Bachelor of Science in Computer Science & Engineering

Developing a Mental Healthcare System for University Students of Bangladesh

by

Abu Bakkar Siddique

ID: 1504110

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

May, 2021

Developing a Mental Healthcare System for University Students of Bangladesh

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Abu Bakkar Siddique
ID: 1504110

Supervised by

Dr. Mahfuzulhoq Chowdhury
Assistant Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**Developing a Mental Healthcare System for University Students of Bangladesh'** submitted by ID: 1504110, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

_______________________________ Chairman

Dr. Mahfuzulhoq Chowdhury
Assistant Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

_______________________________ Member (Ex-Officio)

Dr. Md. Mokammel Haque
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

_______________________________ Member (External)

Dr. Md. Iqbal Hasan Sarker
Assistant Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Abu Bakkar Siddique

Signature of the candidate

Date:

Acknowledgements

I am grateful to Almighty God, who has given me the ability to complete my project and intend to perform the completion of B.Sc. Engineering degree. I am indebted to my supervisor Dr. Mahfuzulhoq Chowdhury, Assistant Professor, Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, for his motivation, proper guidance, constructive criticisms, and endless patience progress of the project. All these things contributed to my humble development as a machine learning analysis practitioner. He supported me by providing books, conference and journal papers, and practical bits of advice. From the very beginning, sir always encouraged me with proper guidelines, so the project never seemed a burden. All my teachers helped me with their invaluable assistance in every phase of my learning process over the course of four years. Finally, I'd like to express my appreciation to my friends, seniors, and the department staff for their valuable suggestions and assistance in completing the project.

Abstract

Depression is a major illness and a growing issue that affects a person's way of life, affecting normal functioning and impeding thought processes while they may be completely unaware of their condition. Depression is particularly common among young people in underdeveloped and developing countries. Youth in Bangladesh face challenges with their studies, careers, relationships, drugs, and family issues, all of which are major or minor contributors to depression. Furthermore, people in Bangladesh are hesitant to talk about this disease and sometimes misunderstand it as madness. This study focuses on gaining useful insights into why university students in Bangladesh, especially undergraduates, suffer from depression, in addition to predicting depression in university undergraduates for the purpose of referral to a psychiatric facility. A survey was used to gather data for this study. The survey was conducted with the help of a Google survey form. After training and testing the dataset with five algorithms, the best methods for predicting depression among Bangladesh undergraduates were discovered. A comparison of various prediction algorithms such as logistic regression, KNN, SVM, random forest, decision tree, including accuracy, precision, recall, specificity, error rate, f-measure, mean average percentage error for analysis was done. We also designed and developed an android mental healthcare mobile application to provide mental support to university students. All these things make people realize the necessity of sound mental health and support.

Keywords— Depression, Machine Learning, Prediction, Evaluation, Mental Healthcare, Android Application.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

With the aid of technology and human abilities, the world is changing at a breakneck rate. Everyone is so busy and materialistic these days that we fail to take time to consider our mental health. People are constantly pushing themselves to keep up with the fast pace of the world, putting themselves under a lot of physical and mental strain, which has a negative impact on their health, especially their mental health.

Depression is a silent killer that can do significant damage to a person if not handled promptly. Depression is a common mental disorder that affects everyone at some stage in their lives. However, due to a massive lack of self-awareness, society's judgment and people's prejudices, this illness is considered taboo in many parts of the world and people mock it and tease those who are diagnosed with it. For example, in Bangladesh, people believe that someone who has mental illness or has been diagnosed with a related illness is mad. It appears to be taboo in many Asian regions, as a result of which people who are experiencing it are less vocal about their problem and attempt to mask their condition. As a consequence, their condition can worsen, resulting in deterioration of their mental and physical health, as well as extreme depression, which may lead to suicidal behavior and self-harm. Furthermore, the majority of people are unaware of their depression problem and do not treat it as seriously as other illnesses, assuming that it will recover on its own given enough time. In the other hand, some people become depressed because they believe they are depressed, although this is not the case.

According to the National Institute of Mental Health [1], depression is the most prevalent mental illness, with 16.2 million people suffering from it. As compared to the general population, students are more likely to suffer from depression [2]. Students in universities, colleges, and medical schools are more vulnerable to depression, according to research conducted in some low and middle-income countries, and this is a worrying signal that should be addressed immediately [3], [4], [5], [2]. Despite the fact that this is one of the world's most pressing issues, few people are speaking out about it, and little research has been done using data mining techniques to uncover the root causes of depression among students, especially in Bangladesh.

Our intention is to make awareness of mental health by focusing on depression. We want to create a model by using that can predict whether a student is depressed or not. Prediction is a machine learning approach that will help analyze datasets collected from students. Again only identifying if a student is depressed or not is not enough as it is an alarming matter. We want to develop an android platform based mobile application that can provide mental healthcare to students who are suffering from depression. By the aid of the mobile application they can address their problems and seek help when they need it. It may save innocent lives by taking necessary precautions.

1.1.1 objective

1. To build a model that can predict whether a student is depressed or not.
2. Collect data from students to create datasets.
3. Visualize data.
4. Use machine learning to analyze the collected data and make prediction on depression.
5. Build an android platform based mobile application to provide mental healthcare.

1.2 Framework/Design Overview

The system overview is quite simple. The whole process starts with collecting data from students to prepare datasets. Collected data will be analyzed for making prediction. An android application will be developed that can provide mental healthcare.

1.3 Difficulties

Several difficulties were faced during project work. While conducting this study, we discovered that the majority of students are not as articulate about their problems; some of them believe they are depressed, which is affecting their performance, but they are not, and some students are unaware that they have moderate to extreme depression symptoms. Another issue we've discovered is that there are some students who are open about their depression problems, while others are unsure and want to know whether they're depressed or not so that they can seek clinical support. The below are some of the most challenging scenarios for making prediction and app development:

1. Identifying the underlying issues or causes that lead to depression.
2. Assuring students that their personal information will be kept confidential.
3. Complex questionnaires in scaling methods and feelings of discomfort to include all personal information tendency among students.
4. Making students attentive while filling the questionnaire as it takes some time and the questionnaire contains large questions.
5. Designing the database schema of the android application
6. Designing the dataflow of the application
7. Failure in getting notification in the mobile application due to issues on Firebase Cloud Messaging (FCM). For development purpose we had to test notification again and again and as a result it was getting blocked temporarily.

1.4 Applications

Since mental healthcare has become a significant concern around the world, it is really necessary to monitor this problem to prevent any kind of harm resulting from it. When accurate knowledge about mental ill and healthcare is accessible, only then people can address these problems and seek help for their own safety. School, college, varsity and other institutions are essential to be under monitoring for mental health related problems. People can test this system or model to monitor their own mental health. There are several application regarding this:

1. School, college level mental healthcare system.
2. University level mental healthcare system.
3. Different organizational health monitoring program.
4. Detection of other mental illness by following similar approach.
5. Creating different cross platform based application to provide mental aid.

1.5 Motivation

Most of the university students of Bangladesh suffers from mental illness for various reasons such as family income, academic result, relationship problems, political involvement, drug addiction, ragging/bullying etc. These problems are true for CUET also. When undergraduate students join here they fall victim to ragging, drug addiction, relationship problems, and political involvement very fast. So to make them interested to academic activities, sports and other activities we need a system that can identify students having psychological disorder such as depression and provide them with necessary mental support and mental healthcare.

1.6 Contribution of the thesis

A specific set of goals is achieved by performing thesis or research work. The areas of contribution of the thesis can be summarized as follows:

1. Designing a model that can predict depression.
2. Detect students suffering from depression.
3. Collecting data that can be used for future research purpose.
4. Developing an android based mobile application that can provide mental healthcare to depressed students.
5. Students who have academic/psychological problems such as family income, low grades, ragging, bullying, sexual harassment etc. can share these problems with respective teachers or doctors through the android application.
6. After good counselling, they might give proper concentration to their studies and daily life.
7. CUET administration might adopt this system to provide mental healthcare to its students and workers.

1.7 Thesis Organization

The following is how the remainder of this study paper is organized:

1. Chapter 2 discusses the previous studies about mental illness prediction related works and different prediction algorithm based works.
2. Chapter 3 illustrates the detailed work of the proposed methodology including designing the model that predicts whether a student has depression or not. It will also illustrate the dataflow and development process of the android mobile application that provides mental healthcare.
3. Chapter 4 describes the working dataset and analysis of the accuracy, precision, recall, F-Measure etc. based on different algorithms.
4. Chapter 5 contains the overall summary of this thesis work and provides some future recommendations.

1.8 Conclusion

In this chapter, an overview is provided. Along with the difficulties, the epitome of this whole system has been described in this chapter. The motivation behind this work and contributions are also stated here. In the next chapter, the background and present state of the problem will be given.

Chapter 2

Literature Review

2.1 Introduction

The aim of the study is to look at the difficulties that are to be faced while building the mental healthcare assistance system that uses machine learning to develop a model that predicts depression. By providing a summary of the previous study, this chapter discusses various applications that are done previously and their limitations and advantages. A detailed description of these methods is provided in this chapter.

2.2 Discussion about Depression

When we tell our family and friends that we are sad, we often use the word depression. Not all sorrow and grief, on the other hand, can be categorized as depression. Depression (major depressive disorder) is a widespread and severe medical condition that affects how we feel, think, and act, according to the American Psychiatric Association [6]. According to them, about one in every fifteen adults (6.7%) is affected by depression at any given time. In addition, 16% of individuals, or one in every six, will experience depression at some stage in their lives. According to studies, people typically develop depression around the age of 20, which is one of the main reasons we decided to conduct our study among undergraduate university students. Women are more likely to experience depression, with research estimating that one-third of women would experience depression at least once in their lives. Sadness and sorrow can strike at any time in one's life, whether as a result of the death of a loved one, issues in one's professional or personal life, or for no apparent cause. Depression, on the other hand, makes people feel worthless, causing them to despise themselves. Our study does not look at general cases of people that are depressed or sad, nor does it look at medical conditions like thyroid disorders, brain tumors, or vitamin deficiencies, which can trigger symptoms that are close to depression. Instead, we examine real-life examples of individuals who have been diagnosed or who need to be diagnosed with psychiatric or clinical depression. Depression manifests itself in a variety of ways, from constant sadness and anxiety, pessimism, loss of interest, and laziness to sleeping problems, weight changes, and suicidal thoughts. Both of these factors were taken into account when designing our survey, with some appearing as features for machine learning classification in the general collection of questions and others in the two depression scales. The paper goes into more detail on these topics later on. These signs must be present for at least two weeks before a person is diagnosed with depression [7]. Only intense sadness that lasts more than two weeks and prevents us from working normally can be a symptom of depression, according to Health Line, a leading source of health information in the United States [8].

2.3 Discussion about depression in university students

According to a study by the Center for Collegiate Mental Health, depression and anxiety are the most common causes of students seeking counseling, and one out of every five university students is depressed. Students at universities are more depressed than the general public [9]. According to another study, the depression rate among undergrad university students is high in both developed and developing countries with low income [5]. The number of college students seeking anxiety and depression counseling is steadily growing. The majority of them attempted suicide or attempted self-harm, and some of them found it difficult to function properly, while others discovered serious anxiety problems.

2.4 Background Study of the Problem

A depression rating scale is a common assessment instrument for analyzing a person's actions to see if further testing is needed to diagnose the person with a depressive disorder. Depression scales, which consist of a series of questions and inquiries to be completed by research participants, are often so reliable that they can detect or predict depression in individuals, despite the fact that their primary goal is to identify people who are at risk of developing the condition. Psychologists or specialists assess whether or not a person should be referred to further observation to determine whether he or she should be diagnosed with depression based on the degree of depression as measured by a depression scale. Depression scales can be divided into two types: those performed by researchers and those completed by patients. The Beck Depression Inventory is one of the most widely used scales by patients (BDI, BDI-1A, BDI-II). The Beck Depression Inventory (BDI) is a commonly used screening tool for determining the degree of depression in adults and adolescents over the age of 13. The Beck Depression Inventory [10] is the most commonly used tool for diagnosing depression.

In the paper [11] the authors used data mining, specifically classification to predict whether people will develop depression in the future if they are not already depressed. To conduct the study, they used synthetic data created with the aid of a Java program. There were 600 instances of data in the training sample and 400 instances in the test collection. Their attribute selection protocol was a lengthy and detailed process that included the analysis of several online surveys and questionnaires. They chose 31 characteristics, with the class variable "May Have Depression" being the final one. The attributes could have values ranging from 0 to 3, with 0 indicating that the attribute was missing in that person and 3 indicating that the person was seriously suffering from that symptom, with 'Mild' (1) and 'Medium' (2) in between. In order to find hidden patterns in the data, they used the famous machine learning and data mining tool WEKA for classification. In the training phase, 555 of the 600 instances were correctly identified, yielding a 92.5% accuracy. There were 263 true positives, 292 true negatives, 11 false positives, and 34 false negatives in their uncertainty matrix. They used accuracy, precision, and recall as classification metrics in their assessment. Their limitations are that they used synthetic data instead of using real data which could have lowered their accuracy by a margin. Their selection protocol was also very lengthy.

In paper [12] the authors used machine learning classifiers to develop the best approach for predicting depression in older people. The research was conducted using WEKA, a data mining method developed by the University of Waikato in New Zealand that can apply various machine learning techniques to problems such as data pre-processing, forecasting, classification, prediction, and regression. After comparing the performance, the best method for predicting depression was determined. By taking into account various socio-demographic variables and co-morbid conditions, the study proposed an automated method to combat depression, which is very common among senior citizens. Their goal was to eliminate the issues that come with manual diagnosis and treat patients as soon as possible. The training dataset was gathered from a slum in Bagbazar, Kolkata, which is served by the Bagbazar Urban Health and Training Centre (UHTC). The Geriatric Depression Scale (GDS) was used to interview 60 senior citizens who were at least 60 years old. Five classifiers, including BayesNet Classifier (BN), Logistic, Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), and Decision Table (DT), were compared using four metrics: accuracy, ROC area, precision, and root mean square error (RMSE). Their shortcomings are that the number of test subjects were very low and they used only senior citizens to gather data. As a result variety in training data is less.

In paper [13] the authors described how they used machine learning to predict whether a twitter user is suicidal or not based on validated self-reports, which are used to calculate suicide rates, and tweet feeds/status of active twitter users. The author's main task is to use their tweeter feeds to predict whether a user is suicidal or not, and the results are then compared to the results of the self-reports that the users completed previously for this study. The data for this study is gathered with great care. All of the participants in the "Survey for Twitter User" were participating Twitter users from the United States. 3 sets of questionnaires were provided to them. To evaluate suicidal rate, authors used the Depressive

Symptom Inventory–Suicide Subscale (DSI-SS), The Interpersonal Needs Questionnaire (INQ), and the Acquired Capability for Suicide Scale (ACSS). There are 85 females and 50 males in the study, with a range of ethnicity, educational income, and Twitter account creation dates. Following that, the article explains how to analyze each user's tweets. Then they explained how to use their tweeter feeds to predict their expected outcome, which is to predict whether a user is suicidal or not. They used a decision tree as their predictive model and used the scikit-learn library to implement the predictive analysis in Python. It correctly identified 9 people as suicidal, and they were actually suicidal in real life, and it correctly identified 115 people as non-suicidal, and they were all non-suicidal in real life. Their limitations are that the factors are not specific, the twitter feed is not the best measure to assess suicidal behavior, and it may call into question the validity and rational methodology of this research paper. Also instead of conducting research on the same age group of people, they conducted research on a variety of age groups.

In paper [14] the authors used clinical variables relevant to suicide and demographic variables for a machine learning approach that can predict whether an individual will attempt suicide or not. They used three algorithms: LASSO, SVM, and RVM, which were all implemented in MATLAB. Overall, the accuracy ranged from 65% to 72%. RVM was the most accurate, with a score of 72%. This research included 144 participants, and the data was collected meticulously to ensure that the consumer had correct information. For this study, a large amount of information about each topic was gathered. The Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID-I) was used to determine demographic backgrounds, Axis-I diagnoses, and clinical characteristics. The Hamilton Depression Rating Scale (HDRS), the Young Mania Rating Scale (YMRS), and the Hamilton Anxiety Rating Scale were used to assess current dimension mood and anxiety symptoms (HARS). As a result, they were able to identify 15 predictive variables that they used to train their model to predict whether an individual will attempt suicide or not. Except for one variable, which is a continuous variable, all of these are categorical variables that are normalized by z-scoring, which marks 0 as no and 1 as yes. Prediction accuracy, sensitivity, precision, positive predictive value (PPV), and negative predictive value (NPV) were used to assess the algorithm's validity in distinguishing individual suicide attempters from non-attempters (NPV). Three models' accuracy ranges from 64.7 to 72%. RVM had the highest result, about 72%, followed by SVM with 64.7% and LASSO with 68%. RVM predicted correctly 103 of the 144 subjects as suicide attempters or non-suicide attempters. Their limitation is that they neglected two important matrices that are F-Measure and Error Rate that are very important in result analysis. The number of participants was also less for such research purpose. They also used a few algorithms to test their data.

2.5 Machine Learning

2.5.1 What is Machine learning?

Learning is the process of transforming personal experience into information or skills. In a machine learning algorithm, training data represents experience and is fed into the algorithm as input to train and learn it, with the output resembling information or expertise gained from the training data.

2.5.2 Algorithms Used

We used 5 algorithms to determine depression. Idea about them has been given.

Decision Tree: Decision tree algorithms graph out data (mined, open access, internal) in branches to demonstrate the possible outcomes of various decisions. Decision trees are easily explainable and accessible to novice data scientists, and they can be used for incomplete data sets. They classify response variables and forecast response variables based on previous choices.

Logistic Regression: This is a method of data processing that employs mathematical analysis. As more data is used, the algorithm's ability to sort and classify data improves, allowing predictions to be made.

SVM: The Support Vector Machine (SVM) can solve classification and regression problems. However, it is most often used to solve classification problems. In this

Table 2.1: Confusion Matrix

		Predicted	
		Positive	Negative
Observed	Positive	TP	FN
	Negative	FP	TN

algorithm, each data object is mapped. Data has been split into both sides and a drawing line has been plotted.

KNN: For the same purpose, the K-nearest neighbor supervised learning algorithm is used. It's been used for both types of analysis. It is primarily used in industry to solve problems of classification and prediction.

Random Forest: From the datasets, random samples were chosen. Following the selection process, a decision will be made, and polling will be performed for each predicted result.

2.5.3 Matrices Used

Confusion Matrix: A binary classification confusion matrix is a two-by-two table generated by counting the number of binary classifier outcomes. They are True Positive(TP), False Positive(FP), True Negative(TN), and False Negative(FN). Table 2.1 shows a confusion matrix

Accuracy: : The number of accurate predictions divided by the total number of predictions in the dataset yields the accuracy (ACC). The highest level of accuracy is 1.0, while the lowest level is 0.0.

$$Accuracy = \frac{TP + TN}{P + N}$$

Precision: The number of accurate positive predictions divided by the total number of positive predictions yields precision (PREC). It's also known as positive predictive value (PPV). The highest precision is 1.0, and the worst precision is 0.0.

$$Precision = \frac{TP}{TP + FP}$$

Recall: The number of accurate positive predictions divided by the total number of positives yields sensitivity (SN). It's also known as the recall rate (REC) or the true positive rate (TPR) (TPR). The highest level of sensitivity is 1.0, while the lowest level is 0.0.

$$Recall = \frac{TP}{TP + FN}$$

Specificity: The number of accurate negative predictions divided by the total number of negatives yields the specificity (SP). It's also known as real negative rate (TNR). The highest level of precision is 1.0, while the lowest level is 0.0.

$$SP = \frac{TN}{TN + FP}$$

Error Rate: The error rate (ERR) is determined by dividing the total number of incorrect predictions by the total number of predictions in the dataset. 0.0 is the highest error rate, while 1.0 is the worst.

$$ERR = \frac{FP + FN}{P + N}$$

F-Measure: A harmonic mean of precision and recall is the F-score.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

Mean Absolute Percentage Error (MAPE): In the sense that it is the percentage of the error relative to the actual value, the mean absolute percentage error is a good tool. This gives a more uniform error measurement. For example, if the error was 10 and the actual value was 100, the percentage would be 10%, whereas if the error was 100 and the actual value was 1000, the percentage would be 10%.

$$MAPE = \frac{\sum \frac{|A_t - F_t|}{|A_t|}}{n} * 100$$

Root Mean Square Error(RMSE): The RMSE squares the difference between the real and forecast values, finds the average of all the squares, and finally calculates the square root. It can appear that squaring and then taking the square root are mutually exclusive operations. This is not right for large errors as larger errors are reprimanded by the RMSE.

$$RMSE = \sqrt{\frac{\sum |A_t - F_t|^2}{n}}$$

2.5.4 Visualization

Data visualization provides a lot of information that data alone cannot provide. Python provides some of the most engaging data visualization tools available. The more basic plot styles are spread through libraries, but some are exclusive to specific libraries. We have used some visualization tools that are Matplotlib and Seaborn.

2.6 Conclusion

All the background study has been discussed in this section. The limitations have also been discussed in this section. Several parameters have been used to develop the system. In the next chapter, details of the working procedure will be illustrated. Finally, Prediction models and it's result analysis will add a new dimension to the work.

2.6.1 Implementation Challenges

A few implementation challenges are preparing datasets, collecting data from students, applying the machine learning algorithms, visualization of the collected data and finally comparing the performance of different algorithms.

Chapter 3

Methodology

3.1 Introduction

This section contains the total overview of how we designed the model that predicts if a student is depressed or not. Detailed steps of data collection and data cleaning have also been described. Then we showed how we designed the android application to provide mental healthcare to the students suffering from depression and help them. Necessary figures and snapshots of the application development have also been attached to understand the total process easily.

3.2 Overview of Framework

3.2.1 Experimental Setup:

The following is a list of methods that have been used to incorporate Mental Healthcare Assistance system:

1. Software

    - Operating System: Windows 10
    - IDE: Android Studio
    - Front End is developed by Java
    - Database: Firebase Real time Database
    - Notification: Firebase cloud messaging (FCM)
    - Python for model prediction
    - Google colab for visualization

3.2.2 System Design for Depression Analysis

The whole procedure starts with questionnaires asked from university students. The questionnaires will have information like family income support, present/past psychological health background, academic result, political involvement, relationship problem, drug addiction, victim to ragging/bullying etc. Then student's data will be collected in a excel sheet for further use. Data cleaning will be done by removing incomplete data and finalize it for data analysis. Then machine learning algorithms such as Logistic Regression, K Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest etc. will be implemented. The reason for using different algorithms is that different algorithms work better in different cases. Some work better with less data, whereas some are effective with huge data depending on the algorithms specialty. We want our module to give us the best accuracy possible to determine if a student is depressed or not. Fig 3.1 shows this whole process below

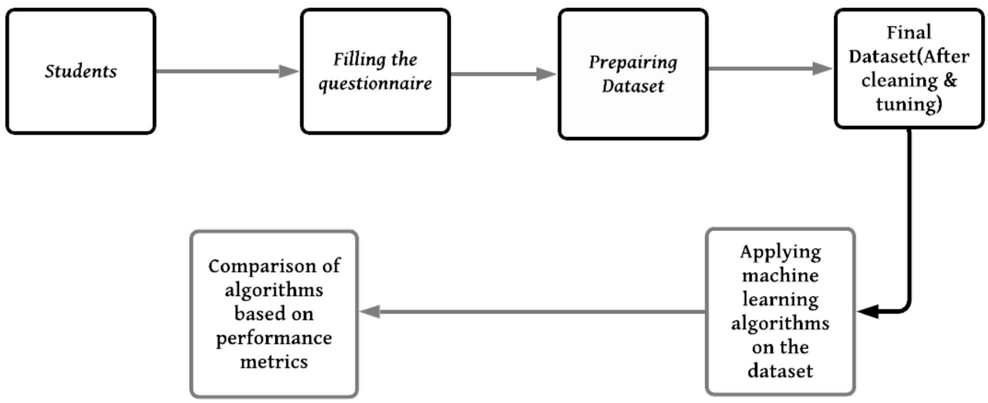


Figure 3.1: Block Diagram for Depression Analysis

3.2.3 System design for Android Application:

After finding the students who are depressed we need to design our android based mobile application to provide mental healthcare to them. Fig 3.2 shows the whole data flow diagram of the android application.

Logging in as a student the first thing a student will find is a page of questions that essential to find the students reason of mental problems. When the student fills the asked questions on the screen, the questions will be analyzed for finding academic and psychological problems separately in two tabs. For academic problems like low CGPA, financial problem, being victim of bullying/ragging/harassment etc. the student has an option of sharing the problems related to academics to a teacher who is an existing user in the application. And for psychological problems like sadness from death or loss, addiction of any sort, recent breakup etc. the student will find an option of sharing his psychological problems to a psychologist who is also an existing user in the application. Students can request for an appointment to the psychiatrist available in the application.

When a teacher logs in to the application he can the student's information such as ID, phone number, department etc. He can also see the student's academic problems along with a personal text from the student. Based on this the teacher can take necessary steps to help the student in improving his academic condition.

Again when a doctor logs in to the application he can see the student's psychological problems along with a personal message of the student. He can then approve students for an appointment and schedule an appointment with a particular student which the student will be notified via notification.

3.3 Detailed Explanation

3.3.1 Dataset Preparation

Our survey paper consists of a series of questionnaires that measure the participants' psychological functioning. It is divided into three parts, the first of which is the consent form, which the individual may accept in order to proceed to the next section. We gave them a short overview of what we were doing in the consent form and stated that the students were participating voluntarily. We also informed them that they will not be monitored at any cost and that they can leave the survey at any time. Fig 3.3 shows the consent form. The questionnaires in the second section consisted of 15 questions. These 15 questions are the most important factors that contribute to a person's depression, as described in Chapter 2. Section 3 also includes the Beck Depression Inventory-II (BDI-II)

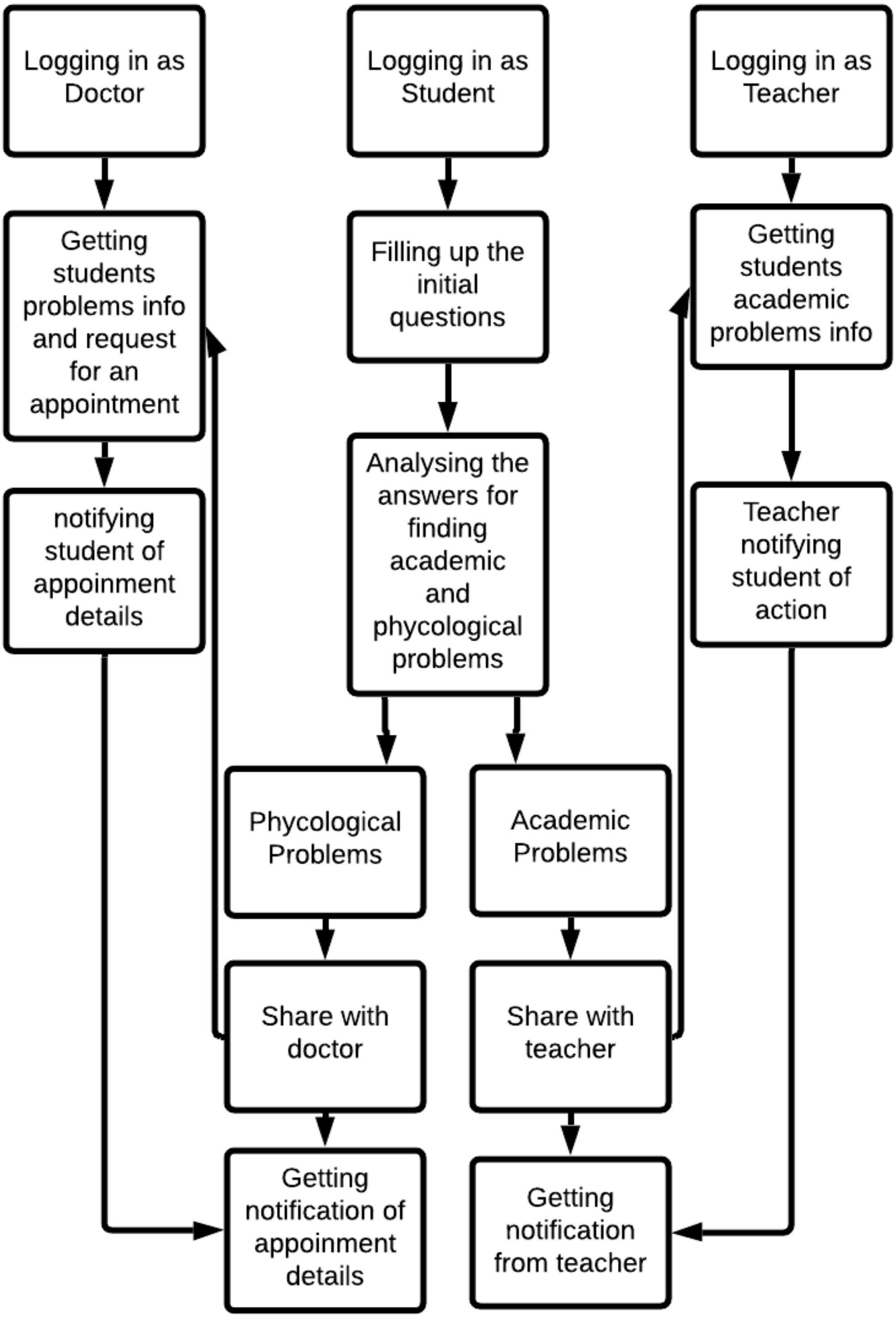
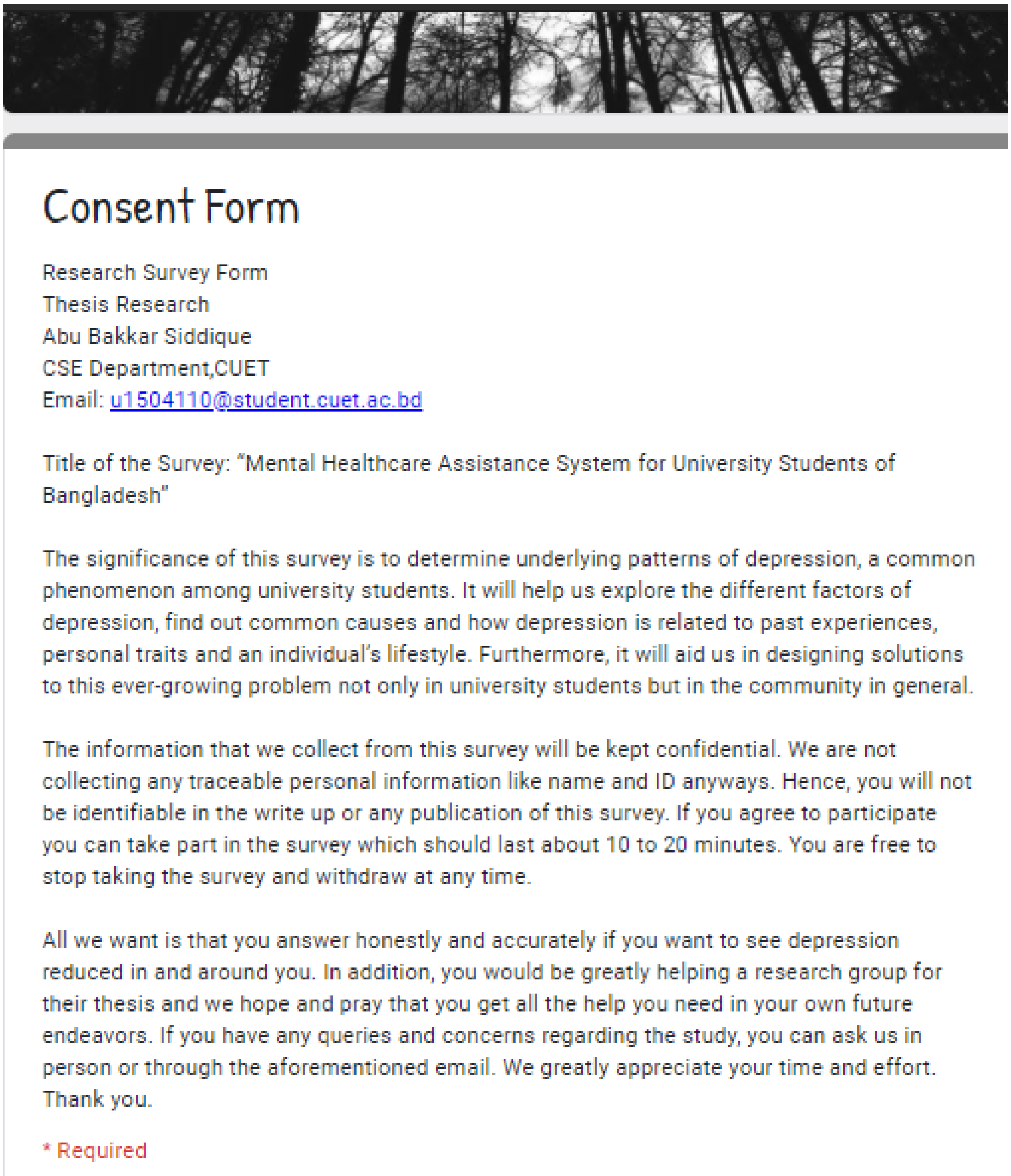


Figure 3.2: Block Diagram for Android Application

scaling method for assessing whether or not a student is depressed.Fig 3.4, fig3.5, fig3.6 shows the beck depression scale.

We conducted comprehensive research to identify the most important factors that



Consent Form

Research Survey Form
Thesis Research
Abu Bakkar Siddique
CSE Department,CUET
Email: u1504110@student.cuet.ac.bd

Title of the Survey: "Mental Healthcare Assistance System for University Students of Bangladesh"

The significance of this survey is to determine underlying patterns of depression, a common phenomenon among university students. It will help us explore the different factors of depression, find out common causes and how depression is related to past experiences, personal traits and an individual's lifestyle. Furthermore, it will aid us in designing solutions to this ever-growing problem not only in university students but in the community in general.

The information that we collect from this survey will be kept confidential. We are not collecting any traceable personal information like name and ID anyways. Hence, you will not be identifiable in the write up or any publication of this survey. If you agree to participate you can take part in the survey which should last about 10 to 20 minutes. You are free to stop taking the survey and withdraw at any time.

All we want is that you answer honestly and accurately if you want to see depression reduced in and around you. In addition, you would be greatly helping a research group for their thesis and we hope and pray that you get all the help you need in your own future endeavors. If you have any queries and concerns regarding the study, you can ask us in person or through the aforementioned email. We greatly appreciate your time and effort. Thank you.

* Required

Figure 3.3: Consent Form

may contribute to depression, which we used to create our first dataset.They're outlined below:

Family history of mental illness: If there is a family history of depression, descendants are more likely to be depressed. A 30-year study of three generations at high and low risk for depression found that grandchildren from two previous generations who had been diagnosed with depression had the greatest risk of being diagnosed with depression [15]. As a result, we can see how a family history of this disease leads to depression in subsequent generations.

Beck's Depression Inventory

This depression inventory can be self-scored. The scoring scale is at the end of the questionnaire.

1.
- 0 I do not feel sad.
- 1 I feel sad
- 2 I am sad all the time and I can't snap out of it.
- 3 I am so sad and unhappy that I can't stand it.

2.
- 0 I am not particularly discouraged about the future.
- 1 I feel discouraged about the future.
- 2 I feel I have nothing to look forward to.
- 3 I feel the future is hopeless and that things cannot improve.

3.
- 0 I do not feel like a failure.
- 1 I feel I have failed more than the average person.
- 2 As I look back on my life, all I can see is a lot of failures.
- 3 I feel I am a complete failure as a person.

4.
- 0 I get as much satisfaction out of things as I used to.
- 1 I don't enjoy things the way I used to.
- 2 I don't get real satisfaction out of anything anymore.
- 3 I am dissatisfied or bored with everything.

5.
- 0 I don't feel particularly guilty
- 1 I feel guilty a good part of the time.
- 2 I feel quite guilty most of the time.
- 3 I feel guilty all of the time.

6.
- 0 I don't feel I am being punished.
- 1 I feel I may be punished.
- 2 I expect to be punished.
- 3 I feel I am being punished.

7.
- 0 I don't feel disappointed in myself.
- 1 I am disappointed in myself.
- 2 I am disgusted with myself.
- 3 I hate myself.

8.
- 0 I don't feel I am any worse than anybody else.
- 1 I am critical of myself for my weaknesses or mistakes.
- 2 I blame myself all the time for my faults.
- 3 I blame myself for everything bad that happens.

9.
- 0 I don't have any thoughts of killing myself.
- 1 I have thoughts of killing myself, but I would not carry them out.
- 2 I would like to kill myself.
- 3 I would kill myself if I had the chance.

10.
- 0 I don't cry any more than usual.
- 1 I cry more now than I used to.
- 2 I cry all the time now.
- 3 I used to be able to cry, but now I can't cry even though I want to.

Figure 3.4: Beck scale

Academic condition: Varsha Srivastava, a Boston University student, talks about how her low grades made her nervous and stressed her out, ultimately leading to depression [16]. As a result, we thought it was important to include this feature in our research.

Drug addiction: According to many studies, drug addiction or alcohol misuse is closely linked to depression. According to the National Bureau of Economic

11.
0 I am no more irritated by things than I ever was.
1 I am slightly more irritated now than usual.
2 I am quite annoyed or irritated a good deal of the time.
3 I feel irritated all the time.

12.
0 I have not lost interest in other people.
1 I am less interested in other people than I used to be.
2 I have lost most of my interest in other people.
3 I have lost all of my interest in other people.

13.
0 I make decisions about as well as I ever could.
1 I put off making decisions more than I used to.
2 I have greater difficulty in making decisions more than I used to.
3 I can't make decisions at all anymore.

14.
0 I don't feel that I look any worse than I used to.
1 I am worried that I am looking old or unattractive.
2 I feel there are permanent changes in my appearance that make me look unattractive
3 I believe that I look ugly.

15.
0 I can work about as well as before.
1 It takes an extra effort to get started at doing something.
2 I have to push myself very hard to do anything.
3 I can't do any work at all.

16.
0 I can sleep as well as usual.
1 I don't sleep as well as I used to.
2 I wake up 1-2 hours earlier than usual and find it hard to get back to sleep.
3 I wake up several hours earlier than I used to and cannot get back to sleep.

17.
0 I don't get more tired than usual.
1 I get tired more easily than I used to.
2 I get tired from doing almost anything.
3 I am too tired to do anything.

18.
0 My appetite is no worse than usual.
1 My appetite is not as good as it used to be.
2 My appetite is much worse now.
3 I have no appetite at all anymore.

19.
0 I haven't lost much weight, if any, lately.
1 I have lost more than five pounds.
2 I have lost more than ten pounds.
3 I have lost more than fifteen pounds.

Figure 3.5: Beck Scale

Research, people who have been diagnosed with depression at least once consume 69% of America's alcohol and 84% of America's cocaine [17]. As a result, there is ample evidence to suggest that there is a connection between drugs and depression.

Relationship/Affair: Couple relationships can be affected by depression, and vice versa. It has been shown that people who have relationship issues are three

20.

0 I am no more worried about my health than usual.
1 I am worried about physical problems like aches, pains, upset stomach, or constipation.
2 I am very worried about physical problems and it's hard to think of much else.
3 I am so worried about my physical problems that I cannot think of anything else.

21.

0 I have not noticed any recent change in my interest in sex.
1 I am less interested in sex than I used to be.
2 I have almost no interest in sex.
3 I have lost interest in sex completely.

INTERPRETING THE BECK DEPRESSION INVENTORY

Now that you have completed the questionnaire, add up the score for each of the twenty-one questions by counting the number to the right of each question you marked. The highest possible total for the whole test would be sixty-three. This would mean you circled number three on all twenty-one questions. Since the lowest possible score for each question is zero, the lowest possible score for the test would be zero. This would mean you circles zero on each question. You can evaluate your depression according to the Table below.

Total Score	Levels of Depression
1-10	These ups and downs are considered normal
11-16	Mild mood disturbance
17-20	Borderline clinical depression
21-30	Moderate depression
31-40	Severe depression
over 40	Extreme depression

Figure 3.6: Beck Scale

times more likely to be depressed than those who do not. University of Waterloo psychologist Uzma Rehman and colleagues (2015) say that people who have clinical depression are not content with their relationships [18]. Breakups may also cause temporary anxiety and sleep problems, but depression can develop only if these and other issues continue for more than two weeks. Both of these considerations were taken into account in a series of questions on our survey form.

Conflict with friends: In a similar way to couple relationships, problems in friendship relationships can lead to depression in people. If a person has frequent conflicts with friends, it will undoubtedly affect their feelings and, in some cases, cause them to be excessively depressed for an extended period of time, leading to depression. Depressed individuals, on the other hand, are more likely to have troublesome friendships because they can be demoralizing and difficult to console [19]. Even though there is little evidence in the literature linking depression to friendship disputes, it is still a factor to consider.

Financial problems in family: Money, or the lack it, dominates most of what we feel and do in an increasingly materialistic world. As a result, it's no wonder that a family's financial problems, which lead to stress and a sense of pressure on children, as well as less money to spend on fun activities, will send a university student into depression. Financial difficulties and concern about debt at university, according to a study conducted by the University of Southampton and the Solent NHS Trust among 454 first-year British undergraduate students [20], increase the risk of mental health problems such as depression and alcohol abuse. As can be shown, there is a well-established correlation between depression and family financial problems.

Violence in family: We made it clear to the questionnaire respondents what we meant by family violence up front. Domestic abuse or violence is described as any threatening or abusive behavior between family members, both physical and emotional, and in particular such misconduct between parents, which can negatively impact a child's or a teenager's well-being and affect them for the rest of their lives. As a major effect of domestic violence, Hiremath Debaje (2014) concludes from a study conducted among adolescents in Mumbai, India, that adolescents may suffer from mild to serious depression [21]. As a result, family abuse has the potential to cause depression in both the short and long term.

Sadness from death or loss: Sadness and sorrow are normal reactions to the death of a loved one that everybody goes through. However, a persistent state of sadness and hopelessness to the point where a person's capacity to work is harmed and they choose solitude may be a symptom of depression [22].

Victim of bullying: Bullying is the attempt to exert control over someone who is perceived to be weaker than the bully or bullies. It entails physically and emotionally harming others by pulling, punching, teasing, offending, and so on. Bullying has been shown to have both short and long term effects on infants, teens, and even adults. Bullied children and teenagers are more likely to develop depression and even commit suicide, according to research [23], [24]. As a result, we used the history of bullying as a criterion for detecting depression in people.

Sexual harassment or abuse: In general, sexual assault takes the form of inappropriate contact, gestures, threats, and, in the most serious case, rape. Particularly from the perspective of Bangladesh, the victim is typically a woman. According to Dr. Colleen Cullen, a licensed clinical psychologist, depression is one of the most common diagnoses of sexual assault victims [25]. According to one study's results, sexual assault is a stressor linked to increased depressive symptoms [26]. The explanation for this is self-evident, as sexual assault is a major social issue that affects people of all cultures and countries, from the poorest to the wealthiest, and disrupts the normal way of life and well-being of more than half of the world's population.

Hours spent on social media: Websites and apps that enable people to communicate with each other virtually and share personal or professional content are referred to as social media. For quite some time, the growth of social media use, especially among younger people, has been a source of concern. According to We Are Social and Hootsuite's joint survey, Digital in 2018, 18% of Bangladesh's population, mainly young people, use social media actively [27]. The number is staggering, at 30 million users, with Facebook being the most popular site and cell phones being the most popular unit. As a result, it is one of the most crucial factors to consider.

3.3.2 Data Collection

The Computer Science and Engineering Department at Chittagong University of Engineering and Technology has accepted the research study described in our proposal report. All respondents gave their consent and had to sign a consent form before taking the survey. Our main concern during this study was the privacy of the participants' data. As a result, we have assured all participants that no personal details such as their name, address, student ID, email address, or phone number will be obtained from them during the survey. We have no means of tracing back and identifying any individual from the information we have gathered. Hence the data privacy of our survey participants has been preserved and maintained. The survey was conducted from July 3rd, 2020 to the present, and data was collected from 493 students. In addition, we are continuing to gather data for our future studies. The majority of our survey was completed by undergraduate students from CUET and IIUC, with a few other universities providing undergraduate students with their own consent. We looked at a lot of previous survey samples and created our own survey questionnaires that followed the standard structure and specifications. Furthermore, we also respected every participant's decisions, such as whether or not they are able to leave the survey at any time. Since this is a psychological survey, we had to make special effort when conducting online surveys because we are not physically present with the participants. We used a Google survey form in which the participants had to fill out Google form questionnaires. We also discussed the guidelines that participants should obey when participating in the survey. The rules are as given:

1. Participants in the survey should make every effort to complete it within 15-20 minutes.
2. Students should respond to the scale based on their most recent feelings, which should be no more than one week old.
3. They do not have to worry about a question's answer twice. They should instead choose the first choice that comes to mind.
4. At any point during the survey, a participant may leave if they do not wish to proceed.
5. The participant must be honest about their feelings and not include false details in the survey.

3.3.3 Data Cleaning:

Finally, after gathering data from 493 students, we decided to begin making predictions based on the information we had gathered. However, before applying the prediction algorithms to the data, we cleaned it to eliminate as many outliers as possible in order to improve accuracy. As a result, we produced different data versions.

Our survey had two parts. The first part contained student's information like family income support, present/past psychological health background, academic result, political involvement, relationship problem, drug addiction, victim to ragging/bullying etc. The second part contained beck scaling which was very important for us. Since our prediction model will be based on this scaling system. Each of the questions in this scaling method had its own unique score. The depression scales' scores add up to provide a final result. Based on the final score, there were many classifications. However, because we're determining whether or not an individual student is depressed. Normal was changed to "No" and all other depression classifications were changed to "Yes". We assigned the appropriate scores to each query in this manner, following the standard format of the scaling method. The cumulative score for the Beck scaling method was then included. A new column was developed to keep track of beck's score. The corresponding score-based classification was then saved in a new column called "Beck Label". Finally, after this move, we had 493 data with a final label of "Yes" or "No" indicating whether or not a student was depressed.

Then we prepared two different datasets. One contained the features based on students personal information related answers. This dataset has 15 features and we named it as F-15. The other dataset contained the standard questions from beck scaling which has 21 different personality trait questions and we named it as F-21. Then we used our five prediction algorithms to see what might unfold. We ran all the algorithms based on the final label and our features on two separate datasets to see how accurate the predictions were.

3.3.4 Android Application Development:

For developing our android application we first designed the dataflow of the android application shown in Fig3.2. We used Android studio which is the official integrated development environment for Google's Android operating system. For creating the front End we used Java. Firstly we created the interface according to our plan which is shown in Fig3.7. Then we created the home layout of the application as shown in Fig3.8. The home will contain information about different mental illness. We also created an in app depression checker which can help students to check their depression level anytime. It is shown in Fig3.9.

For creating the database we used Firebase Real Time Database. Firebase is also a platform developed by google for creating mobile and web applications. The Firebase Real Time Database is a cloud hosted NOSQL database that stores and sync data between users in real time. For giving user notification we used Firebase Cloud Messaging (FCM) multicast that can send message to multiple devices. It contains payload information as well as the list of device registration tokens to which the message should be sent. In our database, data is stored as JSON objects. This database can be visualized as cloud-hosted JSON tree. Unlike a SQL database, there are no tables or records. When a data is added to the JSON tree, it becomes a node in the existing JSON structure with an associated key.

As there are three kinds of user (student, teacher, doctor) in our application we created three objects named Student, Teacher and Doctor each with some associated key to identify them. For students we used department, email, gender, phone number, student ID, university name as associated keys. A snapshot of student instance in the database is shown in Fig3.10. For teachers we used department, DSW, email, phone number, position, university name as associated keys. Fig3.11 shows teacher instance on the database. And for doctors we used degree, email, experience, phone number as associated keys.This is shown in Fig3.12 as doctor instance on the database. After registering using the email every user is given a unique id which is important for checking so that a user can use only one email once for registration.

3.3.5 Implementation

In this section we have implemented the whole setup. With the system design diagrams, we have completed the majority of tasks in this section. We also described the detailed description of data collection and android application development which are two of the most important section of our work.

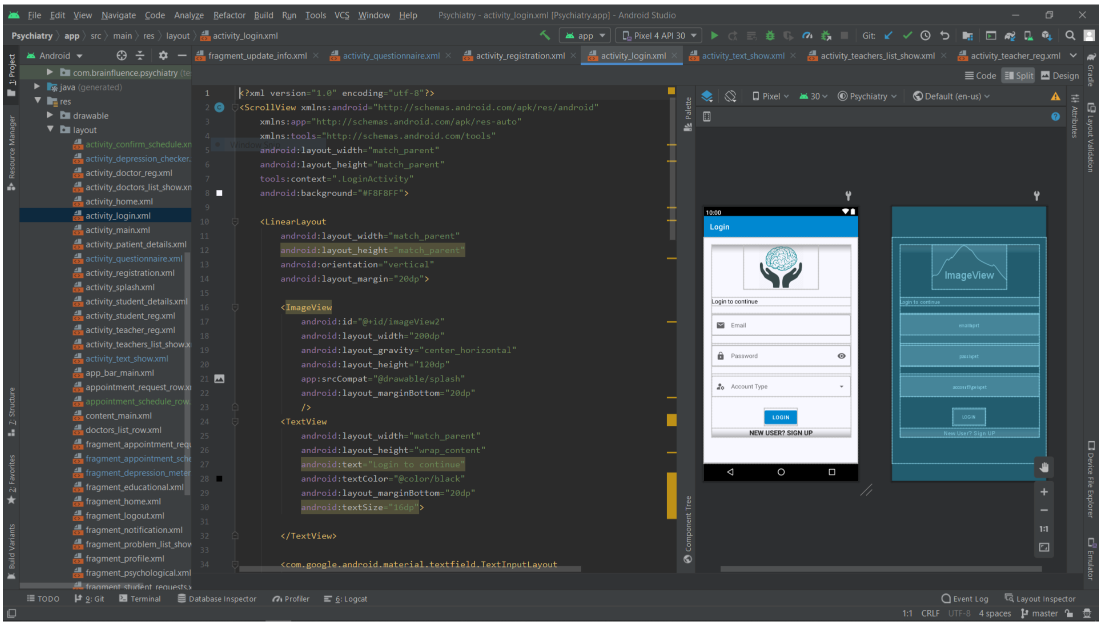


Figure 3.7: Login Layout

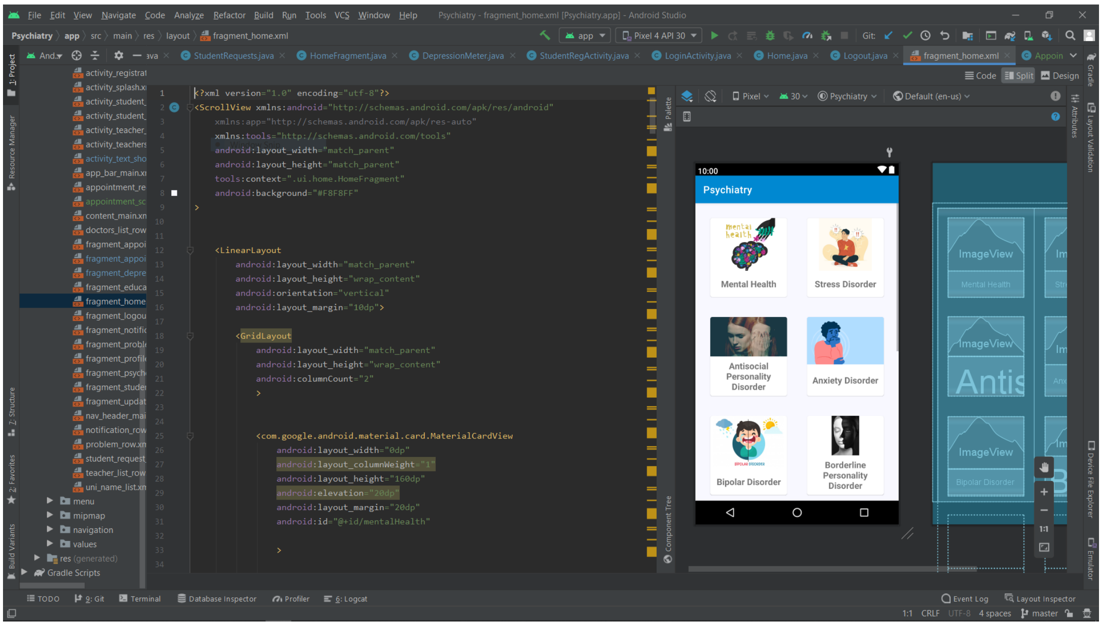


Figure 3.8: Home Layout

3.4 Conclusion

We have provided a brief overview of our Mental Healthcare Assistance System. Development process of the android application have been discussed. We prepared the dataset by collecting the data and cleaned it to apply machine learning algorithms. The whole analysis procedure will be discussed in the next chapter.

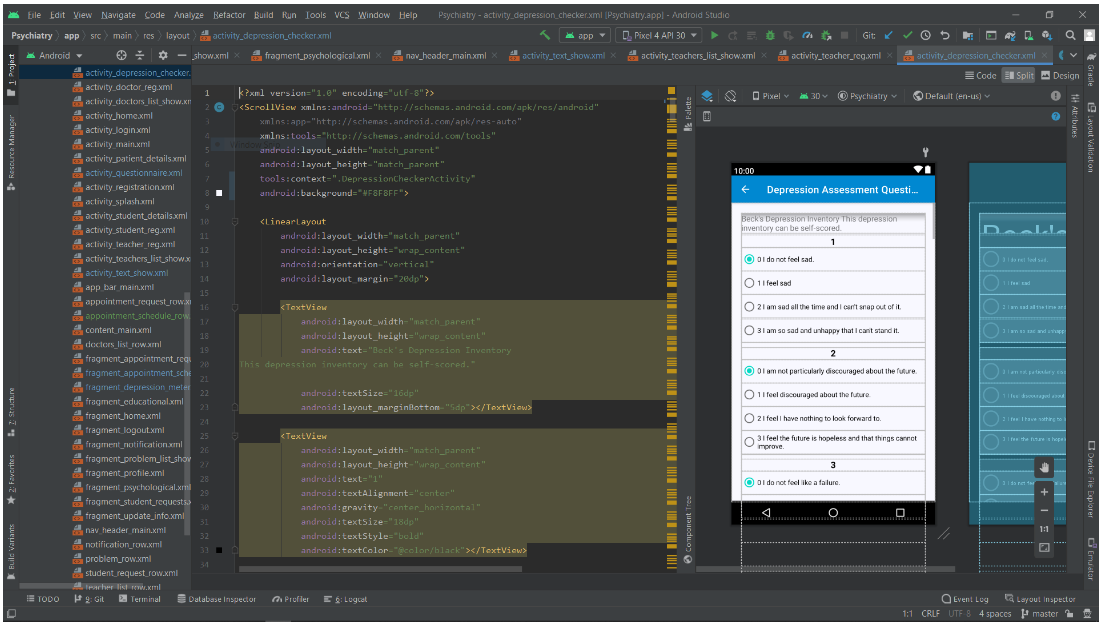


Figure 3.9: Depression Checker Layout

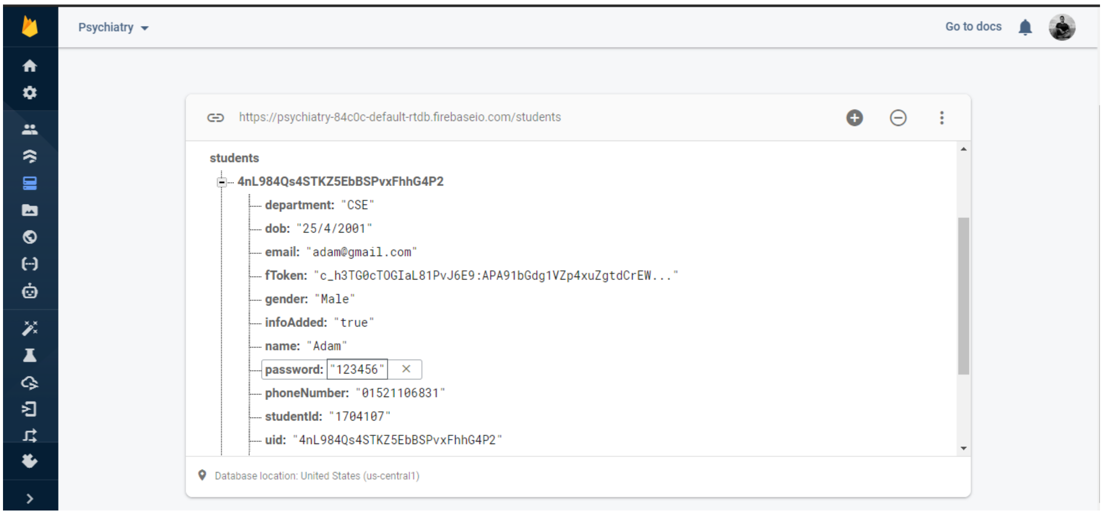


Figure 3.10: Student Instance on Database

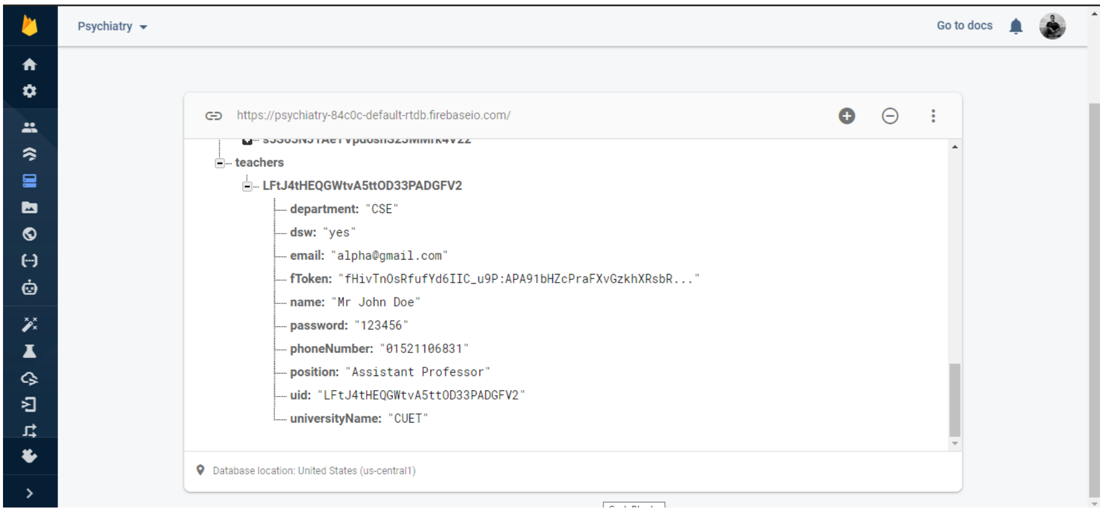


Figure 3.11: Teacher Instance on Database

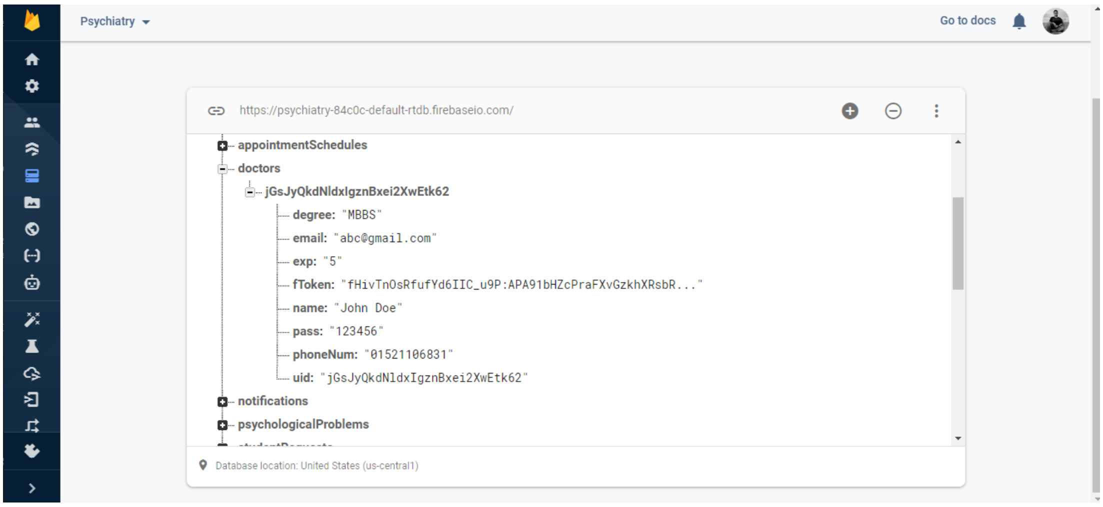


Figure 3.12: Doctor Instance on Database

Chapter 4

Results and Discussions

4.1 Introduction

Two key goals will be discussed in this chapter as we stored data on a dataset. One is to analyze the data. For this, we train the dataset with several algorithms. We consider two different dataset for this analysis part. One of them contains information about students personal life which we have named as F-15 as it contains 15 features. The other one contains beck depression scale which we have named as F-21 as it has 21 features. The second goal is to show the interface and the features of the android mental healthcare mobile application.

4.2 Dataset Description

In this project we analyzed different reasons to determine if a student is depressed or not which we explained in chapter 3. For this we have created two different dataset. The first one contains 15 questions that focuses on student's personal information and life. We named it F-15 as it has 15 features. The second dataset contains 21 question that are collected from beck's depression scale. We named it as F-21. Fig 4.1 shows our whole dataset.

Our prediction model will be based on beck scaling system. Each of the questions in this scaling method had its own unique score. The depression scales' scores add up to provide a final result. Based on the final score, there were many classifications. However, because we're determining whether or not an individual student is depressed. Normal was changed to "No," and all other depression classifications were changed to "Yes." We assigned the appropriate scores to each query in this manner, following the standard format of the scaling method. The

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Timestam	Academic	Medium c	Age [Row	Departme	Are you ha	Are you ac	Are you ha	Did you ha	How ofter	Do you ha	If yes, doe	Violence i	Sadness fr	Have you	If yes doe	Have you	If yes, doe	How many	1 [Row 1]	2 [Rov
2	2020/03/0	4th Year	Bangla	24-26	Business A	No	No	No	Yes	Rarely	Yes	Most of th	Never	No	Yes	Never	No	Never	3 to 6 hou	I do not fe	I am n
3	2020/03/0	4th Year	Bangla	24-26	Engineeri	No	No		No	Rarely	Yes	Often	Never	No	No		No		Greater th	I feel sad	I feel
4	2020/03/0	4th Year	Bangla	21-23	Engineeri	No	No		No	Rarely	No		Rarely	Yes	No		No		Greater th	I feel sad	I feel
5	2020/03/0	3rd Year	Bangla	18-20	Engineeri	Maybe	No	No	No	Rarely	No	Rarely	Rarely	Yes	No	Never	No		Greater th	I am sad a	I feel
6	2020/03/0	4th Year	Bangla	24-26	Bachelor c	No	Yes	No	Yes	All of the	Yes	Most of th	Most of th	No	Yes	Most of th	Yes	Most of th	Greater th	I am so sa	I feel
7	2020/03/0	1st Year	Bangla	18-20	Engineeri	No	No		No	Rarely	No		Rarely	No	No		No		3 to 6 hou	I am sad a	I feel
8	2020/03/0	4th Year	Bangla	24-26	Engineeri	Yes	No	Yes	No	Rarely	No		Never	Yes	Yes	Never	No		1 to 3 hou	I feel sad	I am n
9	2020/03/0	4th Year	Bangla	21-23	Engineeri	No	No		Yes	Rarely	No		Never	Yes	No		No		Less than	I do not fe	I am n
10	2020/03/0	4th Year	English	21-23	Medical	No	No	No	No	Often	No	Most of th	Rarely	Yes	Yes	Never	Yes	Often	Greater th	I feel sad	I feel
11	2020/03/0	4th Year	English	21-23	Medical	No	No	No	No	Rarely	No		Rarely	Yes	Yes	Never	No		1 to 3 hou	I feel sad	I feel
12	2020/03/0	4th Year	Bangla	24-26	Medical	No	No		No	Rarely	No		Never	Yes	No		Maybe	Rarely	3 to 6 hou	I feel sad	I am n
13	2020/03/0	4th Year	English	21-23	Medical	Maybe	No		No	Rarely	Yes	Often	Rarely	Yes	Yes	Rarely	Maybe	Rarely	Less than	I feel sad	I am n
14	2020/03/0	4th Year	Bangla	21-23	Medical	Yes	No	Yes	No	Rarely	No		Never	No	Yes	Never	No		1 to 3 hou	I feel sad	I feel
15	2020/03/0	1st Year	Bangla	21-23	Bachelor c	Maybe	No	No	No	Often	No		Never	Yes	Yes	Most of th	Yes	Most of th	Greater th	I am sad a	I feel
16	2020/03/0	4th Year	Bangla	21-23	Medical	No	No	No	No	Rarely	No		Never	No	Yes	Rarely	Yes	Often	3 to 6 hou	I am sad a	I feel
17	2020/03/0	1st Year	Bangla	18-20	Bachelor c	Maybe	No		No	Rarely	No		Never	No	Yes	Rarely	Maybe	Never	1 to 3 hou	I do not fe	I am n
18	2020/03/0	4th Year	Bangla	24-26	Engineeri	No	No	Yes	No	Rarely	No	Most of th	Rarely	Yes	Yes	Most of th	No	Often	3 to 6 hou	I feel sad	I feel
19	2020/03/0	3rd Year	Bangla	21-23	Engineeri	No	No		No	Rarely	No		Never	Yes	No		Yes	Often	3 to 6 hou	I feel sad	I feel
20	2020/03/0	4th Year	Bangla	21-23	Medical	No	No	No	No	Never	No		Never	No	No		No		3 to 6 hou	I feel sad	I am n
21	2020/03/0	4th Year	Bangla	21-23	Engineeri	No	No	Yes	No	Rarely	Yes	Often	Rarely	Yes	No		No		3 to 6 hou	I do not fe	I feel
22	2020/03/0	2nd Year	Bangla	21-23	Engineeri	No	No		No	Rarely	No		Never	No	No		No		3 to 6 hou	I feel sad	I feel
23	2020/03/0	4th Year	Bangla	21-23	Engineeri	No	No		No	Rarely	Yes	Often	Rarely	No	No		No		1 to 3 hou	I feel sad	I am n

csv

Figure 4.1: Dataset

Table 4.1: Depression Labeling

Beck Score	Status	Label
0-16	Not Depressed	0
17-63	Depressed	1

cumulative score for the Beck scaling method was then included. A new column was developed to keep track of beck's score. The corresponding score-based classification was then saved in a new column called "Beck Label." Finally, after this move, we had 493 data with a final label of "Yes" or "No", indicating whether or not a student was depressed. The table 4.1 shows the details of the labeling

4.3 Impact Analysis

In this project we have tried to build a model that can if a student is suffering from depression or not. But only finding a model that predicts depression is not enough. So we came up with an android platform based mobile application that can provide mental healthcare in time of dire need. This model can be adopted by schools, colleges and even business organizations to keep check on clinical depression. We can say that this project has social and ethical impacts.

4.3.1 Social and Environmental Impact

This project has a very crucial impact on society. People are busy and materialistic these days that they fail to take time to consider mental health. People are constantly pushing themselves to keep up with the fast pace of the world, putting themselves under a lot of physical and mental strain, which has a negative impact on their health, especially their mental health. Depression is a common mental disorder that affects everyone at some stage in their lives. People believe that someone who has mental illness or has been diagnosed with a related illness is mad. It appears to be taboo in many Asian regions, as a result of which people who are experiencing it are less vocal about their problem and attempt to mask their condition. As a consequence, their condition can worsen, resulting in deterioration of their mental and physical health, as well as extreme depression, which may lead to suicidal behavior and self-harm. Our intention is to make awareness of mental health by focusing on depression. By the aid of the depression model people can keep a check on their mental health. By using the mobile application they can address their problems and seek help when they need it. It may save innocent lives by taking necessary precautions.

4.3.2 Ethical Impact

This project has some ethical consequences too. This depression prediction model can be implemented in schools, colleges, university and other institutions for a better youth. If depression or any other similar mental illness increases, this will harm the youth. It can be implemented in industrial or business organizations too. If people became aware of their mental health they can improve their productivity and contribute more to the society we live in.

4.4 Data Visualization

Data visualization is important because it enables people in comprehending the significance of data by visual representation. If we represent data in a text-based system, we can miss important information such as trends and patterns in the data. Visual representation may also reveal important co-relationships between data. As a result, representing data in a visual sense is a perfect way to gain a deeper understanding of data. There are numerous methods for visualizing data. As a result, we will use a histogram to represent our results, which is a technique for representing the distribution of numerical data. In a histogram, the total number of bars is shown on the y-axis, while each bar represents a particular piece of information on the x-axis. As a result, we'll use a histogram to represent each function and try to make sense of the results. The orange bar denotes "Not Depressed," while the blue bar denotes "Depressed." For each possible response, we will count the number of students who are depressed and those who are not depressed.

We can see from the figure 4.2 that there are more students who say they have no family history of mental illness. 427 students out of 493 said no, with 115 being depressed and 312 being unaffected. As a result, it's possible that people who have no problems at home are less depressed than those who answered yes. If there is a family history of mental illness, we can deduce that 30 out of 63 people are depressed.

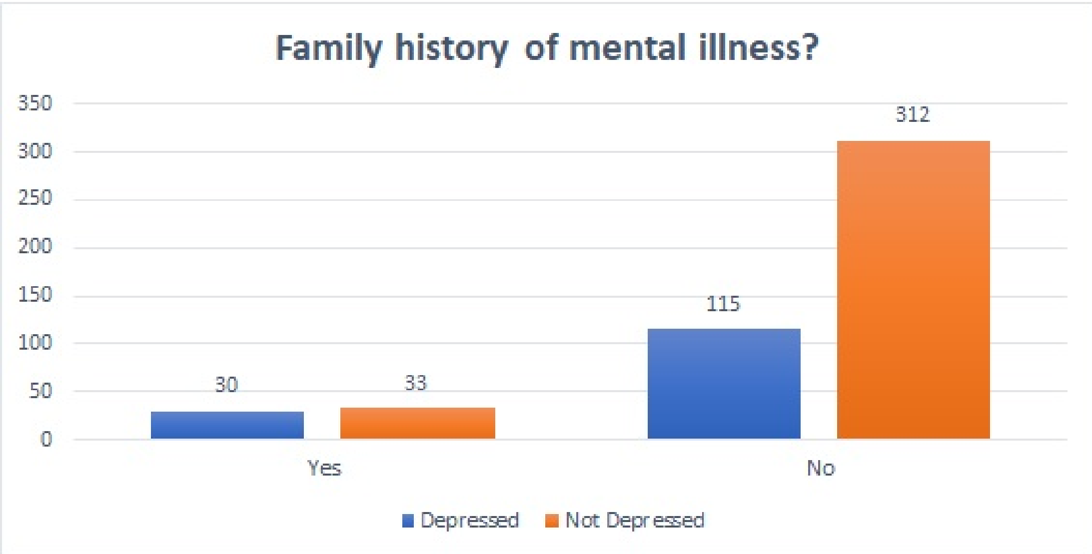


Figure 4.2: Data Comparison for Family history of mental illness?

Figure 4.3depicts the academic situation, which involves performance, family pressure about grades, a large study load, and an awareness of daily classes. We can see from the visual representation that the majority of people who say yes to this question are not depressed. Just 8 students out of 62 are depressed, while 54 are not. Additionally, the number of students who answered "no" when asked if they were satisfied with their academic results demonstrates how strongly the function affects depression among students. When they answer no to this question, about 178 out of 328, are depressed. As a result, we can conclude that this question is one of the most important factors in predicting depression.

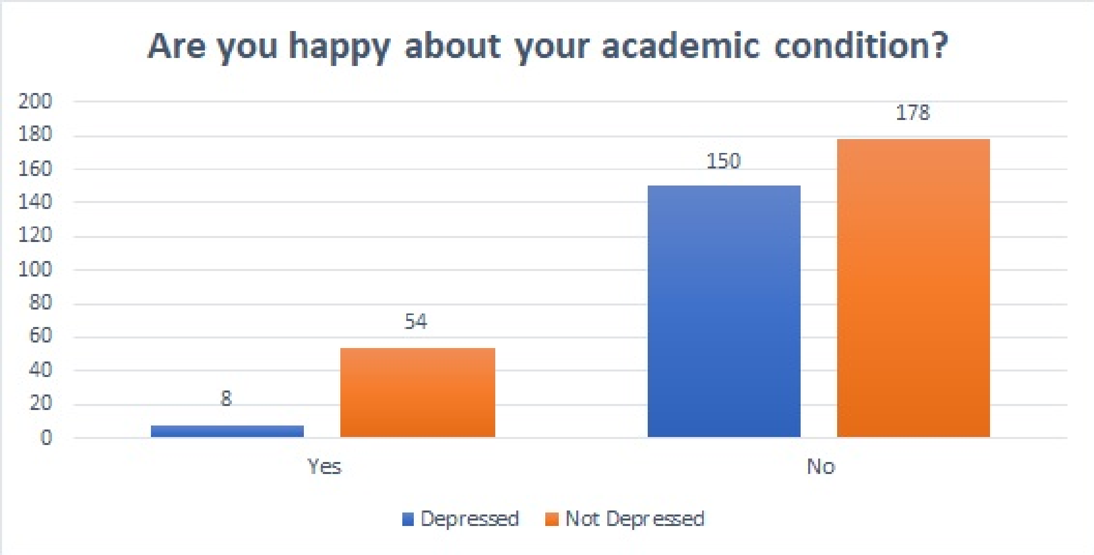


Figure 4.3: Data Comparison for Are you happy about your academic condition?

It's difficult to draw any conclusions based solely on the visual representation of the following data in Figure 4.4. This is due to the fact that we have so little knowledge for the "Yes" answer to this issue. We only have 29 yeses, with 8 being depressed and 21 being unaffected. But we should at least conclude that if someone does not use drugs, he is less likely to suffer from depression. There are 462 students who do not use drugs, with 149 being depressed and 312 being non-depressed.

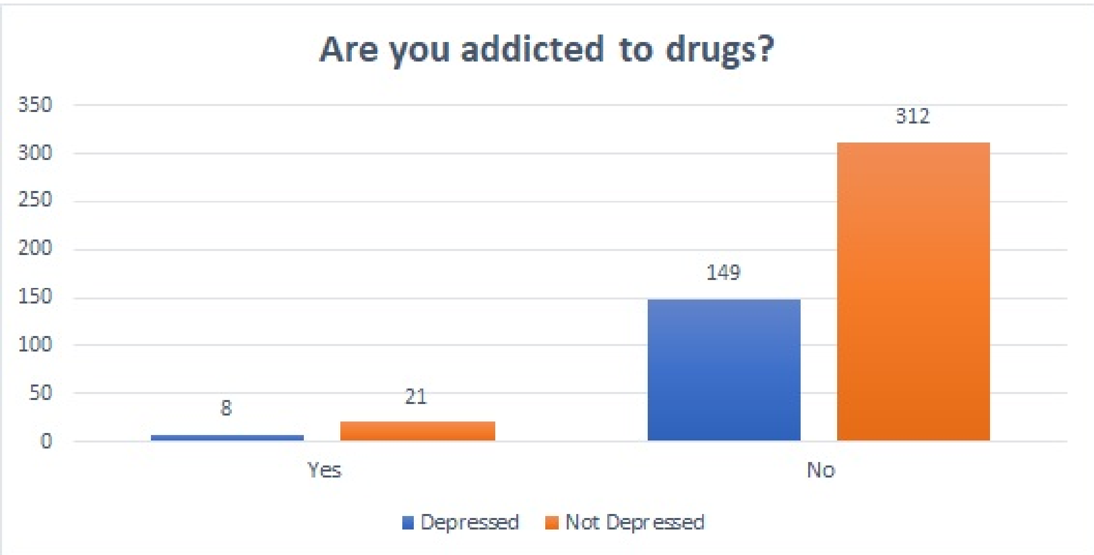


Figure 4.4: Data Comparison for Are you addicted to drugs?

Figure 4.5 shows that even though they are not in a relationship, 223 out of 322 students are not depressed. While 60 out of 168 are depressed when they are in a relationship.

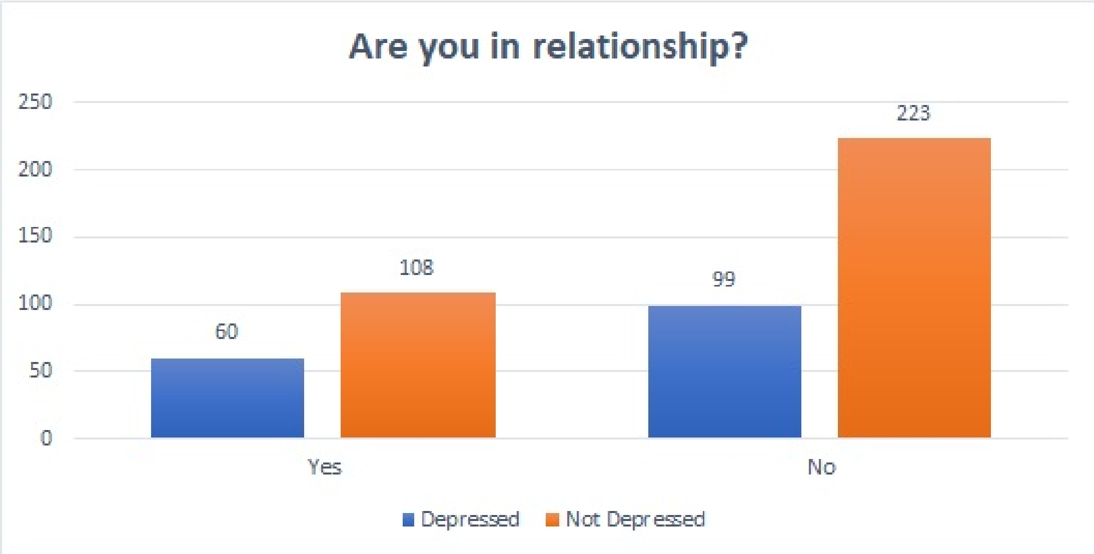


Figure 4.5: Data comparison for Are you in a relationship?

Figure 4.6 depicts whether or not students are depressed after a recent breakup. If a student has recently experienced a breakup, we can see that more 39 out of 76, are sad. We can also see that 273 out of 414 students are not sad if they do not experience a breakup. As a result, we can conclude that a recent breakup has a significant impact on depression risk.

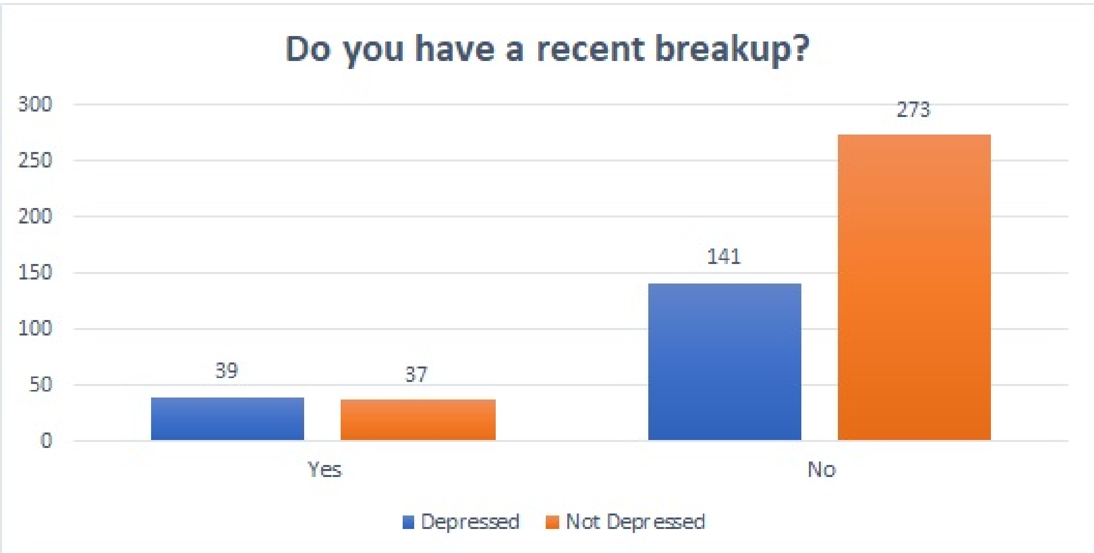


Figure 4.6: Data comparison for Do you have a recent breakup?

The question in Fig 4.7 has a significant impact on depression. We can deduce this by looking at the "Yes" and "No" responses to the following question. Students who answered yes to a family financial issue are more or less depressed. For the situation, 92 out of 194 people, are depressed. However, it can be concluded that students who do not have any financial difficulties are less likely to be depressed. When they do not have any financial problems, 218 out of 296 students are not stressed.

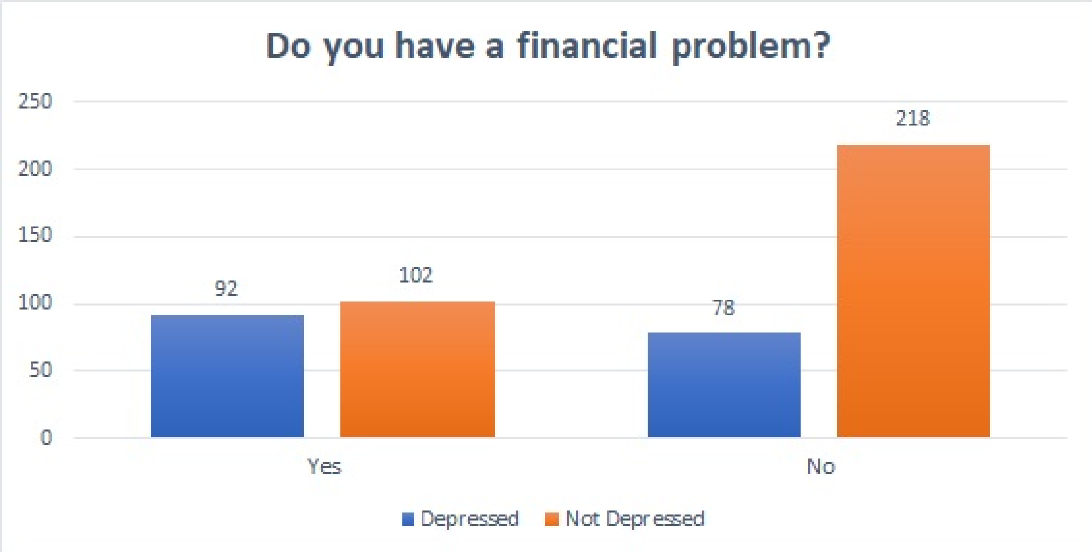


Figure 4.7: Data comparison for Do you have a financial problem?

Members of the family who are physically and verbally violent indicate family violence. Figure 4.8 shows how critical this issue is for predicting depression. 48 out of 52 students are depressed when there is either frequent abuse in the family. Students who rarely or never have violence in their families are often less likely to be depressed. Since there is no family violence, 296 out of 428 people are not depressed. As a result, we can once again consider this function to be a critical factor in depression prediction.

We can't tell whether or not this issue has an effect on the final result just by looking at the pattern in fig 4.9. We can see that the majority of people are not depressed with both "Yes" and "No." Yes, 108 out of 169 people are depressed. Students who have been bullied are more depressed than students who have not been bullied.

If they had ever been sexually assaulted or abused, they were asked as shown in Figure 4.10, has an effect on a student's depression. We can infer from the

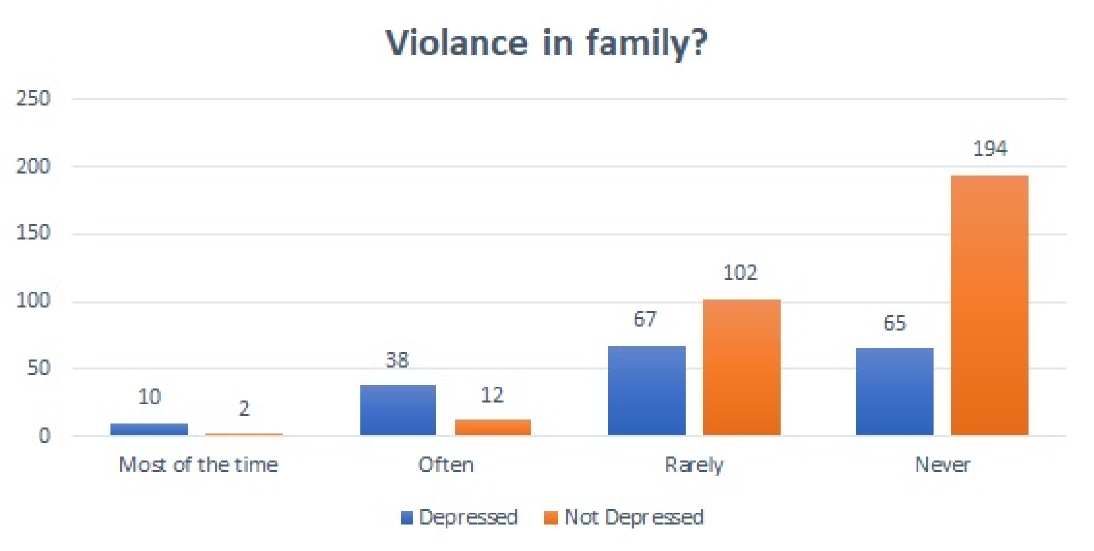


Figure 4.8: Data comparison for Violence in family?

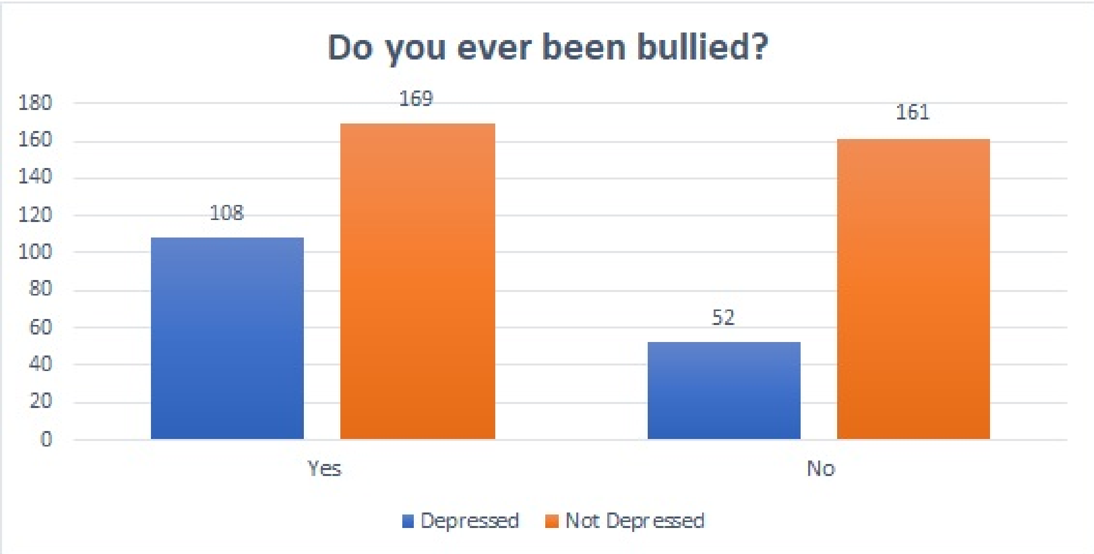


Figure 4.9: Data Comparison for Do you ever been bullied?

number of depressed students that they have been bullied or harassed. When they are not harassed, 52 out of 320 are depressed, while 110 out of 170 are not.

4.5 Evaluation of Framework

Our study aims to determine whether or not a student is depressed. Since our data set is based on binary classification, either of the two outcomes is possible. Our data was also visualized using histograms. We used a couple of the most commonly used algorithms since our data is binary. To determine the accuracy, precision, recall, specificity, error rate, and f-measure of our predictive model, we

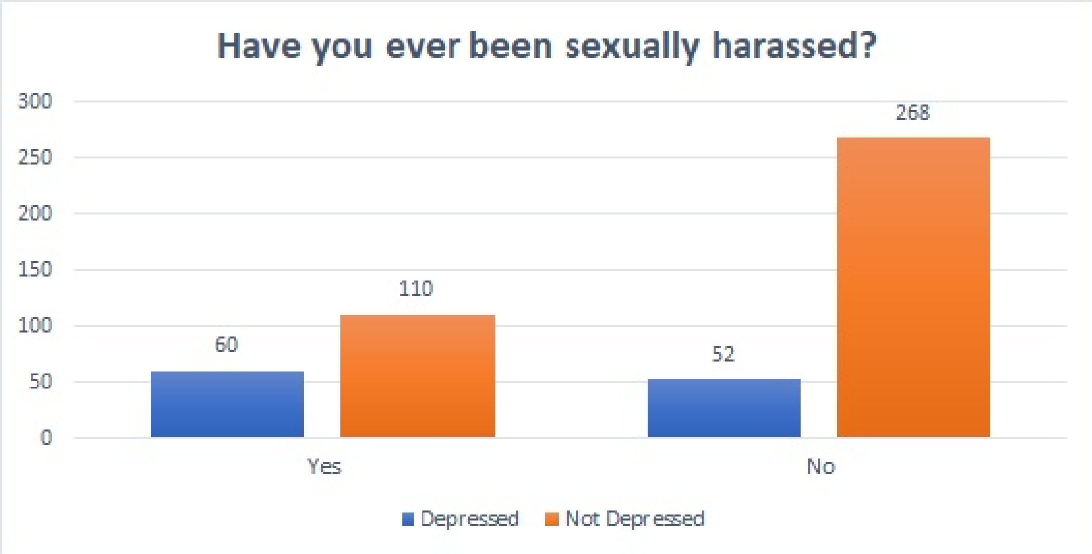


Figure 4.10: Data comparison for Have you ever been sexually harassed?

used 5 algorithms. The method would be better if the accuracy and f-measure are higher. Precision, described as the total number of true positives in all yes predictions, is also an important factor. This means that a student who is genuinely depressed from all of the expected depressed outcomes in our system. Remember another crucial aspect of the prediction system, the number of true positives in actual yes data. This means that a student who is genuinely depressed is selected from among all the actual depressed students in our scheme. Next, we'll compare the accuracy and f-measure of each algorithm to see which one best suits our model. False Negative and False Positive are two more important things that are important to our scheme. The model would be better if the False Negative and False Positive are lower. Often, if both are lower and there is a decent balance between them, a model would be considered good. Since our system predicts depression, it's important to have a low False Negative, because False Negative means the system will mistakenly identify a depressed person as not depressed, while False Positive means the system will mistakenly identify a person who is not depressed as depressed. Furthermore, we would strive for a lower False Negative than False Positive for our scheme, since it is critical to distinguish a depressed person from someone who is not depressed but is being labeled as such. We will also calculate Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) to compare our model. Two datasets have been used for this purpose. One dataset contains 21 features named as F-21 and the other has 15 features which is named as F-15. We have used some of the python libraries to use the model training. These are pandas, scikit-learn, etc. Generally, scikit-learn suitable for numeric data. Our dataset also contains all the numeric data. Scikit-learn also best for classification, regression analysis. Pandas is used for data analysis and manipulation. We have used SVM, KNN, Logistic Regression, Random Forest, and Decision Tree algorithm for model prediction. Using this library, we import the dataset. Dataset has been split into two parts, one part is train data and another one is the test data. Generally, 75% of data has been separated into the train data and another 25% of data has been separated into test data. Model training has been done on different algorithms.

4.5.1 Accuracy for different algorithms with 21 features

4.5.1.1 KNN

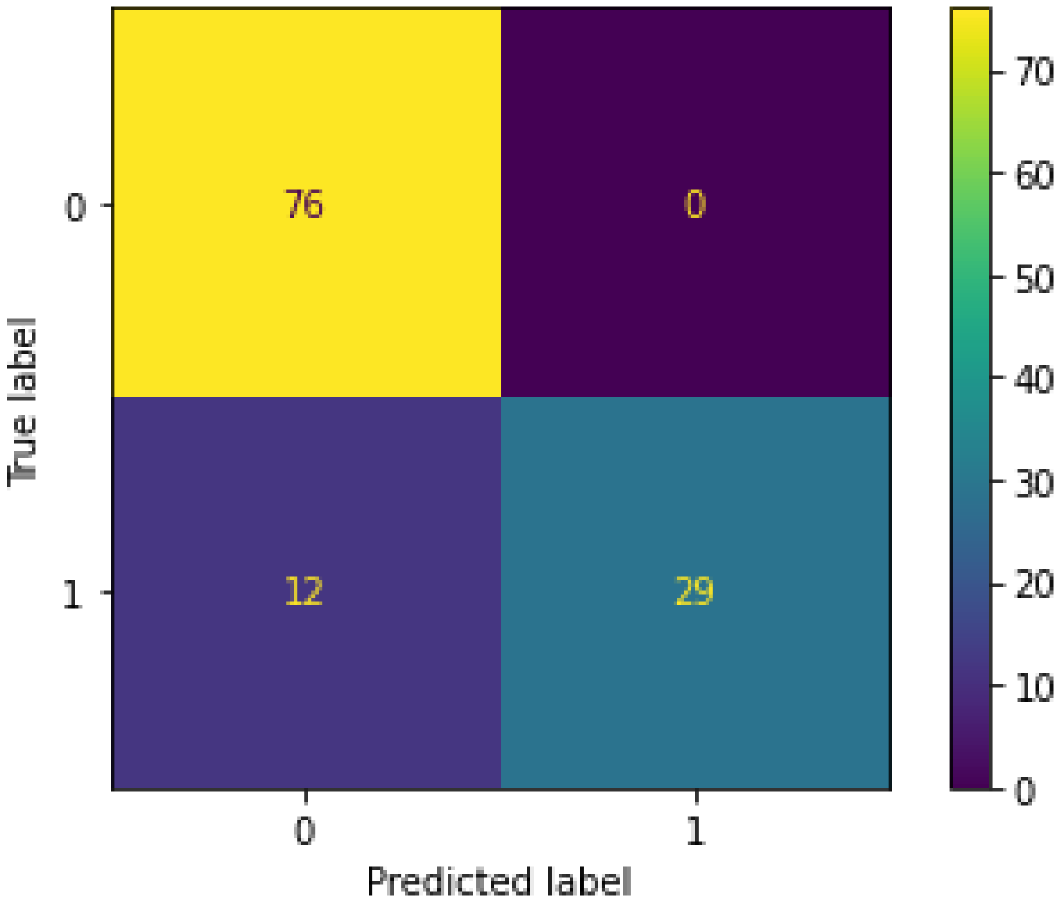


Figure 4.11: Confusion Matrix for KNN for 21 features

From fig 4.11 we can see the confusion matrix which has got 29 true positives, 76 True negatives, 0 false positives and 12 false negatives. From this we can calculate our accuracy which is 89.74% or approximately 90%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 29 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 29. Therefore dividing 29 by 29 and then multiplying it with 100, we get a precision of 100%. Then we look at the recall, which is the number of true positives in actual yes results. There are 29 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 41. Therefore dividing 29 by 41 and then multiplying it with 100, we get a recall of 71% approximately. Specificity or True negative rate is found to be 100%. The number of false negative is 12 and false positive is 0. So 12 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 10% approximately. The F-score which is a harmonic mean of precision and recall is found to be 83% approximately. Lastly we calculated the mean average percentage error which is found to be 15% and root mean squared error which is found to be 8.48.

4.5.1.2 SVM

From fig 4.12 we can see the confusion matrix which has got 34 true positives, 54 True negatives, 2 false positives and 4 false negatives. From this we can calculate our accuracy which is 93.61% or approximately 93%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 34 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 36. Therefore dividing 34 by 36 and then multiplying it with 100, we get a precision of 94%. Then we look at the recall, which is the number of true positives in actual yes results. There are 34 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 38. Therefore dividing 34 by 38 and then multiplying it with 100, we get a recall of 89% approximately. Specificity or True negative rate is found to be 96%. The number of false negative is 4 and false positive is 2. So 6 divided by

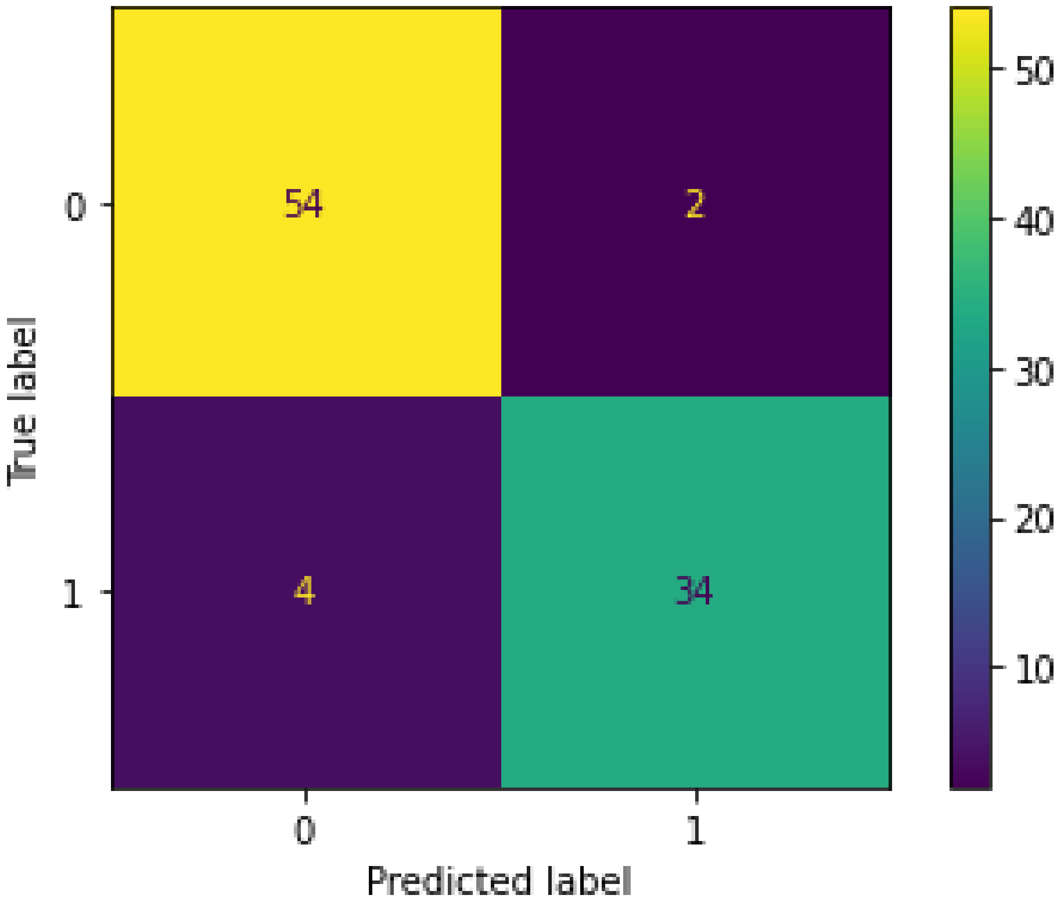


Figure 4.12: Confusion Matrix for SVM for 21 features

total number of predictions multiplied by 100 gives us the error rate that is found to be 6% approximately. The F-score which is a harmonic mean of precision and recall is found to be 91% approximately. Lastly we calculated the mean average percentage error which is found to be 7% and root mean squared error which is found to be 3.16.

4.5.1.3 Decision Tree

From fig 4.13 we can see the confusion matrix which has got 27 true positives, 51 True negatives, 5 false positives and 11 false negatives. From this we can calculate our accuracy which is 82.97% or approximately 83%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 27 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 32. Therefore dividing 27 by 32 and then multiplying it with 100, we

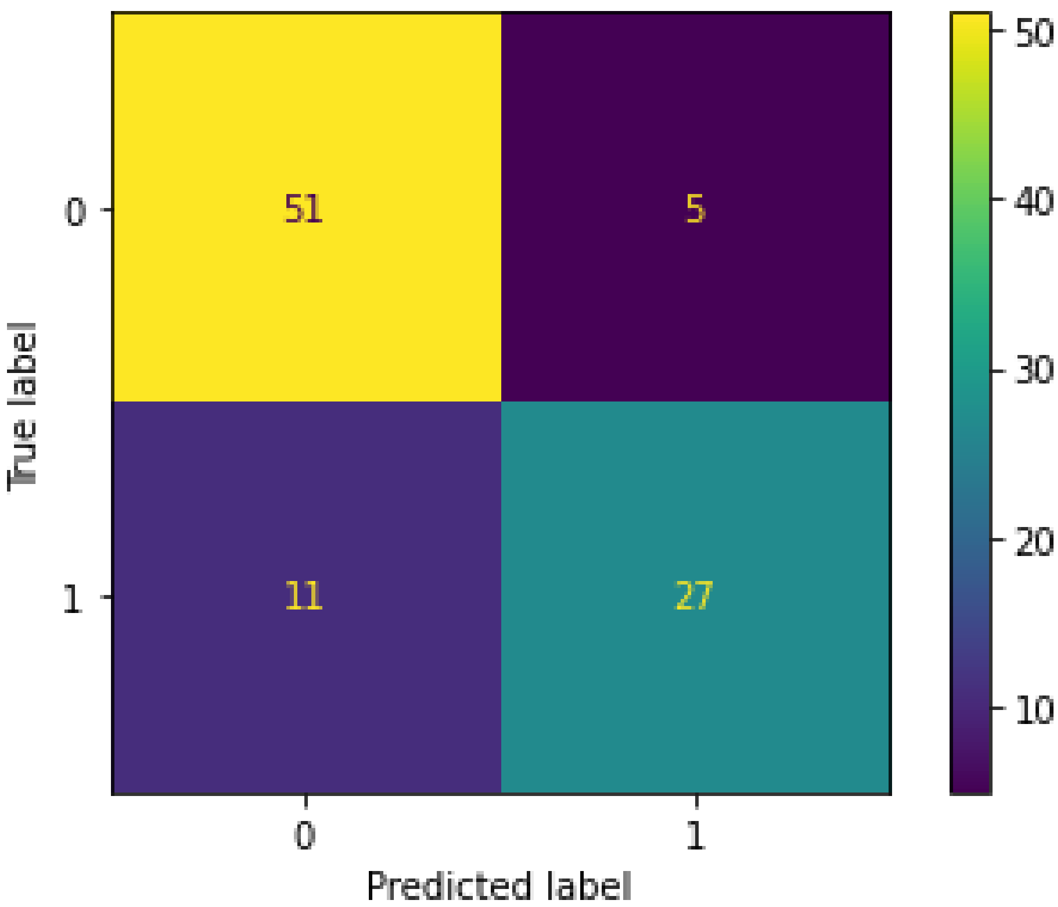


Figure 4.13: Confusion Matrix for Decision Tree for 21 features

get a precision of 84%. Then we look at the recall, which is the number of true positives in actual yes results. There are 27 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 38. Therefore dividing 27 by 38 and then multiplying it with 100, we get a recall of 71% approximately. Specificity or True negative rate is found to be 91%. The number of false negative is 11 and false positive is 5. So 16 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 17% approximately. The F-score which is a harmonic mean of precision and recall is found to be 77% approximately. Lastly we calculated the mean average percentage error which is found to be 19% and root mean squared error which is found to be 8.54.

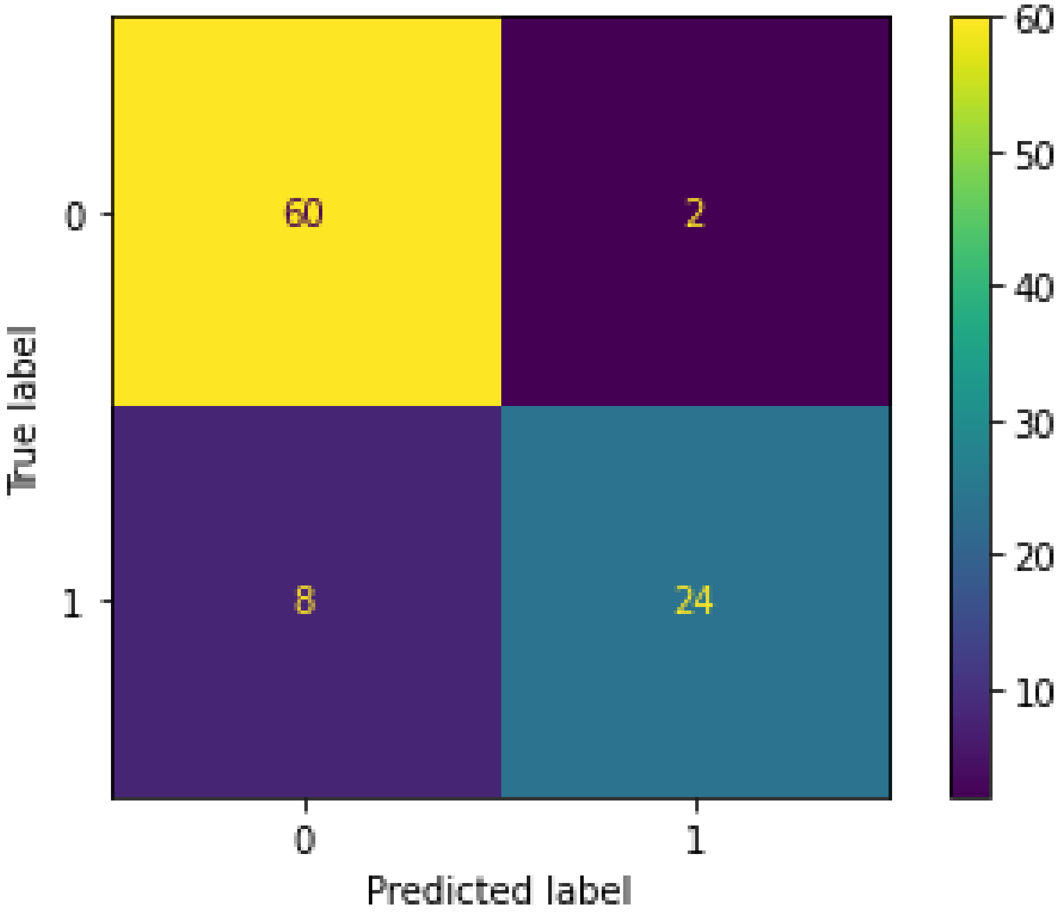


Figure 4.14: Confusion Matrix for Random Forest for 21 features

4.5.1.4 Random Forest

From fig 4.14 we can see the confusion matrix which has got 24 true positives, 60 True negatives, 2 false positives and 8 false negatives. From this we can calculate our accuracy which is 89.36% or approximately 89%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 24 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 26. Therefore dividing 24 by 26 and then multiplying it with 100, we get a precision of 92%. Then we look at the recall, which is the number of true positives in actual yes results. There are 24 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 32. Therefore dividing 24 by 32 and then multiplying it with 100, we get a recall of 75% approximately. Specificity or True negative rate is found to be 97%. The number of false negative is 8 and false positive is 2. So 10 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 11% approximately. The F-score which is a harmonic mean of precision and recall is found to be 83% approximately. Lastly we calculated the mean average percentage error which is found to be 14% and root mean squared error which is found to be 5.83.

4.5.1.5 Logistic regression

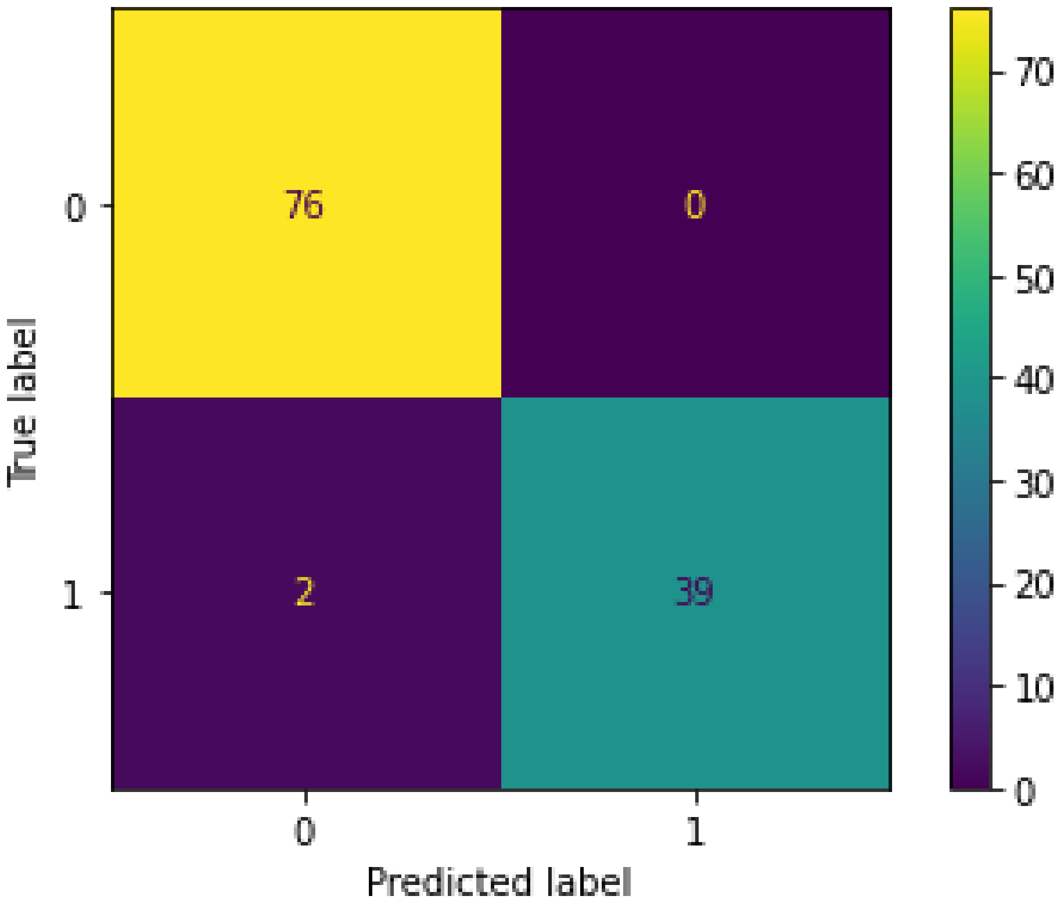


Figure 4.15: Data comparison for Have you ever been sexually harassed?

From fig 4.15 we can see the confusion matrix which has got 39 true positives, 76 True negatives, 0 false positives and 2 false negatives. From this we can calculate our accuracy which is 98.29% or approximately 98%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 39 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 39. Therefore dividing 39 by 39 and then multiplying it with 100, we get a precision of 100%. Then we look at the recall, which is the number of true positives in actual yes results. There are 39 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 41. Therefore dividing 39 by 41 and then multiplying it with 100, we get a recall of 95% approximately. Specificity or True negative rate is found to be 100%. The number of false negative is 2 and false positive is 0. So 2 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 2% approximately. The F-score which is a harmonic mean of precision and recall is found to be 97% approximately. Lastly we calculated the mean average percentage error which is found to be 2% and root mean squared error which is found to be 1.41.

4.5.2 Accuracy for different algorithms with 14 features

Before this our algorithms were giving a little less accuracy for our 15 features dataset. So we calculated the correlation matrix to see if any features were creating any negative correlation. Fig 4.16 shows the correlation matrix.

From the correlation matrix we can see that our feature number 5 labeled as s5 is creating negative correlation. S5 denotes the question “Are you happy about your academic conditions?” So we dropped this column and were left with 14 features. Then we applied our algorithms for these 14 features and the result improved slightly.

4.5.2.1 KNN

From fig 4.17 we can see the confusion matrix which has got 6 true positives, 71 True negatives, 5 false positives and 35 false negatives. From this we can calculate our accuracy which is 65.81% or approximately 66%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 6 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 11. Therefore dividing 6 by 11 and then multiplying it with 100, we get a precision of 55%. Then we look at the recall, which is the number of true

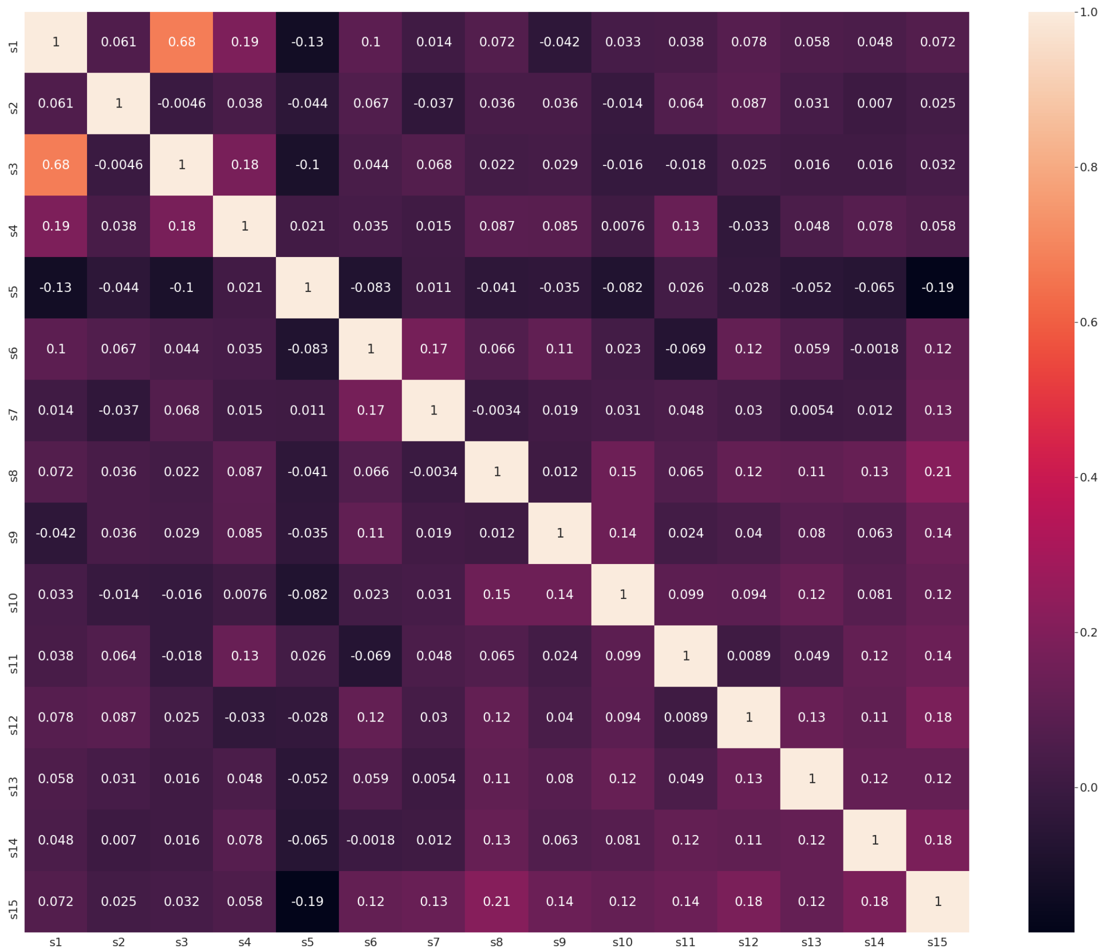


Figure 4.16: Correlation Matrix for 15 features

positives in actual yes results. There are 6 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 41. Therefore dividing 6 by 41 and then multiplying it with 100, we get a recall of 15% approximately. Specificity or True negative rate is found to be 93%. The number of false negative is 35 and false positive is 5. So 40 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 34% approximately. The F-score which is a harmonic mean of precision and recall is found to be 24% approximately. Lastly we calculated the mean average percentage error which is found to be 46% and root mean squared error which is found to be 25.

4.5.2.2 SVM

From fig 4.18 we can see the confusion matrix which has got 5 true positives, 62 True negatives, 1 false positives and 26 false negatives. From this we can

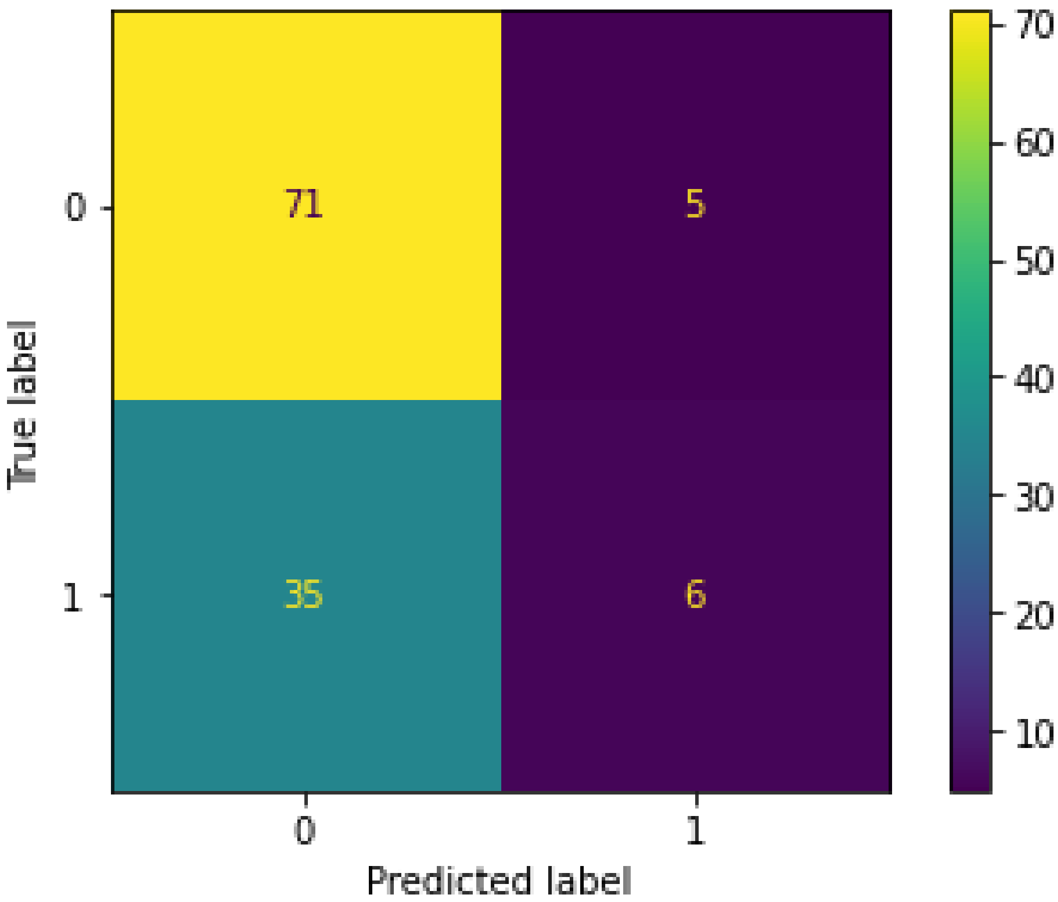


Figure 4.17: Confusion Matrix for KNN for 14 features

calculate our accuracy which is 71.27% or approximately 71%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 5 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 6. Therefore dividing 5 by 6 and then multiplying it with 100, we get a precision of 83%. Then we look at the recall, which is the number of true positives in actual yes results. There are 5 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 31. Therefore dividing 5 by 31 and then multiplying it with 100, we get a recall of 16% approximately. Specificity or True negative rate is found to be 98%. The number of false negative is 26 and false positive is 1. So 27 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 29% approximately. The F-score which is a harmonic mean of precision and recall is found to be 27% approximately. Lastly we calculated the mean average

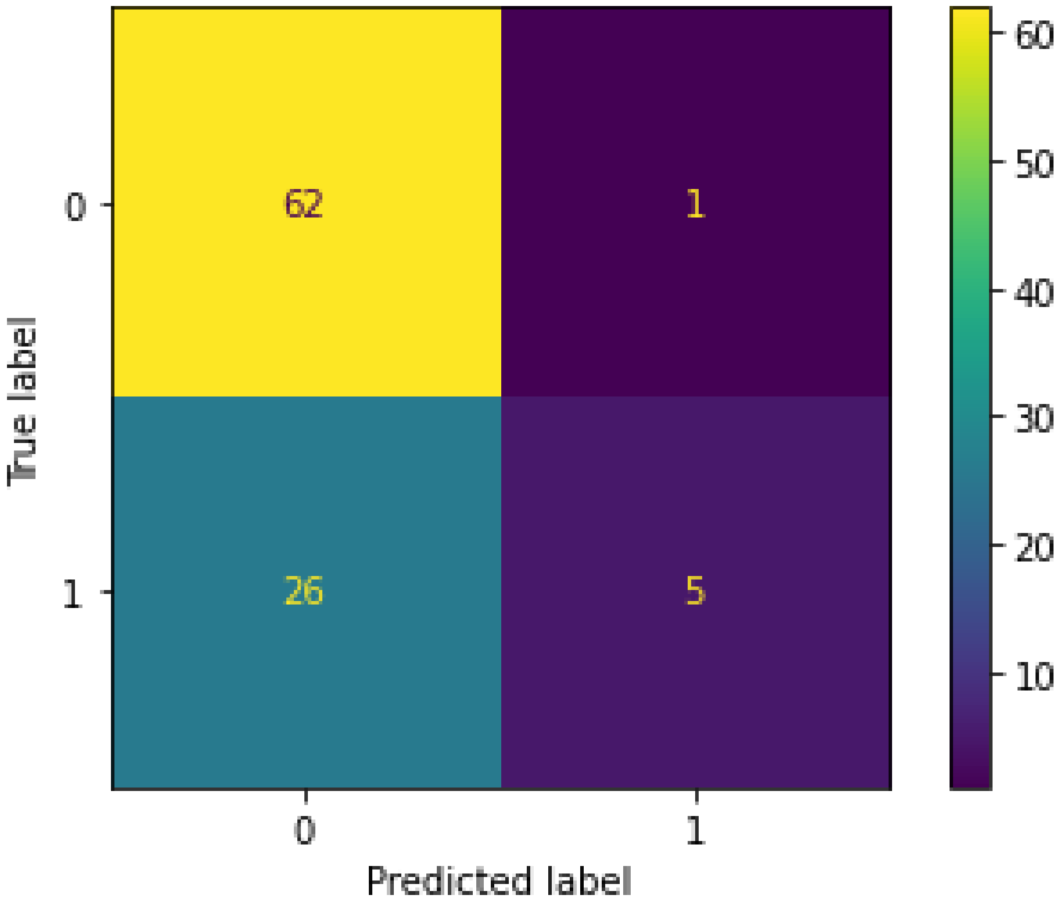


Figure 4.18: Confusion Matrix for SVM for 14 features

percentage error which is found to be 43% and root mean squared error which is found to be 13.5.

4.5.2.3 Decision Tree

From fig 4.19 we can see the confusion matrix which has got 17 true positives, 42 True negatives, 21 false positives and 14 false negatives. From this we can calculate our accuracy which is 62.76% or approximately 63%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 17 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 38. Therefore dividing 17 by 38 and then multiplying it with 100, we get a precision of 45%. Then we look at the recall, which is the number of true positives in actual yes results. There are 17 individuals who were predicted as yes and they are actually depressed which is the true positive and total number

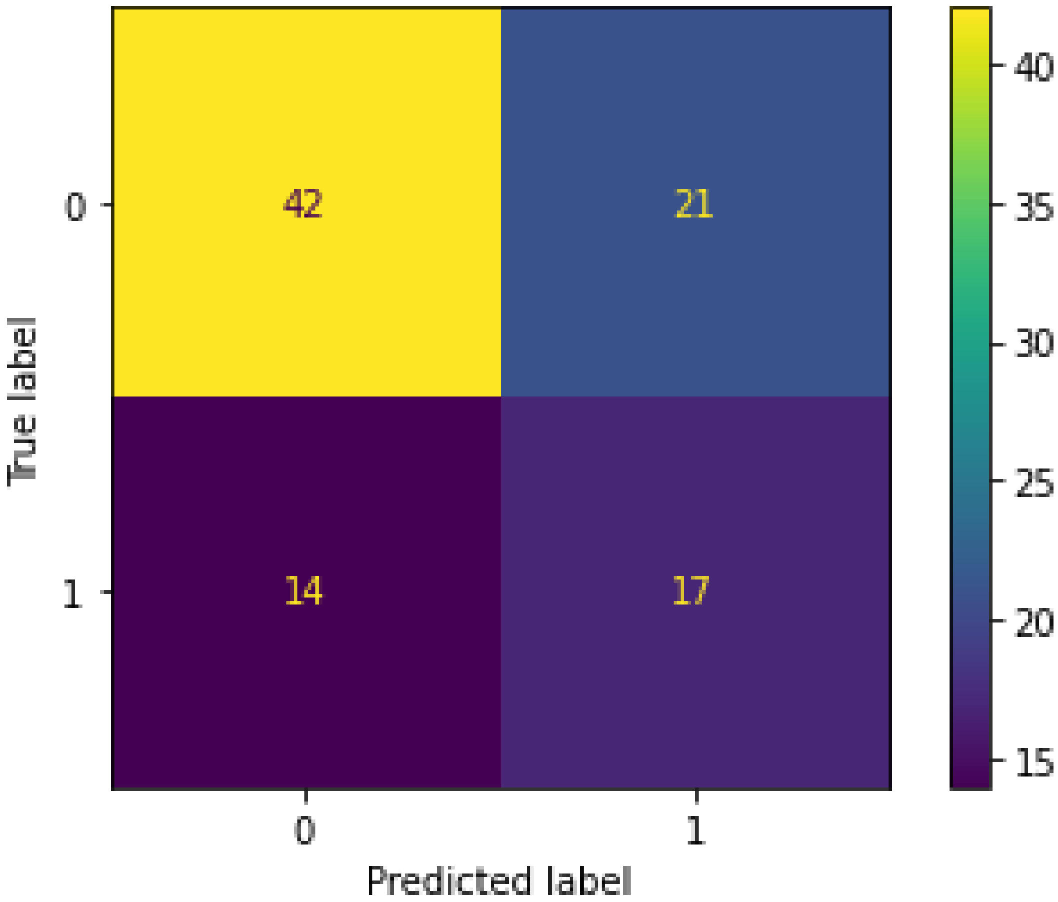


Figure 4.19: Confusion Matrix for Decision Tree for 14 features

of actual yes is 31. Therefore dividing 17 by 31 and then multiplying it with 100, we get a recall of 55% approximately. Specificity or True negative rate is found to be 66%. The number of false negative is 14 and false positive is 21. So 35 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 37% approximately. The F-score which is a harmonic mean of precision and recall is found to be 50% approximately. Lastly we calculated the mean average percentage error which is found to be 39% and root mean squared error which is found to be 17.84.

4.5.2.4 Random Forest

From fig 4.20 we can see the confusion matrix which has got 5 true positives, 57 True negatives, 8 false positives and 24 false negatives. From this we can calculate our accuracy which is 65.95% or approximately 66%. Also if we look at the precision, where precision means the total number of true positive in all the

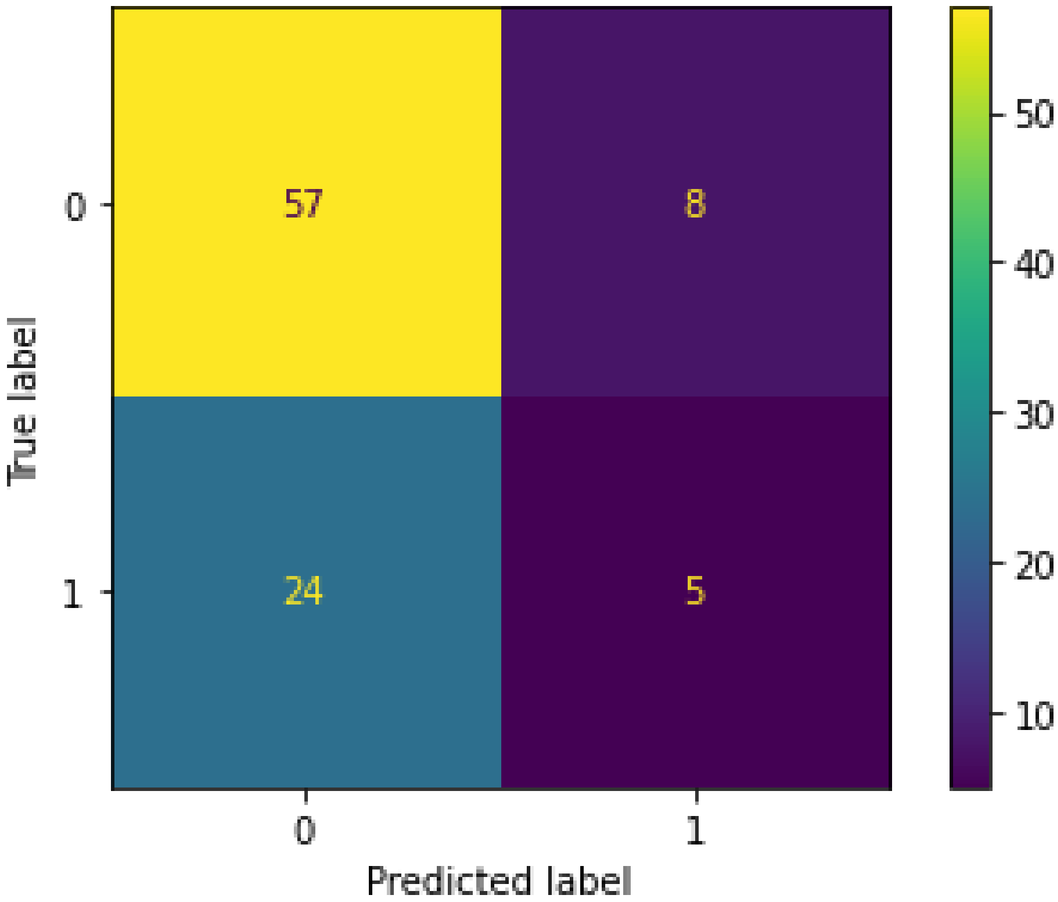


Figure 4.20: Confusion Matrix for Random Forest for 14 features

prediction of yes. There are 5 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 8. Therefore dividing 5 by 8 and then multiplying it with 100, we get a precision of 34%. Then we look at the recall, which is the number of true positives in actual yes results. There are 5 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 24. Therefore dividing 5 by 24 and then multiplying it with 100, we get a recall of 17% approximately. Specificity or True negative rate is found to be 88%. The number of false negative is 24 and false positive is 8. So 32 divided by total number of predictions multiplied by 100 gives us the error rate that is found to be 34% approximately. The F-score which is a harmonic mean of precision and recall is found to be 23% approximately. Lastly we calculated the mean average percentage error which is found to be 48% and root mean squared error which is found to be 17.88.

4.5.2.5 Logistic Regression

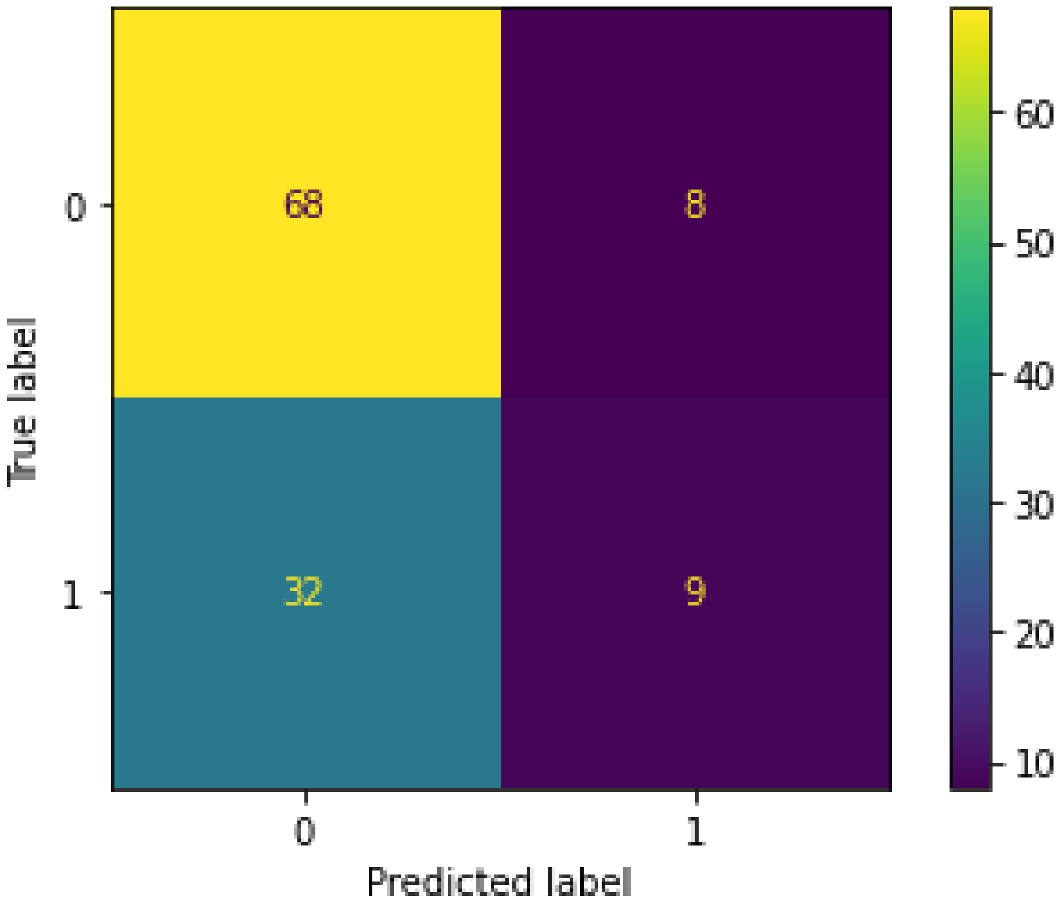


Figure 4.21: Confusion Matrix for Logistic Regression for 14 features

From fig 4.21 we can see the confusion matrix which has got 9 true positives, 68 True negatives, 8 false positives and 32 false negatives. From this we can calculate our accuracy which is 65.81% or approximately 66%. Also if we look at the precision, where precision means the total number of true positive in all the prediction of yes. There are 9 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of prediction of yes is 17. Therefore dividing 9 by 17 and then multiplying it with 100, we get a precision of 53%. Then we look at the recall, which is the number of true positives in actual yes results. There are 9 individuals who were predicted as yes and they are actually depressed which is the true positive and total number of actual yes is 41. Therefore dividing 9 by 41 and then multiplying it with 100, we get a recall of 22% approximately. Specificity or True negative rate is found to be 89%. The number of false negative is 32 and false positive is 8. So 40 divided by

Table 4.2: Comparison among algorithms for 21 features

	Accuracy	Precision	Recall	Specificity	Error Rate	F-Measure	MAPE
K Nearest Neighbor	90%	100%	71%	100%	10%	83%	15%
Support Vector Machine	93%	94%	89%	96%	6%	91%	7%
Decision Tree	83%	84%	71%	91%	17%	77%	19%
Random Forest	89%	92%	75%	97%	11%	83%	14%
Logistic Regression	98%	100%	95%	100%	2%	97%	2%

Table 4.3: Comparison among algorithms for 14 features

	Accuracy	Precision	Recall	Specificity	Error Rate	F-Measure	MAPE
K Nearest Neighbor	66%	55%	15%	93%	34%	24%	46%
Support Vector Machine	71%	83%	16%	98%	29%	27%	43%
Decision Tree	63%	45%	55%	65%	37%	50%	39%
Random Forest	66%	34%	17%	88%	34%	23%	48%
Logistic Regression	66%	53%	22%	89%	34%	31%	44%

total number of predictions multiplied by 100 gives us the error rate that is found to be 34% approximately. The F-score which is a harmonic mean of precision and recall is found to be 31% approximately. Lastly we calculated the mean average percentage error which is found to be 44% and root mean squared error which is found to be 23.32.

4.6 Evaluation of Performance

Previously we discussed about the training part where data has been split for model training. A comparison will be discussed between several algorithms.

Table 4.2 shows the comparison between algorithms based on accuracy, precision, recall, specificity, error rate, F-measure, MAPE, RMSE for 21 features. And fig 4.22 shows the comparison between algorithms for 21 features.

Table 4.3 shows the comparison between algorithms based on accuracy, precision, recall, specificity, error rate, F-measure, MAPE, RMSE for 14 features. And fig 4.23 shows the comparison between algorithms for 14 features.

From fig 4.24 we can see that in case of accuracy Logistic regression for 21 features and SVM for 14 features gives the best accuracy which are 98% and 71% respectively.

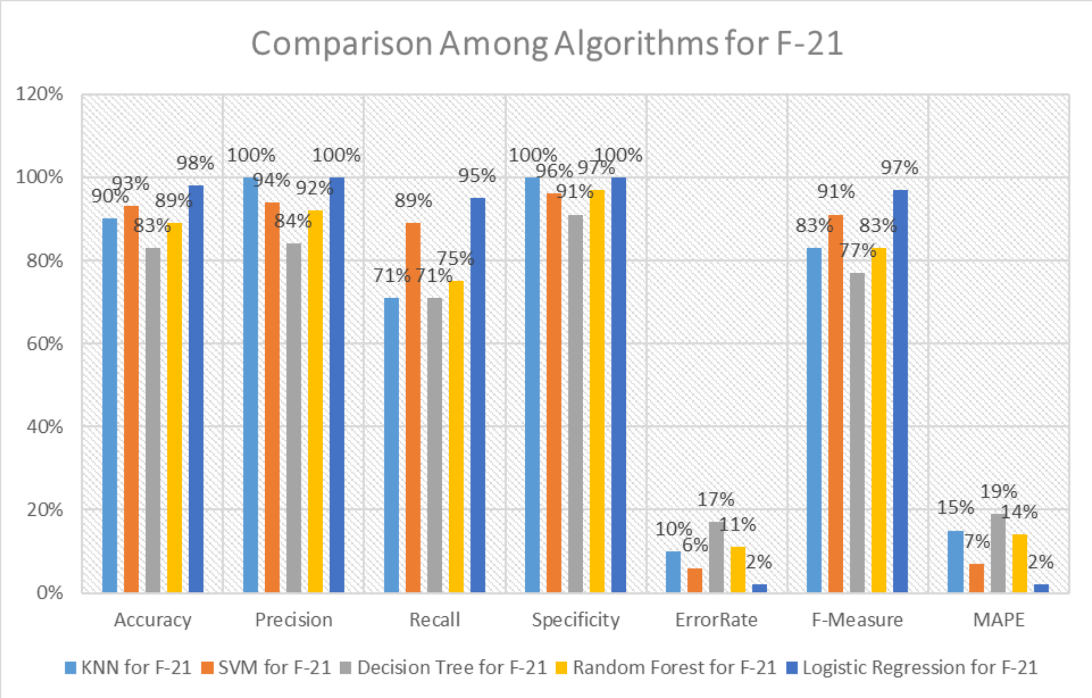


Figure 4.22: Comparison among algorithms for 21 features

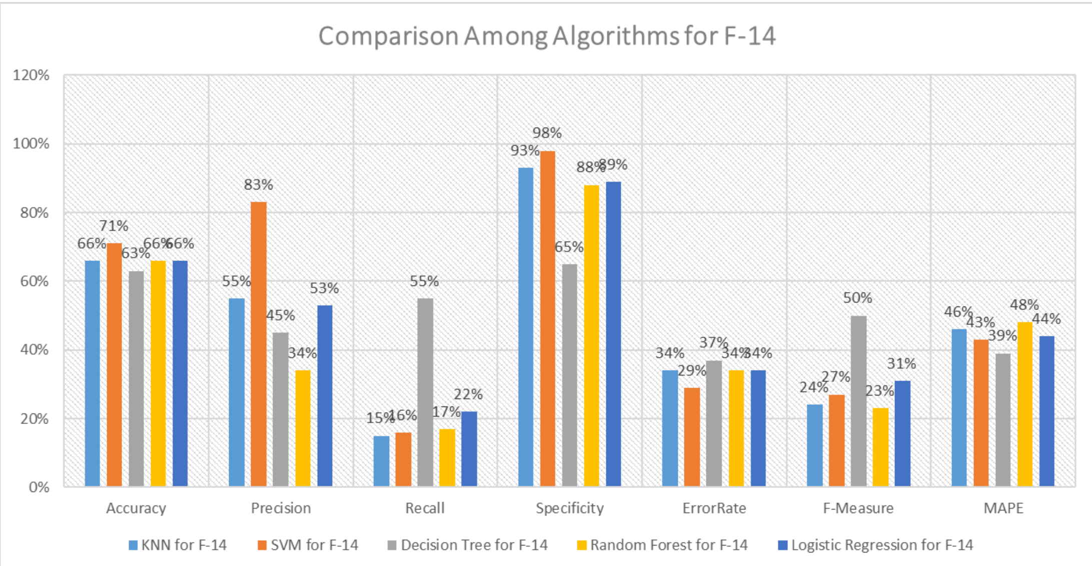


Figure 4.23: Comparison among algorithms for 14 features

From fig 4.25 we can see that in case of precision Logistic regression and KNN for 21 features and SVM for 14 features gives the best precision which are 100%, 100% and 83% respectively.

From fig 4.26 we can see that in case of recall Logistic regression for 21 features and Decision Tree for 14 features gives the best recall which are 95% and 55% respectively.

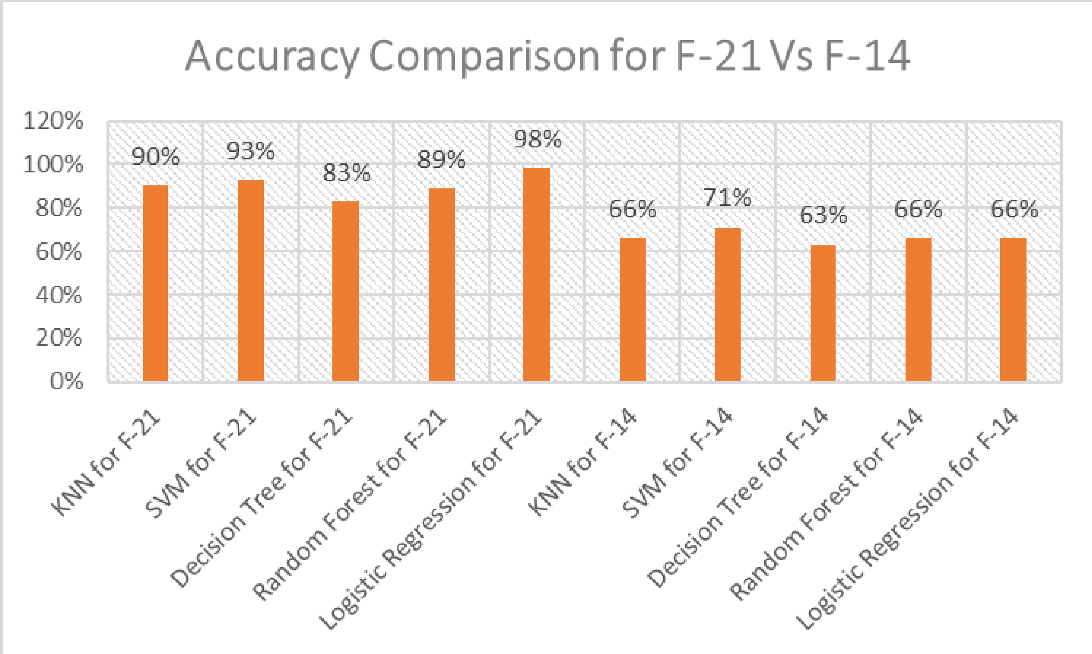


Figure 4.24: Accuracy for F-21 Vs F-14

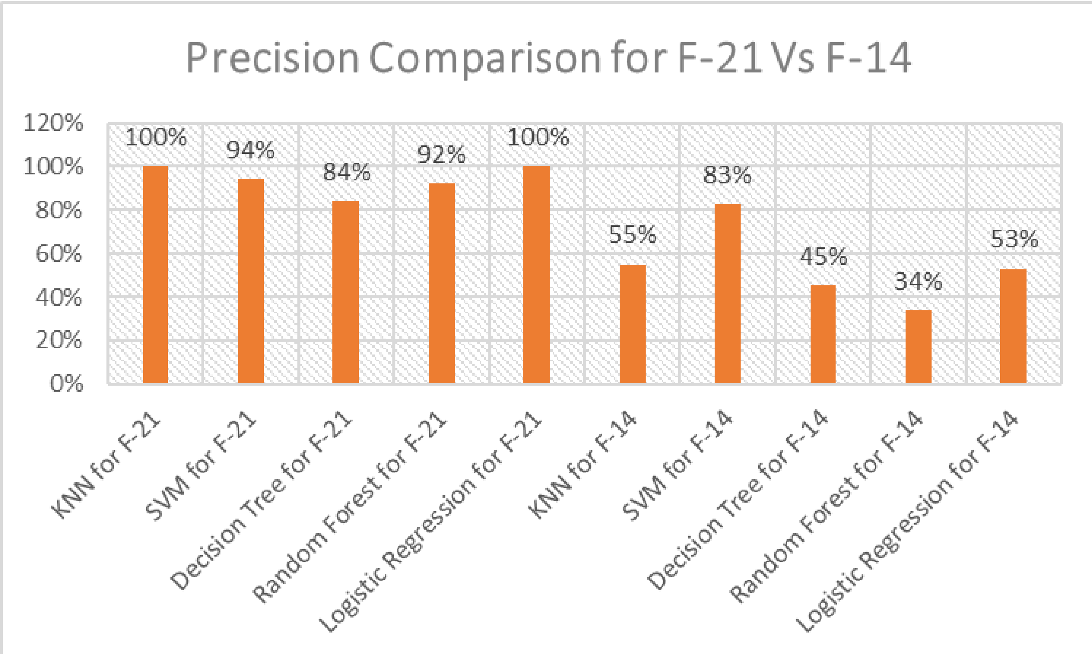


Figure 4.25: Precision for F-21 Vs F-14

From fig 4.27 we can see that in case of specificity KNN and Logistic regression for 21 features and SVM for 14 features gives the best specificity which are 100%, 100% and 98% respectively.

From fig 4.28 we can see that in case of error rate Logistic regression for 21 features and SVM for 14 features gives the lowest error rate which are 2% and 29% respectively.

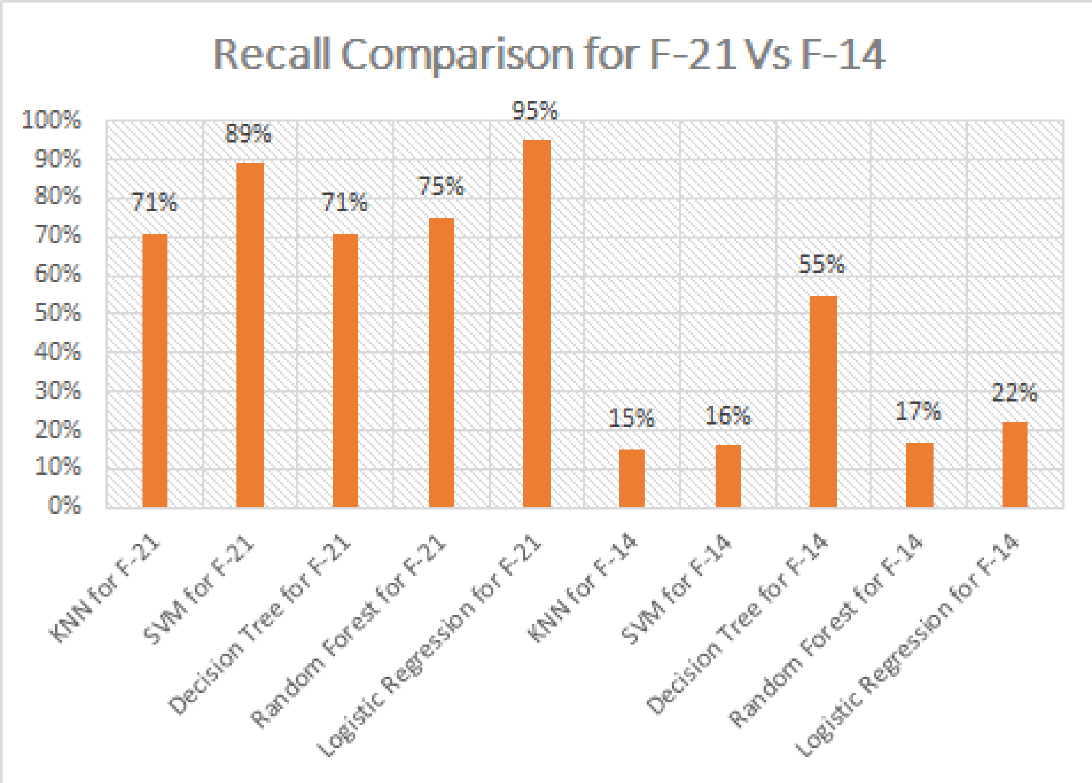


Figure 4.26: Recall for F-21 Vs F-14

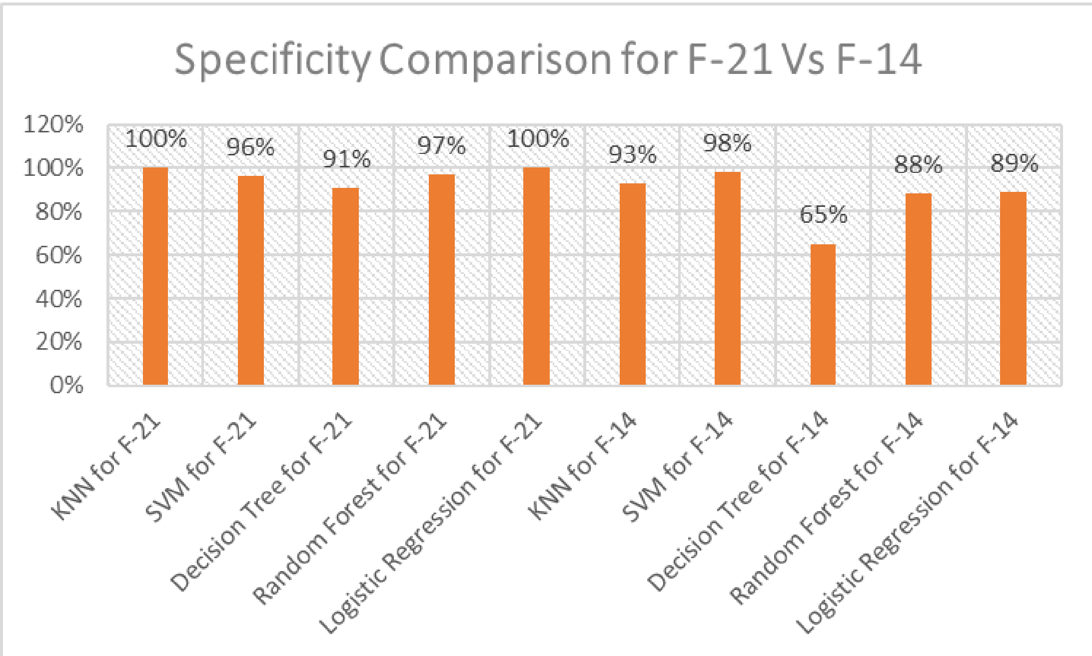


Figure 4.27: Specificity for F-21 Vs F-14

From fig 4.29 we can see that in case of F-Measure Logistic regression for 21 features and Decision Tree for 14 features gives the best F-Measure which are 97% and 50% respectively.

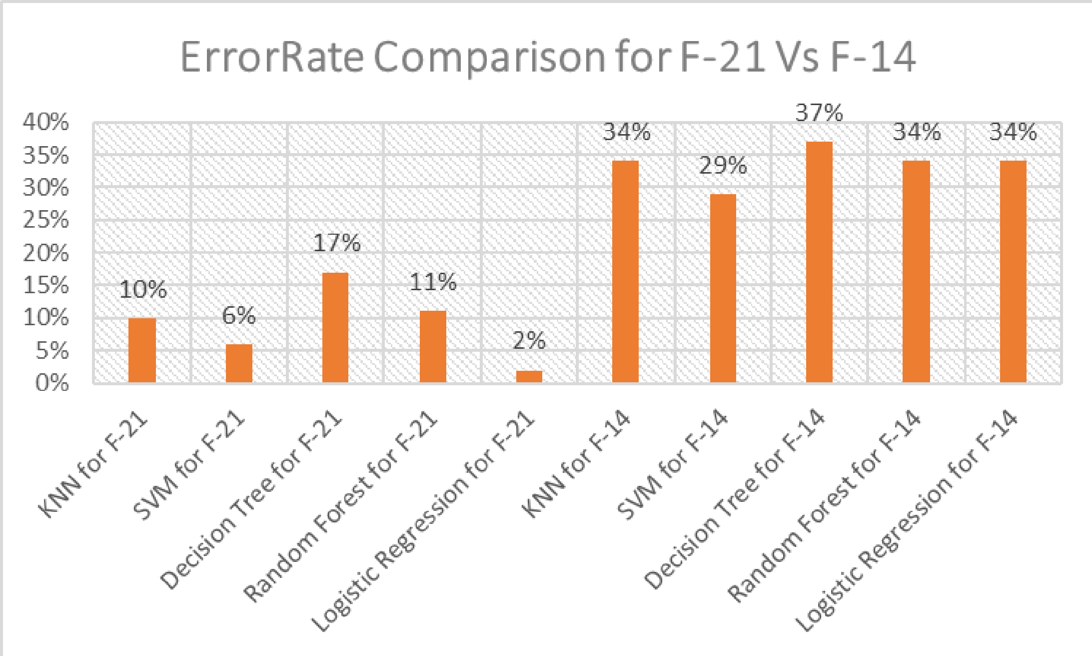


Figure 4.28: Error Rate for F-21 Vs F-14

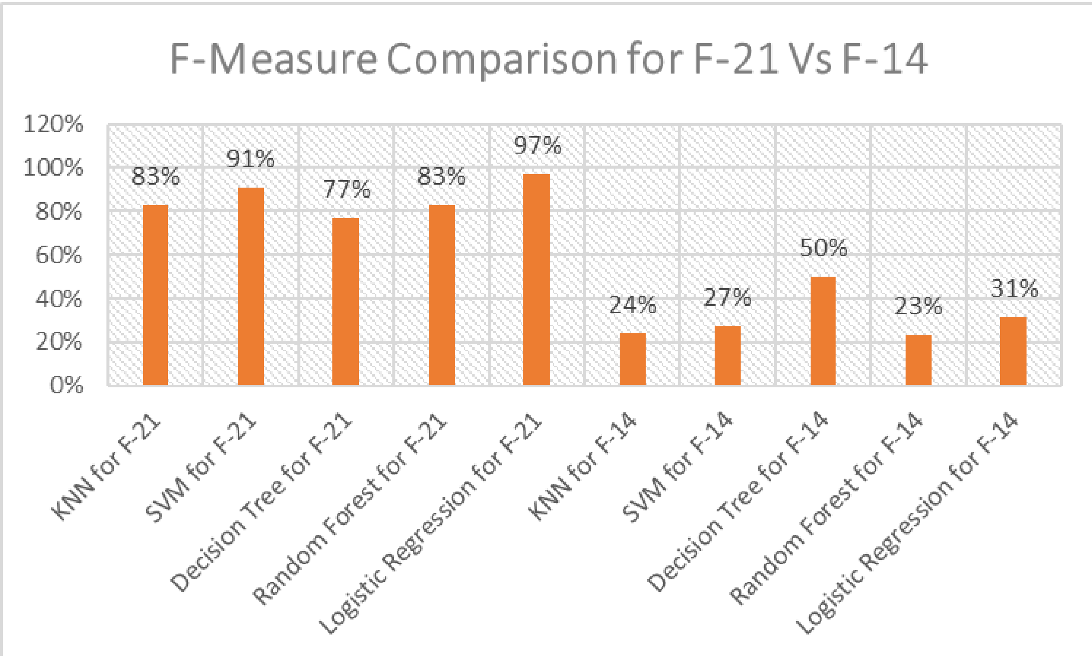


Figure 4.29: F-Measure for F-21 Vs F-14

From fig 4.30 we can see that in case of MAPE Logistic regression for 21 features and Decision Tree for 14 features gives the lowest MAPE which are 2% and 39% respectively.

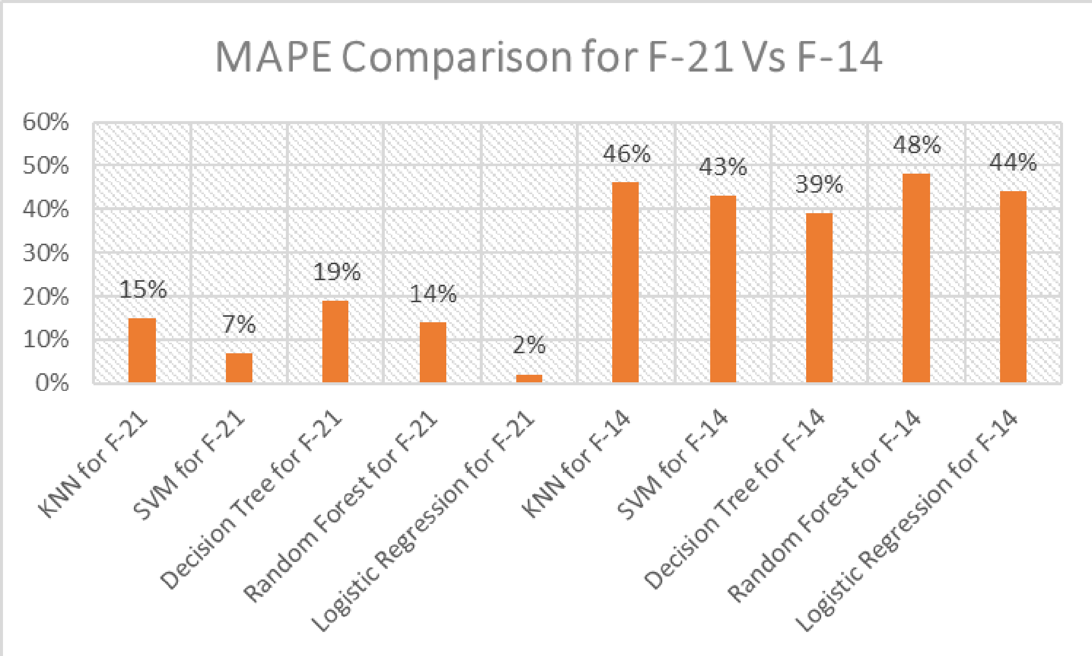


Figure 4.30: MAPE for F-21 Vs F-14

4.7 Mental Healthcare Application Interface & Features

In this section the features and interface of our android based mental healthcare application will be shown and described.

4.7.1 Login

There are 3 types of login. Login as a student/teacher/doctor. Fig 4.31 shows the login. Firstly everyone must register using their email id. One email id can be used once for creating an account. For registering a student has to provide information such as department, phone number, student id etc. A teacher needs to provide information such as department, designation, phone number, association with DSW etc. For doctors they need to provide information like degree, experience, phone number etc.

4.7.2 Student Interface

After login the first thing a student finds is a page filled with some questions that enquires about the student's family history for mental illness, academic

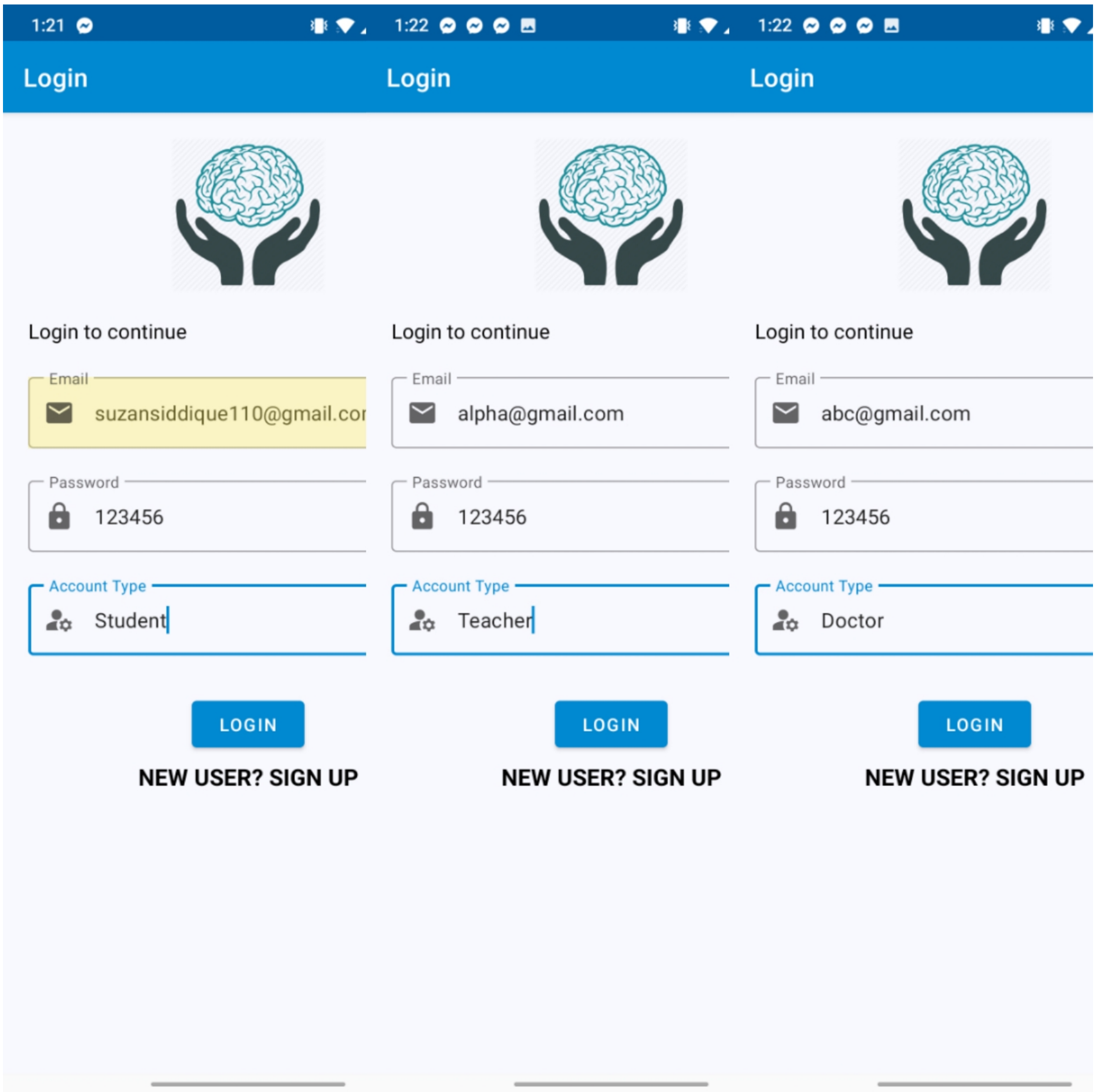


Figure 4.31: Login for student/teacher/doctor

condition, CGPA, any addiction, relationship, relationship with friends, history of being ragged/bullied/harassed etc. It is shown in fig 4.32

After filling those fields the student is taken to the home page of our mobile application. The home page contains information regarding mental health and different kinds of mental illness such as stress disorder, antisocial personality disorder, anxiety disorder, bipolar disorder, clinical depression, obsessive compulsive disorder etc. Each disorder also contains information about the illness, causes, symptoms, prevention and treatment. The left corner tab on the home screen contains different options such as update info, problems list, appointment schedules, depression meter, notifications and logout option. Fig 4.33 and fig 4.34 shows these.

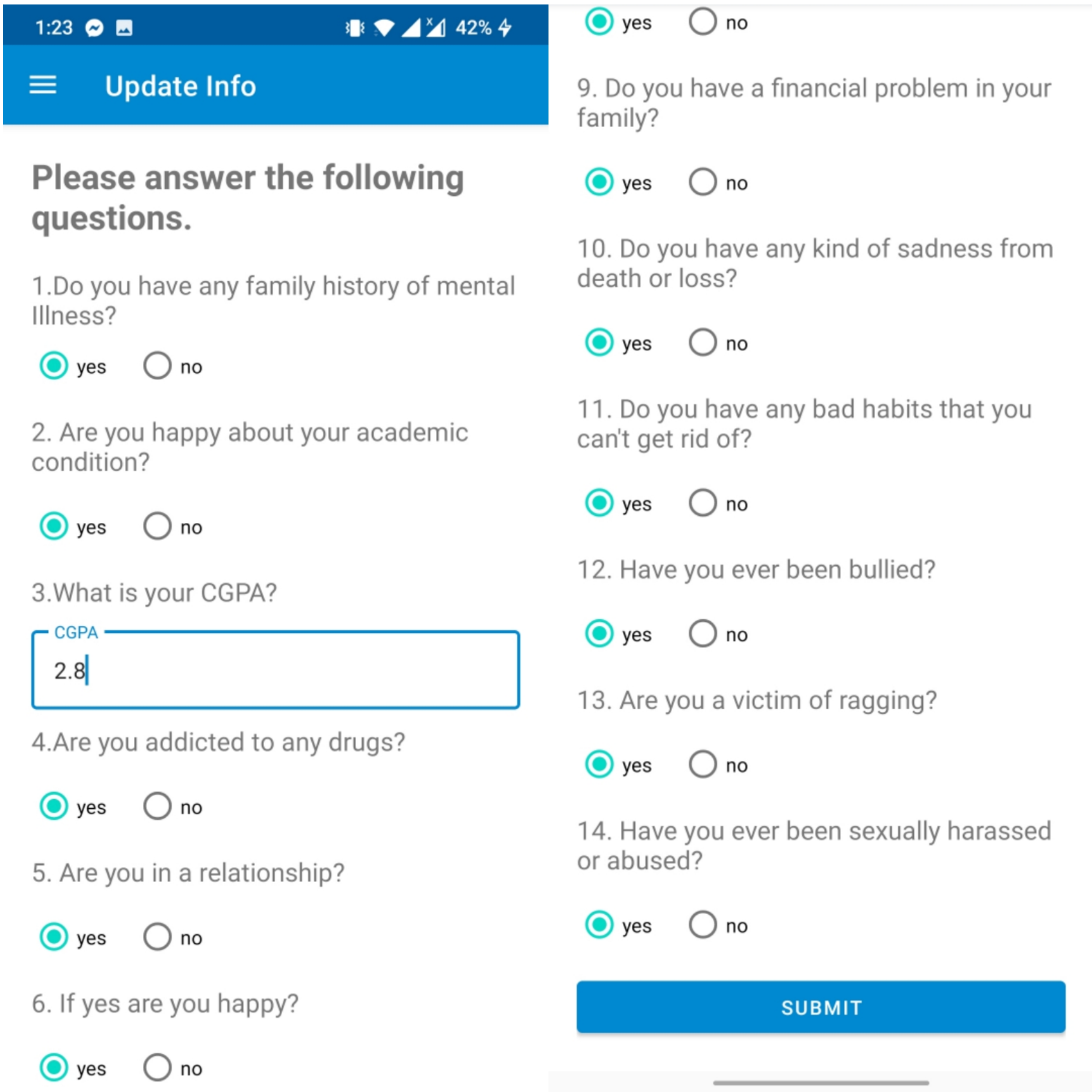


Figure 4.32: Questions for student

The update info feature allows the student to update the answers to the question that were asked to the student on first login. The answers from the questions are then analyzed to find the problem list. The problem list feature contains two tabs. One stores the academic problems and the other stores the psychological problems. It is shown in fig 4.35

The depression meter is a depression checking tool that can find the users level of depression. It asks user some question and based on the answers generates five types of results such as normal ups and downs, mild mood disturbance, borderline clinical depression, moderate depression, severe depression, extreme depression. Normal ups and downs and mood disturbance does not qualify as depression but mild to extreme depression are qualified to be called as depression. These are

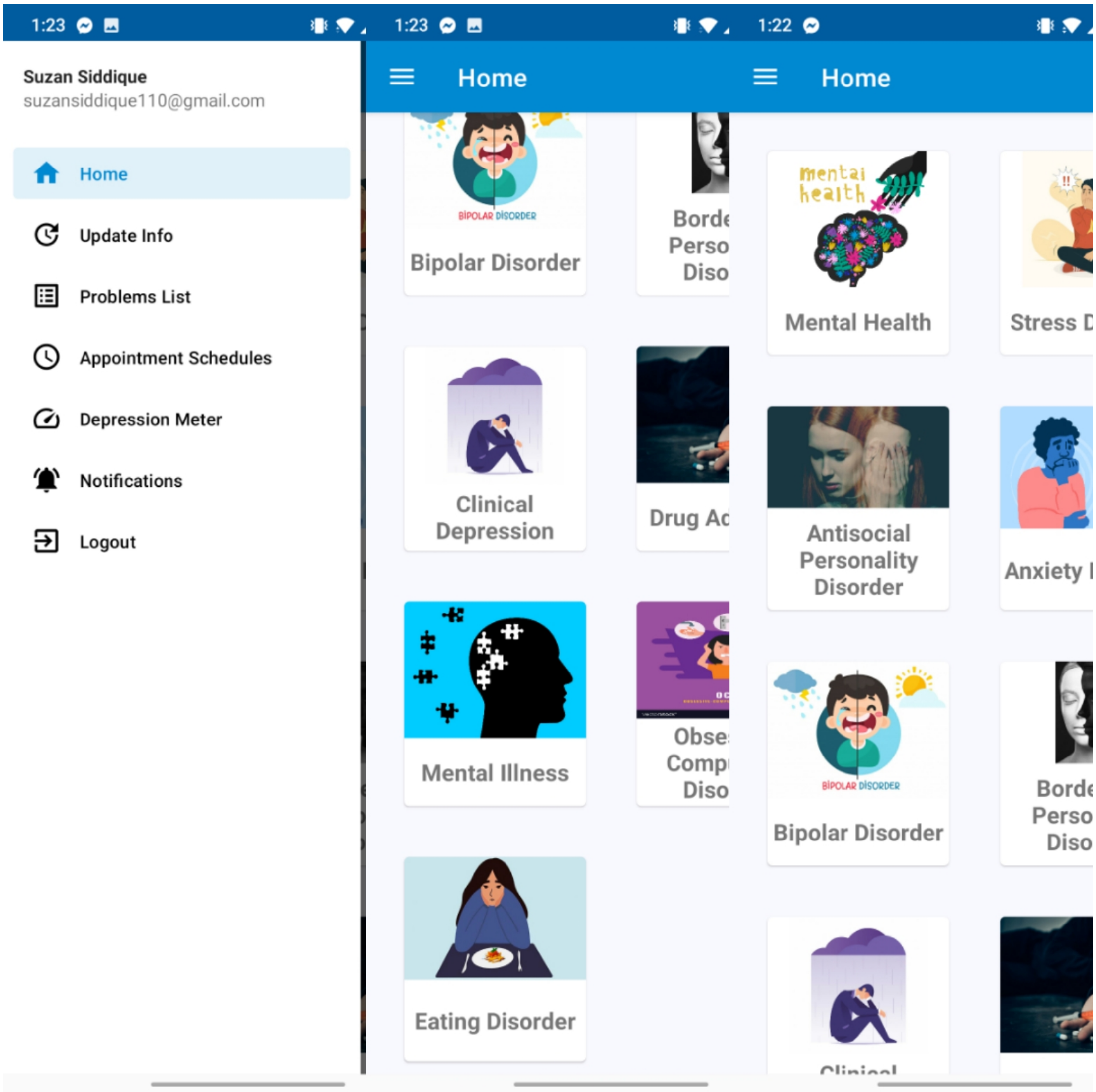


Figure 4.33: Student Home Page

shown in fig 4.36 and fig 4.37

From the problems list feature the student has an option of sharing their problems. A student can share his academic problems with a teacher from the application along with a personal message from the student which is shown in fig 4.38

A student can also share his psychological problems with a doctor along with a personal message and ask for an appointment with the doctor which are stored in appointment schedules which is shown in fig 4.39.

A student gets notified when an action is taken for academic problems or appointment is scheduled with doctor. These are stored in notifications.It is shown in fig 4.40

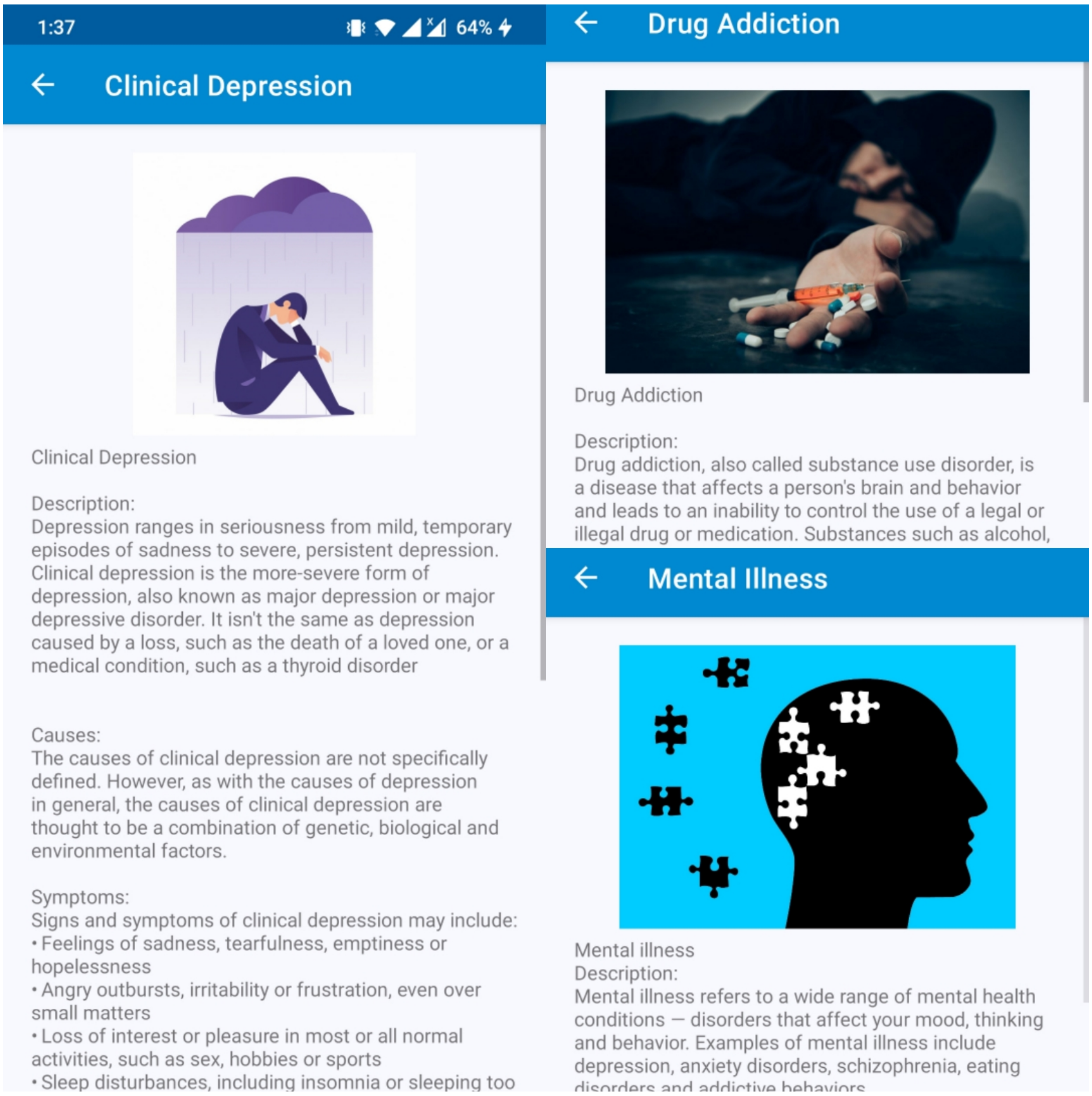


Figure 4.34: Information on different mental illness

4.7.3 Teacher Interface

On the teacher interface the left corner tab contains information of student's request. Problems shared with teacher appear here with student's personal information and personal message from the student. From this teacher can take specified action to help student. Teacher can notify the student that the problem is noted and action will be taken for which student gets a notification.It is shown in fig 4.41

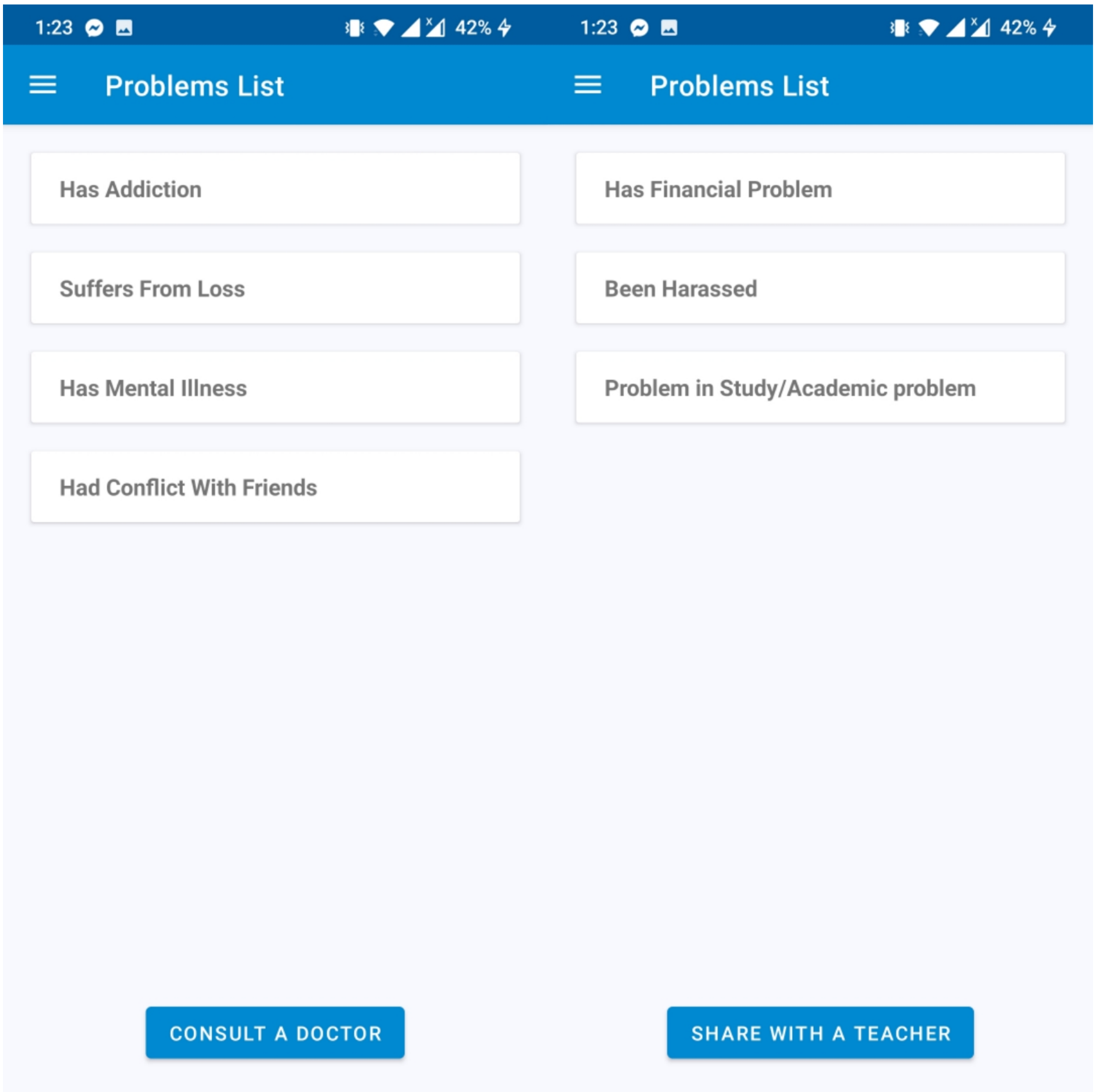


Figure 4.35: Problem list of student

4.7.4 Doctor Interface

On the doctor interface the left corner tab contains appointment request, appointment schedule and notification. Problems shared with the doctor with personal information and message appear in appointment request which is shown in fig 4.42

From here doctor can approve of student's appointment requests and schedule appointments with a date and time of which the student is notified with a notification. Appointment schedule keeps track of appointments given to students to keep things organized.These are shown in fig 4.43 and fig 4.44

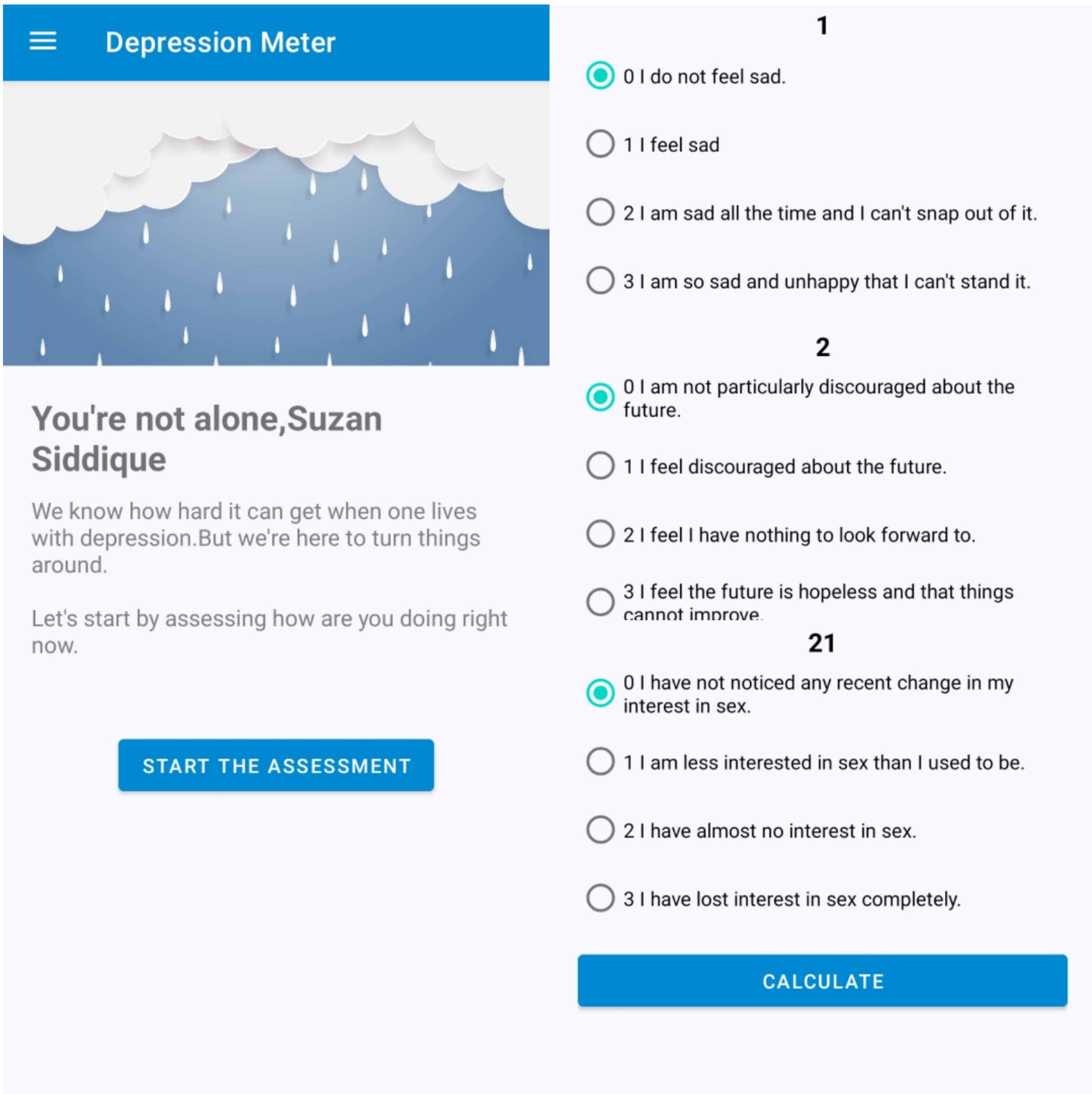


Figure 4.36: depression meter

4.8 Conclusion

firstly it was discovered that depression can be predicted by machine learning approach with incredible accuracy.The designed module for predicting depression produces result based on different matrices such as precision, recall, specificity, error rate, f-measure, MAPE etc. The presented study was related to previous literature. the depression prediction module might be an efficient system for utilization in potential research. A visualization part gives a clear view of the changes of parameters.The android mobile application might come handy to students suffering from mental mental illness. It can provide them necessary help in times of their need. this work can also make a huge impact on our modern society where

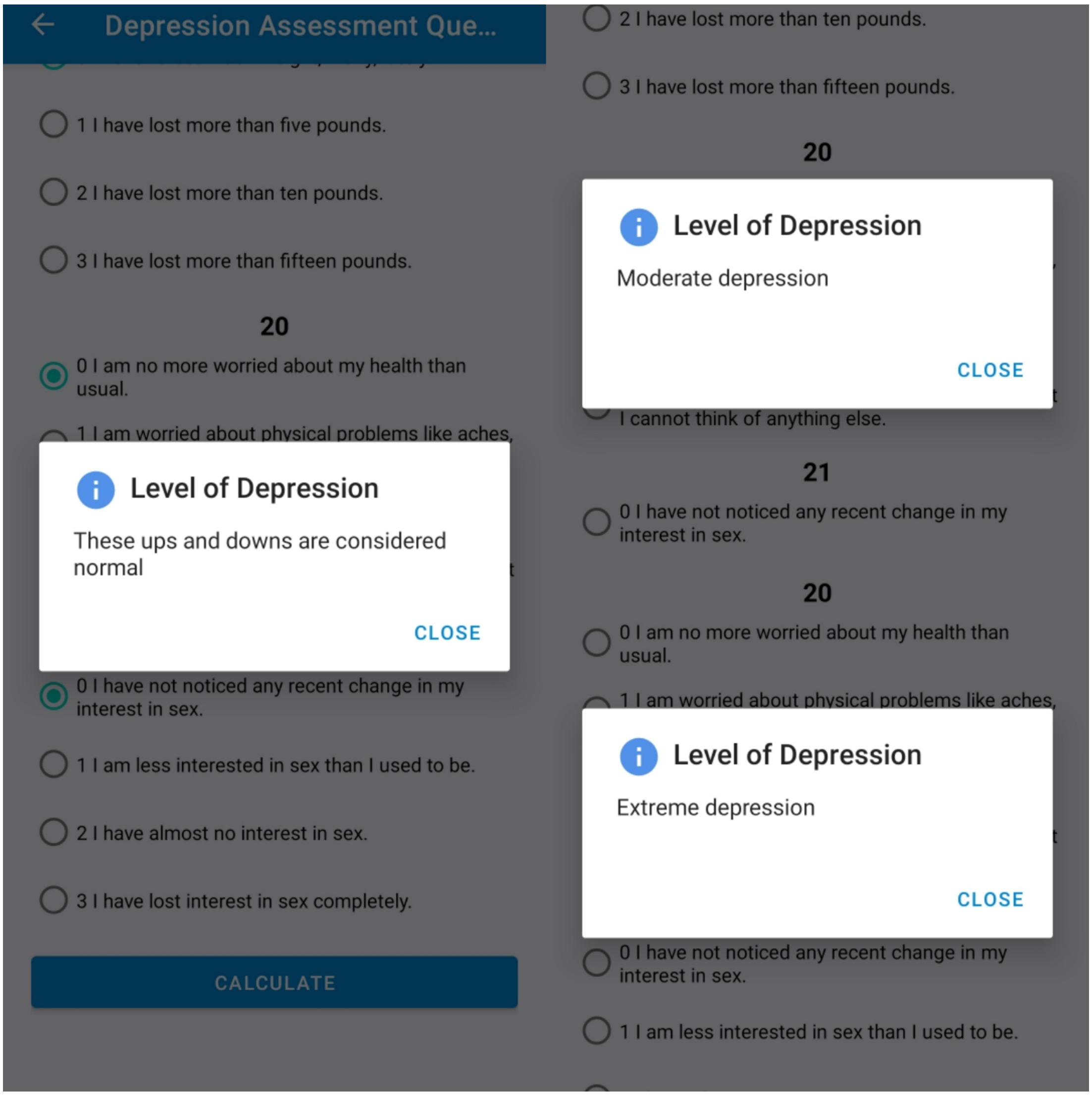


Figure 4.37: Different outputs of depression meter

depression related problems are getting severe day by day. Using these anyone can find their current mental health and take precautions promptly.

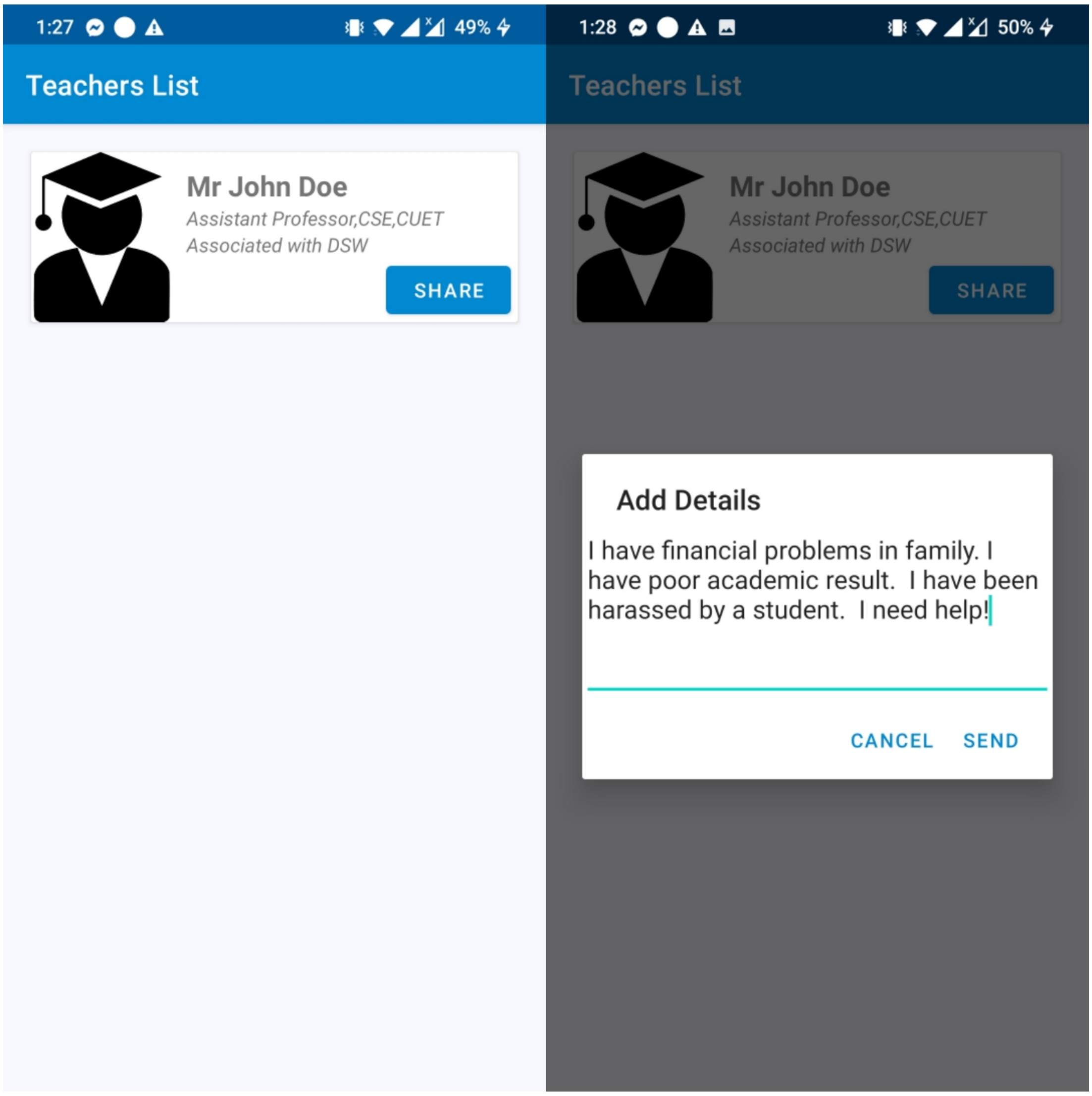


Figure 4.38: Sharing problems with teacher

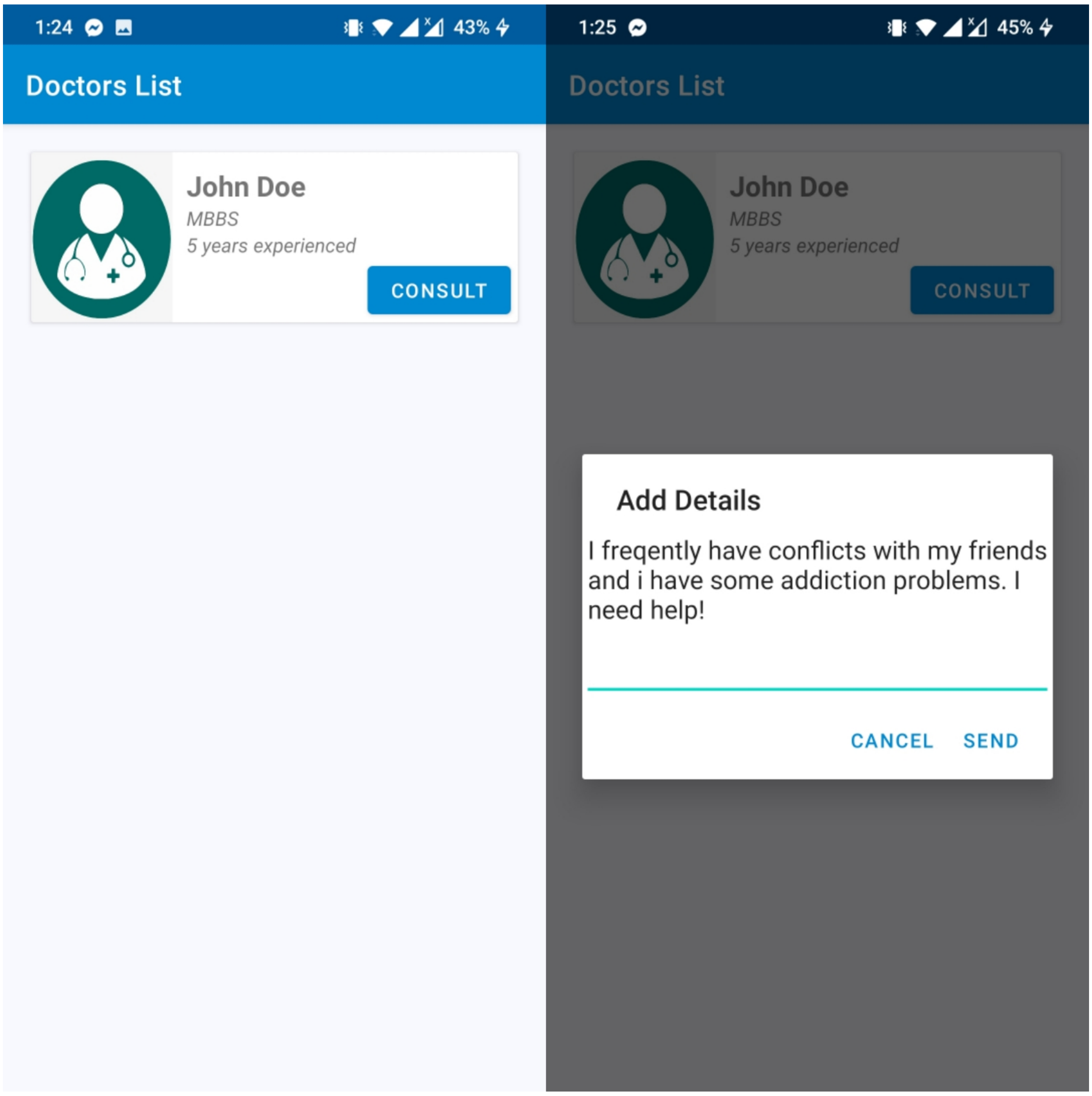


Figure 4.39: Sharing problems with doctor

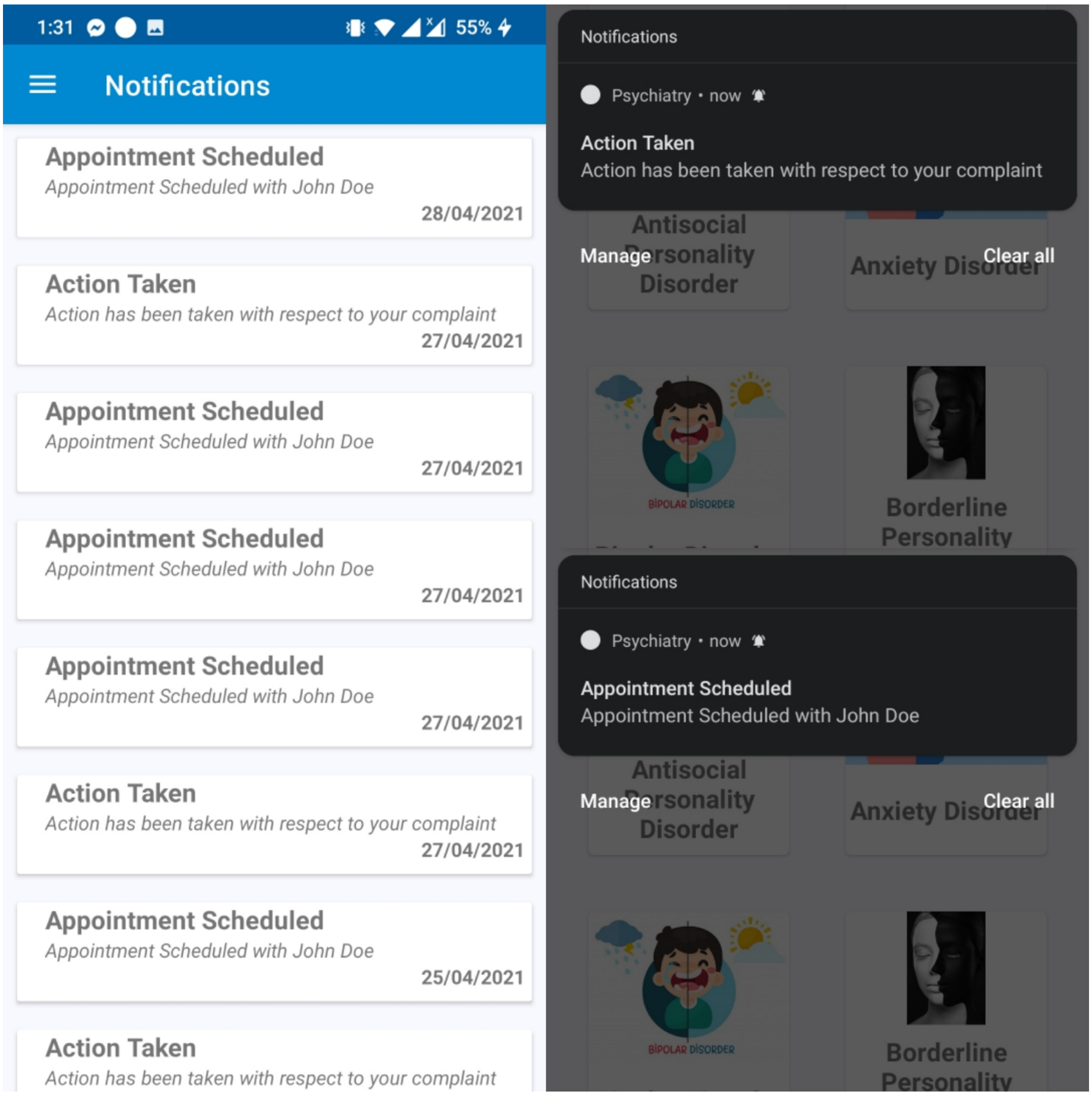


Figure 4.40: Notification to Student

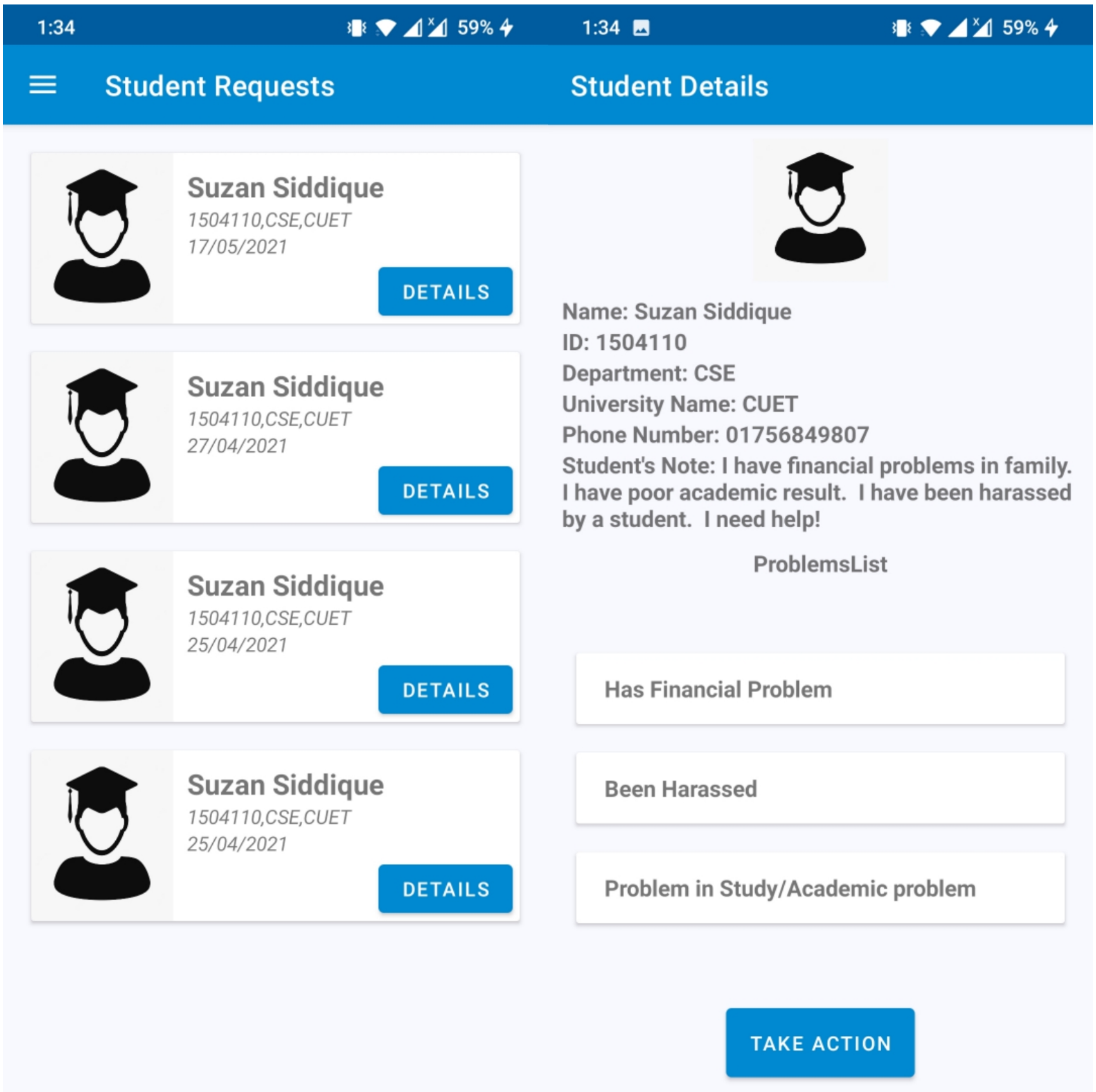


Figure 4.41: Action taken by Teacher

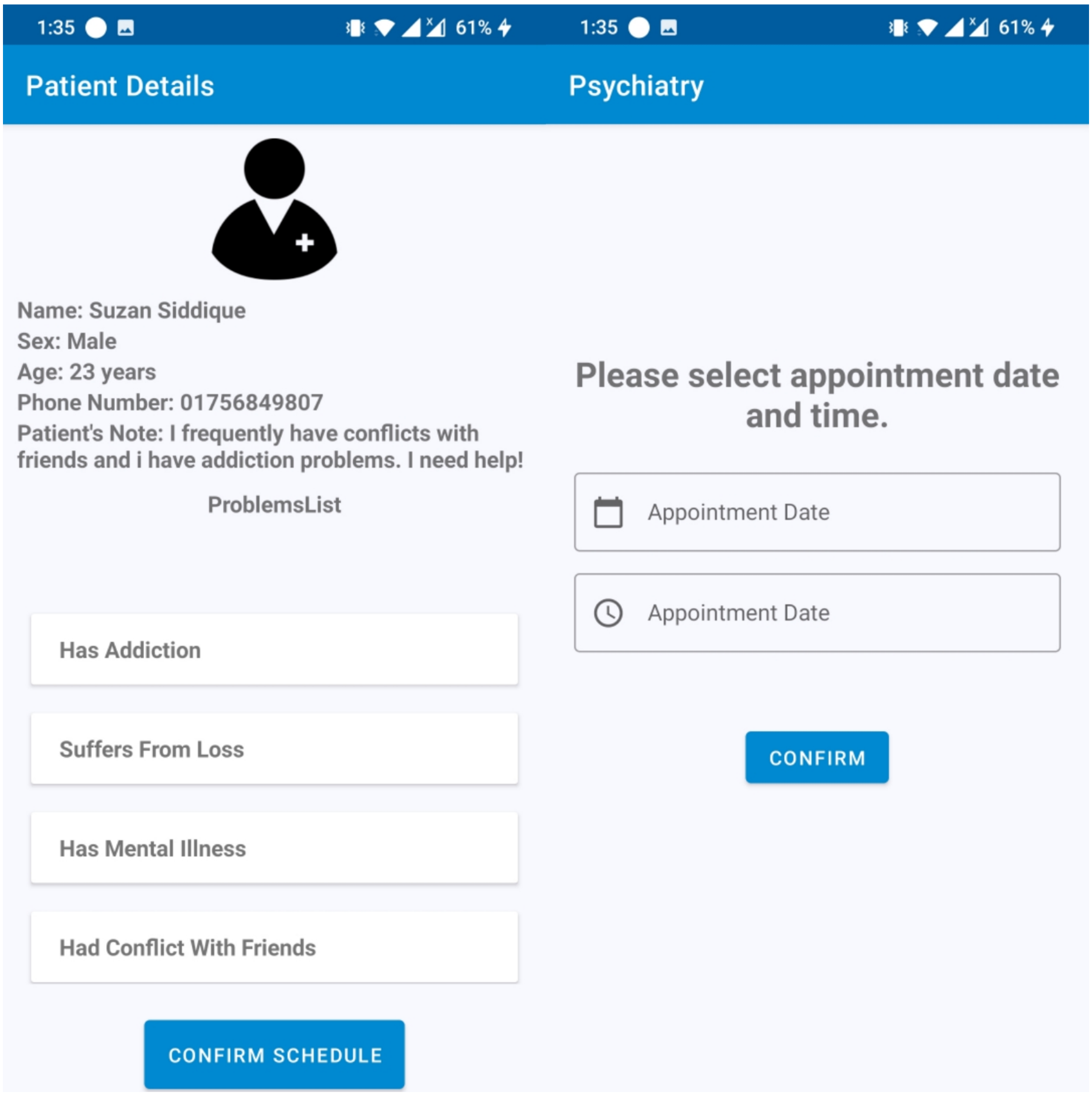


Figure 4.42: Action taken by doctor

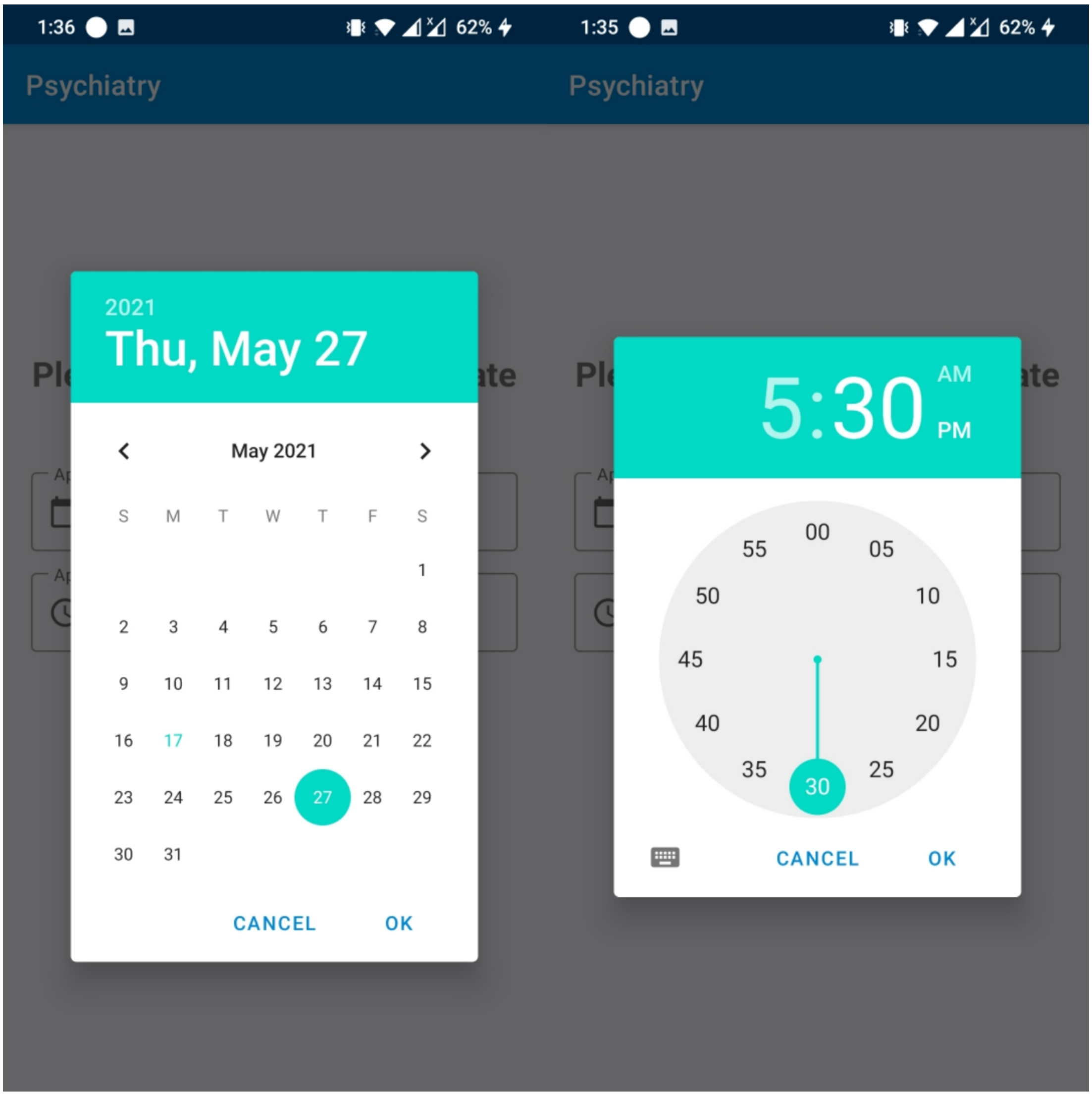


Figure 4.43: Appointment making

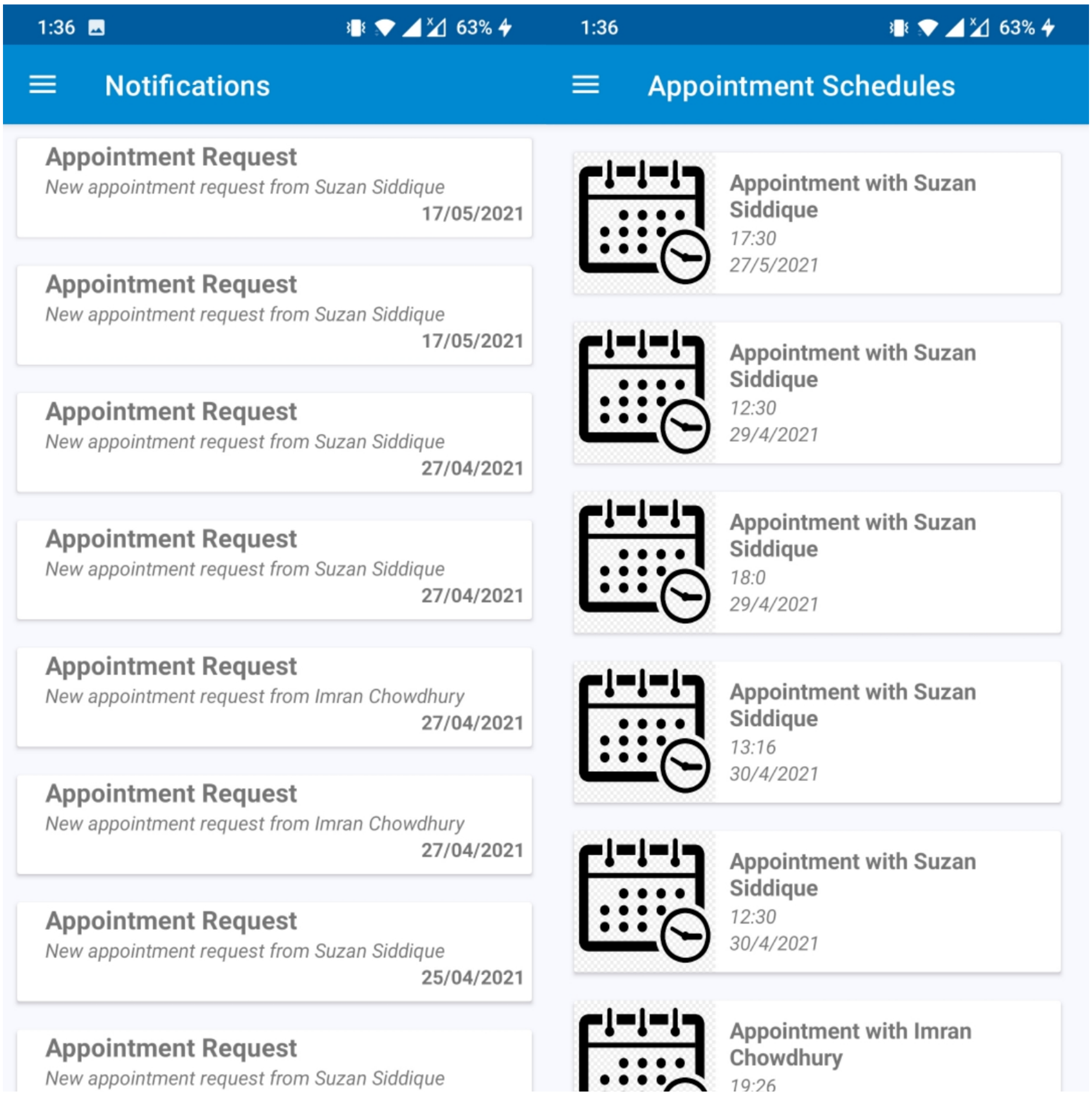


Figure 4.44: Notification and Appointment schedules for doctor

Chapter 5

Conclusion

5.1 Conclusion

Traditional depression checking system is not as efficient as this developed system. Mental health is an important thing that needs to be checked. In this era of modernization, industrialization, the mental health of people has been affected greatly. As a result people are suffering from different mental illness such as depression. For this we need a checking system to know the exact condition of mental health. This may make us a little conscious of our mental health. We designed a model using machine learning approach that can predict whether a student is suffering from depression or not. Visualization of data is also a better part of it. we also have built an android based mobile application to provide mental healthcare. This have a module for depression checking. People who do not understand machine learning can check their depression level from the application. Model training and prediction is the final touch of this project. For prediction we have used several algorithms to make a comparison based on accuracy, precision, recall, specificity, error rate, f-score, MAPE etc.

In summary, a machine learning approach based model was designed to predict depression among university students. From datasets, a machine learning analysis has been done and data was visualized too. To provide necessary mental healthcare to university students, an android based mental healthcare application was developed.

5.2 Future Work

This model can be updated in the future. We have worked with some important parameter or causes that could be crucial for finding depression while creating questionnaire for our datasets. Some other crucial causes could be figured or researched to find out depression with more accuracy and precision. Our dataset participants were also not up to expectations. If the number of participants were higher, our model could provide results with more precision. Our developed android based application could also be improved. Right now it is only available for android users which makes it unavailable for other platform or operating system users. The application could be made available for other platform like iOS or Windows so that those users can also be benefited from the application.

References

[1] *Family history of depression doubles risk for depression*, `https://www.healio.com/news/psychiatry/20160826/family-history-of-depression-doubles-risk-for-depression`, (Accessed on 05/18/2021) (cit. on p. 1).

[2] *1 in 5 college students have anxiety or depression. here's why*, `https://theconversation.com/1-in-5-college-students-have-anxiety-or-depression-heres-why-90440`, (Accessed on 05/18/2021) (cit. on p. 2).

[3] G. S. Kumar, A. Jain and S. Hegde, 'Prevalence of depression and its associated factors using beck depression inventory among students of a medical college in karnataka,' *Indian journal of Psychiatry*, vol. 54, no. 3, p. 223, 2012 (cit. on p. 2).

[4] *Anxiety and depression: More college students seeking help | time*, `https://time.com/5190291/anxiety-depression-college-university-students/`, (Accessed on 05/18/2021) (cit. on p. 2).

[5] J. January, M. Madhombiro, S. Chipamaunga, S. Ray, A. Chingono and M. Abas, 'Prevalence of depression and anxiety among undergraduate university students in low-and middle-income countries: A systematic review protocol,' *Systematic reviews*, vol. 7, no. 1, p. 57, 2018 (cit. on pp. 2, 8).

[6] *What is depression?* `https://www.psychiatry.org/patients-families/depression/what-is-depression`, (Accessed on 05/18/2021) (cit. on p. 7).

[7] *Nimh » depression*, `https://www.nimh.nih.gov/health/topics/depression/`, (Accessed on 05/18/2021) (cit. on p. 8).

[8] *Types of depression: 9 forms of depression and their symptoms*, `https://www.healthline.com/health/types-of-depression`, (Accessed on 05/18/2021) (cit. on p. 8).

[9] A. K. Ibrahim, S. J. Kelly, C. E. Adams and C. Glazebrook, 'A systematic review of studies of depression prevalence in university students,' *Journal of psychiatric research*, vol. 47, no. 3, pp. 391–400, 2013 (cit. on p. 8).

[10] *Bdi-ii beck depression inventory-ii*, `https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Beck-Depression-Inventory-II/p/100000159.html#tab-details`, (Accessed on 05/18/2021) (cit. on p. 9).

[11] K. Daimi and S. Banitaan, 'Using data mining to predict possible future depression cases,' *International Journal of Public Health Science (IJPHS)*, vol. 3, no. 4, pp. 231–240, 2014 (cit. on p. 9).

[12] I. Bhakta and A. Sau, 'Prediction of depression among senior citizens using machine learning classifiers,' *International Journal of Computer Applications*, vol. 144, no. 7, pp. 11–16, 2016 (cit. on p. 10).

[13] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes and C. L. Hanson, 'Validating machine learning algorithms for twitter data against established measures of suicidality,' *JMIR mental health*, vol. 3, no. 2, e21, 2016 (cit. on p. 10).

[14] I. C. Passos, B. Mwangi, B. Cao, J. E. Hamilton, M.-J. Wu, X. Y. Zhang, G. B. Zunta-Soares, J. Quevedo, M. Kauer-Sant'Anna, F. Kapczinski *et al.*, 'Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach,' *Journal of affective disorders*, vol. 193, pp. 109–116, 2016 (cit. on p. 11).

[15] M. Weissman, O. Berry, V. Warner, M. Gameroff, J. Skipper, A. Talati *et al.*, *A 30-year study of 3 generations at high risk and low risk for depression. jama psychiatry. 2016; 73 (9): 970-7* (cit. on p. 20).

[16] *A climb out of depression, doubt and academic failure - the new york times*, `https://www.nytimes.com/2017/06/07/education/one-student-tells-her-story-of-a-climb-out-of-depression-.html`, (Accessed on 05/18/2021) (cit. on p. 21).

[17] *Substance abuse and depression: A dangerous downward-spiral*, `https://www.psycom.net/depression-substance-abuse`, (Accessed on 05/18/2021) (cit. on p. 22).

[18] *How being depressed can affect your relationships | psychology today*, `https://www.psychologytoday.com/us/blog/fulfillment-any-age/201602/how-being-depressed-can-affect-your-relationships`, (Accessed on 05/18/2021) (cit. on p. 23).

[19] *How to help a depressed friend (and when to stop trying): Part 2 | psychology today*, `https://www.psychologytoday.com/us/blog/embracing-the-dark-side/200906/how-help-depressed-friend-and-when-stop-trying-part-2`, (Accessed on 05/18/2021) (cit. on p. 23).

[20] T. Richardson, P. Elliott, R. Roberts and M. Jansen, 'A longitudinal study of financial difficulties and mental health in a national sample of british undergraduate students,' *Community mental health journal*, vol. 53, no. 3, pp. 344–352, 2017 (cit. on p. 24).

[21] R. Hiremath and S. P. Debaje, 'Assessment of prevalence of domestic violence and mental health profile of adolescents exposed to domestic violence in an urban slum in mumbai,' *Int J Res Med Scien [Internet]*, vol. 2, no. 1, pp. 290–92, 2014 (cit. on p. 24).

[22] *Domestic violence and abuse - the impact on children and adolescents | royal college of psychiatrists*, `https://www.rcpsych.ac.uk/mental-health/parents-and-young-people/information-for-parents-and-carers/`

domestic-violence-and-abuse-effects-on-children, (Accessed on 05/18/2021) (cit. on p. 24).

[23] R. Kaltiala-Heino, M. Rimpelä, M. Marttunen, A. Rimpelä and P. Rantanen, 'Bullying, depression, and suicidal ideation in finnish adolescents: School survey,' *Bmj*, vol. 319, no. 7206, pp. 348–351, 1999 (cit. on p. 24).

[24] M. Fekkes, F. I. Pijpers and S. P. Verloove-Vanhorick, 'Bullying behavior and associations with psychosomatic complaints and depression in victims,' *The Journal of pediatrics*, vol. 144, no. 1, pp. 17–22, 2004 (cit. on p. 24).

[25] *The hidden health effects of sexual harassment*, https://www.nbcnews.com/better/health/hidden-health-effects-sexual-harassment-ncna810416, (Accessed on 05/18/2021) (cit. on p. 25).

[26] J. N. Houle, J. Staff, J. T. Mortimer, C. Uggen and A. Blackstone, 'The impact of sexual harassment on depressive symptoms during the early occupational career,' *Society and mental health*, vol. 1, no. 2, pp. 89–105, 2011 (cit. on p. 25).

[27] *Social media users 30 million in bangladesh: Report*, https://thefinancialexpress.com.bd/sci-tech/social-media-users-30-million-in-bangladesh-report-1521797895, (Accessed on 05/18/2021) (cit. on p. 25).