MACHINE LEARNING BASED DEPRESSION DETECTION

BY

MD. MOZAHIDUL ISLAM
ID: 181-15-10995
SAIKAT BISWAS
ID: 181-15-10878
AND
UTPAUL SARKAR
ID: 181-15-11043

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

MD. TAREK HABIB
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. ABDUS SATTAR
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2022

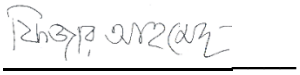
APPROVAL

This research project titled "Machine Learning Based Depression Detection" submitted by Md. Mozahidul Islam, ID: 181-15-10995, Saikat Biswas, ID: 181-15-10878 and Utpaul Sarkar, ID: 181-15-11043 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 04 January 2022.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Dr. Fizar Ahmed
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Nusrat Jahan
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Mohammad Shorif Uddin
Professor
Department of Computer Science and Engineering
Jahangirnagar University

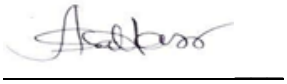
DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Tarek Habib, Assistant Professor, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

SUPERVISED BY:

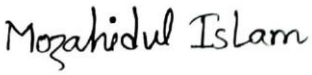
Md. Tarek Habib
Assistant Professor
Department of CSE
Daffodil International University

Co-SUPERVISED BY:



Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Md. Mozahidul Islam
ID: 181-15-10995
Department of CSE
Daffodil International University

Saikat Biswas
ID: 181-15-10878
Department of CSE
Daffodil International University

Utpaul Sarker
ID: 181-15-11043
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First and important my heartfelt and sincere gratitude goes to Almighty Allah who has empowered us to practically whole our thesis. I would first like to thank my thesis advisor **Md. Tarek Habib**, **Assistant Professor**, of the office, was always open whenever we ran into a trouble spot or had a question about our research or writing. He consistently allowed this thesis to be our own work, and also advised us in the right direction.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan, Head of Department of Computer Science and Engineering**, for giving us an opportunity to carry out the research work and also to other faculty members and the staff of the CSE department of Daffodil International University.

Thanks to Daffodil International University for the study opportunity and for the specialized help during the last period of completing this proposal for this thesis.

Finally, I must express my very profound gratitude to my parents and to my friends and for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of master and this thesis. This achievement would not have been conceivable without them.

ABSTRACT

Depression is a common disorder that causes constant mood swings and feelings of sadness. Nowadays It is considered to be a deadly disorder in the world. At present, everyone from young to old is suffering from depression but most of them do not have the right idea about their mental state. It is very important for everyone to have the right idea about their mental state. We will detect depression through machine learning. First, we study some related papers, journals, and online articles then we talk to psychologists and depressed people and then we find some common factors that are related to becoming depressed. Then we collect data based on those factors, such as age, gender, profession, marital status, life satisfaction, feelings, interests, etc. We collect data from both depressed and non-depressed people. We have two outcomes. One is 'Yes' which means depressed and another is 'No' means not depressed. After data collection, we processed all the data and created a processed dataset. Then we applied machine-learning algorithms to our processed dataset. Machine learning, deep learning, and artificial intelligence are used in various predictions, detection, and recognition systems. We use k-nearest neighbor (kNN), logistic regression, Support Vector Classifier (SVC) Linear, naïve Bayes, random forest, adaptive boosting (ADA boosting), decision tree, and Linear Discriminant Analysis (LDA) Classifier. In our work, logistics regression gave the best performance based on accuracy and the accuracy of logistic regression was 93.50%.

TABLE OF CONTENTS

LIST OF TABLES

TABLES **PAGE**

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Depression is a common and serious medical condition that affects people's lives. Depression is a common problem nowadays in the world. Nearly 280 million people have depression in the world [13]. Depression is technically a mental disorder, but it also affects physical health and well-being. People of all ages are suffering from a disease called depression day by day. UNICEF's flagship report found a median of 19% of young people aged 15-24 who often feel frustrated or have little interest in doing anything [17]. On the other hand, a new analysis from the London School of Economics, included in a UNICEF report, estimates that mental illness causes about $ 390 billion a year in damage to the economy, which could lead to disability or death among young people [17]. Depression is a most common illness worldwide, affecting approximately 3.8% of the population, of which 5.0% are adults and 5.7% are over 60 years of age [13]. There can be many 5858 causes of depression including Unemployment problems, Abuse, lack of love, Conflict, lack of family ties, affection, political cataclysm, etc. As a result, people are committing heinous acts like suicide. One of the most common implicit disorders is depression 30% to 70% victims suicide suffer from major or minor depression [22]. Self-awareness is needed to get rid of depression. We use machine learning for this detection of becoming depressed. And also we are planning to create a website through which depressed people can be easily identified by themselves or by doctors.

1.2 Motivation

Depression is a mood-dependent disorder. When a person suffers from depression, he puts himself in solitary confinement. Not giving time to his family, losing focus on work,

avoiding social gatherings became a daily habit. Nowadays depression has become a dangerous fact.

A person can be depressed in any way depending on his life. Depression disorder is a common problem for a human being. But the main problem is that in our country people are not aware of mental health problems. People can't understand that they are depressed. They don't go to the doctor or psychologist.

So based on humans' daily activity or lifestyle it needs to detect whether a person is depressed or not. That's why we think to create a way or system to find a person's mental health condition, where machine learning can perform a great deal.

A lot more research has not been seen by us in this sector as we analyze literature review. A lot of research is badly needed in this field. So we are doing this depression detection work using the popular and widely used machine learning technique.

1.3 Fundamental principle of the study

As we mentioned before, there is less significant work done previously with depression detection from the Bangladeshi perspective. For this reason, we are highly encouraged in working with depression and machine learning Strategies.

Machine learning is a part of artificial intelligence that deals with probabilistic, statistical, and optimization techniques and then allows computers to “learn” from past experience and detect patterns, similarities from massive, noisy, difficult, and complicated datasets.

Nowadays machine learning methods are used across the detection, recognition, and classifying fields. Machine learning is used for diagnosis systems [5], disease recognition [23], Stock Market Analysis [7], Traffic Prediction [8]. A huge range of detection and classification problems are now conducted by using machine learning. Machine learning plays a vital role in this field and is widely used by researchers with a great outcome. So we decided to apply machine learning for our research work.

1.4 Research Questions

- How can we detect depressed people?
- How do we maintain the feature for our data?
- What amount of data do we collect?
- From where and how do we collect the data?
- What is the amount of our train and test set?
- Will our data and machine learning techniques be consistent and be fit?
- Should we use widely used and famous machine learning techniques or build new models?

1.5 Expected Outcome

The goal of our paper is to make depression easily recognizable from human behavior. One of our biggest problems is not being able to identify whether we are suffering from depression. By following our method, a person can easily consider his mental state. As a result, a large part of the population will be relieved of the dilemma of their mental state.

In society, we have to deal with a lot of people. We often notice that some of our acquaintances have wrapped themselves up. They are no longer participating in any work of the society. The number of those people but not less in the society. This paper will help with the detection of considering the mental state of the people.

1.6 Report Layout

The following contents of this research work are given below:

- Chapter 1 describes the introduction of the research with its motivation, fundamental principle of the study, research questions, and expected outcome.
- Chapter 2 contains related previous works, a summary of the research works, the scope of the problem, and challenges we face.
- Chapter 3 explains the working process flowchart of this research, data collection process, data preprocessing, statistical analysis, and feature implementation.
- Chapter 4 contains experimental analysis with some related studies, a summary of accuracy, and the outcome of the research.
- Chapter 5 contains this research's outcome on society.
- Chapter 6 describes the summary, limitations, and future work of this research.

CHAPTER 2
BACKGROUND STUDY

2.1 Introduction

In this part, we will describe previous similar works done by researchers, a summary of their work, the scope of the problem, and the challenges we have faced. In the similar work part, we summarize some research papers, related works, their applied methods, classifiers, and accuracies that are related to our work. In the research summary part, we make a summary of all works, and then for better understanding, we show it on a table. On the scope of the problem part, we discuss how we can help or move forward to this work. And finally, the Challenges contain which type of problem and obstacle we have faced during this research work and how we cope up with that.

2.2 Similar Works

In this literature review section, we are going to discuss similar works done by other researchers on depression detection and recognition and above all machine learning applications in this sector. We have studied and followed their methods, and research progress developed and published by them.

Khan, M. R. H. et al. [9] have done their research on machine learning based sentiment analysis from the Bengali depression dataset. From social media posts and various poems, several novels, and also the quotations of various noble persons data is collected. They use so many algorithms like support vector machines(SVM), multinomial naive Bayes, random forest, k-nearest neighbors(kNN), decision tree, and xg boost. And they got the highest accuracy using Multinomial Naive Bayes which is 86.67%.

Mulay, A. et al. [10] has worked detecting depression level automatically through visual input. They use the FER2013 dataset with a total of 35887 images which is collected from

Kaggle. They use the CNN model for feature extraction, classification generating an output of emotion vector. The highest accuracy gained by the CNN model is 66.45%.

Ding, Y. et al. [11] have shown a method for college students' depression recognition using deep integrated support vector algorithms. They Collect data from social text data (Sina Weibo) from 1000 users. This paper mainly proposes a deep integrated SVM-based depression detection model for Chinese text data. For finding output Deep integrated support vector machine (DISVM) is applied here and they got the highest accuracy 86.15%.

Shukla, D. M. et al. [12] proposed a system for detecting depression in a person using speech or voice signals by drawing strength and statistical features. They identify that a person is depressed or not depressed from the voice signal of that person. Ryerson audio-visual dataset which contains emotional speech and song (RAVDESS) is used here. They used the Multi-Layer Perceptron algorithm. The average accuracy of cross-validation sets and training sets was 81.56%.

Deshpande, M. et al. [16] has done work on detecting depression by emotional artificial intelligence. They detect depression by analyzing Twitter tweets. For generating the training and test dataset 10,000 Tweets were collected from Twitter tweets using Twitter API. For detecting depression SVM and Naive Bayes classification algorithms were used. Support vector machines gain 79% accuracy. But the highest accuracy is gained by the Multinomial Naive Bayes algorithm. The best accuracy was gained 83% in Multinomial Naive Bayes.

Orabi et al. [14] have done work on depression detection of twitter users using deep learning. 1,145 Twitter users' data is combined to build the dataset which are labeled as Depressed, Control, and PTSD. With that each of the users of the dataset is labeled according to their gender and age. The key focus of this work was mainly on recognizing users' capabilities to depression. Recurrent neural networks (RNN) and Convolutional

neural networks (CNN) are used by the researchers. The CNN With Max models using their optimized embedding reported higher accuracy which is 87.95%.

Patel, F. et al. [21] has proposed a system for students which was combating depression using an artificially intelligent chat-bot. They analyze emotion from text data and then recognize depression from that. For this work, they use the ISEAR dataset. Forming Word Vector algorithm is applied for segmentation. RNN,CNN and HAN algorithms are applied for classification purposes. Among those for 15 epochs CNN algorithm has achieved accuracy up to 75% with high consistency. For 15 epochs HAN and RNN have gone up to 70% accuracy. So CNN performed a little better.

Asad, N. A. et al. [15] has worked on social media posts of the users and then detected depression from that. The data is collected from Twitter and Facebook, beautiful soup is used to collect Twitter data, and having permission manually data is collected from Facebook from 150 users. SVM, a Naïve Bayes classification algorithm, is used in this work. Their ways of finding out depression level is made based on the popular BDI-II questionnaire method [1]. It is appraised as non-depressed, appraised normal, Mild and Borderline Depression from 1-55% and above 55% is appraised as depressed. The accuracy in naive Bayes is 74% with a very high precision of 100%.

Dhiraj Dahiwade et al. [27] have mentioned a machine learning-based disease prediction system. They collected the people's personal living habits and checkup information as to their data . The dataset is collected from the UCI machine learning website. Then for preprocessing they deduct commas, punctuations, and space in data. Then they used this dataset as a training dataset. They used the CNN algorithm & the kNN algorithm. In kNN, Hamming distance, Euclidean distance are used as distance calculation matrices. But CNN took less time than kNN. They used two algorithms based on time and accuracy. They found the best accuracy which is 84.5% in CNN.

Md. Tarek Habib et al. [23] has proposed a work on papaya disease recognition system which is based on machine learning. They used different imperfect papayas images. For

preprocessing the images they resize all images into 300 pixels into 300 pixels. Histogram Equalization & Bicubic interpolation and were used here. The total number of imperfect and perfect images is 129. Among popular machine learning techniques Naïve Bayes, BPN, CPN, SVMs, Logistic Regression, kNN, C4.5, Random Forest, and RIPPER are used here. They worked with five common papaya diseases. The highest accuracy is 95.2% which is gained by the SVM algorithm rather than all of the algorithms.

2.3 Comparative Analysis and Summary

Some work has already been done on the detection and recognition of depression with machine learning algorithms and data mining processes. Nowadays, the use of machine learning technology has increased with the detection of depression, the prognosis of alcohol users, and various diseases. This section shows a comparison between these related works. Here, a comparison of various research works with their topics, methods, and results is given in Table 2.1 below.

TABLE 2.1: SUMMARY OF SIMILAR RESEARCH WORKS.

SL No.	Name of the Authors	Methodology	Description	Result
1.	Khan, M. R. H., Afroz, U. S., Masum, A. K. M., Abujar, S., & Hossain, S. A.	Decision Tree, Multinomial Naive Bayes, KNearest Neighbors, Random Forest, Support Vector Machine, and XG Boost	Machine Learning. Based Sentiment Analysis from the Bengali Depression Dataset	Highest accuracy using Multinomial Naive Bayes which is 86.67%.
2.	Mulay, A., Dhekne, A., Wani, R., Kadam, S., Deshpande, P., & Deshpande, P.	CNN	Detecting depression level automatically through visual input	The accuracy of the CNN model is 66.45%.

3.	Ding, Y., Chen, X., Fu, Q., & Zhong, S.	Deep integrated support vector machine Algorithm	Recognizing depression among College Students	86.15% accuracy in DISVM.
4.	Shukla, D. M., Sharma, K., & Gupta, S.	Multi-Layer Perceptron algorithm	detecting depression in a person using speech or voice signals by drawing strength and statistical features.	Multi-Layer Perceptron algorithm got 81.56% accuracy.
5.	Deshpande, M., & Rao, V.	Naive Bayes and support vector machine classification algorithms	Machine learning-based depression detection using emotion artificial intelligence.	Best accuracy was gained 83% in Multinomial Naive Bayes
6.	Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D.	CNN and RNN	Twitter users depression detection using deep learning.	87.95% accuracy in CNN.
7.	Patel, F., Thakore, R., Nandwani, I., & Bharti, S. K.	CNN, RNN and HAN	Finding depression in students by using an intelligent chatBot.	For 15 epochs CNN has achieved accuracy up to 75%
8.	c, Mahmud Pranto, M. A., Afreen, S., & Islam, M. M.	SVM, Naïve Bayes	Machine learning-based depression detection by analyzing social media posts of users.	Accuracy in naive Bayes is 74%
9.	Dahiwade, D., Patle, G., & Meshram, E.	k-nearest neighbors (kNN), CNN.	Machine learning-based general disease prediction.	84.5% accuracy in CNN.

10.	Nuruzzaman, M., Hossain, M. S., Rahman, M. M., Shoumik, A. S. H. C., Khan, M. A. A., & Habib, M. T.	SVM, Logistic regression, RIPPER. Naïve bayes, kNN, random forest, C4.5, CPN & BPN.	Recognition of papaya disease based on machine learning.	SVM got 95.2% accuracy.

For prediction, classification, and detection models recently machine learning, deep learning, and AI is widely used in every sector of data science. Logistic Regression, CNN, ANN, SVM, kNN, and many other famous algorithms are used for detection models. Based on our literature review, It is seen that the kNN, naïve Bayes, SVM, random forest, CNN, and Decision Tree algorithm's usefulness and popularity for prediction, detection, and recognition models are very satisfying. In our research work, we have tried to implement k-nearest neighbor, decision tree, naïve Bayes, logistic regression, adaptive boosting (ADA boosting), SVC Linear, random forest & Linear Discriminant Analysis (LDA) Classifier algorithms to detect the depressed people in Bangladesh and we got 93.50% accuracy in logistic regression.

2.4 Scope of the Problem

Our research work is basically creating a model by analyzing given data and applying machine learning algorithms. Our proposed model can detect depression. This work will have a significant impact on people of the society. There are so many people facing mental disorders in our society. But they don't realize that they are depressed. They keep themself alone and they don't know what they should do. That's why they made the wrong decision at that time. Here he needs a system that will help him to find out his problem and help him to realize whether he is really depressed or not. And finally, give him instructions on what he should need to do. Recently, as machine learning and artificial intelligence are being used for various object detection and disease predictions, the results are quite acceptable. Therefore, we decided that using machine learning, we would create a model of depression detection.

2.5 Challenges

We were facing some problems while doing our research work. The most challenging part was data collection. Normal people and depressed people cannot be easily distinguished. The people who are facing mental health problems, not interested in talking easily and do not want to agree to it. We talked to different people in different classes and read a lot of newspapers, talked to many doctors, but we did not get any proper answer or solution. Nobody can give us any fixed information about depressed people because it varies from person to person and situation to situation. So it was very difficult to fix the research questionnaires also. But people were unwilling to give their personal information. Sometimes we talked Parsons to Parsons directly for the data collection. But due to the covid-19 pandemic situation that was also stopped.

We were not accustomed to featuring engineering and some machine learning algorithms. To learn that, firstly it takes a little time. Then by practicing and with the help of our supervisor we coped up with that and finished our work perfectly.

CHAPTER- 3

RESEARCH METHODOLOGY

3.1 Introduction

This study is for building a model for recognizing depression. On the basis of daily life activity and the response of people, this model is created and analyzed. To build this model various machine learning algorithms are applied. We used Decision Tree, k-nearest neighbor (kNN), Support Vector Classifier (SVC) Linear, logistic regression, Naïve Bayes, adaptive boosting (ADA boosting), Random forest, and Linear Discriminant Analysis (LDA) Classifier algorithms. The algorithms are used for classification purposes which are depressed or not-depressed. We find out total of twenty four features related and directly or indirectly connected to depression. Before final implantation, we processed our dataset as we needed. We find the best accuracy, and also calculate specificity, precision, sensitivity, recall, F1 score, and show the roc-curve of each algorithm to find the best algorithm for the model. We got that logistic regression had the best accuracy.

3.2 Data Collection

The data-set consists of a huge number of features or factors which are directly or indirectly connected with depression. We did not succeed in collecting the required data as we went to the hospital because the authorities said that if they provide patient information that could damage patients' privacy and can affect their rules and they also said that sometimes it is not available as a ready dataset. That’s why we think we need to create our own datasets which are collected from face-to-face questionnaires, google forms online, and a paper with a list of questions. Hopefully, we were successful in collecting 1000 people’s data based on 24 factors which are basically daily activities of a person. After collecting all the data the main challenging part came to focus is data leveling into depressed people and non-depressed people. We consult many doctors to seek help and find patterns to find out. Finally, with the help of one doctor, one physiatrist and one student of psychology the data

were being leveled separately. And combining their three decisions into final leveling outputs with emphasis on the opinion of the majority. Among those data, there are 417 depressing information and 583 non-depressed people's information. All the data we have collected from an Online Survey, Daffodil International University(DIU), Varies Secondary and Higher Secondary Schools and Colleges, Peoples of public areas of different town and rural fields, and some other places also.

Based on the following twenty-four features the data was collected:

- Age.
- Gender.
- Profession
- Marital Status
- Satisfaction in life
- Dropped activities & interests
- Feel Empty
- Get bored
- Feel helpless
- Afraid of something bad happening
- Hopeful about future
- Spend time happily
- Feel full of energy
- Prefer to stay at home
- Avoiding social gatherings and reducing talk
- Memory loss
- Consider worthless
- Cry most of the time
- Sleep activity
- Feeling reduce appetite and losing weight
- Hopeless
- Think most of the people better

- Feeling bad & guilty
- Better off he died

To identify the risk of becoming depressed we overview and find these 24 features. We finalize these features by discussing various physiatrists, Doctors, reading papers and articles [3], [2], [1], and websites[17],[22].

3.3 Research Subject and Instrumentation

In recent years, machine learning algorithms, data mining, and deep learning techniques are immensely acceptable and in vogue for any kind of prediction, recognition, and also detection. We will try to apply several machine-learning algorithms to our collected dataset to see which algorithms will fulfill our will and perform best. We apply several machine learning algorithms which are Logistic Regression, k-nearest neighbor (kNN), naïve Bayes, Support Vector Classifier (SVC) Linear, random forest, adaptive boosting (ADA boosting), Decision Tree, and Linear Discriminant Analysis (LDA) Classifier algorithms. Recently 'Python' is one of the most famous and used programming languages which is mostly used for research purposes by researchers. So, we use 'Python' as our programming language and for data mining tools or platforms, we use 'Google Colab', 'Jupyter notebook', with that 'Microsoft Excel' as our dataset, and weka software as for understanding some data visualization in this research work.

3.3.1 Proposed Methodology

In Figure 3.1. our Steps of the proposed methodology is shown:

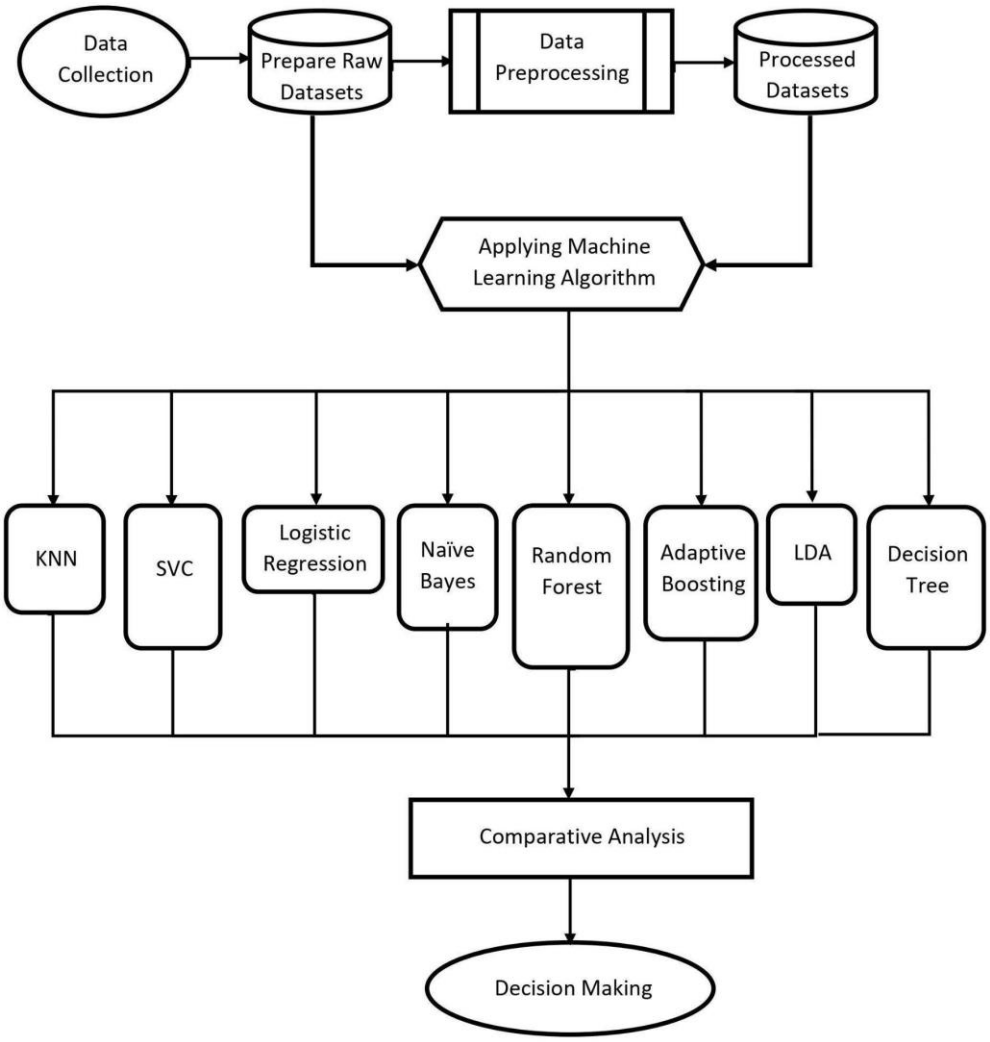


Figure 3.1: Steps of proposed methodology

3.3.2 Data Preprocessing Process

When we successfully end up collecting a sufficient amount of data, we notice that there are some missing values in some of the data, and there are also different types of data like categorical data, numerical data. This kind of data is not suitable for machine learning

algorithms. So we make the decision that we will process our data according to our needs, we will make this data compatible with algorithms. Data processing has the power to convert data into appropriate formats after data collection. Processed data in a specific type helps to best output easily.

Data preprocessing process is shown below here in Figure 3.2.

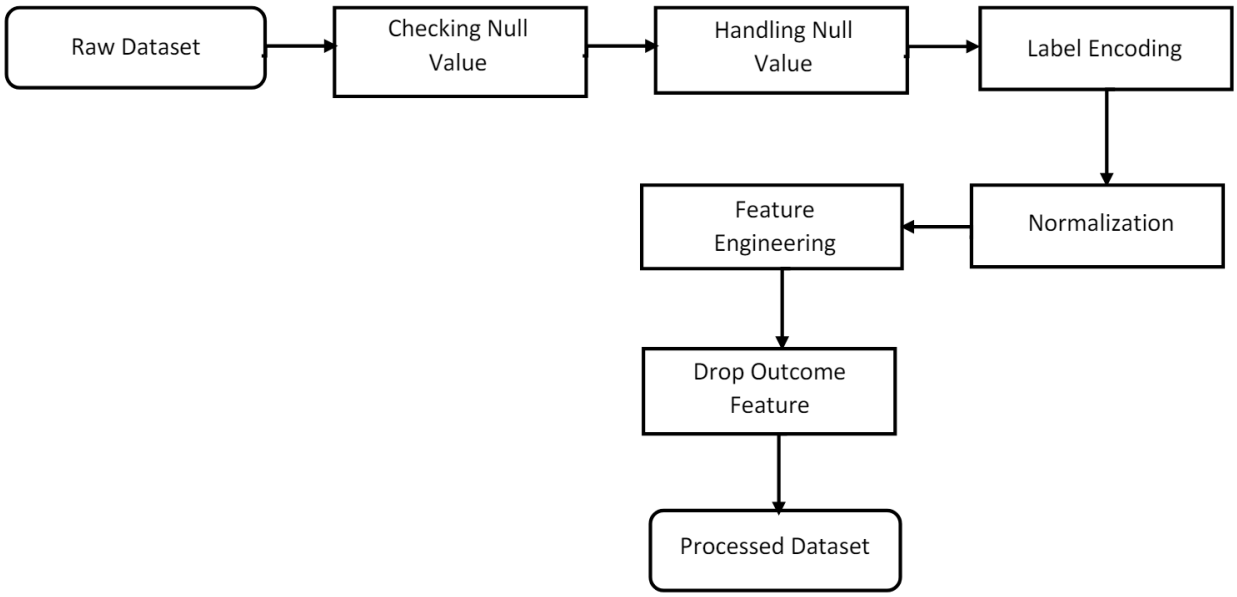


Figure 3.2: Data Preprocessing Process

Firstly, we started our work by collecting and making our raw dataset. Then we focus on data cleaning. We inspect if there is any missing or null value in the data set. We solved the missing value problem using the imputer. Rather than deleting any data, we tried to fill this null value by its relevant possible values. Then we encode the level that transforms all the text or categorical data to its relevant numerical data. Then through normalization, we completed the data transformation. For age data features we use Min-Max Normalization techniques. After that, we explore and examine the correlation matrix, f_clasif for feature engineering purposes. In the correlation matrix, the matrix actually shows how each and every data is connected to each other data. The greater positive value means data is highly connected and similarly negative value means that the data is not so connected and less important. And using f_classif by fit_feature.scores we got the feature values. Considering

all that we drop some of our outcome features, that was, the ‘Avoiding social gatherings and reducing talk’, ‘Occupation’, ‘Marital Status’, ‘Gender’, ‘hopeful about the future’, ‘feel himself bad and guilty’, ‘Sleeping habits’ column. Thus, we finally get the final processed dataset as we wanted. All the work of the data processing process was done using the “Google Colab”.

3.4 Statistical Exploration

We were capable of collecting data from 1000 people in different categories and classes. We collected those data, people from different ages, different occupations, and different districts and so many other categories. In Figure 3.3 we can see in our dataset how much depressed and non-depressed people were. We work and build our model and all further processes are done based on the data of 417 depressed people and 583 non-depressed people. In percentage which is about 41.7% people are depressed and 58.3% are non-depressed.

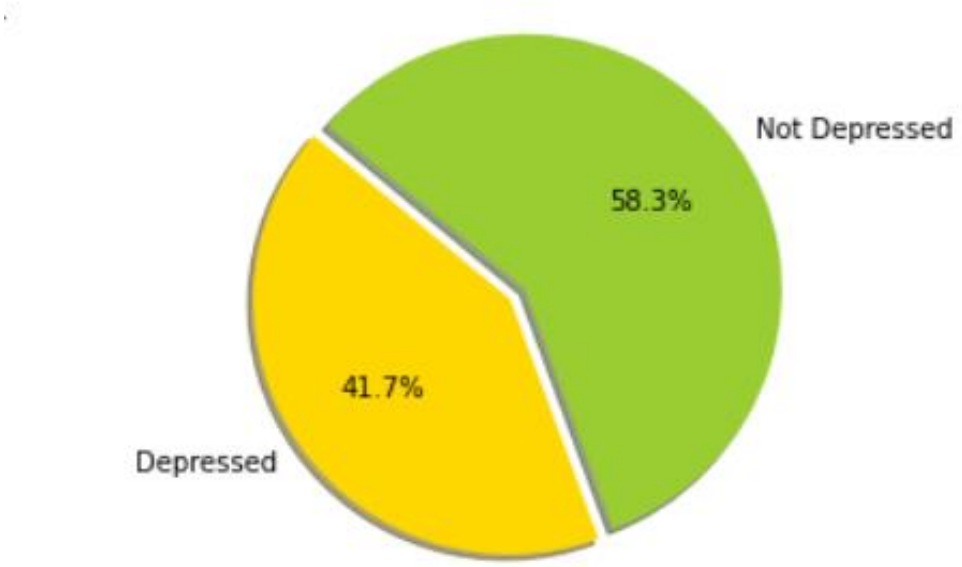


Figure 3.3: Depressed and Non Depressed cases

Figure 3.4 shows how many men and women were depressed. There were 218 males and 199 females who were depressed. Figure 3.4 is shown below.

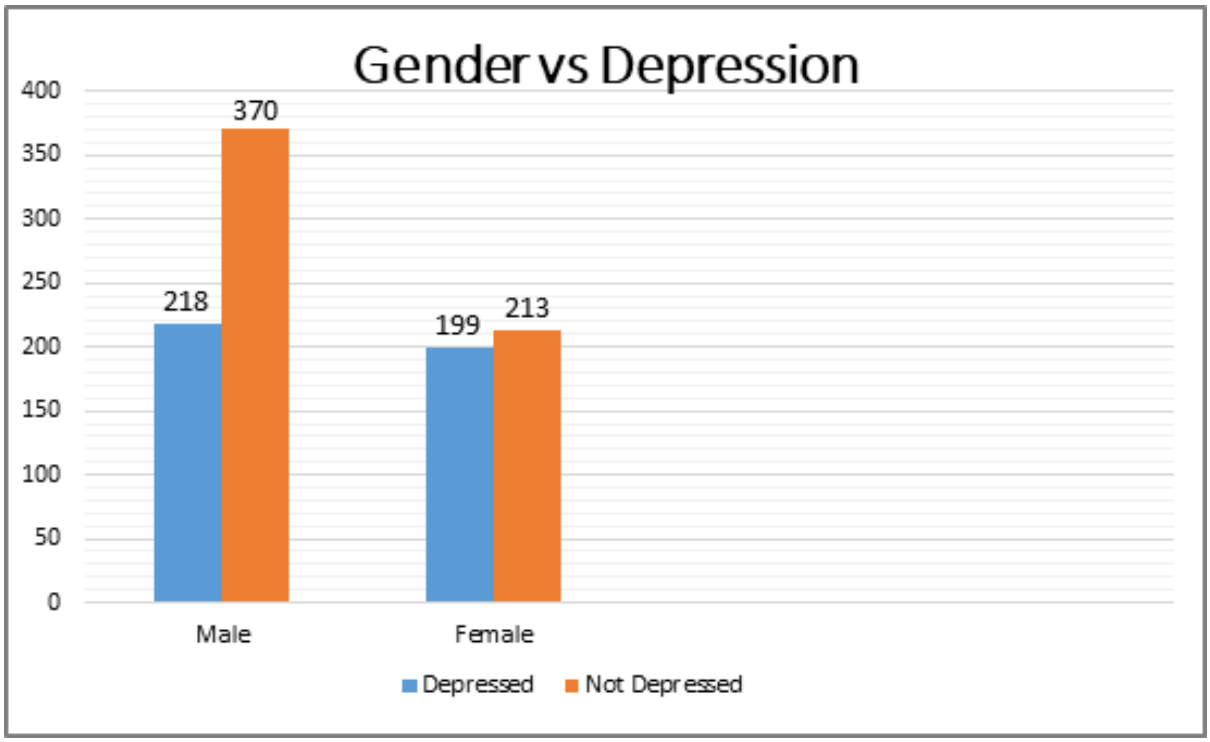


Figure 3.4: Depressed and Gender cases

Figure 3.5 shows the information about people's ages. This figure shows how many people are in which ages. From the picture, it's noticed that most of the data we collected was about young aged people between 13 years to 26 years. Figure 3.5 is shown below.

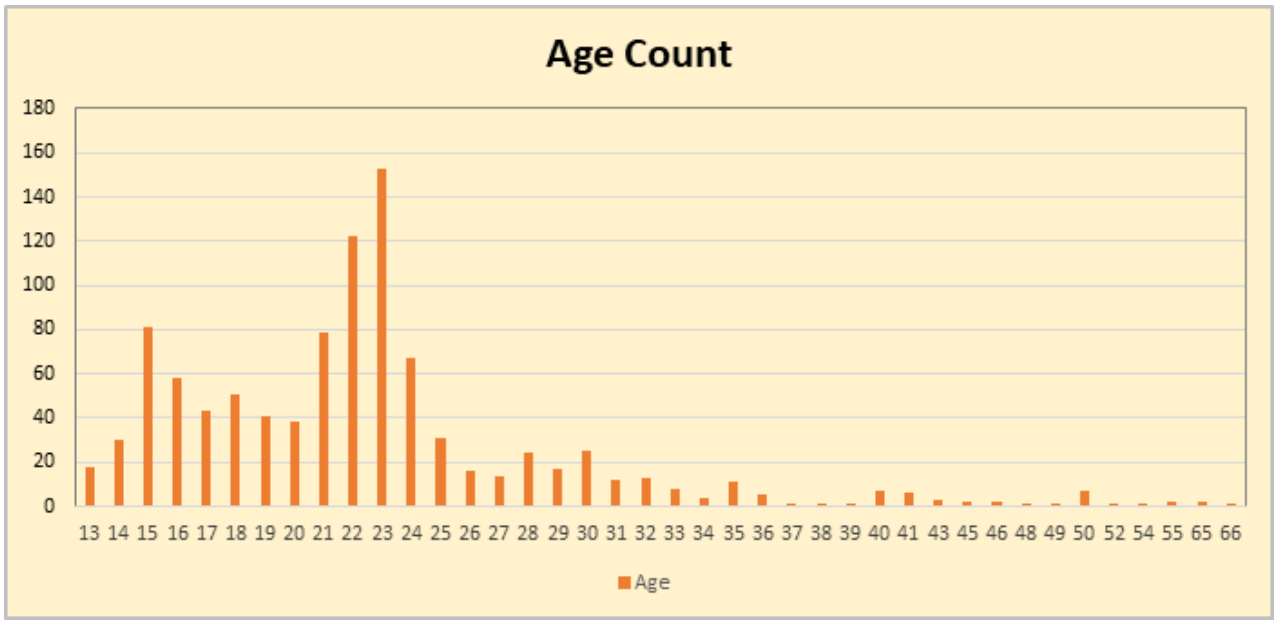


Figure 3.5: Information about Ages

Figure 3.6 shows the professional and depression cases. This picture shows the profession of people and how many of them were depressed. Figure 3.6 is shown below.

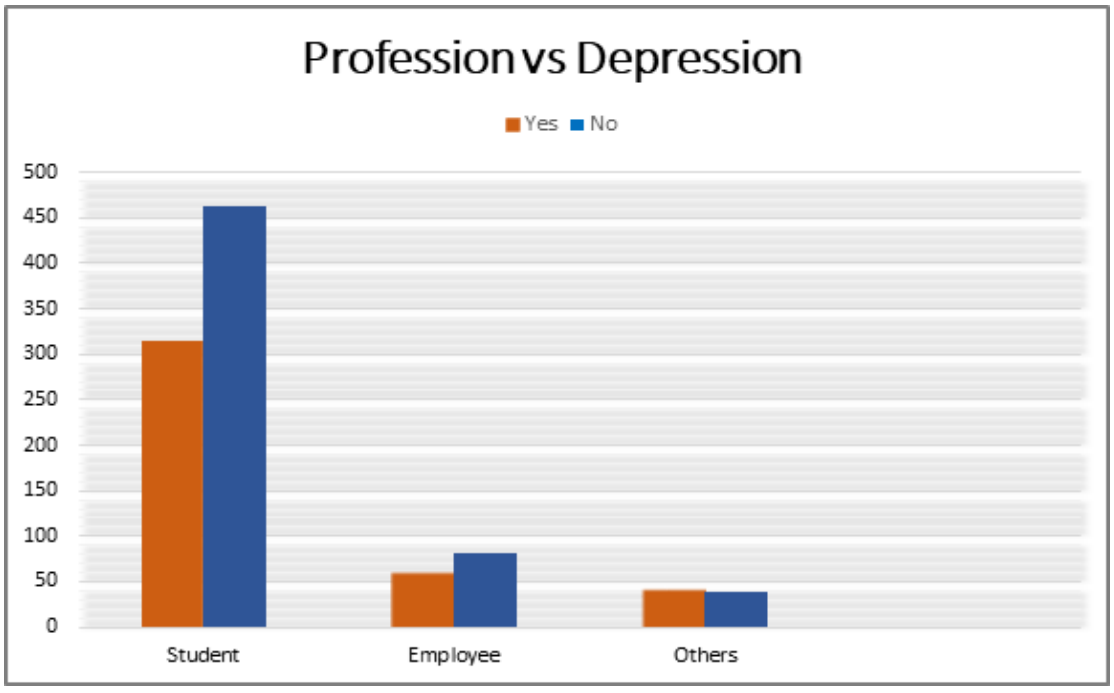


Figure 3.6: Profession vs Depression Cases

Figure 3.7 shows, that the fear of death of the people we have collected our data. Figure 3.7 is shown below.

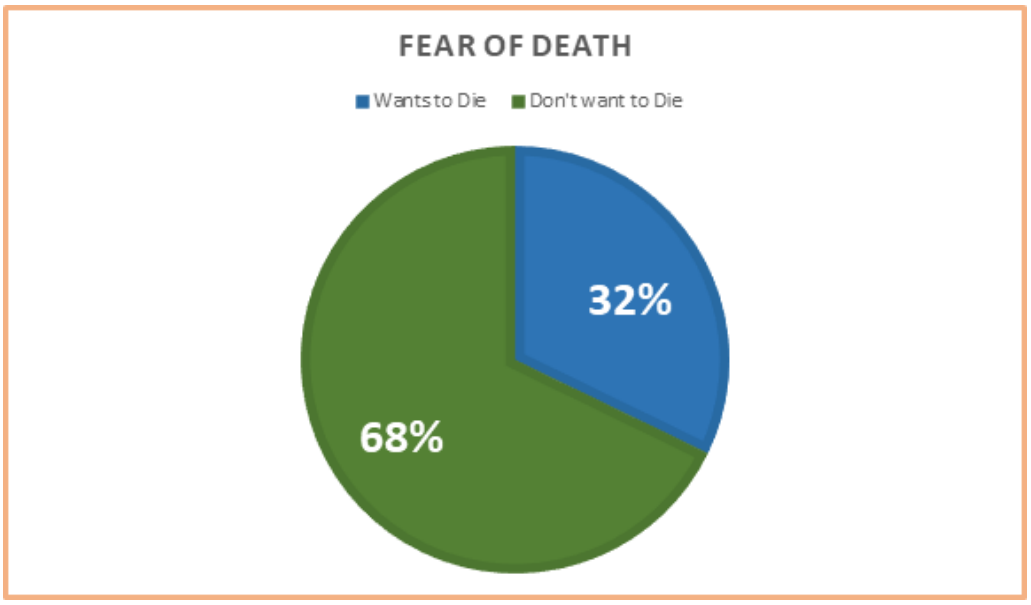


Figure 3.7: Fear of Death

Figure 3.8 shows the correlation between the features. Which helped us to find features near to depression. Figure 3.8 is shown below.

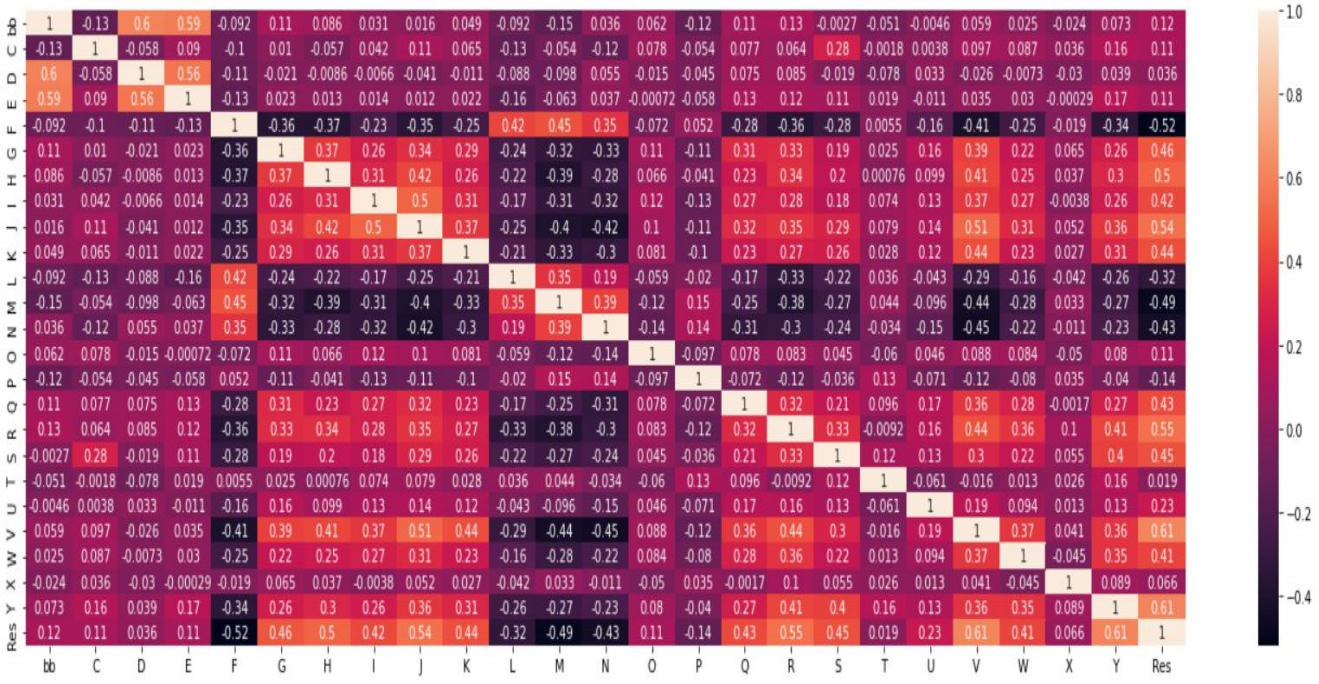


Figure 3.8: Correlation between the features

3.5 Implementation Equipment

To implement this work we must need data storing tools, data processing tools, data mining tools with platforms, and visualization platforms. As our data is collected by google forms and handwritten forms so we use Microsoft Excel for data storing. For understanding data patterns for data visualization sometimes we use the weka platform and google docs. For the data preprocessing and implementation of algorithms, we mainly used “Google Colab” and sometimes “Jupyter notebook”.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

some algorithms are applied on the processed dataset and the results of the algorithm will be discussed in this section. We used Support Vector Classifier (SVC), random forest, k-nearest neighbor (kNN), Linear, naïve Bayes, adaptive boosting (ADA boosting), logistic regression, decision tree, and Linear Discriminant Analysis (LDA) Classifier algorithms. The results help to understand which algorithm gives the best accuracy. Basically, We used two steps to calculate accuracy. The Accuracy is first checked before using feature engineering techniques on unprocessed data and then accuracy is calculated after using preprocessing and feature engineering techniques in processed data. All in all, We collect 1000 data for depressed and not-depressed persons. We basically follow three steps to calculate the best accuracy. Firstly for training data, we used 80 percent of data and for testing data, we used 20 percent of data. Secondly, for training data, we used 90 percent of data and for testing data, we used 10 percent of data. Finally, for training data, we used 70 percent of data and for testing data, we used 30 percent of data. Our dataset is 'Depression-Dataset'.

4.2 Experimental Analysis

We used eight machine learning algorithms. Then we compared them with each algorithm by calculating their sensitivity, precision, confusion matrix, F1 score, accuracy, recall, and specificity.

4.2.1 Experiment for Evaluation

Our dataset consists of 24 features. For processed datasets, we implemented eight machine learning algorithms where feature numbers were 17.

Fig 4.01 shows the accuracy of eight algorithms on an unprocessed dataset with the number of the features was 24, where both SVC and Logistic Regression perform the same with the accuracy of 93%.

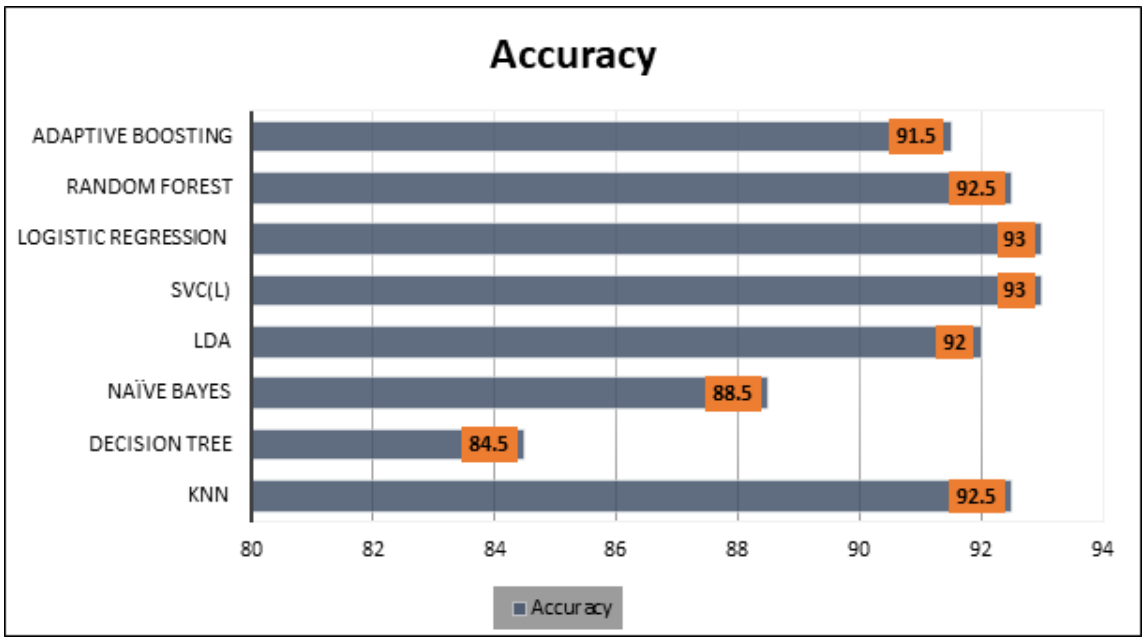


Figure 4.01: Accuracy for Unprocessed-Datasets

Figure 4.02 represents the accuracy of eight algorithms on the processed datasets with the number of features set 17. Then check accuracy by dividing for training data which is 80% and for test data which is 20%. We observe that the accuracy of some algorithms has increased and the accuracy of some algorithms has decreased and the accuracy of some algorithms has remained unchanged. The accuracy of SVC is 93%, the accuracy of Logistic Regression is 93.5%, the accuracy of kNN is 90.5%, the accuracy of Naïve Bayes is 90%, the accuracy of Random Forest is 91%, the accuracy of Adaptive Boosting is 92%, then the accuracy of Decision Tree is 84.5%, the accuracy of LDA is 92%. It appears that before

preprocessing, both SVC and Logistic Regression have achieved 93% accuracy, but after preprocessing Logistic Regression performed best and increased the accuracy to 93.5%.

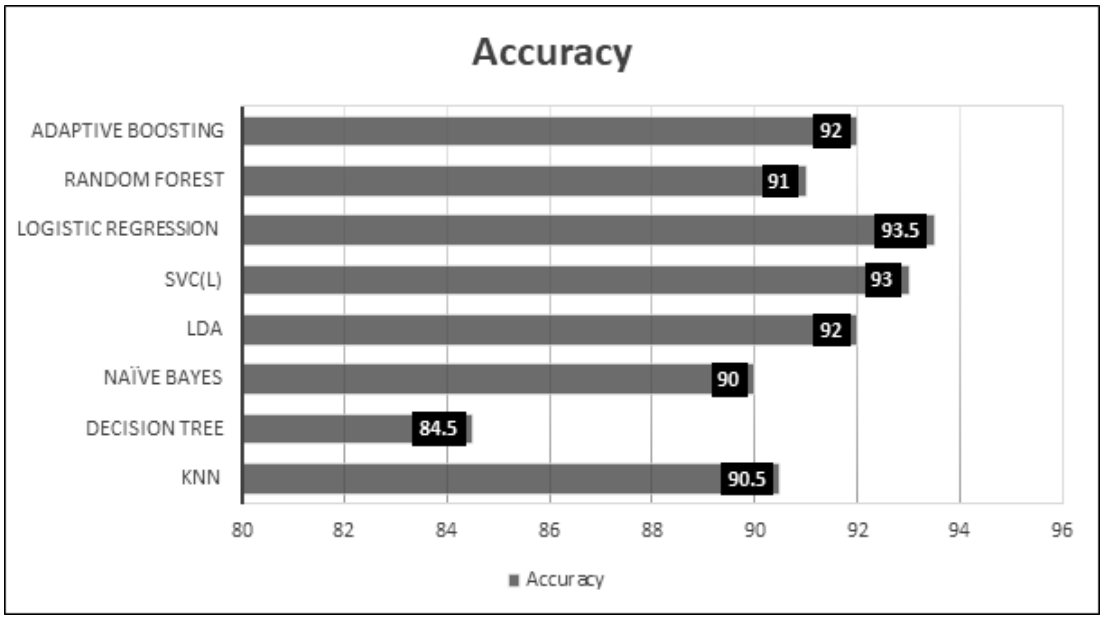


Figure 4.02: Accuracy For Processed Data (80/20)

Figure 4.03 shows the accuracy of eight algorithms for training datasets which is 70% and for training datasets which are 30%. Where the accuracy of kNN is 90.7%, the accuracy of SVC is 91.7%, the accuracy of Logistic Regression is 92.7, the accuracy of Naïve Bayes is 88.3%, the accuracy of Random Forest is 92%, the accuracy of Adaptive Boosting is 91.3%, the accuracy of Decision Tree is 86% and the accuracy of LDA is 92.3%.

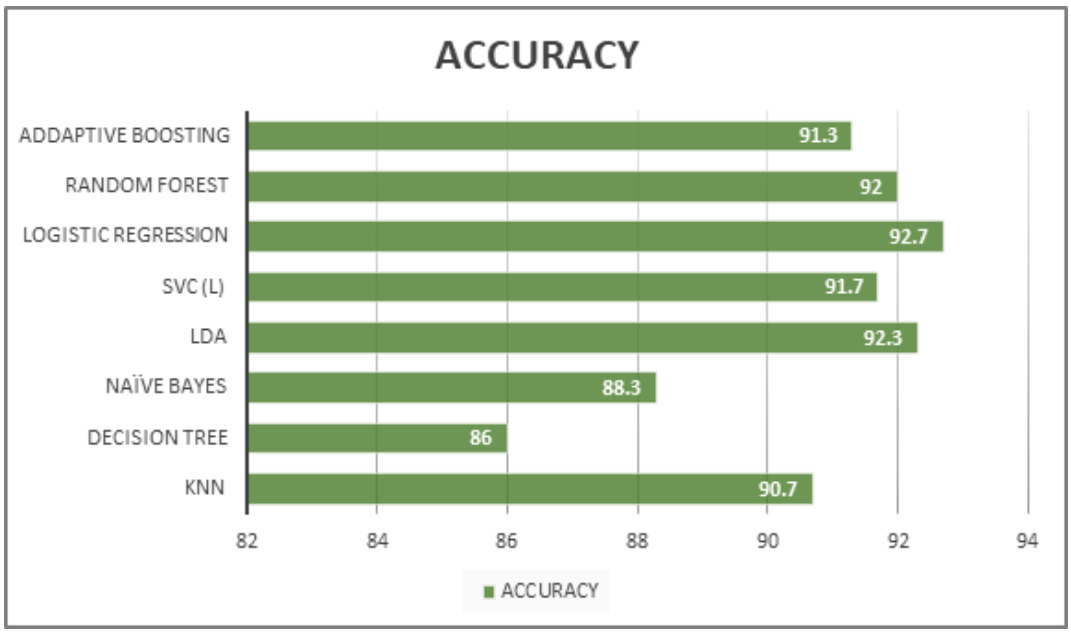


Figure 4.03: Accuracy For Processed Data (70/30)

Figure 4.04 shows the accuracy of eight algorithms for training datasets which is 90% and for training datasets which is 10%. Where the accuracy of kNN is 88%, the accuracy of SVC is 92%, the accuracy of Logistic Regression is 93%, the accuracy of Naïve Bayes is 91%, the accuracy of Random Forest is 89%, the accuracy of Adaptive Boosting is 92%, the accuracy of Decision Tree is 87% and the accuracy of LDA is 90%.

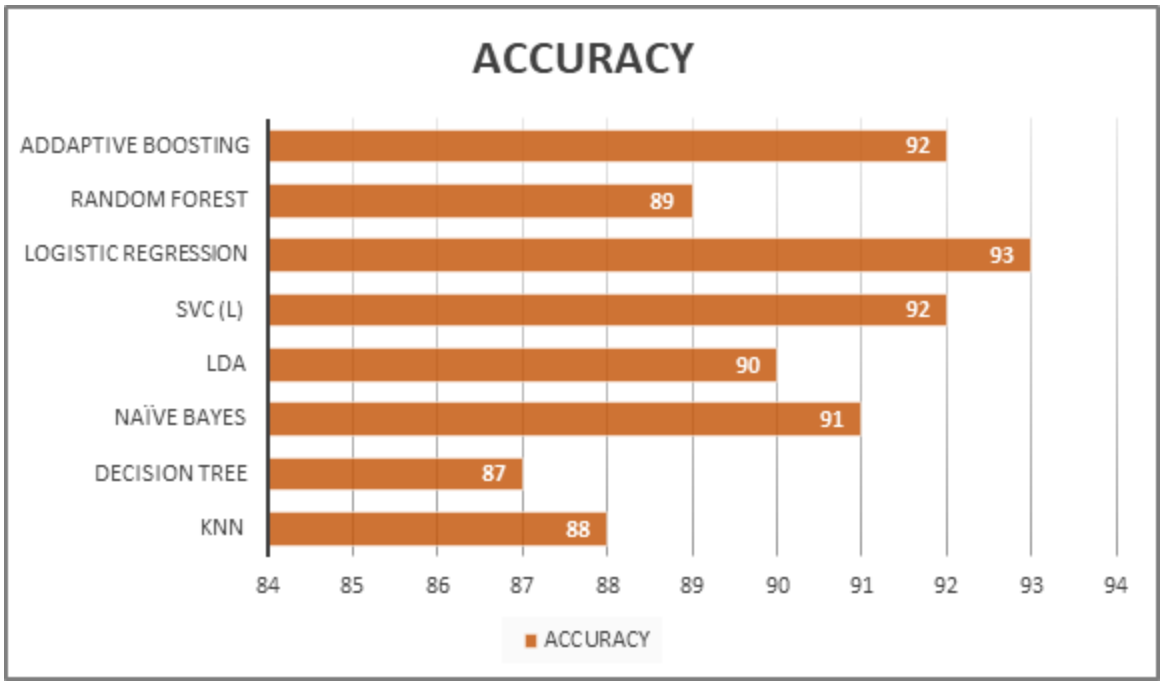


Figure 4.04: Accuracy For Processed Data (90/10)

KNN algorithm is one of the simplest classification algorithms. It is a supervised machine learning algorithm. K Nearest Neighbors stores all available cases and classifies new cases based on similarity measures [25].

Support Vector Machine is a supervised machine learning algorithm. It is capable of doing linear and nonlinear classification, regression problems. But Support vector machines are widely used for classification tasks. Because it uses less computation and gives notable accuracy. It is good because it gives reliable results even if there is less data [25].

Logistic regression is a machine learning algorithm. It is used for classification problems. It's like linear regression. It's a predictive analysis based algorithm and it's concept is probability [25].

The oldest algorithm of machine learning is Naïve Bayes. It is a probabilistic machine learning algorithm based on basic statistics and the Bayes theorem. In supervised learning, the Naïve Bayes classifier algorithm is very useful. Naive bias models are used in conditional Probabilities & Class Probabilities. Gaussian distribution is used in the Naïve Bayes algorithm [26].

Decision trees are some of the most widely used machine learning algorithms. It can be used for both classification and regression. Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of many sub-nodes increases the number of identical sub-nodes. We can also say that this increases the accuracy of the node with the target variable [26].

Adaptive Boosting is a Boosting procedure utilized as an Ensemble Method in Machine Learning. Here the weak learners are formed sequentially in the training phase. This is called adaptive boosting because the higher weights for the incorrectly classified data are re-attached to each data [26].

Linear Discriminant Analysis in 1936 by Ronald A. Was developed by Fisher. This is a very popular level reduction strategy. This helps to minimize high dimensional datasets over a low dimensional space. The goal is to have a decent separation between classes and to do so while reducing resources and computing costs [26].

Random Forest is a supervised machine learning algorithm. The forest that it creates is a group of decision trees and is usually trained in bagging methods. The random forest creates multiple decision trees and combines them to get more accurate and adjusted predictions [26].

From Table 4.1 we get the highest accuracy which is 93.50% for logistic regression. SVC and logistic regression both have achieved the highest accuracy on unprocessed data before

applying preprocessing and feature selection with 93%. On the processed dataset for different training and testing data, we got different accuracy. But considering all the Unprocessed and Processed Data with different training and testing datasets we got the best result on the processed dataset with 80/20 training and testing data with the feature number of 17.

TABLE 4.1: OUTLINE OF ACCURACY

Algorithms	Accuracy for unprocessed dataset (%)	Accuracy for processed dataset (70/30) (%)	Accuracy for processed dataset (90/10) (%)	Accuracy for processed dataset (80/20) (%)
Naïve Bayes	88.5	88.3	91	90
SVC linear	93	91.7	92	93
kNN	92.5	90.7	88	90.5
Logistic Regression	93	92.7	93	93.5
Decision Tree	84.5	86	87	84.5
LDA	92	92.3	90	92
Random Forest	92.5	92	89	91
Adaptive Boosting	91.5	91.3	92	92

4.2.2 Expressive Analysis

We determined the accuracy of many algorithms and also calculated the F-score, accuracy, recall, specificity, roc curve, sensitivity and confusion matrix of every algorithm. Any model selection requires evaluation of that model. In the case of model evolution, specific classifications need to be measured. Classification measurements are made based on test data sets for advanced measurements.

Sensitivity is the ability of a test to correctly identify the true positive rate.

$$\text{Sensitivity} = TP / (TP + FN) \times 100\% \quad \text{(i)}$$

Specificity is the ability of a test to correctly identify the true negative rate.

$$\text{Specificity} = TN / (FP + TN) \times 100\% \quad \text{(ii)}$$

Recall literally is how many of the true positives were found. That is why it is the ratio of true positive value and true positive value.

$$\text{Recall} = TP / (TP + FN) \times 100\ \% \quad \text{(iii)}$$

Precision means how many of the positively classified were relevant. It is all about the ratio of true positive & predicted positive value. A test can cheat and maximize this by only returning positive on one result it's most confident in.

$$\text{Precision} = TP / (TP + FP) \times 100\% \quad \text{(iv)}$$

F1 Score is the weighted average of Precision and Recall. Subsequently, this score considers both false positives and false negatives into account.

F1 score = (2 x Precision x Recall) / (Precision +Recall) ×100 % (v)

Receiver operating characteristics are very useful for the visual comparison of curve classification models. ROC curve made with true positive and false positive rates. The diagonal line represents a random guess. The curve of a model is closer to a random estimate, which is a less accurate model. Below are some ROC curves of our algorithm.

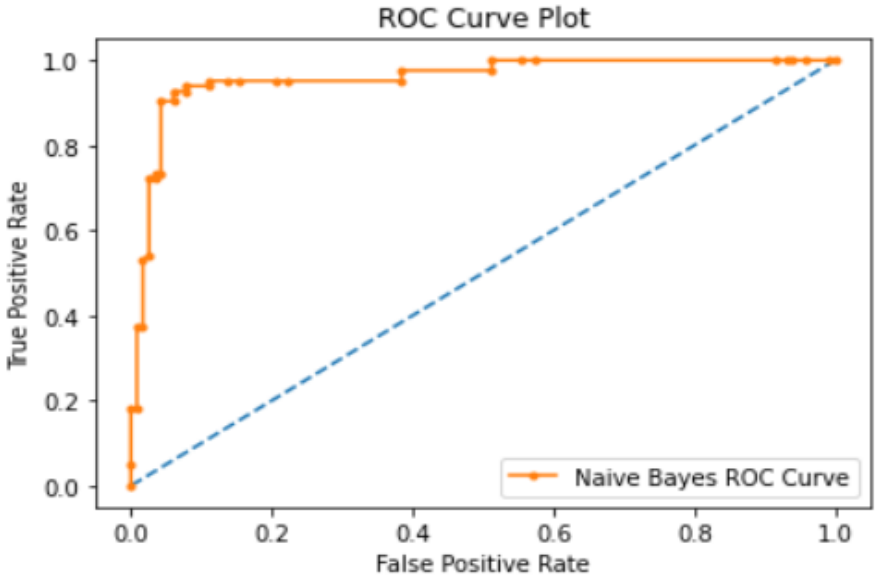


Figure 4.05: ROC Curve of Naive Bayes Algorithm

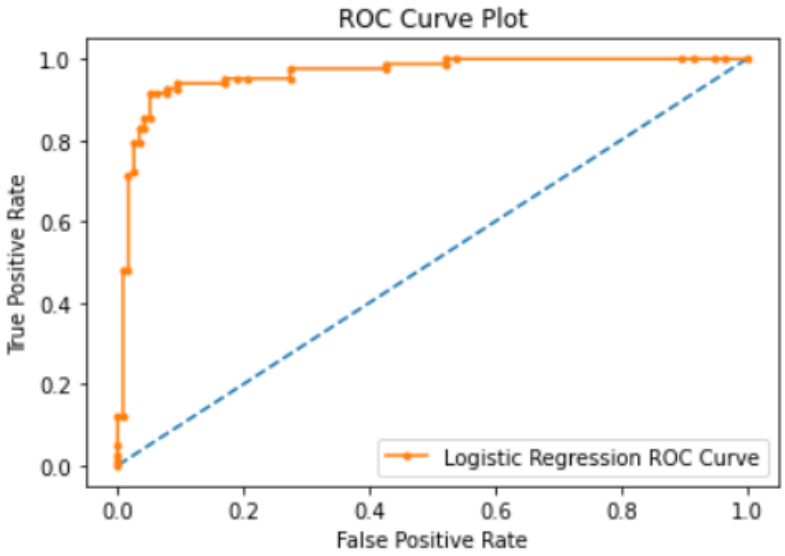


Figure 4.06: ROC Curve of Logistic Regression Algorithm

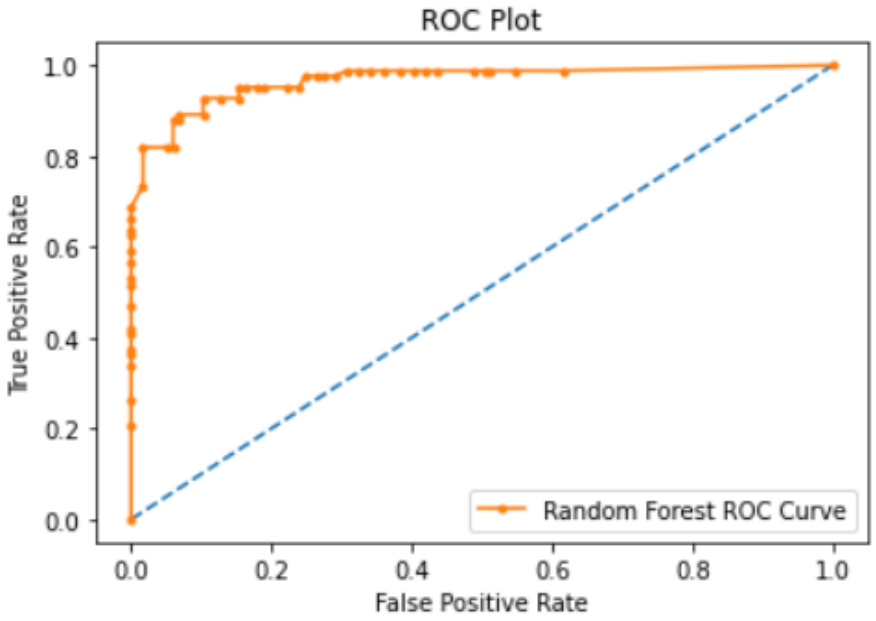


Figure 4.07: ROC Curve of Random Forest Algorithm

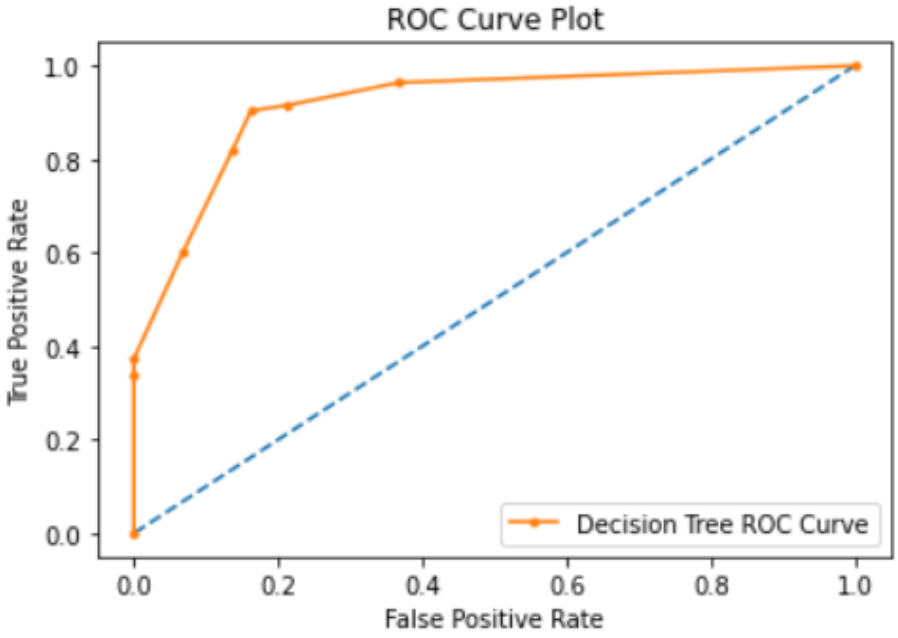


Figure 4.08: ROC Curve of Decision Tree Algorithm

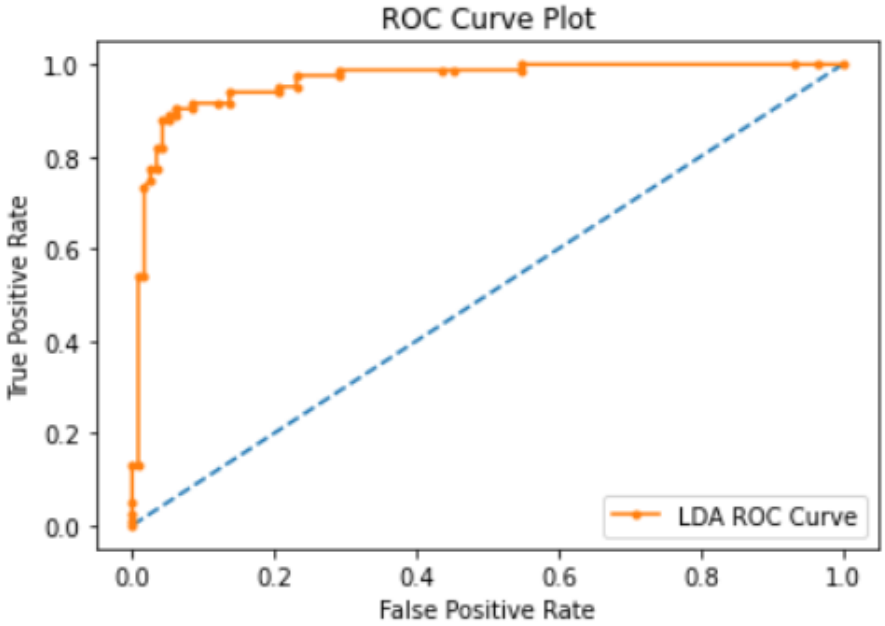


Figure 4.09: ROC Curve of LDA algorithm

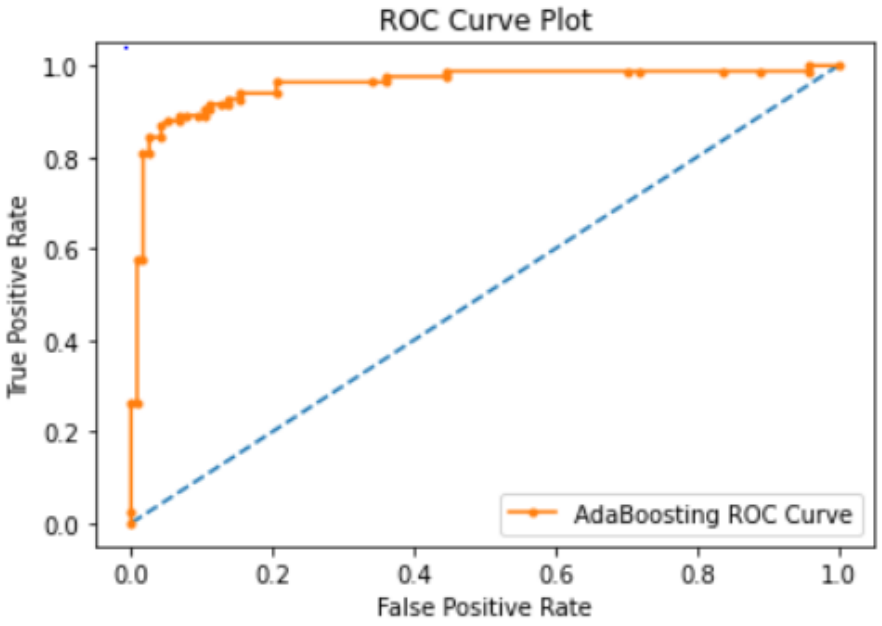


Figure 4.10: ROC Curve of Adaptive Boosting Algorithm

The confusion matrix is a technique of prediction results on a machine learning classification problem. This compares with the actual target values predicted by the machine learning model. It gives us an overview of how good our classification model is working. And also we can find out what kind of errors we have. It plays a vital role in measuring the efficiency of any classifier.

Table 4.2 shows the confusion matrix of the algorithms we have used. An accurate description of each classification is given in the following table.

Table 4.2: Confusion Matrix of all Classifiers.

<table>
<tr><th>Algorithms</th><th>Confusion Matrix</th><th>Algorithms</th><th>Confusion Matrix</th></tr>
<tr><td>KNN</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>105</td><td>12</td></tr>
<tr><td>Yes</td><td>7</td><td>76</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td><td>Logistic Regression</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>111</td><td>6</td></tr>
<tr><td>Yes</td><td>7</td><td>76</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td></tr>
<tr><td>LDA</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>109</td><td>8</td></tr>
<tr><td>Yes</td><td>8</td><td>75</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td><td>SVC</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>109</td><td>8</td></tr>
<tr><td>Yes</td><td>8</td><td>75</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td></tr>
<tr><td>NB</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>101</td><td>16</td></tr>
<tr><td>Yes</td><td>4</td><td>79</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td><td>Decision Tree</td><td>
<table>
<tr><td rowspan="4">True Class</td><td></td><td>No</td><td>Yes</td></tr>
<tr><td>No</td><td>101</td><td>16</td></tr>
<tr><td>Yes</td><td>15</td><td>68</td></tr>
<tr><td colspan="3">Predict Class</td></tr>
</table>
</td></tr>
</table>

Random Forest				Adaptive Boosting			
	True Class		No	Yes	True Class		No	Yes
		No	108	9		No	111	6
		Yes	9	74		Yes	10	73
		Predict Class				Predict Class		

Table 4.3 shows the performance of each algorithm. According to the performances of the algorithms and their accuracy performance, it will be decided which algorithm will fit for our model. Based on this accuracy, specificity, precision it can be surely said that logistic regression is best.

TABLE 4.3: Classifier Performance Table.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F1 score (%)
Naïve Bayes	90	95.18	86.3	83.15	95	89
SVC linear	93	90.36	93.2	90.4	90	90
kNN	90.5	91.6	89.7	86.36	92	89
Logistic regression	93.5	91.6	95	92.7	92	92
Decision tree	84.5	81.92	86.3	81	82	81
LDA	92	90.36	93.2	90.36	90	90
Random forest	91	89.1	92.3	89	89	89
ADA boosting classifier	92	88	95	92.4	88	90

4.3 Comparative Analysis

The motive of our work is to detect depression among Bangladeshi people. In paper [11], recognizing depression with six features and collecting text and emoji data from 1000 college students. In paper [14], depression level detection with 35887 images. In paper [15], detecting depression with text and question based data. In paper [16] and paper [9], detecting depression by using 10000 Twitter data. But In paper [9], sentiment analysis was done by using Bengali depression datasets. In paper [18], analysis of depression from social media datasets in Bangla language but they did not mention the using algorithms. In paper [19], they recognize depression by using facial images but they don't mention using classifiers and accuracy. In paper [20], detecting depression by using feature engineering. In paper [21], creating an application based on student's behavior. In paper [12], depression is identified by using vocal signals. They used 13 features and accuracy was 81.567%. In paper [14], detecting depression by using CNN and RNN algorithms. In paper [24], they create a chatbot that is based on emotion and they don't mention features and accuracy. Table 4.4 shows an overview of our work and other work.

Table 4.4: Comparison of the Result of our work and other works

Method/ Work Done	Object(s)/ Deal with	Problems Domain	Sample size	Feature set	Algorithms	Accuracy
Our work	Depression (Survey based text)	Detection	1000 Data	24	Logistic Regression	93.5%
Ding et al. [11]	Depression (text, emoji)	Recognition	1000 User	NM	DISVM	86.15%
Mulay et al.[10]	Depression (image)	Detection	35887 images	NA	CNN	66.45%
Asad et al. [15]	Depression (text,question)	Detection	150 users	NM	SVM, Naïve Bayes	74%

Deshpande et al.[16]	Depression (text,question)	Detection	10000 tweets	NM	support vector machine and Naive Bayes	NB- 83%, SVM-79%
Khan et al. [9]	Depression (text)	Detection	10000 Tweets	NM	SVM, Multinomial Naïve Bayes	73.33% NB- 86%
Zhou et al.[19]	Depression (Video)	Recognition	490 videos(Dataset AVEC 2013 and 2014)	NM	NA	NM
Stankevich, M [20]	Depression (text)	Detection	887 Reddit users	3	SVM Random Forest	63%
Patel et al. [21]	A chatbot (User text analysis)	Recognition	ISEAR dataset	NM	CNN, RNN, and HAN	75%
Shukla et al. [12]	speech signals of persons	Detection	Ryerson Audio-Visual Database of Emotional Speech and Song	13	Multi-Layer Perceptron algorithm	81.567%.
Orabi et al. [14]	Depression (text)	Detection	1,145 Twitter users	NM	CNN and RNN	87.957%
Ranade, A et al [24]	Chatbot (Based on emotion)	Recognition	NA	NM	NA	NA

[1]*NM*: Not Mentioned

[2]*NA*: Not Applicable

4.4 Discussion

The representation of algorithms, specificity, accuracy, recall, sensitivity, precision, ROC curve, and F1 score are reviewed in this section. The equations of evolutionary models and their effectiveness are also discussed here. We can see that the logistic regression algorithm gives maximum accuracy of 93.50%. Also, the logistic regression algorithm achieved 91.6% sensitivity, 95% specificity, 92.7% precision, 92% recall, and 92% F1 score. Finally, we discover that using logistic regression we can get the best performance in our depression detection model.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Depression detection with a machine learning model will have a positive impact on society. People are social beings so people have to live in society by participating in all social activities. Depression becomes a thorn in the side of walking together. People can get depressed at any time. In the discussion above, we have seen that people suffer from depression when they have trouble sleeping, when they get angry, when they lose energy, when they lose appetite, and when they do not take the stress. The reasons mentioned are common problems in our society which are capable of causing a person more and more difficult mental problems. Many people in our country have committed suicide which is very horrible for our country and society. We have tried to alleviate this condition of the country and society with our model. Everyone in the family should take care of everyone. Parents should take care of their children and pay attention to the activities of their children. Children should have a responsibility to be friendly with their parents and to help them with their work. If you ever feel mentally ill, you will be able to verify your current mental state with some necessary information in our model. In this way, we can make the right decisions at the right time and bring ourselves back to normal life. At the present time, many people in our society are suffering from depression, that is why we think our model will improve the state of mind of those people.

5.2 Ethical Aspects

This depression detection model is not anti-moral and does not violate human rights in such a way. The model does not collect any personal information, name, identity, etc. so there will be no privacy problem. This model does not undermine a person's right to enjoy or use, but rather plays an important role in making a person aware. So using machine-

learning technology, the model of depression detection can be managed without any problems.

5.3 Sustainability Plan

The Sustainability Plan gives us realistic ideas about any project and future plans. The purpose of our model is to find out the tendency of depression. This model should be noticed so that people can easily adjust. And to use this model it is important to make sure that people have the right idea about their position. Psychologists, psychiatrists, and mental health departments can use this model to speed up their work.

CHAPTER 6

SUMMARY, CONCLUSION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary and Conclusions of the Work

We can sum up our work into several parts like Data collection, Data preprocessing, Model implementation, and finally evaluation. From School, College & University, public areas and also online we collected our necessary data. After collecting data, we work on data preprocessing and algorithms implementation using Jupyter Notebook and Google Colab. After data preprocessing, we implement eight machine-learning algorithms and we use logistic regression, adaptive boosting (ADA boosting), k-nearest neighbor (kNN), Support Vector Classifier (SVC) Linear, naïve Bayes, decision tree, and Linear Discriminant Analysis (LDA) Classifier. And all the algorithms' performance with their accuracy, precision, sensitivity, Recall, F1 score is calculated and considered for finding the best model. It is noteworthy that the logistic regression algorithm gives the best performance. So it can be said that the aptitude to be conscious of depression needs to be modeled using logistic regression algorithms for detection.

6.2 Limitations of the Work

Our work is a machine learning algorithm which is a depression detection method. Our work and models have some limitations and flaws. The level of the dataset we used was relatively small, it is better to use it divided into more levels. Due to some limitations, we were not able to collect different classes of people's information. The dataset is not so large also. On data processing and preprocessing many highly advanced, developed methods and models can also be used and finally the model or system can be beautifully presented using a variety of advanced and popular algorithms.

By the help of our described model, it is surely possible to identify the situation of depression. We strongly believe that once this model is fully created, people will be able to use it more easily and realize the importance and value of this model to ensure emotional state. We are optimistic and positive that this model will give people the right idea about depression and that people will be conscious of their condition and will try to solve the problem.

6.3 Implication for Future Work

In recent years Modern technology, Data Science, Artificial intelligence have made our lives faster, easier, and more comfortable in every sector of human life. We want to develop our model in an Android application or web application in the future. In the future, we will try to increase the accuracy of our models. We will create a larger database with a huge amount of people's data and create more classification layers of datasets. In addition, by developing an immersive user-friendly GUI, the website or the mobile application developed the model can be made accessible to all the people and doctors also. Applying new algorithms, adding some different parameters, and adding some more features can make the model much more efficient and useful. A strong dataset can be created by collecting information from more various classes of people according to the district, age, and activity in the future. In addition, the model can be enlarged with the help of the Department of Mental Health.

REFERENCES

[1] OLSEN, L. R., JENSEN, D. V., NOERHOLM, V., MARTINY, K., & BECH, P. (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. Psychological Medicine, 33(2), 351–356.

[2] ZUNG, W. W. K. (1965). A Self-Rating Depression Scale. Archives of General Psychiatry, 12(1), 63

[3] MedicalNewsToday, Available at: https://www.medicalnewstoday.com/categories/depression [Online]. [Accessed 22 October 2021].

[4] Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research, 17(1), 37–49.

[5] Chenn-Jung Huang, ; Ming-Chou Liu, ; San-Shine Chu, ; Chin-Lun Cheng, (2004). [IEEE Fourth International Conference on Hybrid Intelligent Systems (HIS'04) - Kitakyushu, Japan (05-08 Dec. 2004)] Fourth International Conference on Hybrid Intelligent Systems (HIS'04) - Application of Machine Learning Techniques to Web-Based Intelligent Learning Diagnosis System. , (), 242–247.

[6] Cruz, A. Joseph, and D. S. Wishart. “Applications of Machine Learning in Cancer Prediction and Prognosis.” Cancer Informatics, Jan. 2006.

[7] Pahwa, Kunal; Agarwal, Neha (2019). [IEEE 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) - Faridabad, India (2019.2.14-2019.2.16)] 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) - Stock Market Analysis using Supervised Machine Learning. , (), 197–200.

[8] Meena, G., Sharma, D., & Mahrishi, M. (2020). Traffic Prediction for Intelligent Transportation System using Machine Learning. 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE).

[9] Khan, M. R. H., Afroz, U. S., Masum, A. K. M., Abujar, S., & Hossain, S. A. (2020). Sentiment Analysis from Bengali Depression Dataset using Machine Learning.

[10] Mulay, A., Dhekne, A., Wani, R., Kadam, S., Deshpande, P., & Deshpande, P. (2020). Automatic Depression Level Detection Through Visual Input. 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4).

[11] Ding, Y., Chen, X., Fu, Q., & Zhong, S. (2020) .A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm.

[12] Shukla, D. M., Sharma, K., & Gupta, S. (2020). Identifying Depression in a Person Using Speech Signals by Extracting Energy and Statistical Features. 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS).

[13] World EconomicForum,Available at: https://www.weforum.org/agenda/2021/10/brain-implant-could -cure-depression/ [Online]. [Accessed 1 May 2021].

[14] Orabi, A. H., Buddhitha, P., Orabi, M. H., Inkpen, D.(2018). Deep Learning for Depression Detection of Twitter Users, Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pages 88–97.

[15] Asad, Nafiz Al; Mahmud Pranto, Md. Appel; Afreen, Sadia; Islam, Md. Maynul (2019). [IEEE 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON) - Dhaka, Bangladesh (2019.11.28-2019.11.30)] 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON) - Depression Detection by Analyzing Social Media Posts of User. , (), 13–17.

[16] Deshpande, Mandar; Rao, Vignesh (2017). [IEEE 2017 International Conference on Intelligent Sustainable Systems (ICISS) - Palladam, India (2017.12.7-2017.12.8)] 2017 International Conference on Intelligent Sustainable Systems (ICISS) - Depression detection using emotion artificial intelligence. , (), 858–862.

[17] Dhaka Tribune, Available at: https://www.dhakatribune.com/bangladesh/2021/10/05/unicef-battered-by-pandemic-kids-need-mental-health-help[Online]. [Accessed 5 October 2021].

[18] Uddin, Abdul Hasib; Bapery, Durjoy; Arif, Abu Shamim Mohammad (2019). [IEEE 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) - Rajshahi, Bangladesh (2019.7.11-2019.7.12)] 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) - Depression Analysis from Social Media Data

in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique. , (), 1–4.

[19] Zhou, Xiuzhuang; Jin, Kai; Shang, Yuanyuan; Guo, Guodong (2018). Visually Interpretable Representation Learning for Depression Recognition from Facial Images. IEEE Transactions on Affective Computing, (), 1–1.

[20] Stankevich, M., Isakov, V., Devyatkin, D. and Smirnov, I. Feature Engineering for Depression Detection in Social Media. 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018), pages 426-431 ISBN: 978-989-758-276-9

[21] Patel, Falguni; Thakore, Riya; Nandwani, Ishita; Bharti, Santosh Kumar (2019). [IEEE 2019 IEEE 16th India Council International Conference (INDICON) - Rajkot, India (2019.12.13-2019.12.15)] 2019 IEEE 16th India Council International Conference (INDICON) - Combating Depression in Students using an Intelligent ChatBot: A Cognitive Behavioral Therapy. , (), 1–4.

[22] Mental Health America, Availableat: https://www.mhanational.org/conditions/suicide [Online]. [Accessed 23 February 2021].

[23] Nuruzzaman, M., Hossain, M. S., Rahman, M. M., Shoumik, A. S. H. C., Khan, M. A. A., & Habib, M. T. (2021). Machine Vision Based Potato Species Recognition. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS).

[24] Ranade, Advait Gopal; Patel, Maitri; Magare, Archana (2018). [IEEE 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) - Solan Himachal Pradesh, India (2018.12.20-2018.12.22)] 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) - Emotion Model for Artificial Intelligence and their Applications. , (), 335–339.

[25] Stuart J.Russell, Peter Norvig, Artificial Intelligence a Modern Approach, 3rd Edition, Upper Saddle River, NJ : Prentice Hall, 2010,pp. 725-744.

[26] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concept and Technique, 3rd Edition, Morgan Kaufmann, 2012,pp. 332-398.

[27]. Dahiwade, D., Patle, G., & Meshram, E. (2019). Designing Disease Prediction Model Using Machine Learning Approach. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC).

[28] Arif, M. A. I., Sany, S. I., Sharmin, F., Rahaman, M. S., Habib, M. T. (2021). Prediction of addiction to drugs and alcohol using machine learning: A case study on Bangladeshi population : International Journal of Electrical and Computer Engineering (IJECE), Vol. 11,No.5, pp. 4471-4480.

APPENDICES

Abbreviation

k-NN = k-nearest neighbors.

ML= Machine Learning.

SVM = Support Vector Machine.

Reflections of Research

Toward the start of this research work, we had almost no thought regarding Artificial Intelligence and machine learning consciousness discovery and acknowledgment. Our supervisor was extremely kind and earnest. He gave us significant direction and aided us a great deal. In this entire season of exploration, we learned new algorithm procedures, learned new data, figured out how to utilize calculations, and how to work with various techniques. We likewise found out with regards to the Jupyter notebook, Google Colab, and Python language. At first, there were issues working with these, yet step by step we turned out to be increasingly more acquainted with Google Colab and Jupyter notebook and Python.

At last, by doing the research we have acquired mental fortitude and been roused to accomplish more later on.

Plagiarism Report

ORIGINALITY REPORT

28%	24%	16%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	10%
2	librarysearch.aut.ac.nz Internet Source	1%
3	Umme Sanzida Afroz, Rafidul Hasan Khan, Sharon Akter Khushbu, Abu Kaisar Mohammad Masum. "Refinement of Bengali Obscene Words using sequence to sequence RNNs", 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021 Publication	1%
4	Submitted to University of Westminster Student Paper	1%
5	etds.lib.ncku.edu.tw Internet Source	1%
6	link.springer.com Internet Source	1%
7	www.knowledgehut.com Internet Source	1%