AN EFFICIENT APPROACH TO DETECT DEPRESSION THROUGH PREDICTIVE ANALYSIS

A Thesis Submitted in Partial Fulfillment of the Requirements
For theDegree of

MASTER OF TECHNOLOGY
In
Software Engineering

By

Km. Sakshi Rastogi
(University Roll No: 1200449006)

Under the supervision of

Dr. Gaurav Kumar Srivastava
Assistant Professor
Department of Computer Science & Engineering
School of Engineering, BBD University, Lucknow
&
Mr. Sunil Kumar Vishwakarma
Assistant Professor
Department of Computer Science & Engineering
School of Engineering, BBD University, Lucknow

CSE Department BBDU, Lucknow

to the

SCHOOL OF ENGINEERING
BABU BANARASI DAS UNIVERSITY
LUCKNOW
May 2022

CERTIFICATE

It is certified that the work contained in this thesis entitled **"AN EFFICIENT APPROACH TO DETECT DEPRESSION THROUGH PREDICTIVE ANALYSIS"** by **Km. Sakshi Rastogi** (Roll No 1200449006), for the award of **Master of Technology** from Babu Banarasi Das University has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

Signature.....
Dr. Gaurav Kumar Srivastava
Assistant Professor
Department of CSE
School of Engineering BBDU
Lucknow

Date: .....

Signature.....
Mr. Sunil Kumar Vishwakarma
Assistant Professor
Department of CSE
School of Engineering BBDU
Lucknow

Date: .....

AN EFFICIENT APPROACH TO DETECT DEPRESSION THROUGH PREDICTIVE ANALYSIS

Km. Sakshi Rastogi (Roll No. 1200449006)

Abstract

Nowadays in this world when Machine is learning by algorithms and performing according to the inputs Python has its unique importance. Depression is said to be a feeling in which a person feels having a low mood and a state of strongly not liking someone/something. It has many effects on the body of the person. The symptom which is recognized as the core of depression is not feeling interested in the works or not feeling pleasure in the things that give joy to them earlier. The goal is to firstly write a review paper on this topic and to give a brief review on how the Depression can be predicted in the body of the person and also present a review on tool that is needed for the system. The primary intention of this research is to carry out a comparison amongst the different Algorithms based on accuracy, precision, sensitivity, F1 score, and Confusion Matrix to find which algorithm gives the best performance on depression datasets. The final goal of this thesis is to provide a Machine Learning Model that will predict depression in the human body.

ACKNOWLEDGEMENT

I would like to place on record my deep sense of gratitude to Dr. Gaurav Kumar Srivastava, Assistant Professor, Department of Computer Science and Engineering, Babu Banarasi Das University, Lucknow, India for his generous guidance, help and useful suggestions.

I express my sincere gratitude to Mr. Sunil Kumar Vishwakarma, Assistant Professor, Department of Computer Science and Engineering, Babu Banarasi Das University, Lucknow, India for his stimulating guidance, continuous encouragement and supervision throughout the course of present work.

I also wish to extend my thanks to Dr. Manuj Darbari, Mrs. Akanksha Singh, Miss. Hina Rabbani, Mrs. Upasana Dugal and other faculty members, colleagues for attending my seminars and for their insightful comments and constructive suggestions to improve the quality of this research work.

I am extremely thankful to Dr. Praveen Kumar Shukla, HOD, Department of Computer Science and Engineering, Babu Banarasi Das University, Lucknow, India for providing me infrastructural facilities to work in, without which this work would not havebeen possible.

Km. Sakshi Rastogi

Babu Banarasi Das University

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ran- Random Forest Classifier

mlp- Multi- Layer Perceptron

et- Extra Trees Classifier

ada- Ada Boost Classifier

dec- Decision Trees Classifier

svc- Support Vector Classifier

cm- Confusion Matrix

acc- Model Accuracy

prec- Precision of Model

rec- Sensitivity of Model

f1- F1 Score of Model

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Depression

Depression is said to be a feeling in which a person feels having a low mood and a state of strongly not liking someone/something. The symptom which is recognized as the core of depression is not feeling interested in the works that give joy to them earlier. This can result in the person having a state of sadness, thinking difficulty, problems in paying attention. It can also lead to an increase or decrease in the diet and sleeping time of the person. In this condition person also experience the feeling of dejection, hopelessness and suicidal thoughts.

Fig. 1.1 Depression

1.1.1 Symptoms of Depression

- Problem in Sleeping
- Loss of Interest
- Increment in Fatigue
- Emotions that are Uncontrollable
- Appetite of a Person Changes
- Weight of a Person Changes

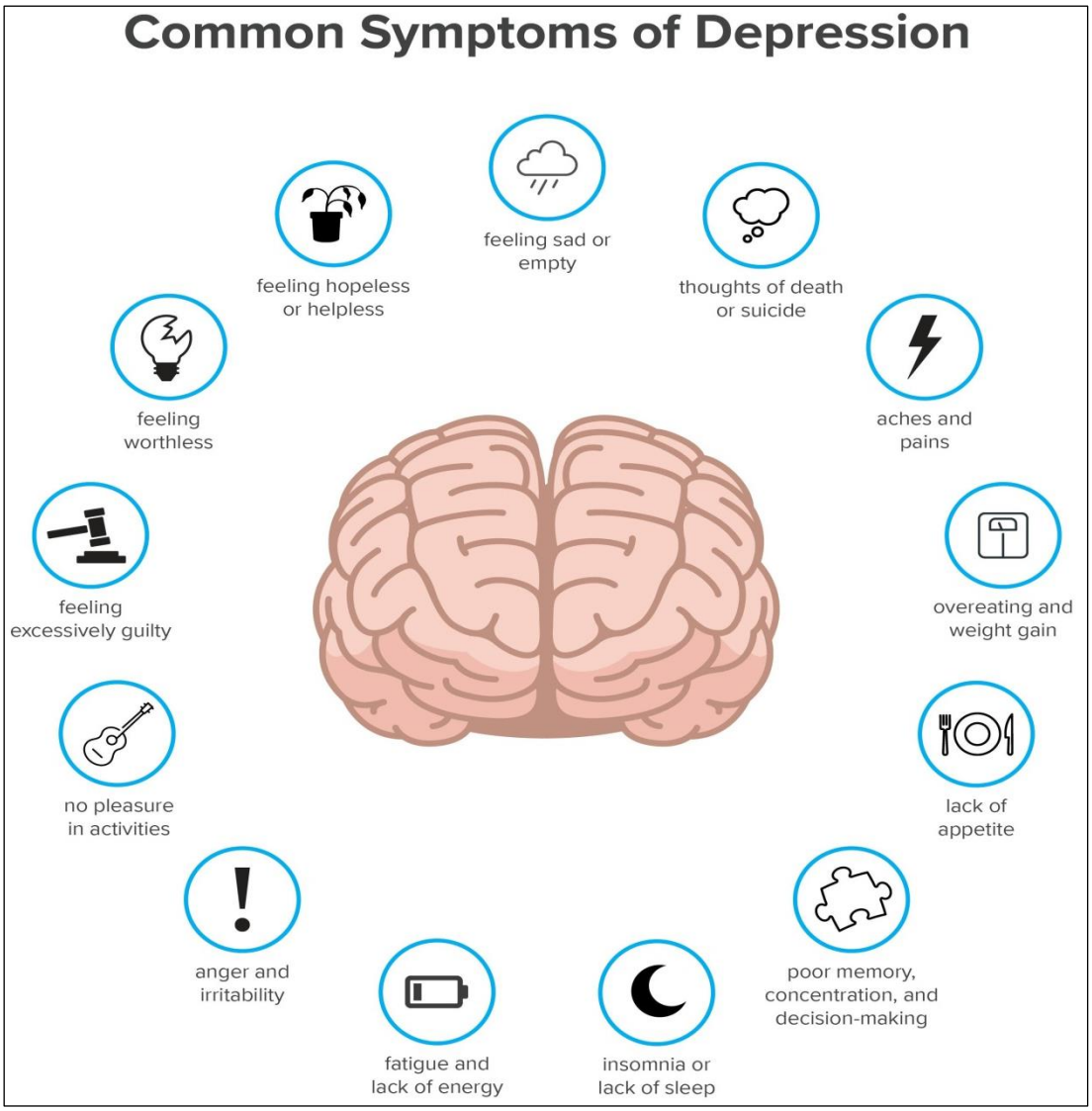

Fig. 1.2 Symptoms of Depression

1.1.2 Causes of Depression

- Genetics
- Brain Chemistry
- Trauma & Abuse
- Personality
- Drugs
- Loneliness
- Chronic Illness
- Social Media
- Lifestyle

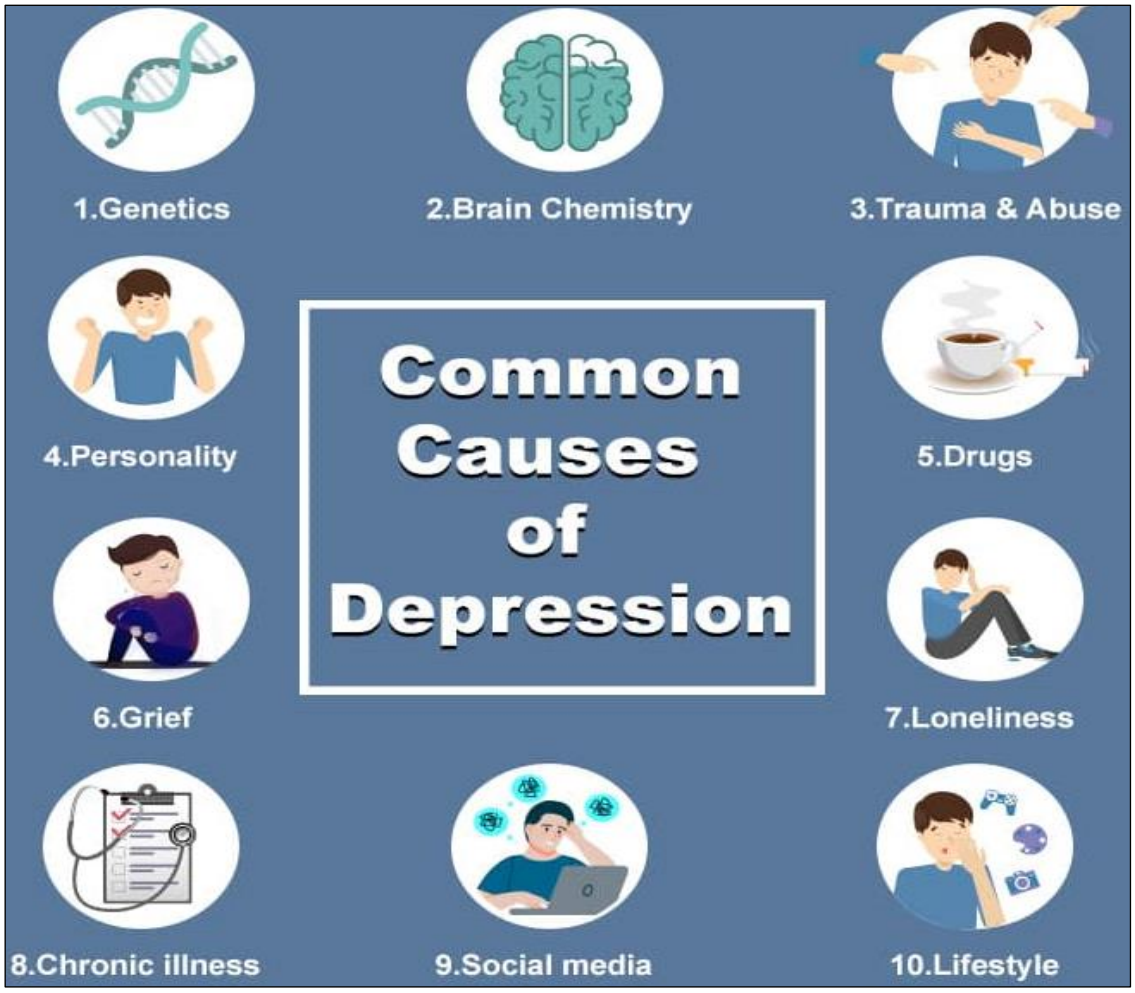

Fig. 1.3 Causes of Depression

1.2 Predictive Analytics

Predictive Analytics is said to be one branch of advanced analytics that is used in making predictions about outcomes of future. It uses historically collected data that is combined with statistical modeling, techniques of data mining and machine learning. It is sometimes associated with data science and big data.

Fig. 1.4 Predictive Analytics

1.2.1 Advantages of Predictive Analytics

- Gaining of Competitive Advantage
- Finding of Opportunities for New Revenue
- Improvement in Fraud Detection
- Increase Utilization of Asset
- Improves Capacity and Quality of Production
- Reduce Risks

1.2.2 Applications of Predictive Analytics

- Used in Accounting
- Used in Business
- Used in Child Protection
- Insurance
- Pharmaceuticals
- Marketing
- Banking
- Healthcare

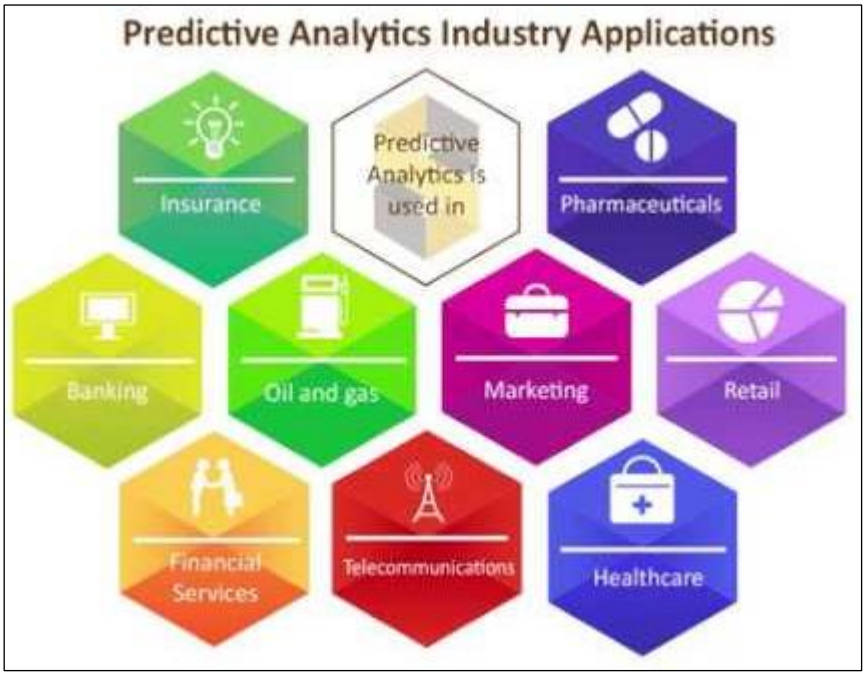

Fig. 1.5 Applications of Predictive Analytics

1.2.3 Limitations of Predictive Analytics

Everything in this world has its own advantages, disadvantages and limitations in the same way Predictive Analytics also have some limitations that are given below-

- The data collected may be incomplete
- If the data is collected from any survey it can't be seen as fully authenticated
- Data from various sources can be different

1.3 Machine Learning

It could be stated in the form of capacity of a machine by which it can learn and respond on the data that is provided to them. This is performed with the help of Artificial Intelligence, Algorithms and Models.

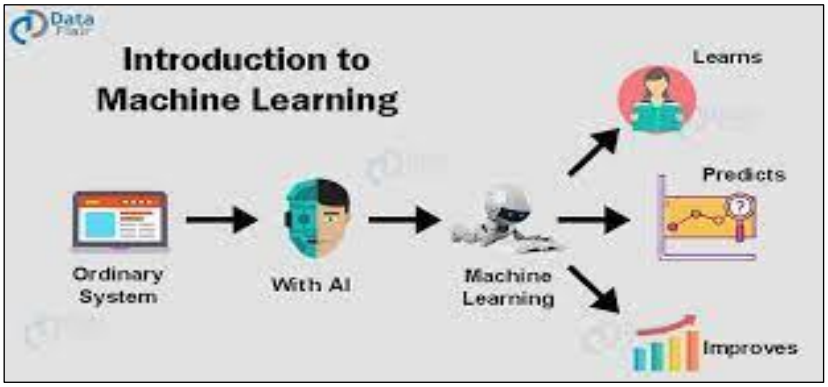

Fig. 1.6 Process of Learning of Machine

Approaches from which Machine can learn are categorized into three parts:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

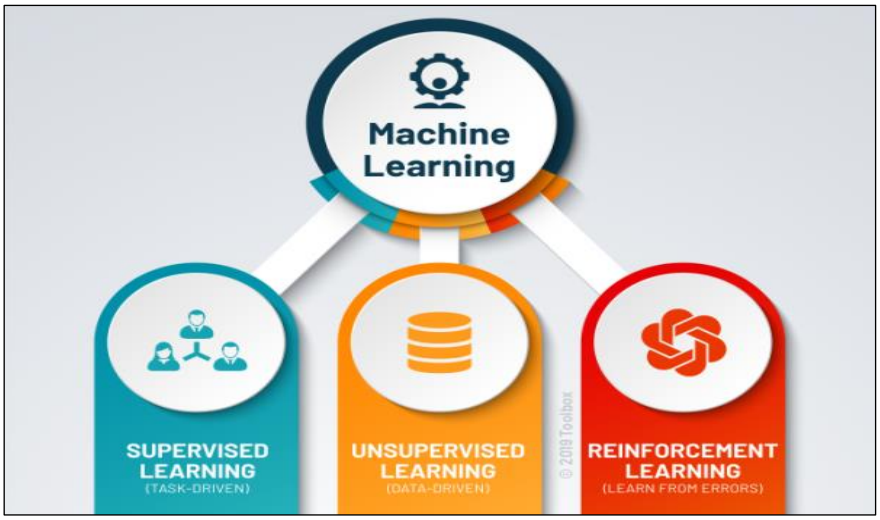

Fig. 1.7 Types of Machine Learning

1.3.1 Supervised Learning

As from name we can easily figure that the Supervised Learning relies upon supervision. This implies that in this type of technique the machines are trained using labeled datasets and further based on this machine gives the required output. The main aim of this learning is mapping of variable that are used for input to the variable that are used for output.

Some of the Application of this type of technique are-

- Image Segmentation
- Medical Diagnosis
- Fraud Detection
- Spam Detection
- Speech Recognition

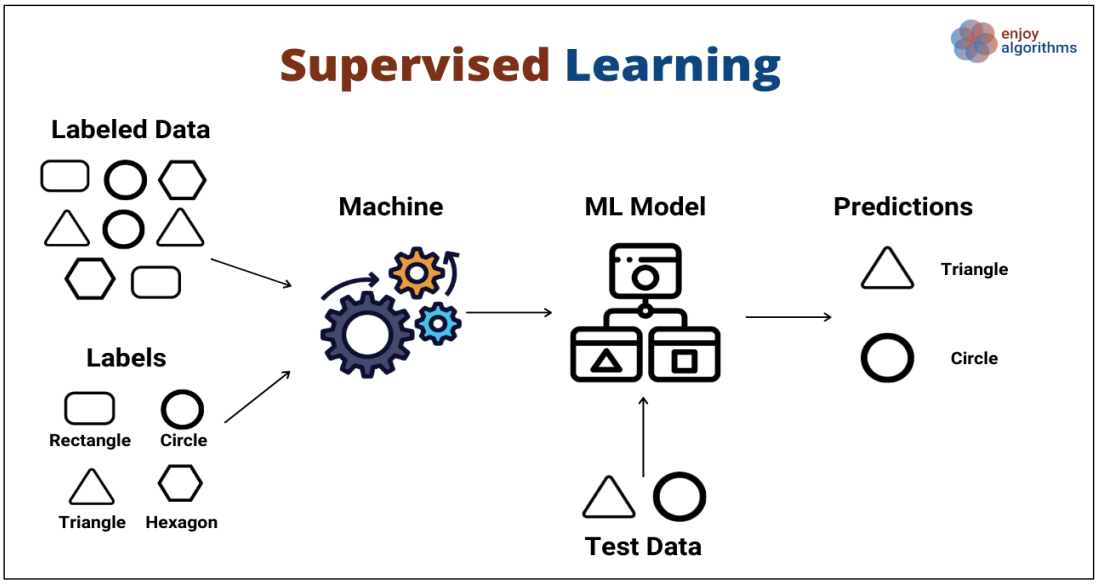

Fig. 1.8 Supervised Learning

This technique could be further divided into two types of problems that are-

- Classification
- Regression

Classification

These types of Algorithms can be utilized in solving problems of classification for which the output variable is in the form of "0" or "1", "Yes" or "No" etc. These types of Algorithms are used in predicting of categories that are given in the dataset that are provided. Some of the applications of Supervised learning that are used in real-world are- Spam Detection, Email Filtering etc.

Regression

Regression Algorithms are used in solving the problems related to regression. In this of Algorithms there is a linear type of relationship in variables that are used for input and output. This type of algorithm is used in predicting the variable that are used for output that is continuous. Some of the applications of Supervised learning that are used in real-world are- Market Trends, Weather Prediction etc.

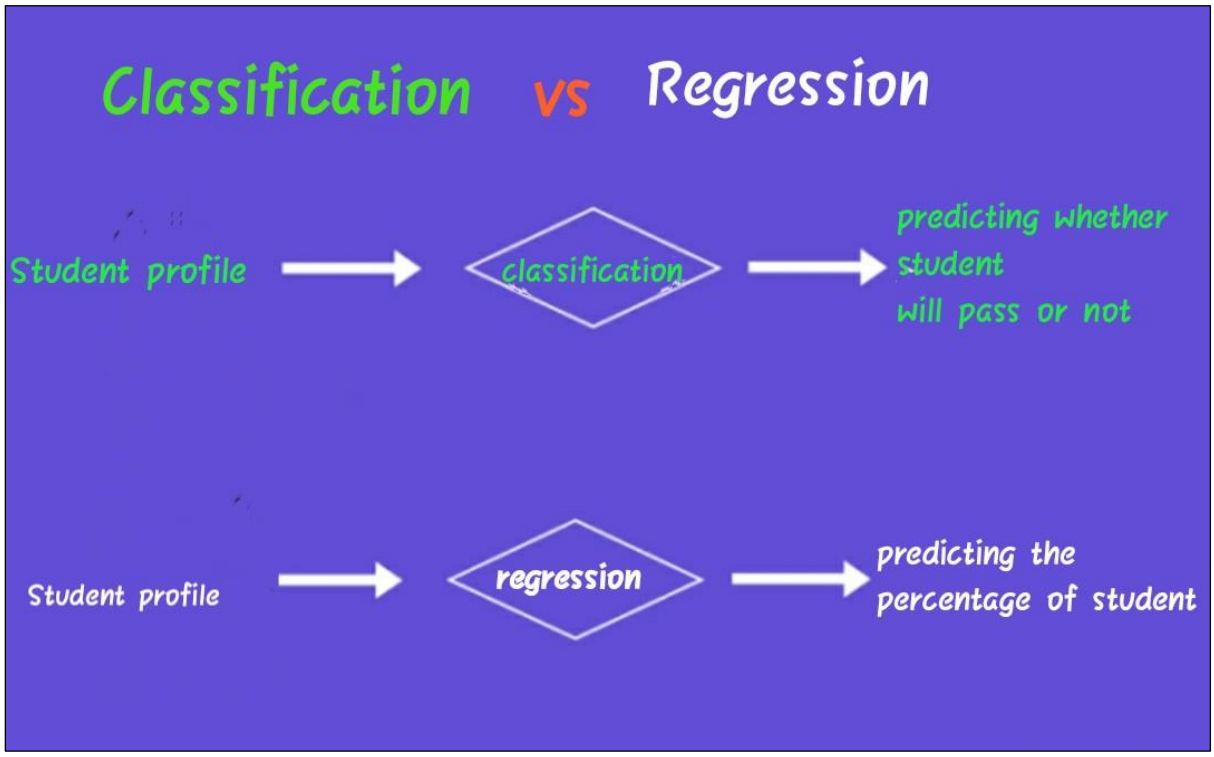

Fig. 1.9 Classification vs Regression

1.3.2 Unsupervised Learning

As from the name it is clearly figured that the Unsupervised Learning has no relation with the supervision. This implies that in this type of algorithms the training of the machine is done with the help of dataset that is unlabeled and further the machine gives the prediction on the same without any requirement of supervision. The main aim of these types of algorithms is to combine the datasets on the basis of certain similarities, patterns and differences that are found in the datasets.

Some of the applications of this technique are-

- Analysis of Network
- Recommendation Systems
- Detection of Anomaly
- Singular Value Decomposition

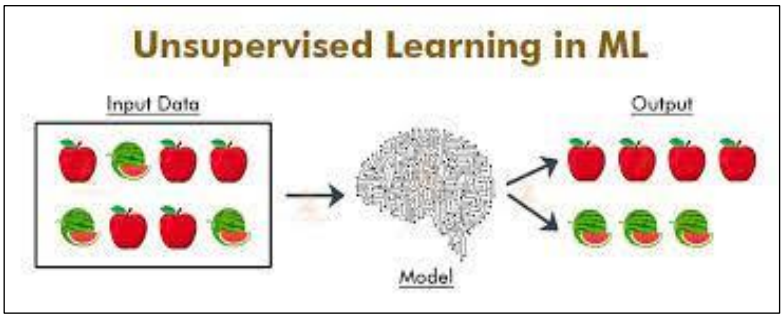

Fig. 1.10 Unsupervised Learning

This technique could be further divided into two types of problems that are-

- Clustering
- Association

Clustering

When we want to find the inherent groups out of the data we use clustering techniques. It is a way in which the objects from the data are grouped into cluster. It is done in the way that the objects that have most similarities exist in one group and the objects having very few or no similarities exist in the other groups.

The real-world example of this type is- Grouping of the Customers on the basis of their behavior of purchasing.

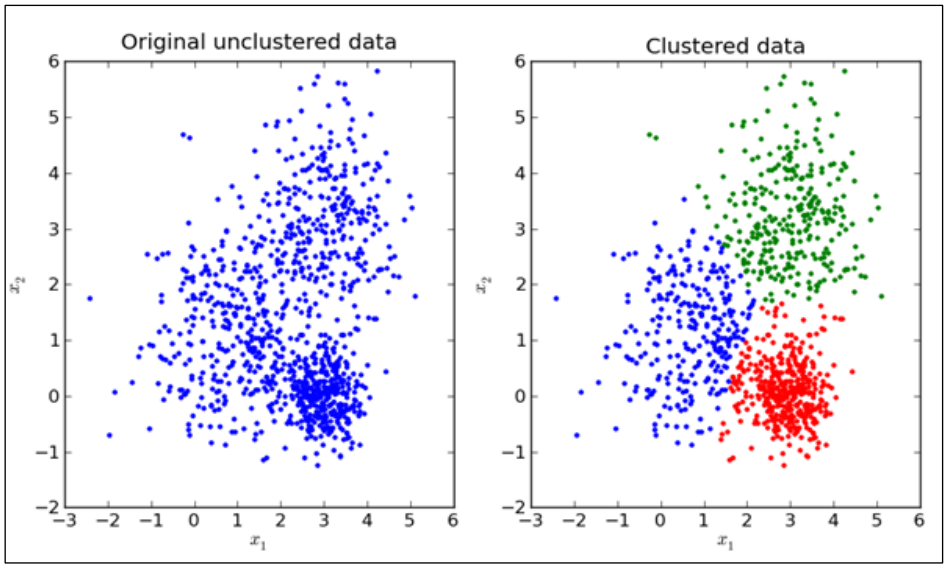

Fig. 1.11 Clustering

Association

Association is the type of Unsupervised Learning Techniques, in which interesting relations are found amongst variables that are in dataset that is of large amount. The final goal of Association Technique is collecting the dependency of one item on another item and mapping them in the manner that they can give the profit up to the maximum level. Some of the applications of Supervised learning that are used in real-world are- Market Basket Analysis, Web Usage Mining etc.

According to this research, Unsupervised Learning Algorithms are not advisable for the Predictions.

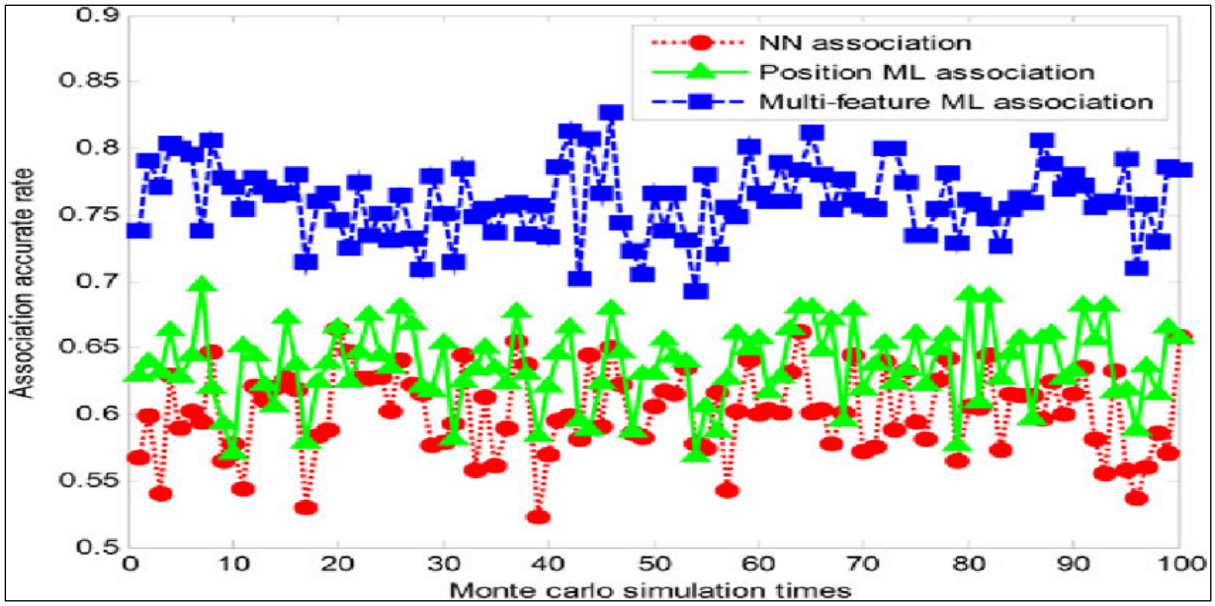

Fig. 1.12 Association

1.3.3 Reinforcement Learning

Job of above kind of learning is to work on feedback-based process. In this an agent of AI by default explore around its surroundings by hitting & trailing, learning from experiences, takes actions and further improves its performance. In this type of learning the agent who is working gets the reward for every action that is good and also gets the punishment if the action is bad.

The categories of Reinforcement Learning are as follows-

- Positive Reinforcement Learning
- Negative Reinforcement Learning

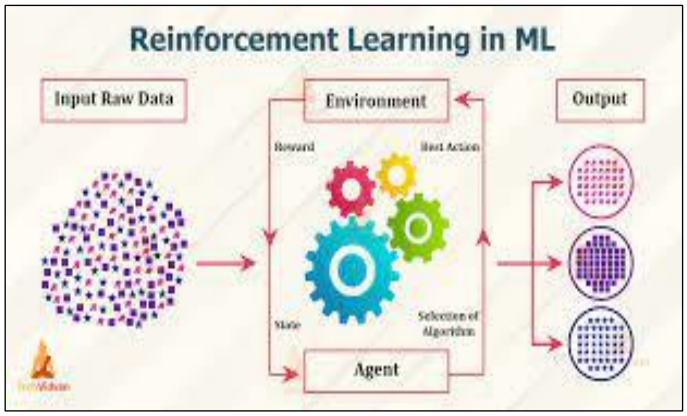

Fig. 1.13 Reinforcement Learning

The real-world use cases of Reinforcement Learning are-

- Video Games
- Resource Management
- Robotics
- Text Mining

According to this research, Reinforcement Learning Algorithms are not advisable for the Predictions.

1.4 Statement of the Problem

Depression is the most growing problem nowadays. On an estimate it affects the 3.8% of the total world person's. Approximately, in the world 280 million people are suffering from depression. When depression comes to its worst stage it leads to suicide. Every year, Over 700000 person are dying from Depression.

There are many Machine Learning Algorithms that can be used in predicting the depression from the pre-defined datasets. The main problem arises when we start doing predictions are that which algorithm will give the best predictions.

1.5 Objectives of Study

The primary scope of this research is to do comparative analysis amongst the different Machine Learning Algorithms like Random Forest Classifier, Extra Trees Classifier, Ada Boost Classifier, Decision Tree Classifier and Multi-Layer Perceptron, Support Vector Classifier on the basis of accuracy, precision, sensitivity, F1 score and Confusion Matrix with the aim to find that which algorithm gives the best performance on my depression data. The final goal of this research is to design a Machine Learning Model that will be used in predicting the Depression in the human body.

1.6 Limitations of Study

- It will only work on pre-defined datasets.
- Applications are need to be trained specially.
- They do not learn incrementally.
- Required lengthy training of batches.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In first part of thesis, we have performed the literature for the work that deals with the study related to the problem we were working that is Depression. There were many works that are done on this work. We have used most the work as a reference for our study. This gives the motivation for performing more work on this topic.

In the first section of this chapter, we have discussed all the literature review that was done for the writing review paper. In the second section of this chapter, we have discussed all the literature review that was done for writing the implementation paper.

2.2 Importance of Review Related Literature

The literature review is the most important step in performing any type of thesis. It also helped us in many ways that are as-

- Establishing the Research Context
- Identifying the Theoretical Framework
- Identifying the Quality of Research that is performed previously
- Comparing the Studies that are done
- Identifying the Problems on which we can do study
- Provides a Map for Future Work
- Designing New Studies
- Understanding the New Topics

2.3 Related Study

Researcher's objective was to design a model that can be used in predicting the Depression from 4-17 years old aged children and adolescents. Also, they say to evaluate the results in order to determine which performs the best. Further there aim was to associate the related issues of families. Their conclusion said that Random Forest based prediction model performs more accurately. Some authors have provided the general introduction of various Techniques that are based on Machine Learning which are used in the field of logistics to give help to the researchers who further use it. From their research Supervised Learning can be considered as the most prominent way in which the problems of Machine Learning can be solved. Some have proposed an approach in which they use Long Short- Term Memory based Recurrent Neural Network for the purpose of identifying symptoms of Depression. They performed on the dataset which was of young age people of Norway. The conclusion given by them was that if we use their proposed approaches we get 98% and 99% mean prediction performance. Some have used the data from 2014 and 2016 Korea National Health and Nutrition Examination Surveys. They have used 2014 dataset for training purposes and 2016 dataset for testing purposes. The result they got was that when they use 37 variables they got 0.822 accuracy and when they use 13 variables they got 0.828 accuracy. Some author's research revolves around evaluating different models that are used in predicting outcomes of psychotherapy for Depression. In the conclusion they said that models they used for prediction did not perform accurately for treatment outcomes of Depression. Some authors said that Machine Learning helped in many ways like medical research and also to doctors in the whole world. Machine Learning has helped them in Pattern Recognition that comes in the patients that can be helped in early detection of any disease. According to them, K-Nearest Neighbor gives accuracy as 89.06%. They have discussed different text analytical approaches

for predicting Depression in users who shared their ideas over websites. They have presented summary of results that they obtained from some researchers literature review.

This section is concerned with all the previous works that are done in this field. It presents a study that gives the comparison between different Algorithms of Machine Learning. They have used algorithms as- Logistic Regression, Random Forest, XG Boost, Support Vector Machine, Ada Boost, K-NN and Decision Tree. They have performed their research on the prediction of liver disease at an early stage. They have compared the above Machine Learning Algorithms on the basis of Accuracy, Precision, Recall, F1 Score, an area under curve and Specificity. They have collected their datasets from UCI Machine Learning Repository. There result in states that Random Forest Algorithm performs the best in terms of accuracy with 83.70%. Random Forest also performs well in the terms of other parameters too. So, they concluded Random Forest as the best algorithm that can be used in predicting Liver Disease. It describes about the machine learning techniques principles and he also described the use of them in the domains of real-world applications. They further describe the challenges and potential they need to perform in their research. On the basis of their goal, they shortly discussed how the methods of machine learning are being used in providing an appropriate way in solving the problems of the real world. The conclusion was that machine learning is built upon the data that is provided to the algorithms for learning purposes and the performance provided by them. It has the algorithms of Machine Learning for doing the predictions on anxiety, depression and stress in their paper. They have gathered their data by making the questionnaire related to their topic. This consists of the data of several cultures and communities which are employed and unemployed. They realize in their research that classes they made were imbalanced at the time when they start making confusion matrix. So they measure f1 score to identify the best accuracy model. They find that Random Forest Classifier as the best model. The conclusion was that the f1 score is the important aspect in

finding the best accuracy model. It describes the use of various kinds of machine learning. It also merges the results of the analysis that comes from all the algorithms that were used for performing their research. Their main purpose was to increase the awareness of Machine Learning among the persons. Their conclusion presents that it is necessary for the Machine Learning model to continuously grasp from the past doing that come from countries that are developed, set up algorithms of machine learning mostly for the making enterprises in domestic areas and providing help of the economy in developing industry. It presents about the survey on how machine learning can be used for providing investigation on depression. The methods which they use in their systems are based on the method of detection via posts on social media, syntax and semantic analysis of the person's emotion in order to predict the depression levels of different age groups. Some have performed comparative research on four Algorithms of Machine Learning. For the purpose of reducing attributes, they used CFSSubsetEval. They have collected their datasets from OASIS-Brains.org. They finally concluded as J48 is the best algorithm for the purpose of detecting Dementia.Some have conducted their research on various Algorithms of Machine Learning with the aim of finding the effectiveness. The datasets that were used in this research were from different types of clinics. In these datasets are small, medium and large that can be accessed publicly. The comparison between algorithms was done on the basis of the requirement of accuracy and time in training and testing algorithms. The result implies that K-Nearest Neighbor performed well amongst all the algorithms used. It also presents that social network data gives the opportunity to work on the user's moods and attitudes when they convey messages with the use of social media. The data for the analysis was on the Facebook data that they collected from an online public source. They gave their analysis on 7146 Comments on Facebook. They got the conclusion as 54.77% depressive person who conveys between mid-night to mid-day & 45.22% depressive person who conveys between mid-days to mid-night.

Some have done comparative research amongst some popular Algorithms of Machine Learning. They have used two datasets in order to provide the best efficiency. They have collected all information including datasets from UCI Machine Learning storehouse. There first dataset contains 6500 rows & 13 columns & second dataset contains 1055 rows & 13 columns. There result shows that Support Vector Machine performs the best accuracy of 99.38%.

CHAPTER 3: PRESENT WORK

3.1 Introduction

In this chapter the overview of Methodology of Research is given. It also presents the review on how this research is performed from the first step to last step. It describes the overview of the tools and libraries that we used.

3.2 Research Methodology

The model that we have proposed for this research is described in the Figure 3.1.

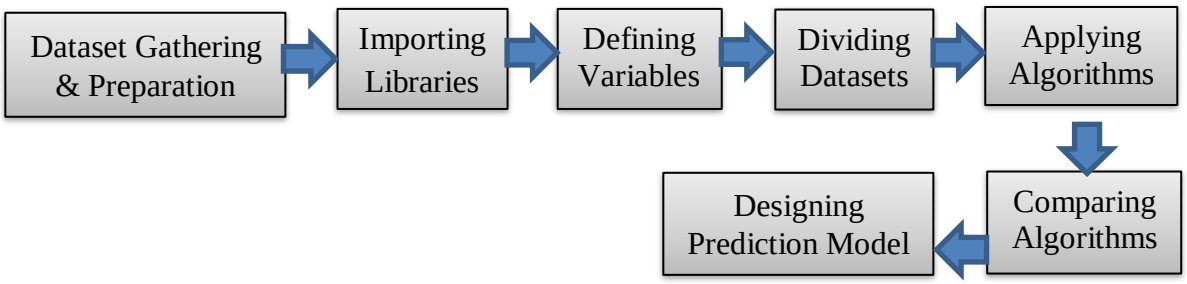

Fig. 3.1 Proposed Methodology

3.3 Dataset Used

In developing a Machine Learning Model or performing any type of research in Machine Learning the first and most important step is to gather the datasets. We have taken the Depression datasets from an online portal. There were 1290 rows and 27 columns in the original dataset. Table 2 describes the original Depression Dataset.

Table 3.1 The Original Dataset

1.	Timestamp	Age	Country		obs_consequence	Comment
2.	8/27/2014 11:29	37	United States	-----	No	NA
3.	8/27/2014 11:29	44	United States	-----	No	NA
⋮	⋮	⋮		⋮	⋮	⋮
1259.	11/30/2015 21:25	46	United States	-----	No	NA
1260.	2/1/2016 23:04	25	United States	-----	No	NA

Table 3 describes the final dataset that we have used for performing this research. In the column ‘Gen’, 0 represents Female and 1 represents Male and in all the other columns 0 stands for No and 1 stands for Yes.

Table 3.2 The Final Dataset

1.	Age	Gen	family_history		phys_health_consequence	Target
2.	37	0	0	--------	0	1
3.	44	1	0	--------	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1248.	32	1	0	--------	0	0
1249.	36	1	0	--------	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1259.	46	0	0	--------	0	1
1260.	25	1	1	--------	0	1

3.4 Classifiers Used

3.4.1 Random Forest Classifier

The first algorithm that was used for this was created in 1995 by Tin Kam Ho. An extended version of this algorithm was further created by Leo Breiman and Adele Cutler which was registered as “Random Forests” in 2006. For the tasks of classification, the output is given in the form of class that is selected by most of the trees. For the tasks of regression, the output comes in the form of mean or average prediction that is returned by the every individual tree. This type of algorithm is most frequently used for black-box models in businesses because they provide reasonable kind of predictions.

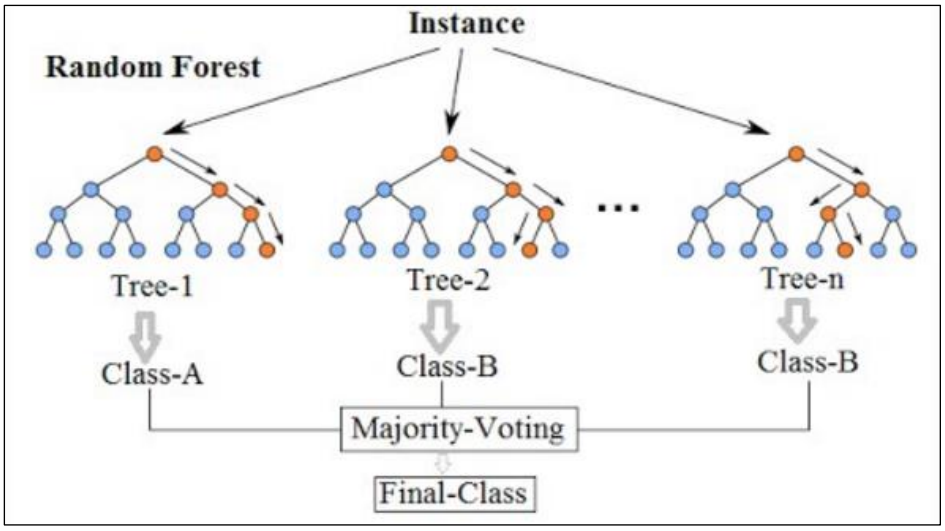

Fig. 3.2 Random Forest Classifier

3.4.2 Multi- Layer Perceptron

Multilayer perceptron (MLP) is a class that contains feed forward Artificial Neural Network. This term is sometimes used for any feed forward ANN and sometimes it is used for the networks that are combination of multiple layers of perceptron's. It consist of at least three types of layers- input layer, hidden layer & output layer. This Algorithm use Supervised type of learning Technique that is called back propagation for the purpose of training.

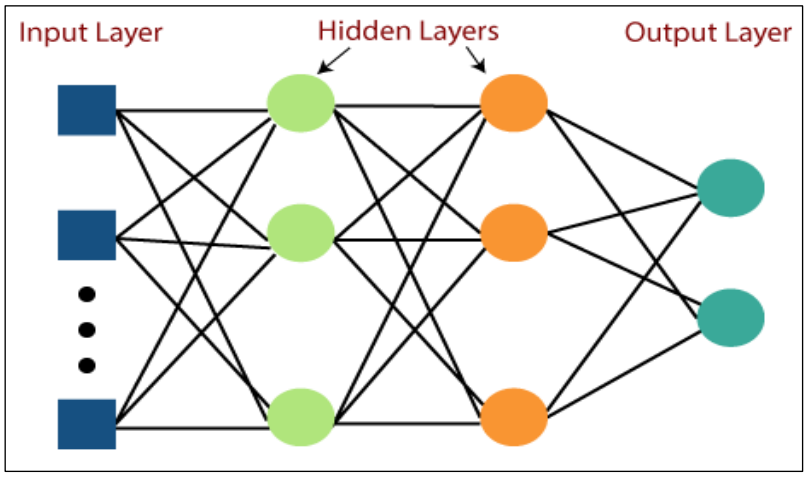

Fig. 3.3 Multi-Layer Perceptron

3.4.3 Extra Trees Classifier

This classifier is often called as Extremely Randomized Trees Classifier. Extra Trees Classifier is a method in which results of different kinds of co-related decision trees are placed in the forest in order to provide the result of their classification. In this type of Classifier each and every decision tree is being created from the dataset that is given originally. After this at each node of testing every tree is given a random sample of k features amongst all the features provided and further each decision has to select the feature that is best suited in a way of data splitting.

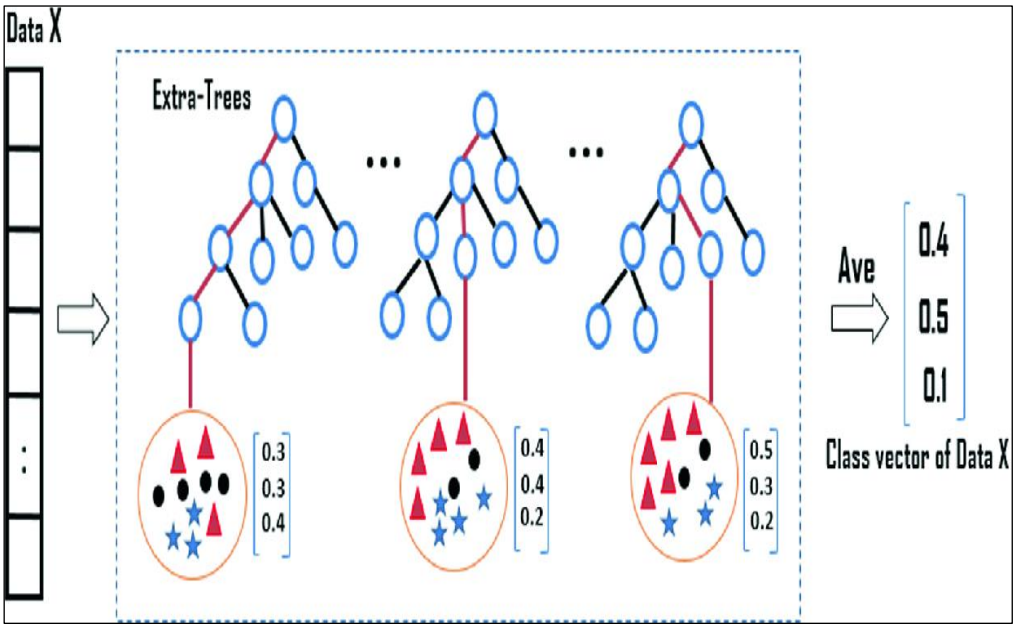

Fig. 3.4 Extra Trees Classifier

3.4.4 Ada Boost Classifier

It is one of boosting ensemble classifier that was given by Yoav Freund and Robert Schapire in 1996. It mixes different type of multiple classifiers in a way of increasing the classifier's

accuracy. This classifier sets weight for classifiers and training in each iteration in order to ensure the predictions accurately. Ada Boost has to meet the following conditions:

- Each classifier has to be trained interactively for all the different weighing examples of training.
- For every iteration, this classifier try to provide a fit that is excellent for these examples through minimizing the errors of training.

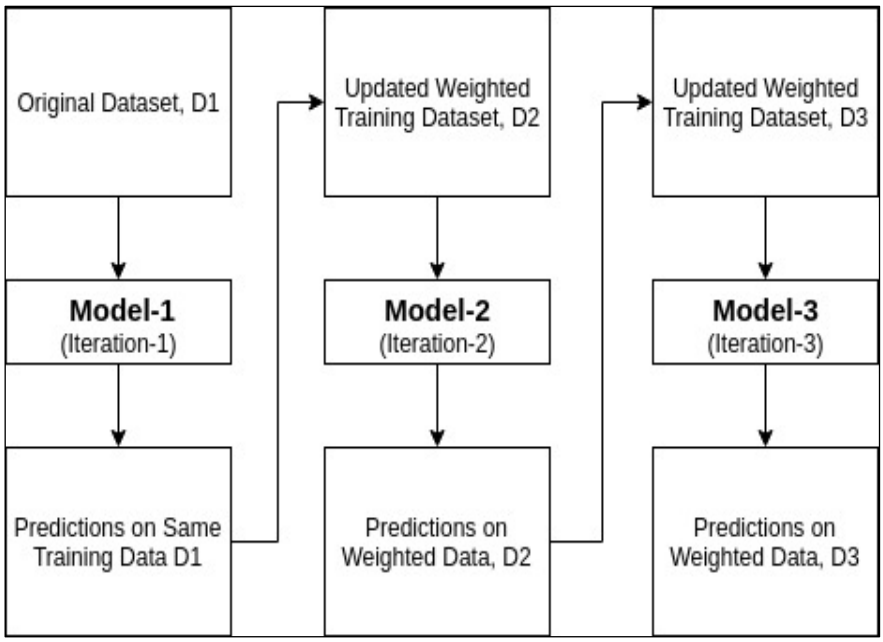

Fig. 3.5 Ada Boost Classifier

3.4.5 Decision Trees Classifier

It is one kind of Supervised Learning which could be used in both of the purposes Classification and Regression but most of the times it is used in solving different types of problems that are related with Classification. It can be said as the tree which is structured in which the internal nodes imply the dataset features, branches implies rules of decision and leaf node implies the outcomes. For building of a tree, CART (Classification and Regression

Tree) Algorithm is being used. It simply works on the answer (Yes/No), and after that it splits into sub trees.

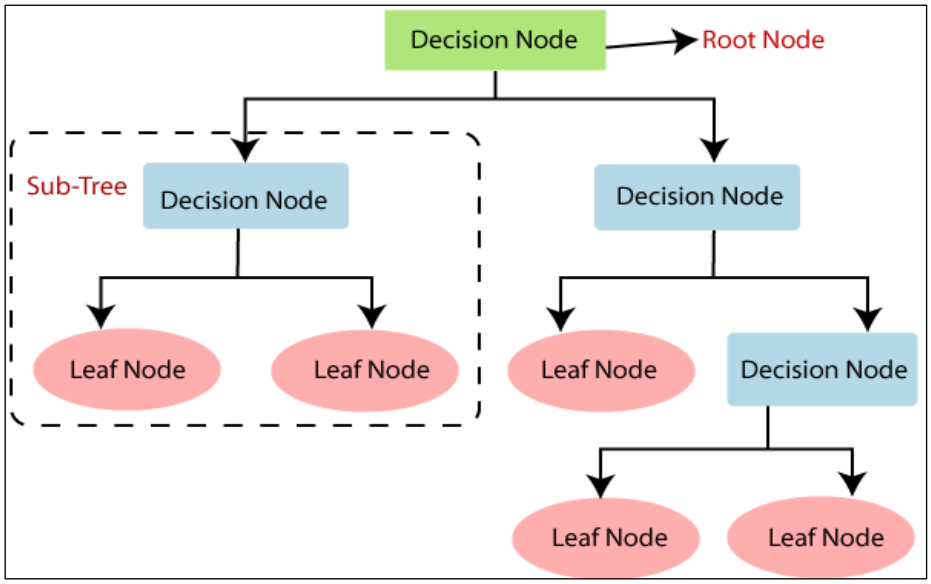

Fig. 3.6 Decision Tree Classifier

3.4.6 Support Vector Classifier

Support Vector Classifier is one kind of Supervised Learning which could be used for both the purposes Classification and Regression. The final aim of the above classifier is to create the line that is best or boundary of decision that splits n-dimensional space into different classes in order to put new data point in category that is correct for the future use. It chooses the points/vectors that are extreme in making the hyperplane.

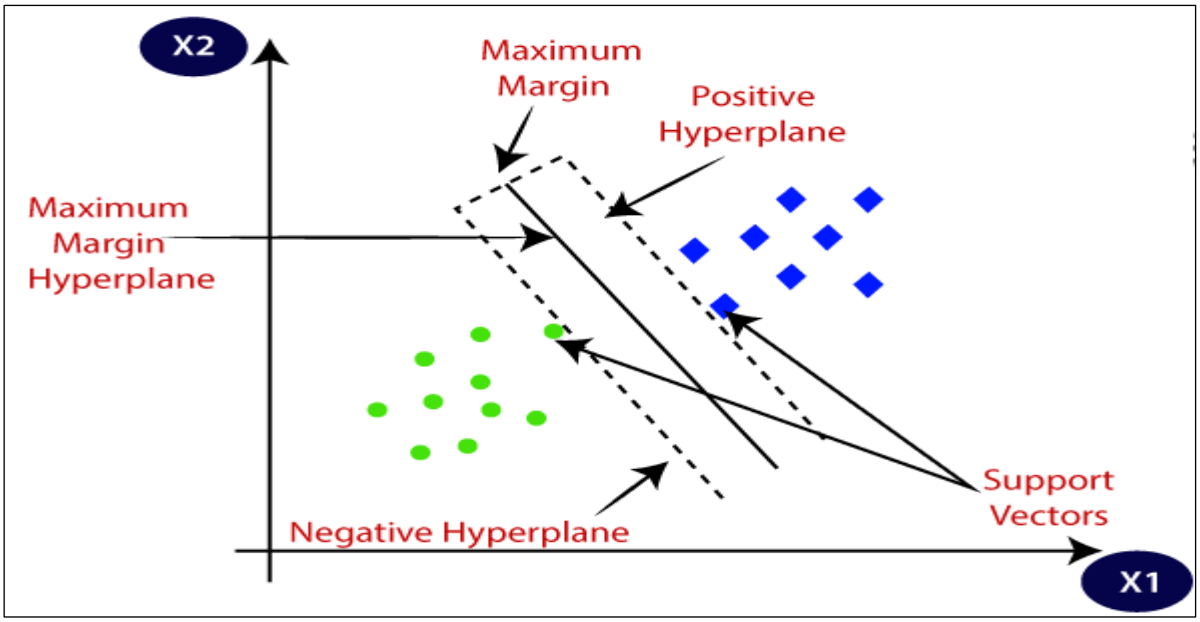

Fig. 3.7 Support Vector Classifier

3.5 Parameters Used

The parameters that I have used for this research for doing Comparative Analysis are- Accuracy, Precision, Sensitivity, F1 Score and Confusion Matrix. The formulas of all these are presented in the further description.

3.5.1 Accuracy

Accuracy is the measurement that is used for the evaluation of classification.

$$\text{accuracy} = \frac{True\ Positive + False\ Positive}{N}$$

N= True Positive+ False Positive+ True Negative+ False Negative

3.5.2 Precision

Precision is the recognition of the positive answers that are correct.

$$\text{precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

3.5.3 Sensitivity

Sensitivity is the measure of the ability of classifier in the terms of guessing the greatest possibility of all positive responses.

$$\text{sensitivity} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

3.5.4 F1 Score

It is defined as the average that is calculated between Precision and Sensitivity.

$$\text{f1 score} = \frac{2(Precision * Sensitivity)}{Precision + Sensitivity}$$

3.5.5 Confusion Matrix

It is defined as table-like layout which is used for viewing the performance of the algorithm. It consists of two dimensions- Actual Value and Predicted Value. It also consists of two rows and two columns that are defined as True Positive, False Positive, False Negative and True Negative.

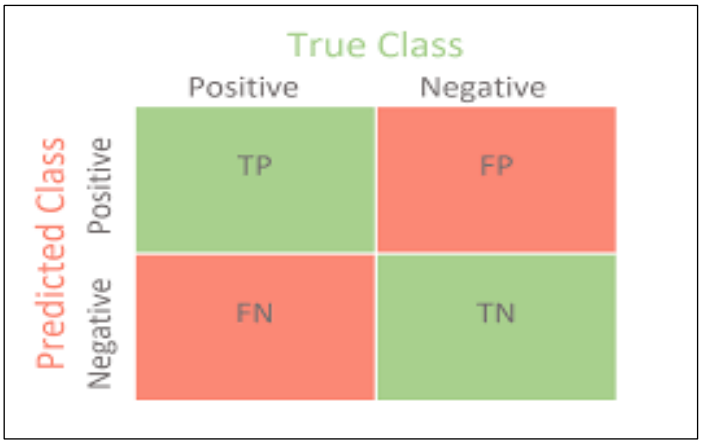

Fig. 3.8 Confusion Matrix

3.6 Tool Used

The tool that we have used for this research is Jupyter Notebook. It was formed in February 2015. The main aim of Project Jupyter is to make software's that are open-sourced, open-standards and services for computing interactively amongst different programming languages. It provides the reference for three core programming languages that are Julia, Python and R. The version on which we have performed our research is Jupyter Notebook 6.1.4.

Fig. 3.9 Jupyter Notebook

We have performed all of our work on Python 3. We have used many Python Libraries for doing the work on this research. Python is a general purpose interpreted, interactive, object oriented and high level programming language. It was designed by Guido Van Rossum and was released in 1991.

Fig. 3.10 Python

3.6.1 Uses of Python

- Web Development
- Software Development
- Mathematics
- System Scripting etc.

3.6.2 Advantages of Python

- Ease in Reading, Learning and Writing
- Productivity Improved
- It is Interpreted Language
- It is Typed Dynamically
- It is available as Free and Open- Source
- It has vast Library
- It is Portable

3.6.3 Disadvantages of Python

- Speed is Slow
- Non Efficient Memory
- Database Access
- Errors in Runtime

3.7 Libraries Used

In normal world, A library is said to be a composition of books or is a kind of room where we can place or store many kinds of books. Similarly in the world of programming, a library is a composition of codes that are pre-compiled that can be used in the further programs. Other than these codes, library may contain various types of documentation, data configuration, templates of messages etc.

We have used various Libraries of Python that were required for this research. The libraries and their discussions are given as follows-

3.7.1 pandas

They are an important library that is required by data scientists. It gives flexible type of data structure and various analysis tools. It gives its best in Analysis of Data, Manipulation of Data and in Data Cleaning.

Fig. 3.11 Pandas Library

3.7.2 numpy

The name stands for **Numerical Python.** It is the library that is used commonly. It is one of the popular library of Machine Learning that supports matrices of large kinds and data that is multi-dimensional. The libraries like TensorFlow use this library internally for performing several kinds of operations on tensors.

Fig. 3.12 Numpy Library

3.7.3 matplotlib

This library is totally responsible for plotting of numerical data and this is the reason this library is highly used in analysis of data. It is also an open- source library and plots various figures like pie-charts, histograms, scatterplots, graphs etc.

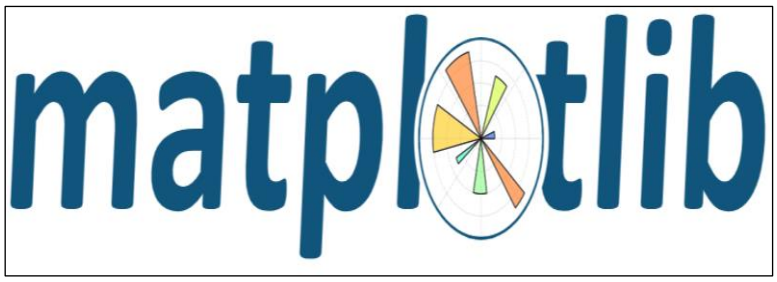

Fig. 3.13 Matplotlib Library

3.7.4 seaborn

Seaborn is the important library of Python. It is one of amazing library that is used for the visualization of statistical graphics. It consists of most amazing styles and color palettes by default for the purpose of making statistical plots look more attractive. It is built on the features of matplotlib and is very closely related to data structures that are from pandas.

Fig. 3.14 Seaborn Library

3.7.5 Scikit-Learn

It is the most useful and one of the robust libraries of Machine Learning that is being used in Python. It gives a selection of efficient tools for the purpose of Machine Learning and Statistical Modeling. This library is written largely in Python and is built upon Numpy, Scipy and Matplotlib.

Scikit Learn has many modules that can be used for variety of purposes. The modules of Scikit Learn that we have used are as follows-

- sklearn.model_selection
- sklearn.ensemble
- sklearn.neural_network
- sklearn.tree
- sklearn.neighbors
- sklearn.metrics

Fig. 3.15 Scikit-Learn Library

3.7.6 warnings

They are given to warn the developer of problems that are not expected necessarily. Usually, it occurs at the time when some functions or keywords is not being used at the time. A warning that comes in the program is very different from an error.

3.8 Working of the Study

3.8.1 Importing Libraries

Python has plenty of libraries that can be used by the developers for performing different kinds of research like on Machine Learning Models, Robotics etc. All the libraries are predefined and are easily available we just have to import them and use them according to our requirement. We have done the same we have imported the libraries like pandas, matplotlib, sklearn etc. We have used the following libraries-

1. Pandas- We used this library for reading the datasets.
2. Matplotlib, Seaborn- We have used the seaborn library for plotting heat map and matplotlib library for plotting the labels, giving the titles etc. around the heat map.
3. Numpy- We have used this library for designing the prediction model.
4. Sklearn- We have used many modules of this library in this research paper. We have used the modules like model_selection, ensemble, neural_network etc.

3.8.2 Defining Variables

Defining the variables that are included in the creation of the Model is also the most important step as the Model fits only on variables.

3.8.3 Dividing Datasets

It is necessary to divide the datasets into Training Module and Testing Module so that we can perform the works on Training Module and make predictions on Testing Module. We have divided the whole datasets into Training and Testing Module. We have divided the dataset

into 80-20 ratio that means our Training Module contains 80% and Testing Module contains 20% respectively.

3.8.4 Applying Algorithms

The main step in this research is to apply algorithms. We have worked on some of the Machine Learning Algorithms like Random Forest Classifier, Extra Trees Classifier, Ada Boost Classifier, Decision Tree Classifier and Multi-Layer Perceptron.

3.8.5 Comparing Algorithms

The pre-final step in this research is to compare the above-mentioned models. We have compared all the algorithms on the basis of accuracy, precision, sensitivity, F1 score and Confusion Matrix to find the best algorithm amongst all the models.

The another parameter that we are giving for comparison is the Confusion Matrix. Confusion Matrix is divided into two rows and two columns. It is plotted using matplotlib and seaborn libraries. We have done this in Three Steps. We have plot the Confusion Matrix of Two Algorithms at a time and followed by the other Algorithms.

3.8.6 Designing Prediction Model

The final step in the study is to evaluate a model that tries to predict depression in the human body. This prediction model works on the dataset that is provided to the model. This model has been made with the help of library named as numpy. The input contains the data in the same manner that is defined in the dataset. It implies that the 1st value is the Age of the Person, 2nd is the Gender in the form of 0 and 1, 3rd defines that if the person is having any

Family History of Depression?, 4th is asking that if the person is working in any tech company?, 5th is asking that if the person is going through any treatment?, 6th is asking if the person is working remotely? 7th is asking if the person has mental health consequence? 8th is asking if the person has physical health consequence? and finally 9th is the target variable that predicts the depression in human body.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter gives the overview of results and the discussions that comes from the research. It gives the comparison table, analysis of Confusion Matrix etc.

4.2 Results and Discussions

The application of all above mentioned machine learning methods with all parameters- accuracy, precision, sensitivity, F1 score & Confusion Matrix is described in Table 3. In the case of Confusion Matrix, the first value represents False Negative value, second represents False Positive value, third represents True Negative value & fourth represents True Positive value respectively. From the below table, it has been clearly shown that Ada Boost Classifier and Multi-Layer Perceptron both have shown the good performance in some parameters but Support Vector Classifier has shown the best performance in all the parameters. In the case of Confusion Matrix, Multi-Layer Perceptron has performed good as it predicts 78 True Positives but Support Vector Classifier has again performed the best as it has given the total 88 True Positives. So, it has been cleared that Support Vector Classifier is the best algorithm amongst all the presented algorithms so we have used the same algorithm for performing the final step of research i.e., designing of the Prediction Model. According to dataset which was given the prediction model performs effectively as it can clearly predict depression in human body.

The Prediction Model that has been designed from this research uses numpy library of Python. It takes the results of Comparative Analysis and gives the prediction in the form of '0' and '1' in which 0 implies that "Person is not having Depression" and 1 implies that "Person is having Depression".

4.3 Comparative Table

On comparing the Random Forest Classifier, Multi-Layer Perceptron, Extra Trees Classifier, Ada Boost Classifier, Decision Tree Classifier and Support Vector Classifier on the basis of accuracy, precision, sensitivity and F1 Score we present he data in the form of comparative table that is given as below-

Table 4.1 Result Analysis

Sr. No.	Model	Accuracy	Precision	Sensitivity	F1 Score
0	Random Forest	0.480159	0.488722	0.507812	0.498084
1	Multi-Layer Perceptron	0.503968	0.506224	0.953125	0.661247
2	Extra Trees Classifier	0.476190	0.482143	0.421875	0.450000
3	Ada Boost Classifier	0.523810	0.533333	0.500000	0.516129
4	Decision Tree Classifier	0.476190	0.482456	0.429688	0.454545
5	Support Vector Classifier	0.551587	0.549669	0.648438	0.594982

4.4 Calculation of Confusion Matrix

Another parameter on which we have performed the comparison between all the above mentioned classifiers is- Confusion Matrix. The figures that come from the analysis of dataset are presented in the form of Heatmap that are given below-

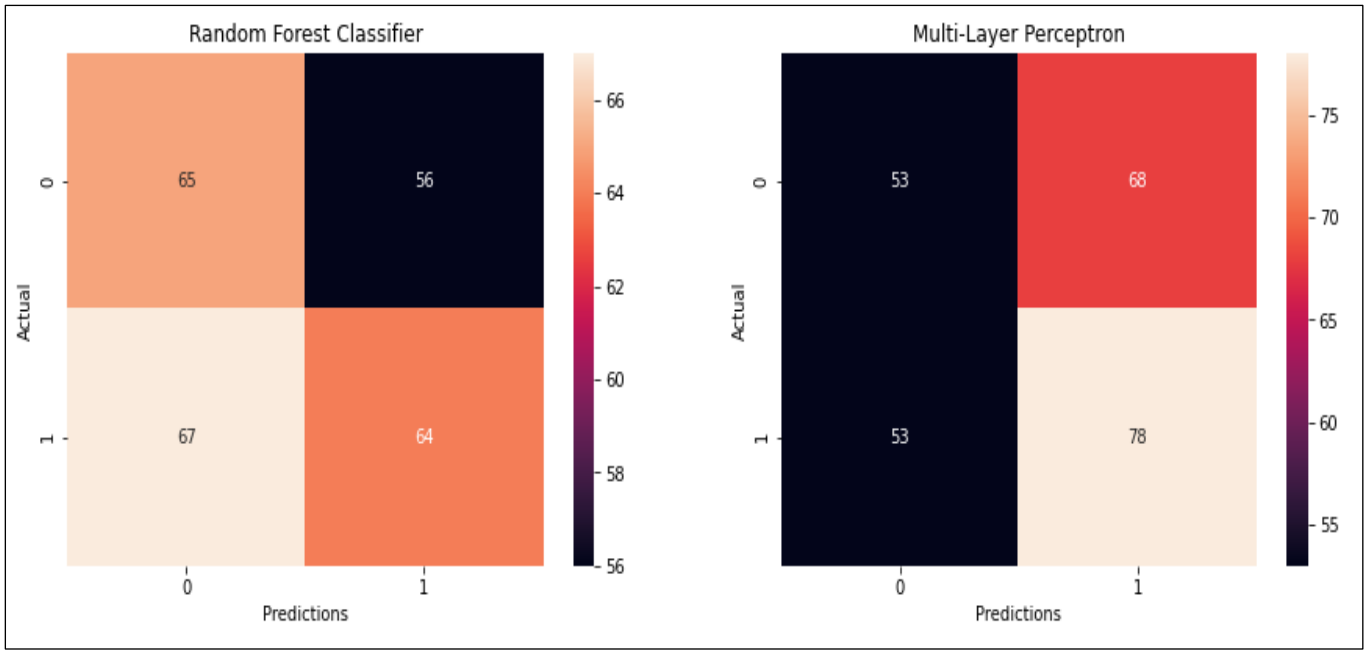

Fig. 4.1 Confusion Matrix of Random Forest Classifier and Multi-Layer Perceptron

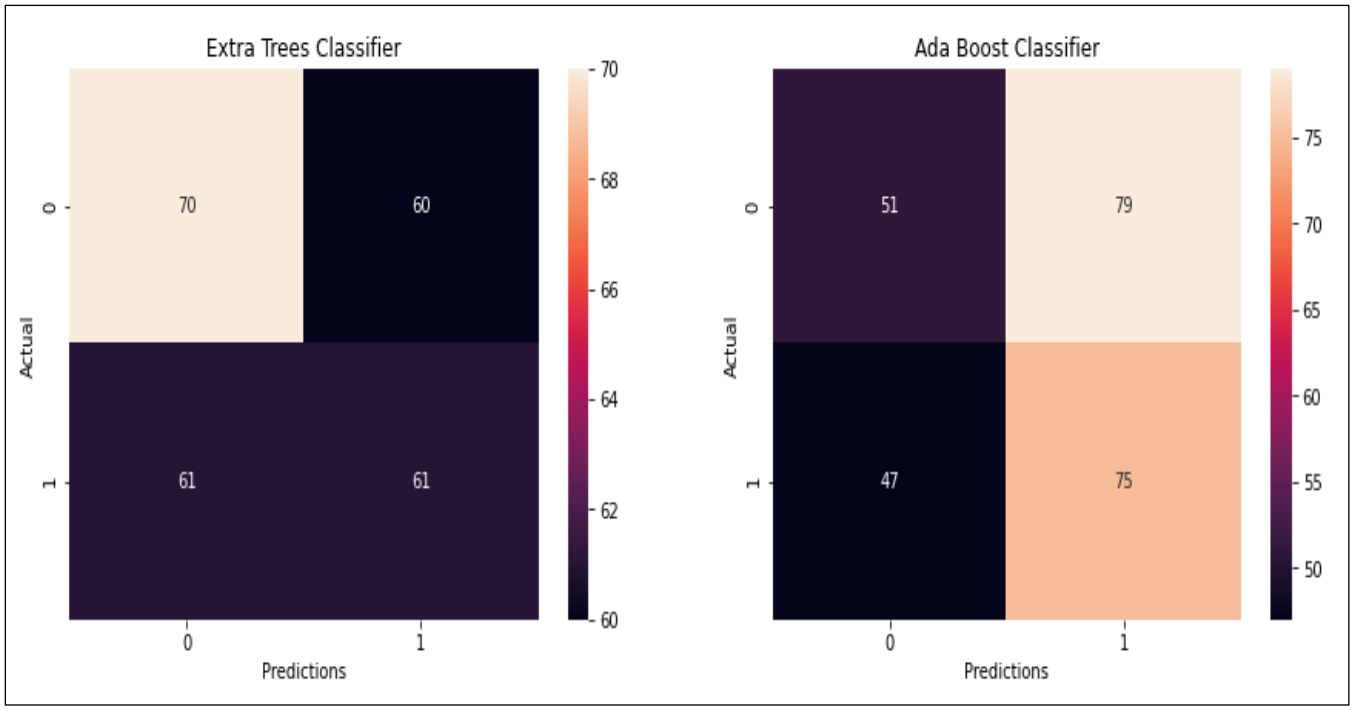

Fig. 4.2 Confusion Matrix of Extra Trees Classifier and Ada Boost Classifier

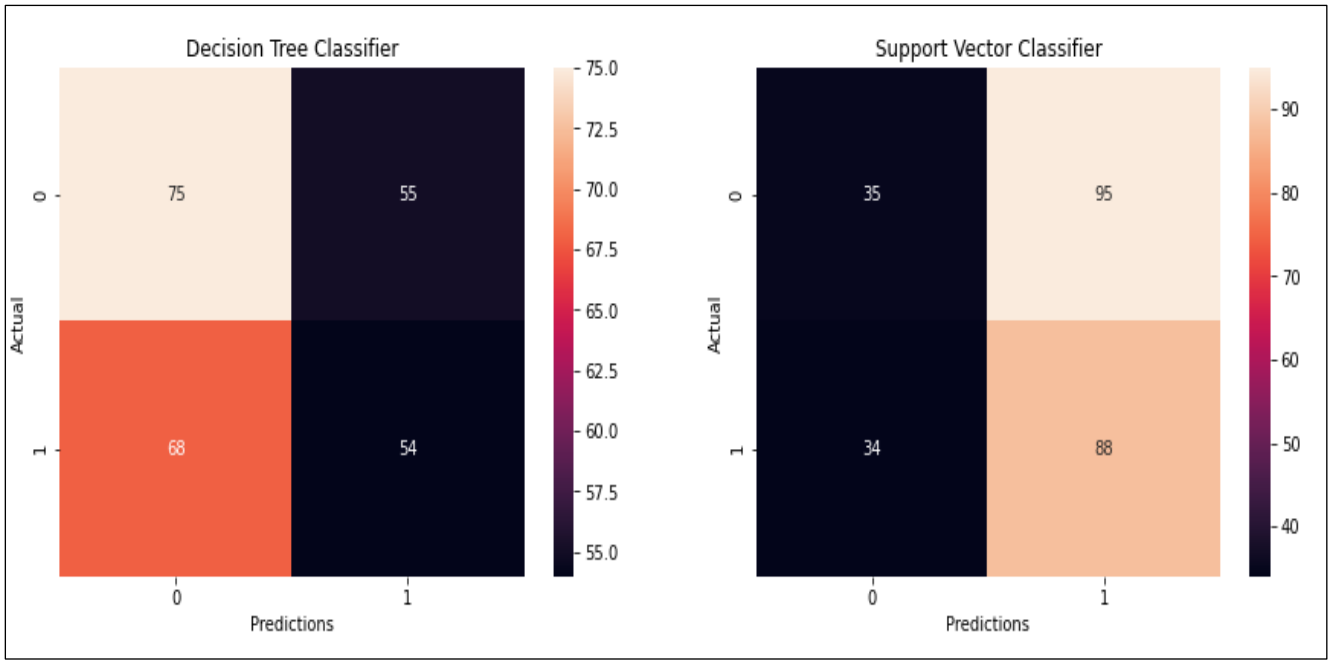

Fig. 4.3 Confusion Matrix of Decision Tree Classifier and Support Vector Classifier

CHAPTER 5: CONCLUSIONS AND FUTURE SCOPE

5.1 Introduction

This chapter gives the overview of all the conclusions that comes from both review and implementation paper. It also presents the future scope that is related to both the papers.

5.2 Conclusions

The major problem is that which algorithm is to be selected for predicting Depression. In this we have seen that Support Vector Classifier gives the best results so it should be chosen for the purpose of prediction. This paper provides the brief review of Algorithms that are related with Machine Learning which can be used for prediction of Depression. This paper clearly suggests the use of Support Vector Classifier and Ada Boost Classifier for prediction of depression. The tool that can be used for prediction is also presented in this paper.

This paper has shown a comparative study between multiple Machine Learning Algorithms very well. The dataset used in this research is completely authenticated and unique as the proper analysis and editing is done on the dataset. According to the results provided by this research the Support Vector Classifier has performed really well in all the parameters. So, this algorithm can be used for future usage.

This paper has also shown a Model that will predict Depression in the human body according to the given dataset. The model predicts in the form of 0 and 1 i.e., 0 implies that the person is not having Depression and 1 implies that the person is having Depression. In this research we have used the results of the comparative study that is done as the primary objective of this research.

5.3 Future Scope

In future this research will be helpful in the following aspects which are given as follows-

- It will be helpful for researchers to perform study in Predictive Analytics.
- It will be helpful for further research on Depression.
- It will provide an overview on how to predict using Machine Learning.
- It will be helpful in selecting the best Machine Learning Model if the aim is to determine Depression at an initial stage.
- It will be helpful in designing the Prediction Model using Machine Learning.
- It will be helpful in predicting Depression from the Human Body.

REFERENCES

[1] U. Haque, E. Kabir, R. Khanam, "Detection of child depression using Machine Learning Methods", *PLOS ONE,* 16 December, 2021, DOI: 10.1371/journal.pone.0261131.

[2] A. Singh, M. Wiktorsson, J. Hauge, "Trends in Machine Learning To Solve Problems in Logistics", *Procedia CIRP 103 (2021)* , October 2021, Pg. No. -67-72, doi:10.1016/j.procir.2021.10.010.

[3] Md. Uddin, K. Dysthe, A. FoIstad, P. Brandtzaeg, "Deep Learning for prediction of depressive symptoms in a large textual dataset", *Neural Computing and Applications (2022),* 34:721-744, 27 August 2021, doi: 10.1007/s00521-021-06426-4.

[4] S. Cho, Z. Geem, K. Na, "Predicting Depression in Community Dwellers Using a Machine Learning Algorithms" , *Diagnostics 2021, 11,* 1429, 7 August 2021, doi: 10.3390/diagnostics11081429.

[5] M. Diwakar, P. Singh, and A. Shankar, "Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain," *Biomedical Signal Processing and Control,* vol. 68, July 2021, doi: 10.1016/j.bspc.2021.102788.

[6] R Coley, J. Boggs, A. Beck, G. Simon, "Predicting outcomes of psychotherapy for depression with electronic health record data", *Journal of Affective Disorders Reports 6 (2021) 100198,* 18 July 2021, doi: 10.1016/j.jadr.2021.100198.

[7] A. Pandita, S. Vashisht, A. Tyagi, Prof. S. Yadav, "Review Paper on Prediction of Heart Disease using Machine Learning Algorithms", *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* ISSN: 2321, Volume 9 Issue VI, June 2021, doi:10.22214/ijraset.2021.35626.

[8] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science,* vol. 2, no. 3, May 2021, doi: 10.1007/s42979-021-00592-x.

[9] M. Ghosh *et al.*, "A comparative analysis of machine learning algorithms to predict liver disease," *Intelligent Automation and Soft Computing,* vol. 30, no. 3, pp. 917–928, 2021, doi: 10.32604/iasc.2021.017989.

[10] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," in *Procedia Computer Science*, 2020, vol. 167, pp. 1258–1267. doi: 10.1016/j.procs.2020.03.442.

[11] W. Jin, "Research on Machine Learning and Its Algorithms and Development," in *Journal of Physics: Conference Series*, Jun. 2020, vol. 1544, no. 1. doi: 10.1088/1742-6596/1544/1/012003.

[12] S. Raschka, J. Patterson, C. Nolet, "Machine Learning in Python : Main Developments and Technology Trends in Data Science, Machine Learning and Artificial Intelligence", *Information(Switzerland) 11(4):193,* 4 April 2020

[13] Meenakshi, “Machine Learning Algorithms and their Real-life Applications: A Survey”, *International Conference on Innovative Computing and Communication (ICICC 2020)*

[14] K. Sethi, A. Gupta, G. Gupta, and V. Jaiswal, “Comparative Analysis of Machine Learning Algorithms on Different Datasets,” 2019. [Online]. Available: www.ccsarchive.org.

[15] G. Srivastava, S. Kumar, H. Pandey, G. Kumar Srivastava, S. Kumar, and H. Pandey, “Modelling of an offline and online software for normalization of microarray data of gene expression by Perl, Bioperl and PerlTk and Perl-CGI,” 2019.

[16] S. Vishwakarma, B. Sharma, and S. Qamar, “Digital Watermarking for Image Authentication using Spatial-Scale Domain based Techniques,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 4, pp. 2334–2341, Nov. 2019, doi: 10.35940/ijrte.d8215.118419.

[17] Ravil Muhamedyev et. al., “Comparative Analysis of Classification Algorithms”, *Research Gate*, October 2015
Uploaded Full Paper on- 30 July 2019

[18] Mylapalle Yeshwanth et. al., “Comparative Study of Machine Learning Algorithms for Rainfall Prediction”, *International Journal of Trend in Scientific Research and Development*, Volume 3, Issue 3, eISSN- 2456-6470, Mar-Apr 2019
Paper ID- IJTSRD22961

[19] I. Hammad, K. El-Sankay, J. Gu, "A Comparative Study on Machine Learning Algorithms for the Control of a Wall Following Robot", *IEEE International Conference on Robotics and Biomimetics (ROBIO)-2019*

[20] D. Ramalingam, V. Sharma, P. Zar, "Study of Depression Analysis using Machine Learning Techniques", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN:2278-3075, Volume-8, Issue-7C2, May 2019

[21] V. Mishra, T. Garg, " A Systematic Study on Predicting Depression using Text Analytics", *Journal of Fundamental and Applied Sciences*, 1 May 2018, ISSN 1112-9876, doi: 10.4314/jfas.v10i2.21.

[22] V. Kumar, "Python Libraries, Development Frameworks and Algorithms for Machine Learning Applications", *International Journal of Engineering Research & Technology*, ISSB: 2278-0181, Vol. 7 Issue 04, April-2018

[23] D. Bansal, R. Chhikara, K. Khanna, and P. Gupta, "Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia," in *Procedia Computer Science*, 2018, vol. 132, pp. 1497–1502. doi: 10.1016/j.procs.2018.05.102.

[24] Giorgio Biagetti et. al., "A Comparative Study on Machine Learning Algorithms for physiological Signal Classification", *22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, Procedia Computer Science 126 (2018) 1997-1984

[25] Md. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health Information Science and Systems*, vol. 6, no. 1, Dec. 2018, doi: 10.1007/s13755-018-0046-0.

[26] P. Singh, M. Diwakar, X. Cheng, and A. Shankar, "A new wavelet-based multi-focus image fusion technique using method noise and anisotropic diffusion for real-time surveillance application," *Journal of Real-Time Image Processing*, vol. 18, no. 4, pp. 1051–1068, Aug. 2021, doi: 10.1007/s11554-021-01125-8.

PUBLICATION CERTIFICATE: JOURNAL

"A Review on Predicting Depression" authored by **"Km. Sakshi Rastogi"** was reviewed by experts in this research area and accepted by the Editorial Board of **"International Conference on Smart Innovations for Society – Convergence 2022"** and is in the process of publication in the same.

"An Efficient Approach to Detect Depression through Predictive Analysis" authored by **"Km. Sakshi Rastogi"** was reviewed by experts in this research area and accepted by the Editorial Board of **"International Conference on Intelligence System (ICIS-2022)"** and is in the process of publication in the same.

Appendix 1


UTTARANCHAL UNIVERSITY
ACCREDITED BY NAAC A+
Uttaranchal Institute of Technology
CERTIFICATE
OF PARTICIPATION
This certificate is awarded to Sakshi Rastogi
for participating in International Conference on Intelligence System
(ICIS-2022) during 11-12th March, 2022.
ORGANIZED BY:
UTTARANCHAL INSTITUTE OF TECHNOLOGY
UTTARANCHAL UNIVERSITY
DEHRADUN, UTTARAKHAND
Prof.(Dr.) S.D. Pandey,
Dean-UIT, Convener ICIS-2022
Uttaranchal University, Dehradun

AN EFFICIENT APPROACH TO DETECT DEPRESSION THROUGH PREDICTIVE ANALYSIS

Sakshi Rastogi[1], Gaurav Kumar Srivastava[2], Sunil Kumar Vishwakarma[3]
Student[1], Assistant Professor [2, 3]
Department of Computer Science & Engineering[1-3]
Babu Banarasi Das University, Lucknow[1-3]
E-Mail ID- sakshirastogi.2607@gmail.com[1], gaurav18hit@bbdu.ac.in[2], sunilvishwakarma83@gmail.com[3]

Abstract.

Nowadays in this world when Machine is learning by algorithms and performing according to the inputs Python has its unique importance. Depression is said to be a feeling in which a person feels having a low mood and a state of strongly not liking someone/something. It has many effects on the body of the person. The symptom which is recognized as the core of depression is not feeling interested in the works or not feeling pleasure in the things that give joy to them earlier. The primary intention of this research is to carry out a comparison amongst the different Algorithms of Machine Learning based on accuracy, precision, sensitivity, F1 score, and Confusion Matrix to find which algorithm gives the best performance on depression data. The final aim of this research paper is to provide a model that will predict depression in the human body.

Keywords. Random Forest Classifier, Extra Trees Classifier, Multi-Layer Perceptron, Support Vector Classifier, F1 Score, Confusion Matrix, Predictive Analytics

1. INTRODUCTION

Depression is said to be a feeling in which a person feels having a low mood and a state of strongly not liking someone/something. The symptom which is recognized as the core of depression is not feeling interested in the works that give joy to them earlier. This can result in the person having a state of sadness, thinking difficulty, problems in paying attention. It can also lead to an increase or decrease in the diet and sleeping time of the person. In this condition person also experience the feeling of dejection, hopelessness and suicidal thoughts.

1.1 Symptoms of Depression

- Problem in Sleeping
- Loss of Interest
- Increment in Fatigue
- Emotions that are Uncontrollable
- Appetite of a Person Changes
- Weight of a Person Changes

2. LITERATURE SURVEY

This section is concerned with all the previous works that are done in this field. It presents a study that gives the comparison between different Algorithms of Machine Learning. They have used algorithms as- Logistic Regression, Random Forest, XG Boost, Support


International Conference on
Smart Innovations for Society
"ICSIS-2022"
May 06-07 2022
Scopus
Clarivate Analytics
WEB OF SCIENCE
in conjunction with
Convergence 2022
"The Smart City"
Role of Engineers in Innovation and Technological Advancement
Certificate of Presentation
Organized by
DEPARTMENT OF APPLIED SCIENCE
POORNIMA
INSTITUTE OF ENGINEERING & TECHNOLOGY
This is to certify Ms. Sakshi Rastogi, Student of Babu Banarasi Das University Lucknow has presented a paper with Title A Review on Predicting Depression in the International Conference on Smart Innovations for Society (ICSIS-2022) held on 06-07 May 2022 at Poornima Institute of Engineering & Technology, Jaipur.
Dr. Rekha Rani Agarwal
Convener
ICSIS-2022
Dr. Bhanu Pratap
Program Co-Chair
Convergence 2022
Dr. Sama Jain
HOD I Year, PIET, Jaipur
Program Chair (ICSIS-2022)
Dr. Dinesh Goyal
Director & Principal, PIET
Patron Convergence 2022

A Review on Predicting Depression

Sakshi Rastogi[1,a)], Gaurav Kumar Srivastava[2,b)], Sunil Kumar Vishwakarma[3,c)]

[1]Student, [2,3]Assistant Professor

[1-3]Department of Computer Science & Engineering, Babu Banarasi Das University, Lucknow, Uttar Pradesh, India

E-Mail ID- [a)] sakshirastogi.2607@gmail.com, [b)] gaurav18hit@bbdu.ac.in, [c)] sunilvishwakarma83@gmail.com

"Abstract" - In this growing world the problem of Depression has taken the huge space in the life of a person. Depression is the feeling of not having interest in the things which provides joy to the person earlier. The symptoms of which are- Problem in Sleeping, Loss of Interest, Increment in Fatigue and many more on the body of the person. The main goal of writing the review paper on this topic is to give a brief review on how the Depression can be predicted in the body of the person. This paper also presents a review on tool that is needed for the system.

Keywords- Predictive Analysis, Depression, Machine Learning Algorithms, Classification

INTRODUCTION

Depression is the feeling of not having interest in the things which provides joy to the person earlier. The symptoms of which are- Problem in Sleeping, Loss of Interest, Increment in Fatigue and many more on the body of the person. According to "Institute of Health Metrics and Evaluation, Global Health Data Exchange (GHDx)" It is a common type of problem that has its effect on 3.8% of the population which includes 5.0% of adults and 5.7% of the person's older than 60 years. The worst effect of Depression that can occur on the person is Suicide.

Predictive Analytics is said to be one branch of advanced analytics that is used in making predictions about outcomes of future. It uses historically collected data that is combined with statistical modeling, techniques of data mining and machine learning. It is sometimes associated with data science and big data.

LITERATURE REVIEW

Researcher's objective was to design a model that can be used in predicting the Depression from 4-17 years old aged children and adolescents. Also, they say to evaluate the results in order to determine which performs the best. Further there aim was to associate the related issues of families. Their conclusion said that Random Forest based prediction model performs more accurately. Some authors have provided the general introduction of various Techniques that are based on Machine Learning which are used in the field of logistics to give help to the researchers who further use it. From their research Supervised Learning can be considered as the most prominent way in which the problems of Machine Learning can be solved. Some have proposed an approach in which they use Long Short- Term Memory based Recurrent Neural Network for the purpose of identifying symptoms of Depression. They performed on the dataset which was of young age people of Norway. The conclusion given by them was that if we use their proposed approaches we get 98% and 99% mean prediction performance. Some have used the data from 2014 and 2016 Korea National Health and Nutrition Examination Surveys. They have used 2014 dataset for training purposes and 2016 dataset for testing purposes. The result they got was that when they use 37

Appendix 2

Importing Libraries

```
import pandas as pd

from matplotlib import pyplot as plt

import seaborn as sns

import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.ensemble import AdaBoostClassifier

from sklearn.ensemble import ExtraTreesClassifier

from sklearn.neural_network import MLPClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.svm import SVC

from sklearn.metrics import precision_score, f1_score

from sklearn.metrics import recall_score

from sklearn.metrics import confusion_matrix

from sklearn.metrics import accuracy_score

import warnings

warnings.filterwarnings('ignore')
```

Defining Variables

```
X = data.drop(columns='target', axis=1)

y = data['target']

print(X.shape,y.shape)
```

Dividing Datasets

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2)

print('Distribution of Feature variable')

print(X_train.shape,X_test.shape)

print('Distribution of target variable in Training Set')

print(y_train.value_counts())

print('Distribution of target variable in Testing Set')

print(y_test.value_counts())
```

Applying Algorithms

```
#Random Forest Classifier

model=RandomForestClassifier()

model.fit(X_train,y_train)

y_pred_ran=model.predict(X_test)

#Multi-Layer Perceptron

mlp=MLPClassifier()

mlp.fit(X_train,y_train)

y_pred_mlp=mlp.predict(X_test)

#Extra Trees Classifier()

et=ExtraTreesClassifier()

et.fit(X_train,y_train)

y_pred_et=et.predict(X_test)

#Ada Boost Classifier

ada=AdaBoostClassifier()

ada.fit(X_train,y_train)

y_pred_ada=ada.predict(X_test)

#Decision Tree Classifier

dec=DecisionTreeClassifier()
```

```
dec.fit(X_train,y_train)

y_pred_dec=dec.predict(X_test)

#Support Vector Classifier

svc=SVC(kernel='linear',gamma='auto',probability=True)

svc.fit(X_train,y_train)

y_pred_svc=svc.predict(X_test)
```

Comparing Algorithms

```
acc=accuracy_score(y_test,y_pred_ran)

prec=precision_score(y_test,y_pred_ran)

rec=recall_score(y_test,y_pred_ran)

f1=f1_score(y_test,y_pred_ran)

model_results=pd.DataFrame([['Random Forest',acc,prec,rec,
f1]],columns=['Model','Accuracy','Precision', 'Sensitivity','F1 Score'])

data={

    'Multi-Layer Perceptron':y_pred_mlp,

    'Extra Trees Classifier':y_pred_et,

    'Ada Boost Classifier':y_pred_ada,

    'Decision Tree Classifier':y_pred_dec,
```

```
    'Support Vector Classifier':y_pred_svc
}
models=pd.DataFrame(data)
for column in models:
    acc=accuracy_score(y_test,models[column])
    prec=precision_score(y_test,models[column])
    rec=recall_score(y_test,models[column])
    f1=f1_score(y_test,models[column])
    results=pd.DataFrame([[column,acc,prec,rec,f1]], columns=
['Model','Accuracy','Precision','Sensitivity','F1 Score'])
    model_results=model_results.append(results,ignore_index=True)
model_results
```

Confusion Matrix

```
plt.figure(figsize=(15,5))

CM=confusion_matrix(y_test,y_pred_ran)

plt.subplot(1,2,1)

sns.heatmap(CM, annot=True)

plt.title('Random Forest Classifier')

plt.xlabel('Predictions')

plt.ylabel('Actual')

Cm=confusion_matrix(y_test,y_pred_mlp)

plt.subplot(1,2,2)

sns.heatmap(Cm, annot=True)

plt.title('Multi-Layer Perceptron')

plt.xlabel('Predictions')

plt.ylabel('Actual')

plt.figure(figsize=(15,5))

CM_et=confusion_matrix(y_test,y_pred_et)

plt.subplot(1,2,1)

sns.heatmap(CM_et, annot=True)

plt.title('Extra Trees Classifier')
```

```
plt.xlabel('Predictions')

plt.ylabel('Actual')

CMada=confusion_matrix(y_test,y_pred_ada)

plt.subplot(1,2,2)

sns.heatmap(CMada, annot=True)

plt.title('Ada Boost Classifier')

plt.xlabel('Predictions')

plt.ylabel('Actual')

plt.figure(figsize=(15,5))

CMdec=confusion_matrix(y_test,y_pred_dec)

plt.subplot(1,2,1)

sns.heatmap(CMdec, annot=True)

plt.title('Decision Tree Classifier')

plt.xlabel('Predictions')

plt.ylabel('Actual')

CMsvc=confusion_matrix(y_test,y_pred_svc)

plt.subplot(1,2,2)

sns.heatmap(CMsvc, annot=True)

plt.title('Support Vector Classifier')
```

```
plt.xlabel('Predictions')

plt.ylabel('Actual')
```

Designing Prediction Model

```
input=(56,0,0,1,0,1,1,0)

input_numpy=np.asarray(input)

input_reshape=input_numpy.reshape(1,-1)

prediction=svc.predict(input_reshape)

print(prediction)

if(prediction[0]==0):

    print("Patient doesn't have Depression")

else:

    print("Patient have Depression")
```

Thesis_File.docx

ORIGINALITY REPORT

7%	4%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	docplayer.net Internet Source	1%
2	Submitted to University of Wales Institute, Cardiff Student Paper	1%
3	Anna Koroleva, Sanjay Kamath, Patrick Paroubek. "Measuring semantic similarity of clinical trial outcomes using deep pre-trained language representations", Journal of Biomedical Informatics, 2019 Publication	1%
4	dokumen.pub Internet Source	<1%
5	Nakayiza Hellen, Ggaliwango Marvin. "Explainable AI for Safe Water Evaluation for Public Health in Urban Settings", 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), 2022 Publication	<1%
	www.coursehero.com	

6	Internet Source	<1%
7	www.mdpi.com Internet Source	<1%
8	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2022 Publication	<1%
9	en.wikipedia.org Internet Source	<1%
10	jcdronline.org Internet Source	<1%
11	Submitted to Gitam University Student Paper	<1%
12	Qiong Wang, Min Yang, Bo Pang, Mei Xue, Yicheng Zhang, Zhixin Zhang, Wenquan Niu. "Predicting risk of overweight or obesity in Chinese preschool-aged children using artificial intelligence techniques", Endocrine, 2022 Publication	<1%
13	Shraddha Suratkar, Elvin Johnson, Karan Variyambat, Mihir Panchal, Faruk Kazi. "Employing Transfer-Learning based CNN architectures to Enhance the Generalizability of Deepfake Detection", 2020 11th	<1%

	International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	
14	en.ru.is Internet Source	<1%
15	techscience.com Internet Source	<1%
16	Submitted to University of Witwatersrand Student Paper	<1%
17	assets.researchsquare.com Internet Source	<1%
18	lib.buet.ac.bd:8080 Internet Source	<1%
19	sersc.org Internet Source	<1%
20	Mesut Toğaçar, Burhan Ergen, Zafer Cömert. "Tumor type detection in brain MR images of the deep model developed using hypercolumn technique, attention modules, and residual blocks", Medical & Biological Engineering & Computing, 2020 Publication	<1%
21	Rifat Jahan Lia, Abu Bakkar Siddikk, Fahim Muntasir, Sheikh Shah Mohammad Motiur Rahman, Nusrat Jahan. "Chapter 9 Depression	<1%

	Detection from Social Media Using Twitter's Tweet", Springer Science and Business Media LLC, 2022 Publication	
22	www.sciencegate.app Internet Source	<1%
23	eprints.usq.edu.au Internet Source	<1%
24	ijece.iaescore.com Internet Source	<1%
25	synapse.koreamed.org Internet Source	<1%
26	www.diva-portal.org Internet Source	<1%
27	www.ijert.org Internet Source	<1%
28	"Emerging Research in Computing, Information, Communication and Applications", Springer Science and Business Media LLC, 2019 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

BABU BANARASI DAS UNIVERSITY, LUCKNOW

CERTIFICATE OF FINAL THESIS SUBMISSION

(To be submitted in duplicate)

1. Name: **Km. Sakshi Rastogi**
2. Enrollment No.: **12004490677**
3. Thesis Title: ***"An Efficient Approach to Detect Depression through Predictive Analysis"***
4. Degree for which the thesis is submitted: **M.Tech. (SE)**
5. School (of the University to which the thesis is submitted):

School of Engineering

6. Thesis Preparation Guide was referred to for preparing the thesis.	**YES**/NO
7. Specifications regarding thesis format have been closely followed.	**YES** /NO
8. The contents of the thesis have been organized based on guidelines.	**YES** /NO
9. The thesis has been prepared without resorting to plagiarism.	**YES** /NO
10. All sources used have been cited appropriately.	**YES** /NO
11. The thesis has not been submitted elsewhere for a degree.	**YES** /NO
12. All the corrections have been incorporated.	**YES** /NO
13. Submitted 2 hard bound copies plus 2 CD.	**YES** /NO

Signature: .....

Name: Dr. Gaurav Kumar Srivastava

Assistant Professor

Department of Computer Science & Engineering

.....

(Signature of Candidate)

Name: **Km. Sakshi Rastogi**

Enrollment No.: **12004490677**

Signature: .....

Name: Mr. Sunil Kumar Vishwakarma

Assistant Professor

Department of Computer Science & Engineering