Nationality	PlaceofBirth	mean			AnnouncementsView			Discussion			VisitedResources			raisedhands		
				Bad	Good	All	Bad	Good	All	Bad	Good	All	Bad	Good	All	Bad	Good	All
F	Egypt	Egypt	60.5	60.50000	60	80.00000	80	80.00000	67	67.00000	67	57.5	57.50000					
		Iran	7	7.00000	55	55.00000	55	55.00000	9	9.00000	9	2	2.00000					
		Iraq	37.3333	37.33333	19	19.00000	19	19.00000	90.5	90.50000	90.5	62.8333	62.83333					
Jordan	Egypt	95	95.00000	90	90.00000	80	80.00000	100	100.00000	100	100.00000	100						
		Jordan	25.087	44.75	38.380282	53.4783	40.9375	45.000000	36.2609	84.2083	68.676056	31.6087	58.9583	50.096952				
		Kuwait	40	40.00000	10	10.00000	88	88.00000	88	88.00000	88	88.00000	87	87.00000				
KW	Palestine	Palestine	68	68.00000	81.5	81.50000	81	81.00000	81	81.00000	81	77.5	77.50000					
		USA	93	93.00000	43	43.00000	82	82.00000	82	82.00000	82	60	60.00000					
		Iraq	50	81.5	71.00000	70	86	80.666667	75	86	82.333333	100	75	83.333333				
USA	Kuwait	Kuwait	24.1739	37.5172	31.615385	50.9565	56.7931	54.211538	38.4783	56.2759	48.403846	38.3043	47.3103	43.326923				
		USA	50	95	72.50000	40	70	75.00000	80	80	80.00000	60	70	65.00000				
		Morocco	Morocco	73	73.00000	66	66.00000	65	65.00000	65	65.00000	65	72	72.00000				
Palestine	Jordan	27.8333	27.83333	21.8333	21.83333	21.8333	21.83333	21.8333	21.83333	21.8333	79.1667	79.166667	79.166667					

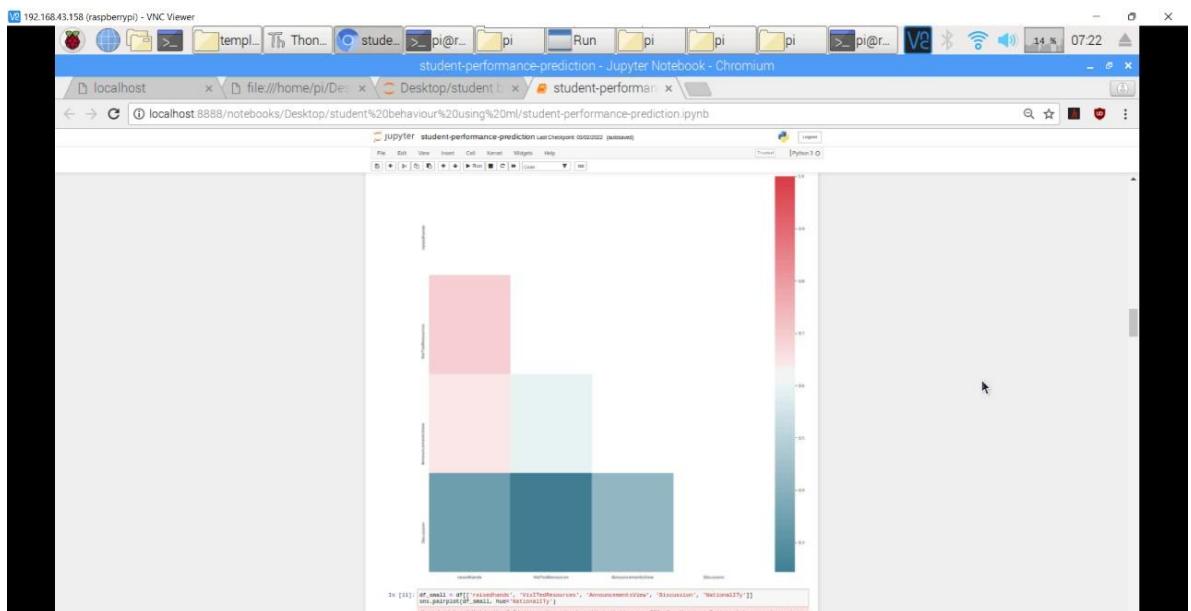
7. Colours coding scale of behaviour of students



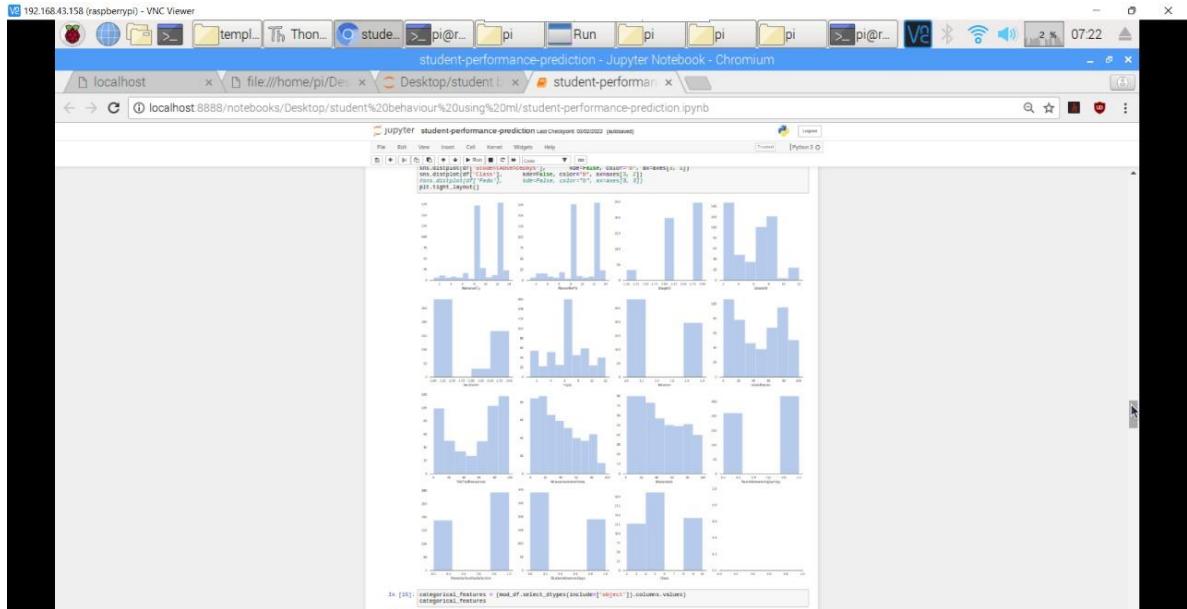
8. Graph plotting using dots and waves of geographic area



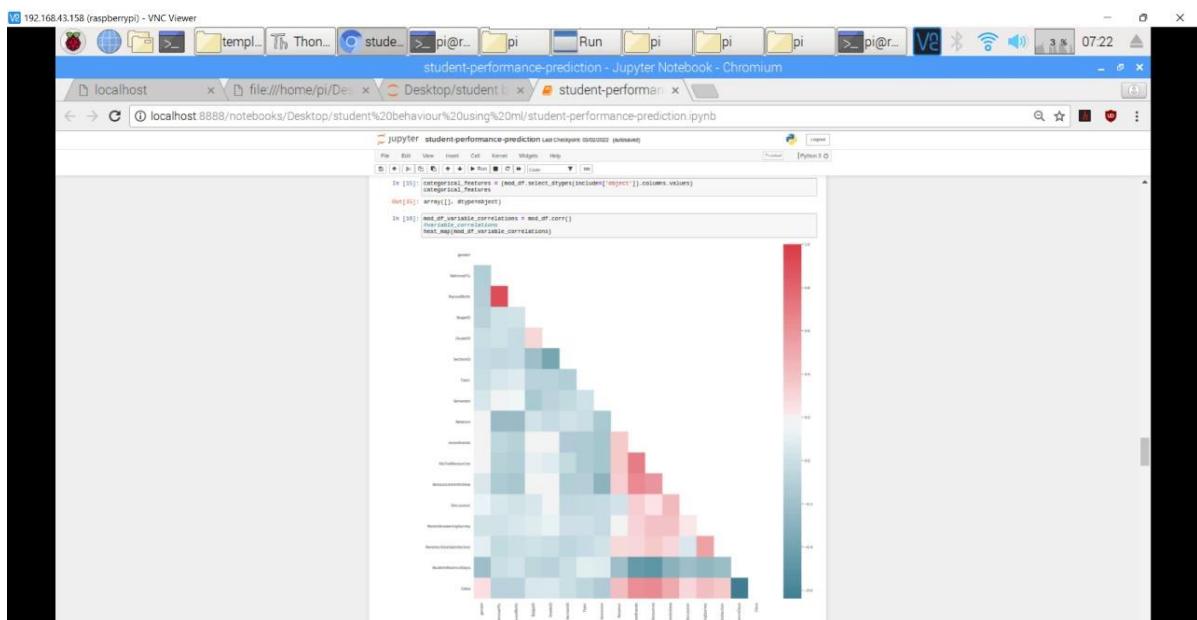
9. Colour coding system for the Data



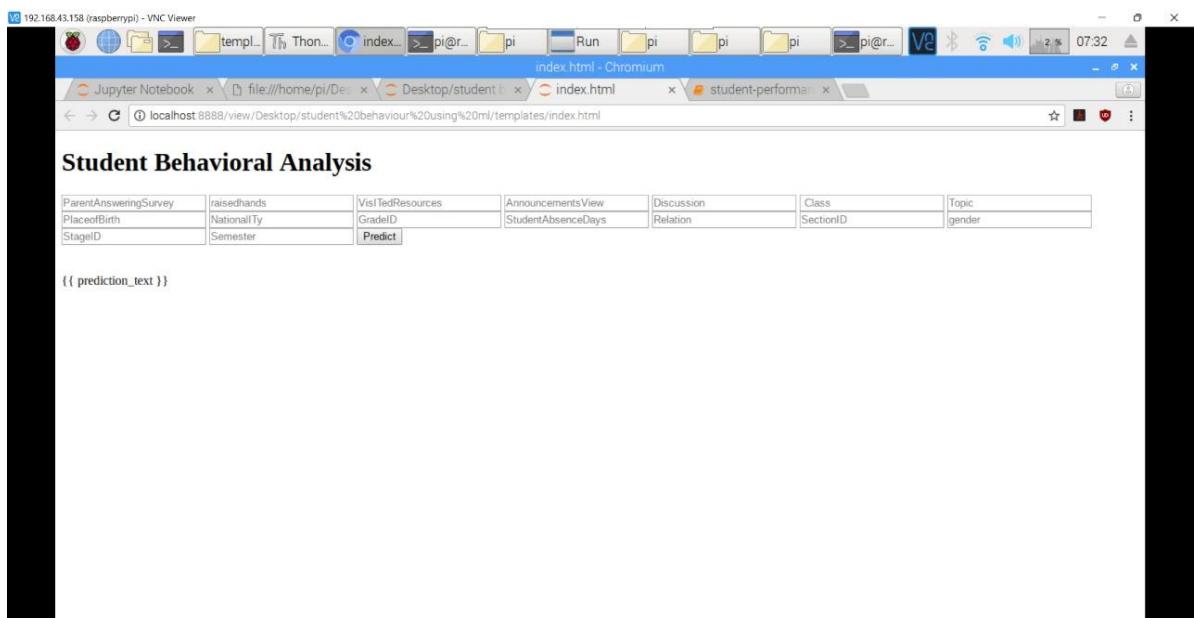
10. Plotted graphs based on individual data



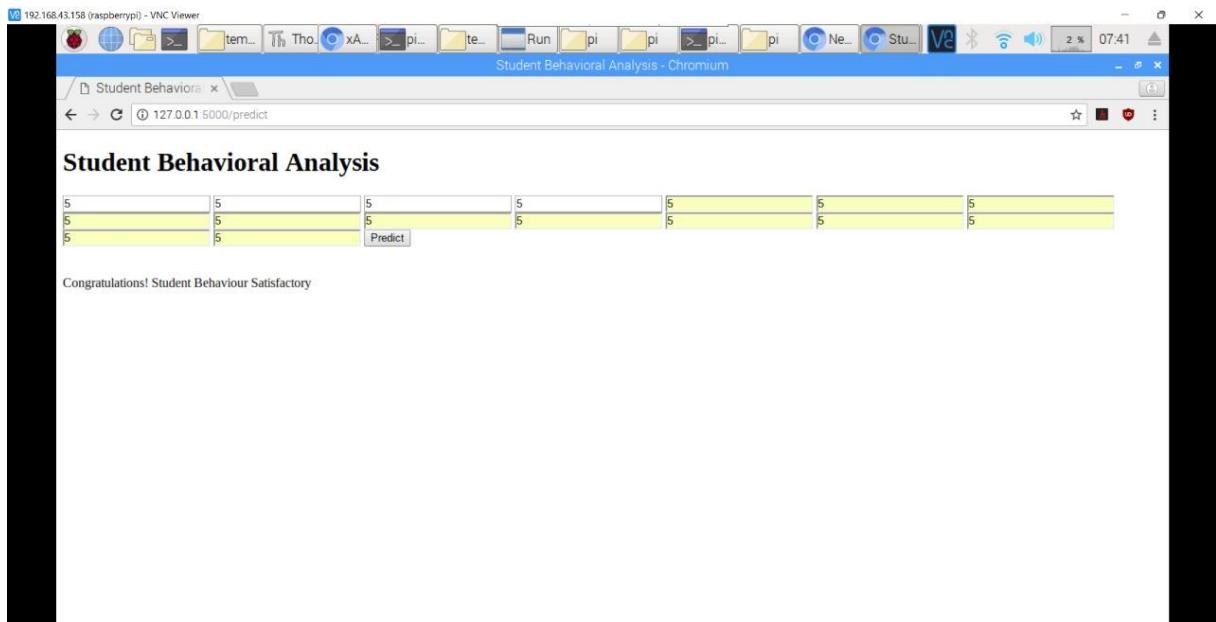
11. Data segregated according to the colour segregation



12. Run the IP address in the device with Raspberry Pi and VNC viewer in browser and the following HTML page is opened.



13. After the values are given, the following is displayed. Hence, this is the final result.



REFERENCES

- [1] A.H. EliassonC .J. Lettieri and A.H. Eliasson, "Early to bed, early to rise! Sleep habits and academic performance in college students," *Sleep Breath.*, vol. 14, no. 1, pp. 71–75, 2010, doi: 10.1007/s11325-009-0282-2.
- [2] X.D. KeatingD. Castelli and S.F. Ayers, "Association of Weekly Strength Exercise Frequency and Academic Performance Among Students at a Large University in the United States," *J. Strength Cond. Res.*, vol. 27,no. 7, pp. 1988–1993, 2013,doi: 10.1519/JSC.0b013e318276bb4c.
- [3] M. Valladares, E. Duran, A. Matheus, S. Duran-Agueero, A. Maria O-bregon and R. Ramirez-Tagle, "Association between Eating Behavior and Academic Performance in University Students," *J. Am. Coll. Nutr.*, vol. 35,no. 8, pp. 699–703, 2016, doi: 10.1080/07315724.2016.1157526.
- [4] J. FilippouC. Cheong and F. Cheong, "Modelling the impact of studybehaviours on academic performance to inform the design of a persuasivesystem," *INFORMATION & MANAGEMENT*, vol. 53, no. 7SI, pp. 892–903, 2016, doi: 10.1016/j.im.2016.05.002.
- [5] S. Ghosh and S.K. Ghosh, "Exploring the association between mobilitybehaviours and academic performances of students: a context-aware traj-graph (CTG) analysis," *Progress in Artificial Intelligence*, vol. 7, no. 4, pp.307–326, 2018, doi: 10.1007/s13748-018-0164-6.
- [6] Z.Y. Yang, X.Y. Mo, D.X. Shi and R. Wang, "Mining relationshipsbetween mental health, academic performance and human behaviour,"in *SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, San Francisco,CA, 2017, pp.1–8
- .[7] T. PhanS.G. McNeil and B.R. Robin, "Students' patterns of engage-ment and course performance in a Massive Open Online Course,"*COMPUTERS & EDUCATION*, vol. 95, pp. 36–44, 2016, doi:10.1016/j.comedu.2015.11.015.
- [8] I. JO, Y. Park, J. Kim and J. Song, "Analysis of Online Behavior andPrediction of Learning Performance in Blended Learning Environments,"*Educational Technology International*, vol. 15, no. 2, pp. 71–88, 2014.
- [9] G. Kostopoulos, S. Kotsiantis, N. Fazakis, G. Koutsonikos and C. Pier-rakeas, "A Semi-Supervised Regression Algorithm for Grade Predictionof Students in Distance Learning Courses," *International Journal On Artificial Intelligence Tools*, Vol. 28, No. 19400014si,2019, Doi: 10.1142/S0218213019400013

Student Performance Analysis Behavior Using Unsupervised Approach

M. Srikanth¹,Guntaka Sai pramidha²,Gurram Svetalina², Kuppa Naga Soumya²

¹Professor, Department of ECE, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India.

²UG Scholar, Department ofECE, Malla Reddy Engineering College for Women, Hyderabad, Telangana,India.

Abstract:

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. This system developed to analyze and predict the student's performance only. The proposed framework analyzes the students' demographic data, study related and psychological characteristics to extract all possible knowledge from students, teachers and parents. Seeking the highest possible accuracy in academic performance prediction using a set of powerful data mining techniques. The framework succeeds to highlight the student's weak points . The realistic case study that has been conducted on 200 students proves the outstanding performance of the proposed framework in comparison with the existing ones.

Existing System:

The previous predictive models only focused on using the student's demographic data like gender, age, family status, family income and qualifications. In addition to the study related attributes including the homework and study hours as well as the previous achievements and grades. These previous work were only limited to provide the prediction of the academic success or failure, without illustrating the reasons of

this prediction. Most of the previous researches have focused to gather more than 40 attributes in their data set to predict the student's academic performance. These attributes were from the same type of data category whether demographic, study related attributes or both, that lead to lack of diversity of predicting rules.

Disadvantage:

- As a result, these generated rules did not fully extract the knowledge for the reasons behind the student's dropout.
- Apart from the previously mentioned work, there were previous statistical analysis models from the perspective of educational psychology that conducted a couple of studies to examine the correlation between the mental health and the academic performance.
- The type of the recommendations was too brief, they missed illustrating the methodologies to apply them.

Proposed System:

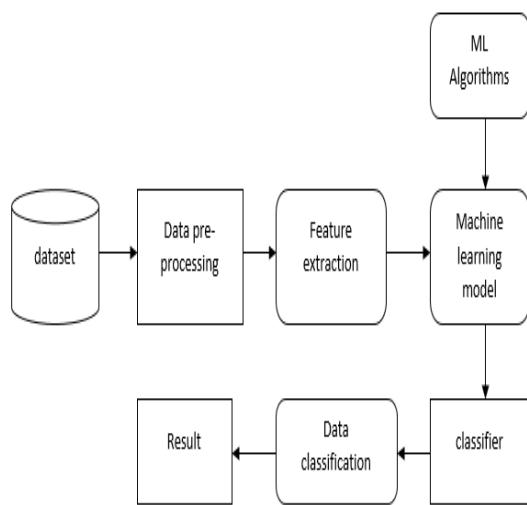
The proposed framework firstly focuses on merging the demographic and study related attributes with the educational psychology fields, by adding the student's psychological characteristics to the previously used data

set (i.e., the students' demographic data and study related ones). After surveying the previously used factors for predicting the student's academic performance, we picked the most relevant attributes based on their rationale and correlation with the academic performance.

Advantage:

- The proposal aims to analyze student's demographic data, study related details and psychological characteristics in terms of final state to figure whether the student is on the right track or struggling or even failing. In addition to extensive comparison of our proposed model with the other previous related models.

System Architecture:



System Requirements:

Software:

- Python

- Anaconda navigator

Hardware :

- Windows 7,8,10(64 bit)
- RAM 3GB
- Raspberry pi 3 B+

Raspberry Pi

Raspberry Pi is a small single-board Computer developed in UK by Raspberry Pi foundation to promote the teaching of computer science in schools and in developing countries.

Original model become far more popular than anticipated sealing outside of its target market, for uses such as robots.

Processor

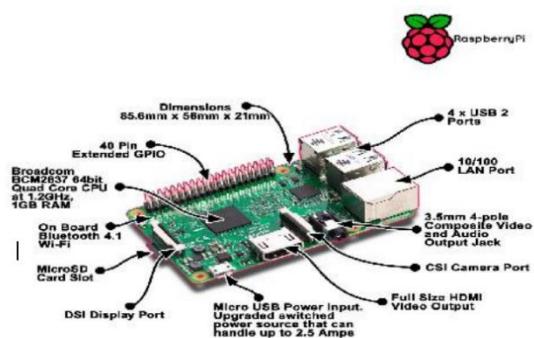
The processor at the heart of the Raspberry Pi is a Broadcom BCM28XX.

This is the Broadcom System on Chip (SOC) chip use in the Raspberry Pi. The processor from first to third generations include:

- Raspberry Pi 1: Broadcom BCM2835 SOC with 700MHz CPU speed, L2 cache of 128kb with ARM compatibility AR1176JZF-S (ARMv6) 32-bit RISC ARM.
- Raspberry Pi 2: Broadcom BCM 2836 SOC with 900MHz CPU speed, L2 cache of 256kb with 32-bit quad-core ARM cortex-A7 (ARMv7).

- Raspberry Pi 3: Broadcom BCM2837 SOC with 1.2GHz 64-bit quad-core –A53 with 512 kb shared L2 cache (64-bit instruction set ARMv8).

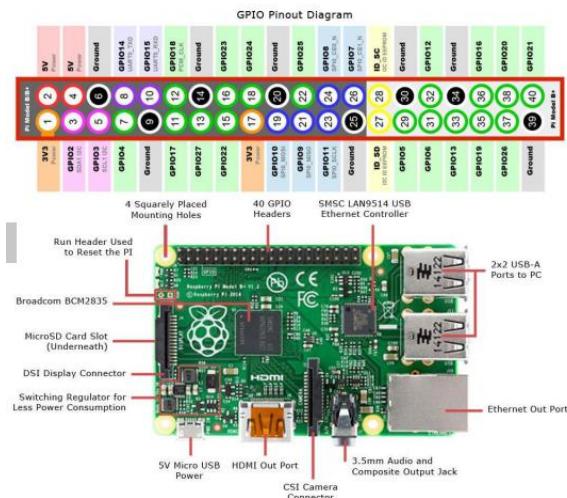
Raspberry Pi 3



In this project we are using latest version of Raspberry i.e. Raspberry Pi 3. The processor at the heart of the Raspberry Pi 3 is a Broadcom BCM2837, and the later models of the Raspberry Pi 2. The underlying architecture of the BCM2837 is identical to the BCM2836. The Only significant difference is the replacement of the ARMv7 quad core cluster with quad-core ARM Cortex A53 (ARMv8) cluster.

The ARM cores run at 1.2GHz, making the device about 50% faster than the Raspberry Pi 2 with a on board memory of 1GB RAM. The videocoreIV runs at 400MHz.

One powerful feature of the Raspberry Pi is the row of GPIO (general purpose input/output) pins along the top edge of the board. These pins are a physical interface between the Pi and the outside world. At the simplest level, you can think of them as switches that you can turn on and off (output). Of the 40 pins, 26 are GPIO pins and the others are power or ground pins (plus two ID EEPROM pins).



Modules

- DATA COLLECTION
- DATA PRE-PROCESSING
- FEATURE EXTRACTION
- EVALUATION MODEL

DATA COLLECTION

Data used in this paper is a set of student details in the school records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called *labelled data*.

DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

1. **Formatting**
2. **Cleaning**
3. **Sampling**

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely.

Sampling: There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

FEATURE EXTRACTION

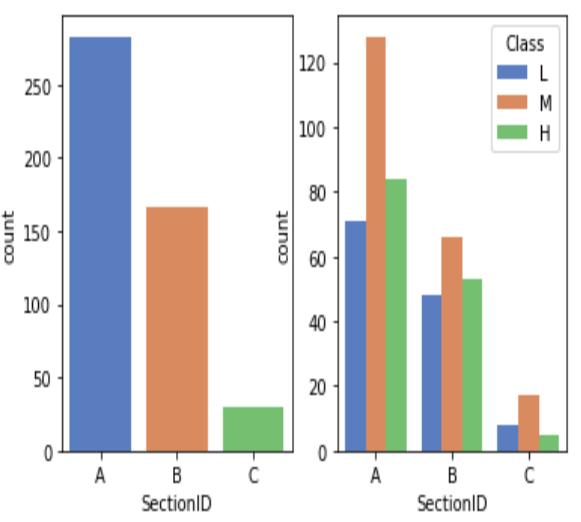
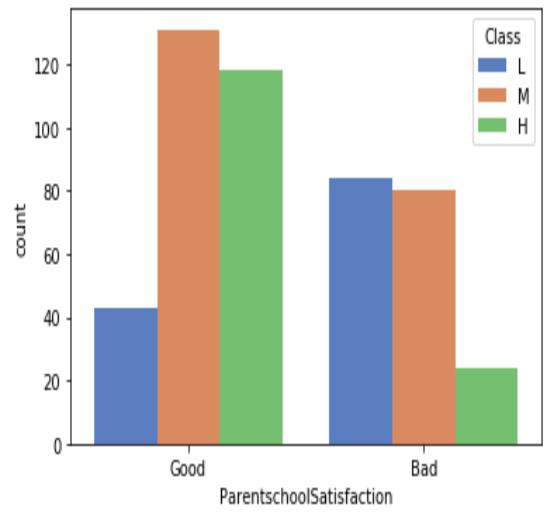
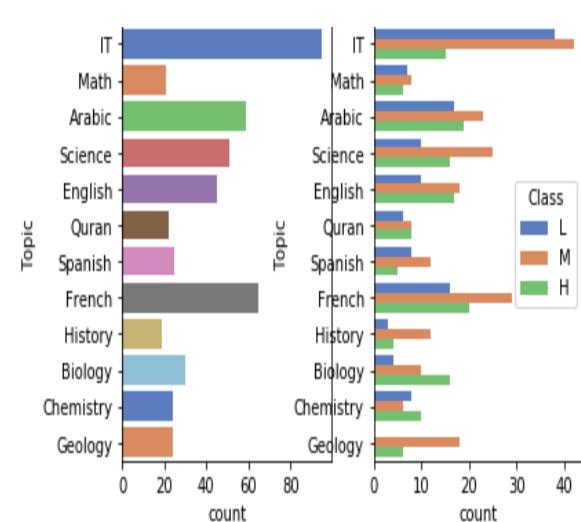
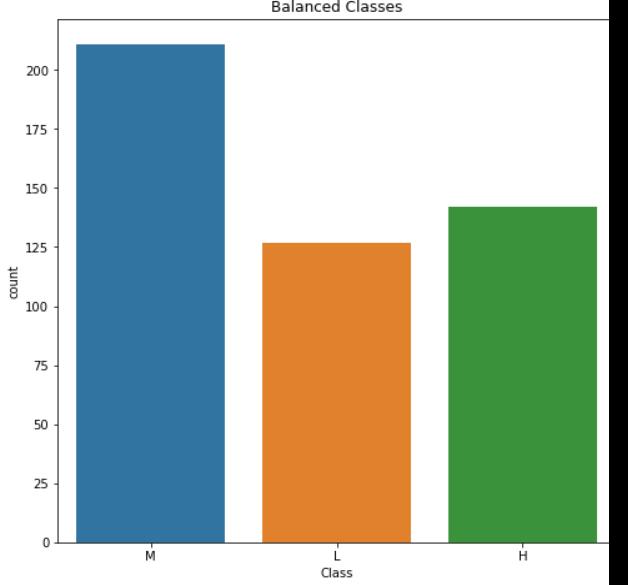
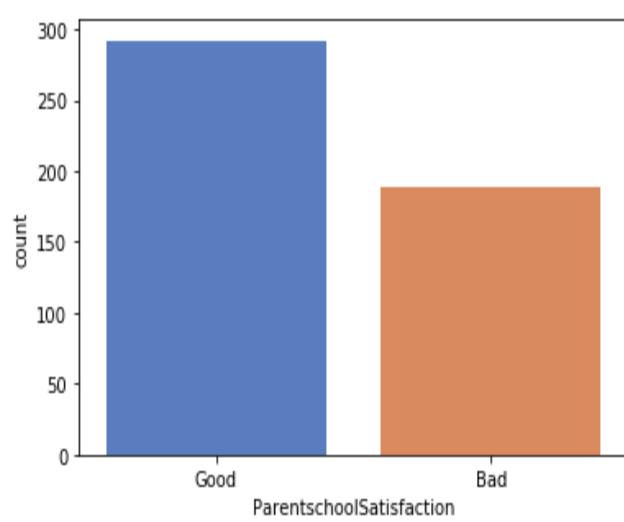
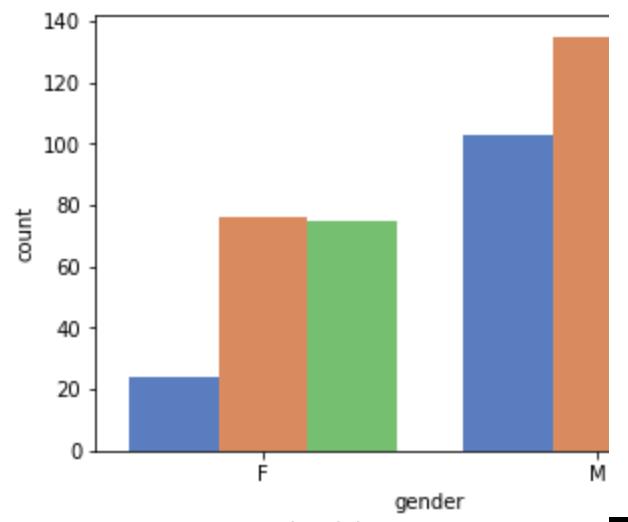
Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

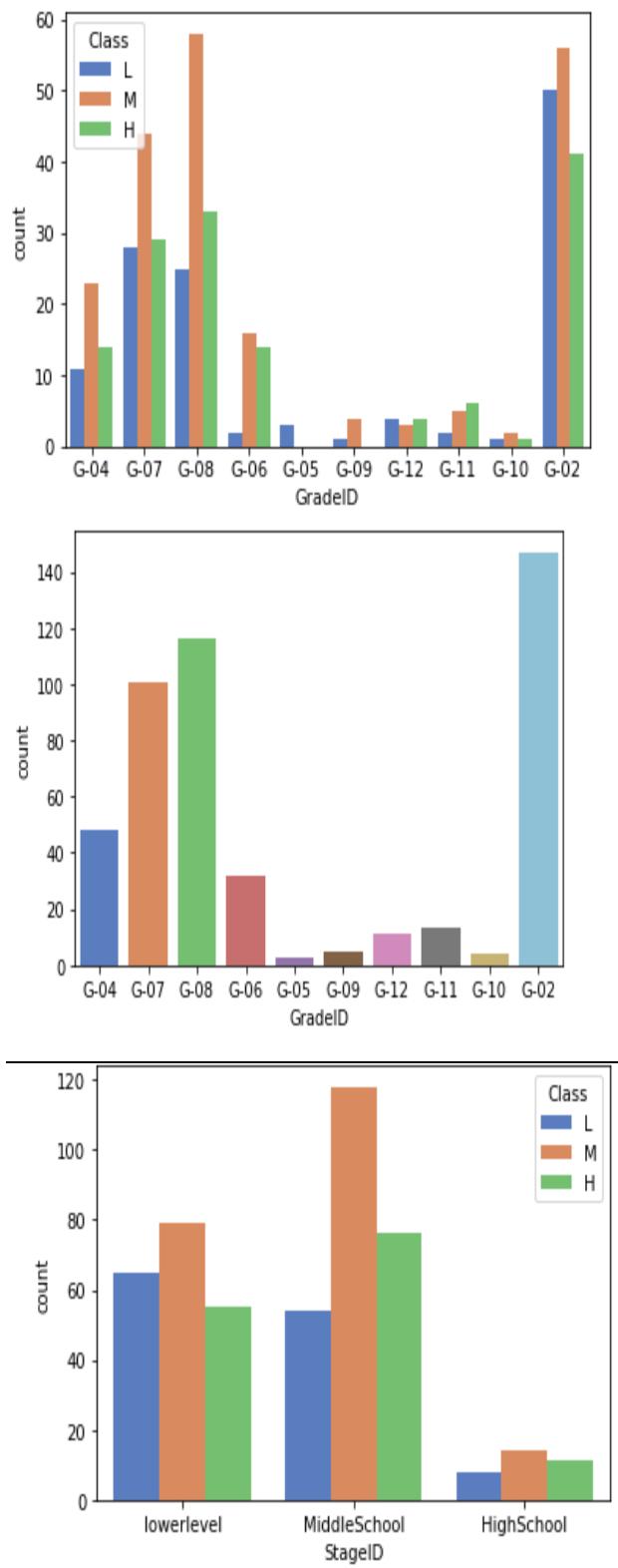
methods of evaluating models in data science, Hold-Out and Cross-Validation to avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

Screen Shots:

EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two





Conclusion

Finally, performance analysis for students are a major problem. It is important that they are countered. The work reported in this thesis indicates the machine learning techniques with supervised learning algorithms to understand the performance of algorithm with respect to student records where we analyses the performance of student and categorized it into three classes as high , average, low with the accuracy of 79% .

Future Work

In the future we provide some technical solution by improve the efficiency of student performance .The user interaction model could be derived for giving the record of student dynamically and it could give staff an alert message about those students who are having low performance . We could build the prediction using Neural Network and can expect improvised results. We can add non- academic attributes along with academics attributes.

References :

- [1] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From data mining to knowledge discovery databases.,” AI Magazine, pp. 37-53, 1996.

- [2] "ÍndiceBrasscom de Convergência Digital," 2015. [Online]. Available: www.brasscom.org.br. [Accessed: 14-May-2016].
- [3] H. Witten, E. Frank and M. A. Hall, Data mining: Practical Machine Learning Tools and Techniques. Burlington: Morgan Kaufmann, 2011.
- [4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," in ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, Jun. 2009.
- [5] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance," in International Journal of Advanced Computer Science and Applications – IJACSA, vol. 2, no. 6, pp. 63-69, 2011.
- [6] J. R. Quinlan, C4.5, San Francisco: Morgan Kaufmann Publishers Inc., 1993. [7] E. L. Ferreira, H. Rausch, S. Campos, A. Faria-Campos, E. Pietra and L. da Silva Santos, "Medical data mining: A case study of a Paracoccidioidomycosis patient's database," e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on, Natal, 2014, pp. 275-280.
- [8] F. H. D. Araujo, A. M. Santana and P. A. S. Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process," in IEEE Latin America Transactions, vol. 13, no. 1, pp. 299306, Jan. 2015.