Fighting Depression: Psychological Approaches Among Bangladeshi University Students

by

Sanjib Kumar Sen
19301101

Md. Shifatul Ahsan Apurba
19301086

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
May 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sanjib	Apurba
Sanjib Kumar Sen 19301101	Md. Shifatul Ahsan Apurba 19301086

Approval

The thesis/project titled “Fighting Depression: Psychological Approaches Among Bangladeshi University Students” submitted by

1. Sanjib Kumar Sen (19301101)
2. Md. Shifatul Ahsan Apurba (19301086)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 29, 2022.

Examining Committee:

Supervisor:
(Member)

Tawhid

Md. Tawhid Anwar
Lecturer
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Noor

Jannatun Noor
Lecturer
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

In recent years, mental health deterioration cases have increased exponentially indicates that this issue needs our attention. Stressful situations of our daily life such as excessive study load, relationship predicaments, domestic abuse, sexual harassment, and many other reasons cause depression which is often prevalent and ends up causing physical or mental harm. The psychiatrists, psychoanalysts, and counselors are having a tough time dealing with large numbers of cases. They could only help some of the patients as most of them do not have access to them. Moreover, some people cannot bear the cost or feel hesitant to open up to them. In addition, the overall process takes a long time to understand the patient's condition, and sometimes patients hide information from the counselors that lead to wrong assessment. Sometimes, it is too late to diagnose and treat their depression. As a result, they reach an extremely vulnerable stage and choose the path of self-harm that contributes to the increasing rate of suicide. Analyzing their history and then taking proper measurements can contribute to the treatment of depression. However, the challenge is that human behavior is ambiguous and inconsistent. Therefore, we propose methodologies for diagnosing their mental health conditions by tracking the probable cause of their depression. With the help of deep learning and machine learning, our goal was to analyze large data sets for observing patterns such as age, gender, the causality of depression, the delta of behavior changes, and many other things related to students and excavating things efficiently to help patients. When it comes to making predictions about depression and offering advice, the survey data that we have gathered over the course of this project has been of great assistance. According to the findings of our research, the Random Forest Classier Algorithm is capable of accurately predicting depression with an accuracy of 87%, an f-measure of 86%, and this model is also the best model. In comparison to the other algorithms that we used, such as K-Nearest Neighbor, Support Vector Machine, Gaussian Naive Bayes, Artificial Neural Network, Gradient Boost, and Decision tree, this one performed far better. The recommendation model that was built by us as part of this research project is our novel contribution to this discussion. We will prognosticate to assist the students with mobile application in the near future, so that they feel better with the help of Machine Learning and Deep Learning by inspecting and examining those patterns.

Keywords: Machine Learning; Depression; Mental Health; Prediction; Random Forest; Gaussian Naive Bayes; Neural Network; Decision tree; Support Vector Machine; Deep Learning

This research is dedicated to all those who suffered from depression and were unable to receive adequate assistance, as well as those who committed suicide due to depression...

Acknowledgement

To begin with, we praise God and are grateful to him for allowing us to finish this thesis without severe disruptions. We traversed a long path and encountered several obstacles, but by his grace we have triumphed over them all.

Secondly, we express our greatest respect to our respected supervisor Mr. Md. Tawhid Anwar and co-supervisor Ms. Jannatun Noor. Without them, this thesis may not have been possible in the first place. Mr. Md. Tawhid Anwer helped us with every step from the start. Without hesitance, he backed us and inspired us to be involved. Ms. Jannatun Noor, on the other hand, actively advised us on how to approach, what to accomplish, and how to preserve the quality of this thesis. She also inspires us. We thank both of you, you have been the most helpful guides on our adventure.

Moreover, we would like to thank everyone who participated in the survey process by filling out the questionnaire. We are grateful to Akib Hossain, Tanima Pal, Nowshin Shahid, and the other friends, classmates, seniors, and juniors who assisted us in distributing the survey and collecting data. Without their co-operation, this study would not have been accomplished.

Finally, we would like to thank our parents and other family members for providing us with everlasting love and support.

All of them have our appreciation for helping us develop our research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we will provide an overview of the study, followed by a statement of the problem, research objectives, our contributions, and an outline of the research.

1.1 Overview

In the current era, depression has become a significant problem among university students. According to studies, psychologists examining university students with mental health-related issues often get perplexed because mental health has several dimensions. The primary dimension that causes depression can be academic quandaries, social dilemmas, relationship related problems, financial crisis, Etc. This situation does not come around certainly; instead, it is like the butterfly effect. Meaning, when a person is about to start university, he certainly faces an environmental change, i.e., leaving his school friends behind and getting introduced to new responsibilities with new hope to shape life. All those certain events sometimes work as a fear mechanism for them. The word "depression" covers a far-flung area of suffering. Not only does it come from a specific condition, but it also can be a result of possible traumas. Somebody can even feel depressed over nothing. It can additionally originate from frustration, tiredness, stress Etc. Most guardians in Bangladesh keep high expectations from their children to do better in academic activities and get high-paid professions. So, newly admitted university students feel immersive pressure. The scenario absurdly looks unrealistic to them, and they fail to cope with the new surroundings, consequently making them the victim of depression. Machine Learning and Deep Learning are two proven tools that accommodate the analysis of large amounts of data. Previously, in many interoperability models, machine learning and deep learning were used successfully—for example, Disease Identification and Diagnosis, Medical Imaging Diagnosis, Patient Support Tasks Etc. In fact, with the help of Deep Learning, a complex form of Machine Learning is used for mimicking human brain functions that help a lot in radiology and medical imaging (University of Illinois Chicago, 2021)[33]. It can be helpful in such a complicated sphere of our physical ecosystem, and it may also be very effective in determining possible patterns from our input data as human brain functions are also related to mental activities. So, we believe machine learning and deep learning can be a weapon to demolish depression with frequently improved technologies. Unlike other diagnoses, we can now diagnose whether a person is depressed or not and assist them by presenting a possible pathway to resolve.

1.2 Problem Statement

Depression has become one of society's most severe problems and has been identified as a threat to the future generation. While the people of Bangladesh have adequate access to the internet and educational possibilities, yet they cannot respond appropriately to mental health problems. Mental health problems, on the other hand, are taboo in this environment. Although there are many social awareness initiatives, this problem remains unresolved. There are a variety of stigmas associated with mental illness. Most of the time, a person suffering from depression does not open up to family, friends, or other close relatives since they do not want to be associated with such a severe disease at first. Instead, they treat them strangely and in a different way than they should [4]. In real life, they are often referred to as "attention seekers." This type of mentality limits them and pushes them to the side; so that they feel marginalized and unimportant. Aside from that, the constant fear of being judged prevents people from speaking up [22].

Furthermore, in Asian nations, it is typical for individuals to think that a narrative such as time would heal everything. Some individuals still consider that depression is a non-existent condition that everyone should avoid at all costs. People who feel sad just because they are thinking about depression are in the minority. In real life, the situation may be different.

Furthermore, several of the students seemed to be concerned about whether or not they were depressed. Most often, people are unsure whether they need to see a therapist or psychiatrist or avoid seeing one altogether. Many people are unaware that therapeutic concourses exist and may be of assistance [21]. As a result, the situation continues to deteriorate, and young people are increasingly unable to cope with the circumstances described above and resort to self-harm, including suicide. According to recent research, 69.5 percent of students from private and public institutions in Bangladesh suffer from depression, compared to 24 percent in 2015 [26]. Consequently, about 3.7 percent of depressed college students attempted suicide during the Covid-19 period [36]. These alarming figures suggest that it needs immediate attention and action. Depressive disorders may have a negative impact on even those who are not suicidal because they cause people to delay, which results in a reduction in total production for both individuals and the nation. Additionally, it indicates a waste of abilities and capabilities.

Additionally, some victims might become so emotionally unstable that they lose their sense of morality and begin to harm others physically. Violence in the home, forced assault on children, and various other issues are examples of what is outlawed. However, it is crucial to remember that it is a health issue and that we should treat it as such, just as we would other health problems and illnesses. Otherwise, most individuals would lose hope in their ability to live their lives, which will result in severe unemployment or issues with creativity. Nonetheless, as we have stated, depression is a far more significant subject matter. Anyone can get depressed for any reason; we have no way of knowing for sure. It is challenging to come up with a remedy for their despair. Because they have not discovered a solution, they have been unable to dedicate enough time to their education, skills, and other options. As a result, the depression problem among university students contributing to the nation's future might lead to a dreadful situation in terms of the individual, the family, and the larger community [2].

Because the number of depression cases is increasing, we may assume that professional psychotherapists and counselors will be unable to effectively treat such instances within a few years due to the increased demand. It is high time to scrutinize more effective methods of detecting and resolving the issue. With this in mind, the question could be asked whether or not depression can be detected using machine learning and deep learning techniques. If yes, do we find a method to fix the issue or not? "Yes," is the response to the first of the two questions. The use of machine learning for the detection of depression has been the subject of many studies. We have seen depression prediction techniques in [21] and [29], among other publications. There is also more study being done in this area. Meanwhile, the answer to the second question is, regrettably, a resounding "No."

We have set the following research questions (RQ) for our thesis:
RQ1: In what ways may depression be detected?
RQ2: How can we develop a support system for depressed individuals?

As mentioned earlier, machine learning models have not been used yet for determining solutions for depression. So, determining the right tool, which in that case is the most efficient machine learning classifier, would be our first objective. We have used some popular machine learning classifiers to test prediction and suggestion accuracy, including- Support Vector Machine (SVM), K-nearest Neighbor (K-NN), Gaussian Naive Bayes (GNB), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Gradient Boosting (GB). After determining the right tool, we have focused on our following objectives: prediction of depression and the suggestions to overcome that.

A secondary source of worry is that the prediction outcome may fall short of our expectations. This is because depression may have various underlying reasons, and the majority of them cannot be resolved via technological means. Our main goal is not overcoming or curing depression; instead, we want to help the individual feel better. It is impossible to anticipate our models to heal the century-old disease dramatically since human emotions and sentiments operate in complicated ways that require biological study and external medicine to treat difficult clinical situations.

Given this, we would want to explore our possibilities of helping individuals feel better during their difficult times, despite our severe reservations. We are hopeful that our study will be fruitful and provide answers to the issues we have raised.

1.3 Research Objective

Our objectives have been split into two broad categories based of the effect, usefulness, issue analysis, and potential solutions and social impact that were discussed before.

1.3.1 Core Objectives

This research aims to build a system that can detect the depressed state of an individual, analyze and find out the root causes of being in a depression, and suggest tasks or activities that make the individual feel better.

So, in brief, our target for this research is to:

1. Determine if a student is in depression or not using Machine Learning Model.
2. Analyze and find the root causes of his/her depression with the help of Data Science, Machine Learning
3. Test, Determine and optimize a machine learning model/classifier for this research.
4. Give suggestions to the student based on his mental, physical, career, personal and environmental states to make him feel better using the Machine Learning model.
5. Receive Feedback from the volunteer students and calculate the accuracy of the model.
6. Determine if it is possible to reduce anxiety, sadness using the suggestions made by machine learning and deep learning models.
7. Determine the patterns of depression among Bangladeshi Students and how they overcome this disorder.
8. Determine if a student's mental state can be improved upon specific activities based on his lifestyle.
9. Make a user-accessible tool for public usage so the bad effects and depression causing suicidal attempts can be reduced.

1.3.2 Additional Focus

Since we have noticed that, mental illness or depression is still a stigma here in Bangladesh, we are also focused on certain supplementary field:

1. Educate students on problems related to mental health, and monitor their progress to ensure that they are receiving treatment as soon as possible.
2. Help students overcome the self-doubt and the humiliation that they feel because of the depression they are experiencing.
3. To find a solution to the issue that is leading to depression, the culture of self-isolation has to be challenged, and people need to learn to communicate more freely with one another.
4. Create a community in which individuals are more willing to discuss these issues openly and support one another.

1.4 Contributions

1. In order to categorize the results of our survey, we used a multi-class classifier approach, which is a technique that is not only innovative but also very effective. Instead of only getting an overarching picture of what's happening, we were able to get the specific feedback or survey data by using these strategies. Our results suggested a higher degree of precision and accuracy as a consequence of these factors.
2. Our data set for the survey was expanded to include a new component that consisted of a specific question that distinguished between students who are now suffering from depression and students who had successfully dealt with their depression. This not only allowed us to acquire data from students with healthy mental states but also expanded the scope of our study.
3. Using specific questions and multi-class classifier approach we became able to gather a lot of data from the students' community from different diversity, institution and areas. We also collected data from the student who were able to came out of the depression. Combining and analyzing these data set we were able to generate suggestions based on previous experience, environment, characteristics, various physical, mental and abstract conditions.

1.5 Thesis Outline

1. In the first chapter, we illustrated the problem statement and our aims for this study to solve the difficulties that were shown earlier. The chapter began with an overview section. In addition to that, our contributions and this general outline have been included into this chapter. The first chapter covers, more or less in its entirety, all of the goals that we had hoped to accomplish as a result of our study.
2. In the second chapter, as the name 'Literature Review' suggests, we will acquire a fundamental understanding of the terminologies associated with depression, as well as the machine learning environment and the algorithms that

we have used for this research. We will also become familiar with some of the related works that have been done in the field of depression.

3. In the third chapter, we went into more detail about our research methodologies, including data collection process, how we get the data ready for testing, how we split the data up, and how we classify it, as well as the algorithms that are used to make machine learning a more effective enhancement mechanism for the system.

4. In the fourth chapter, we discussed about the interview questions, described the data set that we have collected, described the common problems that cause depression, showed the method that we used to clean our data, and finally, we visualized the important data to compare the conditions. All of these topics were covered in the chapter.

5. In the fifth chapter, the experimental results and the explanation of those results are presented. Here, we demonstrate the accuracy, as well as the precision and the recall, of the different algorithms that we apply. In addition to that, it compares the outcomes of the various algorithms. In the end, we presented the strategy for suggestion distribution that we had recommended.

6. In chapter six, we have described the limitations, the future scope, and the work that is now being done.

7. In the seventh chapter, we have presented our final observations, which include a restatement of the overall stance of the study.

Chapter 2

Literature Review

Depression is the most burdensome disease in terms of health impact, affecting people from all walks of life, regardless of where they inhabit or how wealthy they are. Depressive disorders often begin in youth, interfere with everyday activities, and eventually become chronic. Because of the pervasive nature of the problem, depression therapy and other mental health disorders are in great demand among those suffering from them.

The World Health Organization [34] estimates that about 280 million individuals suffer from depression at any one time across the world. They claim that more than 700,000 individuals commit suicide each year due to severe depression, with the majority of them being between the ages of 15 and 29. Unfortunately, countries such as Bangladesh cannot offer therapy to 75% of depression sufferers, which prompted us to consider this worldwide issue and begin working on it.

Firstly, we will discuss the fundamentals of depression, machine learning, and deep learning before moving on to the topic of related research and its implications.

2.1 Depression

Depression is the most commonly diagnosed psychotic disorder in psychiatric practice. When they feel depressed, they remain gloomy for a lasting period. According to UC Berkeley's University Health Service [40], clinical depression is a severe medical condition that impairs mood, thinking, and behavior. Clinical depression impairs an individual's ability to operate normally. Additionally, as per American Psychiatric Association [6], clinical depressive illnesses are distinguished by the persistent presence of specific physical and cognitive abnormalities, as well as a sad, empty, irritated, or anhedonia mood. Depression is very different from sadness or bereavement. During our studies we have seen many students consider themselves depression due to breakup or death of their loved ones. According to [30], sadness may be debilitating, it is not the same as depression. In some ways, the mourning process resembles depression in that it is normal and individual to each person going through it. Grief and depression may both cause a person to retreat from their daily routines with feelings of deep sorrow. So it is marked that, depression and bereavement may live side by side. When a loved one dies, someone loses a job or is a victim of a physical assault or severe catastrophe, it is possible for that person to suffer depression. Bereavement is more intense and can last when it is accompanied by depression. Understanding the difference between bereavement and depression may help individuals obtain the support and prophylaxis they need. The few common depression types are Major depression, Persistent depressive disorder, Bipolar disorder, Seasonal affective disorder (SAD) [27]. As it is related to mental health, we cannot consider this a static situation to ensure what kind of depression a person is going through.

2.2 Machine Learning

According to the Nvidia blog article [12], at its most fundamental level, machine learning is the method of understanding data, learning from it, and then producing a judgment or prediction about anything in the world based on that information. Instead of manually programming software routines with specific instructions to accomplish a task, the machine is trained to utilize enormous quantities of data and algorithms to complete the task. Many studies have analyzed Computer Vision, Artificial Intelligence-driven projects, and a variety of other topics using machine learning techniques. Many algorithms are available to boost machine learning, and new algorithms are being created to enable a more precise decision-making process. Deep Learning is a more sophisticated machine learning that we refer to as “Deep Learning.”

2.3 Deep Learning

As previously stated, deep learning is a more advanced and powerful form of machine learning. Deep learning significantly improved the performance of machine learning. According to IBM [24], deep learning is a subset of machine learning characterized by a three or more layered neural network as its main component. Through the use of vast quantities of data, these neural networks aim to mimic the function of the human brain. While a single-layer neural network may still make approximate predictions, additional hidden layers allow the network to be tuned and adjusted to improve its overall accuracy. We use deep learning to power a variety of applications, including chatbots, extensive data analysis, and image recognition.

2.4 Supervised Learning Method

The supervised learning approach has been applied in our study. For example, it is possible to use a data set that already includes both the input and the output. It means that the system already knows the answer and has made a choice in light of that information. According to the [25], in the field of artificial intelligence, supervised learning, commonly known as supervised ML, is a subclass. It is characterized by the use of labeled data sets to train algorithms that properly categorize data or predict outcomes. The model modifies its weights when input data is fed into it as a way of the cross-testing process until the model is reasonably fit. Many genuine issues are solved at scale using supervised learning.

2.5 Describing Used Algorithms

We used several algorithms to ensure our thesis had more accurate results than we could get via other means. In total we have used 7 algorithms.

2.5.1 Support Vector Machine (SVM)

It is one of the popular algorithm among the data scientist. It can produce significant accuracy with less power consumptions. The purpose of the support vector machine method is to locate a hyperplane in an N-dimensional space that can categorize the data points in a distinguishing manner [18]. The method known as the Support Vector Machine selects the data points that are the most extreme in order to generate a hyperplane [39]. The following figure presents a decision boundary, also known as a hyperplane, that is used to classify two distinct categories.

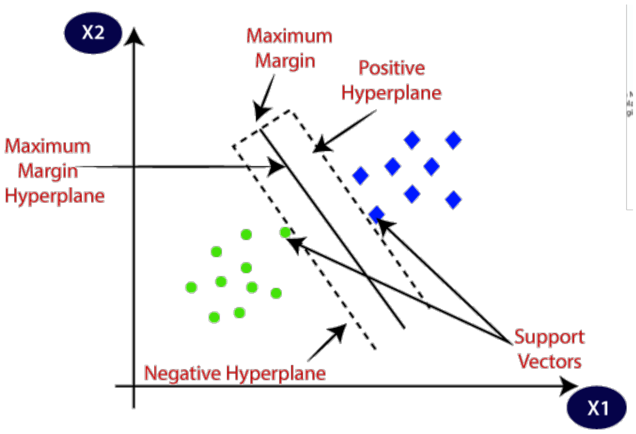

Figure 2.1: Support Vector Machine Algorithm [18]

2.5.2 Artificial Neural Networks (ANN)

The neural network found in the human brain served as inspiration for the Neural Network algorithm. Artificial neural networks, which are very similar to human brains, are made up of a large number of linked processors that are referred to as neurons. Neurons are responsible for the summing function. The weights that are attached to the connections are where the information is kept. In a nutshell, an artificial neural network (ANN) is made up of a group of unit cells, also known as artificial neurons, that are organized into an input layer, one or more hidden layers, and an output layer. Adaptive weights are used to establish connections between each neuron and the neurons in the layers that surround it [15].

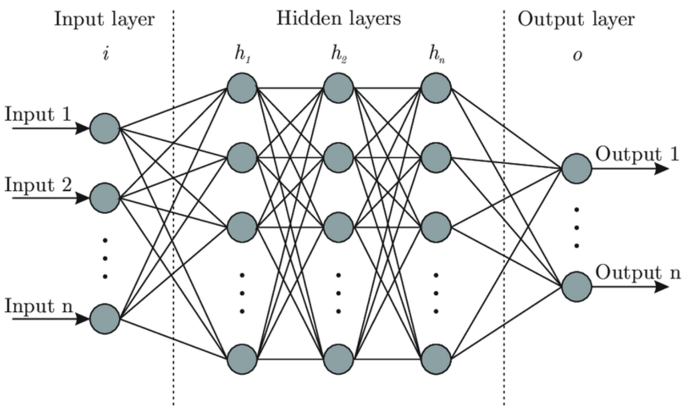

Figure 2.2: Artificial Neural Networks Architecture [15]

2.5.3 Decision Tree

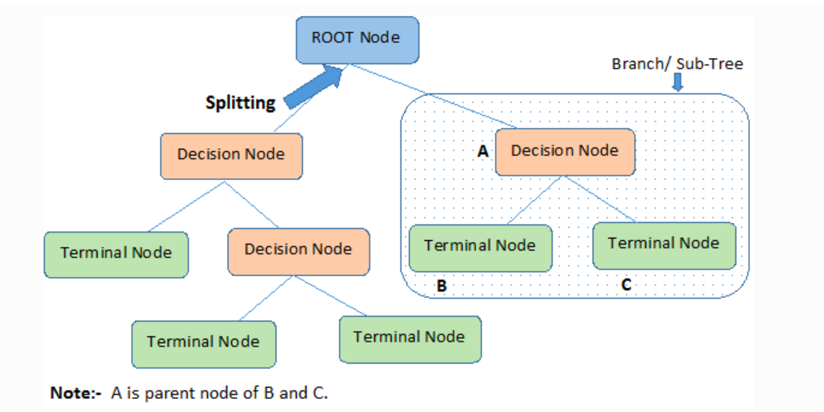

Figure 2.3: Decision Tree Algorithm [38]

Another important supervised learning algorithm is the Decision Tree algorithm. A training model that can be used to predict the class or value of the target variable by learning basic choice rules inferred from past data may be created with the help of a Decision Tree. This purpose of employing a Decision Tree is intended to be accomplished. When using Decision Trees to forecast a class label for a record, we begin at the tree's root to make our predictions. It compares the values of the root attribute and the attribute of the record. Following the path indicated by the comparison, we go to the next node after following the branch that corresponds to the value in question [38].

2.5.4 Random Forest

The Random Forest classifier is an example of an ensemble approach. It involves training many decision trees simultaneously via bootstrapping, which is then proceeded by aggregation, often known as bagging. When using decision trees, we may come up against the issue of over-fitting, which keeps the question of accuracy open. Random Forest classifiers are being used to accomplish this task since they assure generalization and combine each decision tree. As a consequence of this, its accuracy is much improved [28].

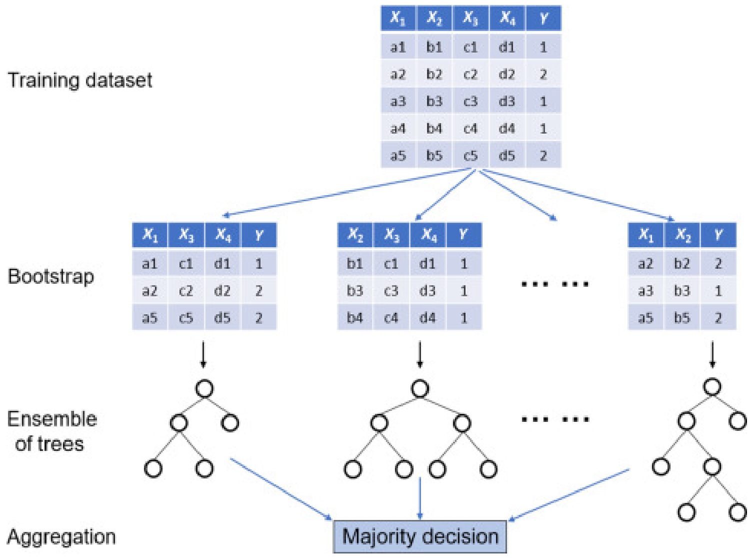

Figure 2.4: Random Forest Classifier [28]

2.5.5 K-Nearest Neighbor (KNN)

K-Nearest Neighbor performs a distance calculation between the test data and all of the training points in an effort to determine which class the data being evaluated should be assigned to. The K number of points that are most closely related to the test data should then be chosen. The K-Nearest Neighbor method works by determining the likelihood of the test data belonging to each of the K classes of the training data. It then chooses the class that has the greatest probability. It does this by calculating the Euclidean distance between each pair of data points using the data set. After that, the data that is closest to the neighbor is the data that is allocated to that neighbor group [32].

The figure 2.5, 2.6, 2.7 and 2.8 demonstrate more about this algorithm.

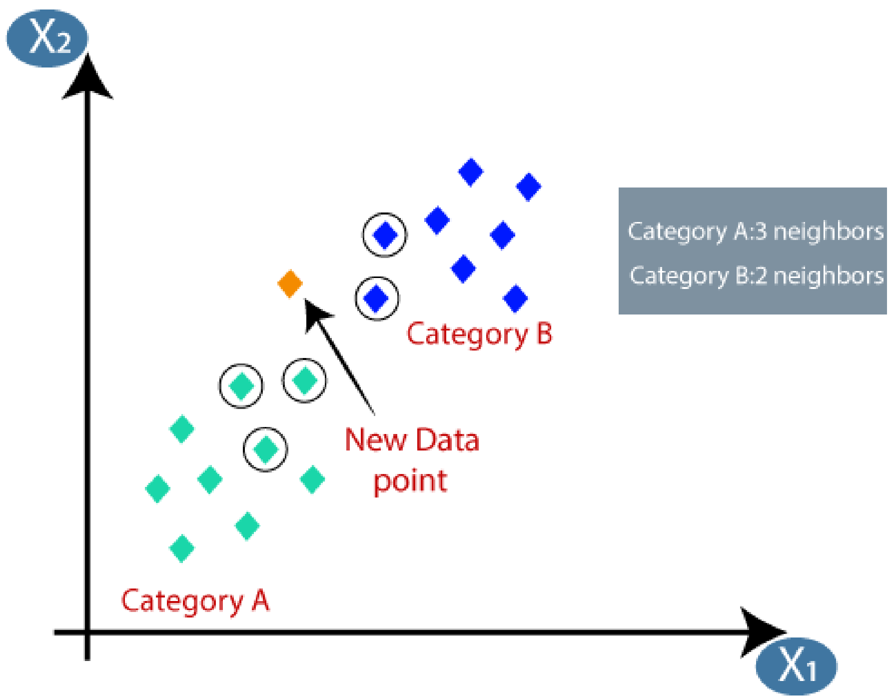

Figure 2.5: Before Applying K-NN [32]

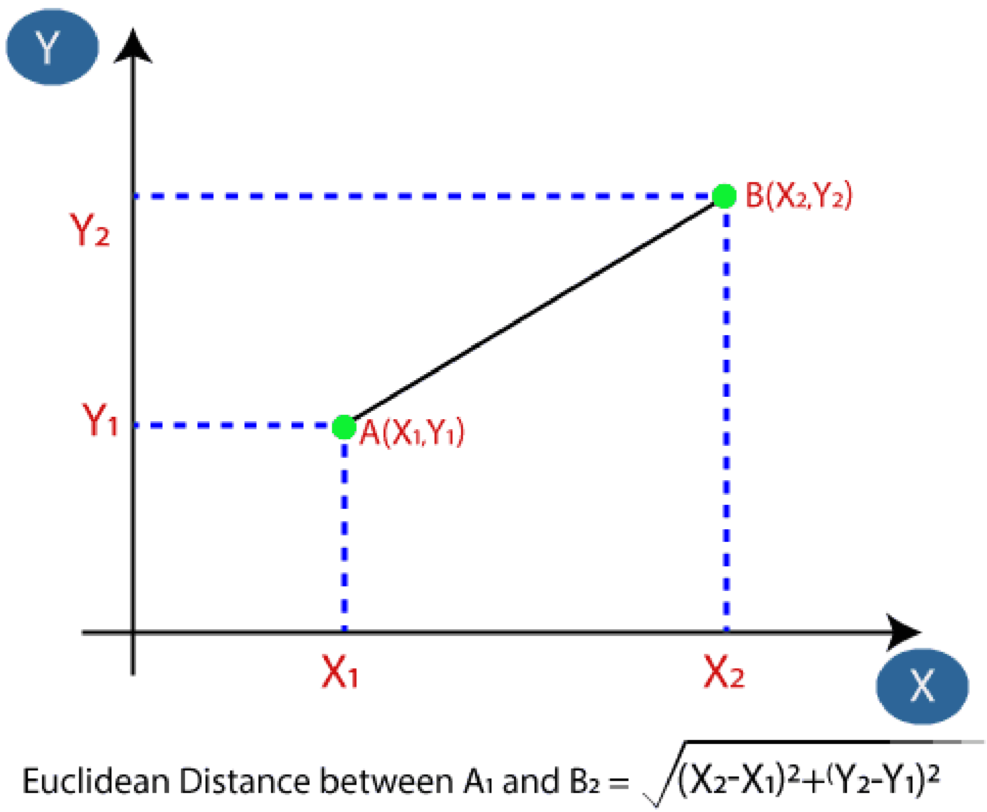

Figure 2.6: Euclidean Distance[32]

Figure 2.7: After Applying K-NN[32]

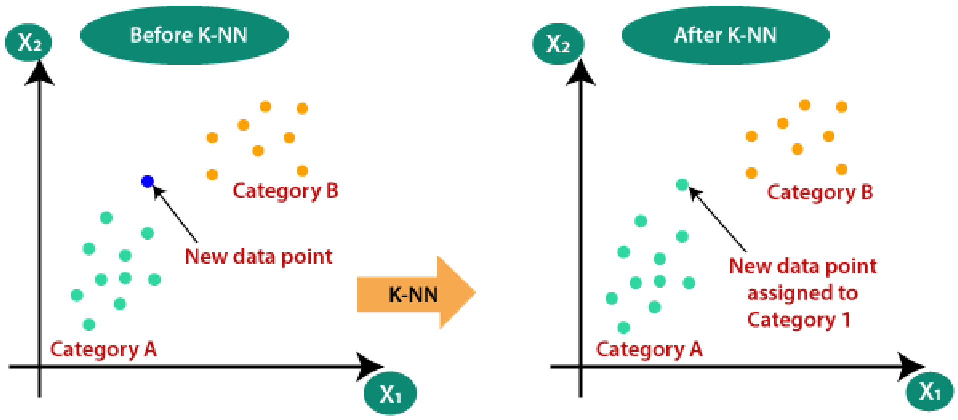

Figure 2.8: Assignment Completed by K-NN [32]

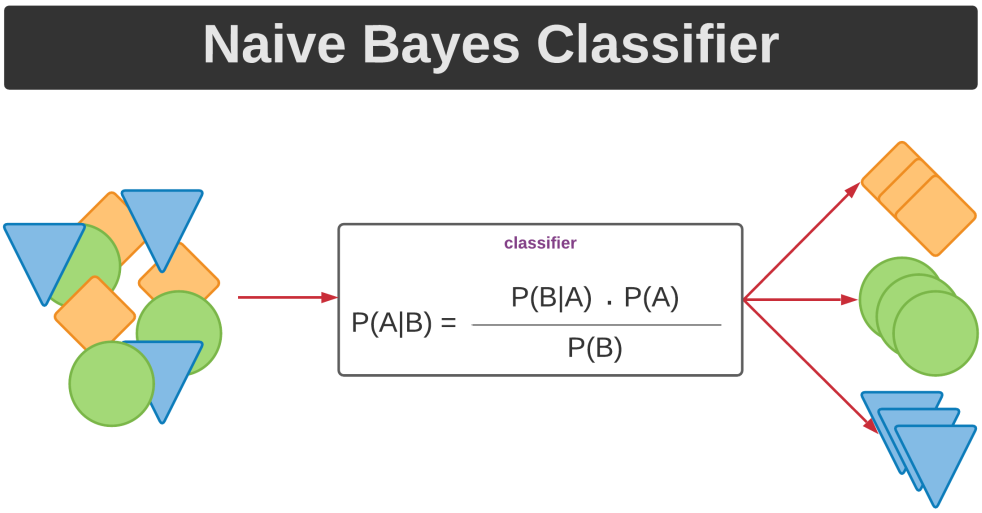

Figure 2.9: Naive Bayes Algorithm [37]

2.5.6 Naive Bayes Algorithm

Naive Bayes is a probabilistic classification method that is one of the simplest Supervised Learning algorithms. It is based on the Bayes Theorem, which is a mathematical proposition. In particular, for huge data sets, the Naive Bayes classifier is an approach that is not only fast but also accurate and reliable [37].

2.5.7 Gradient Boosting

Figure 2.10: Simulated Data of x Input and y Output [16]

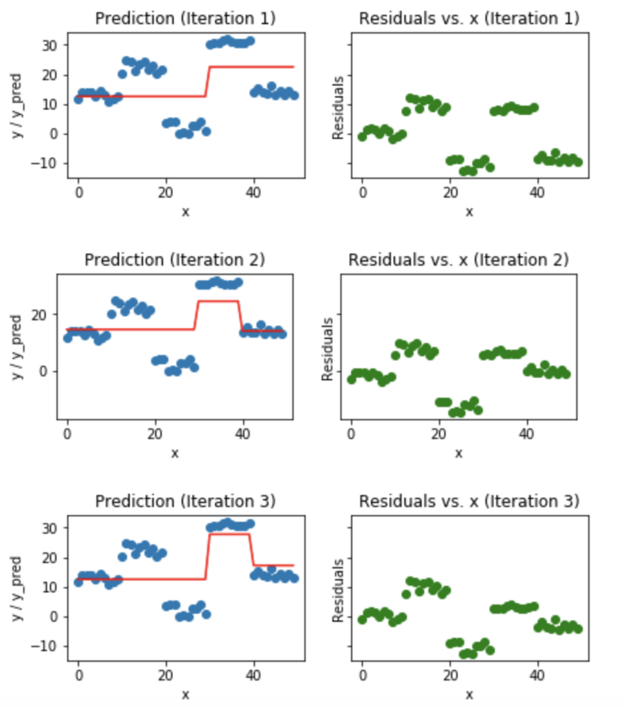

Figure 2.11: Visualization of Gradient Boosting Predictions After 4 Iteration

Gradient Boosting is very productive machine learning algorithm for classification and regression problems. From comparatively weak assumption by continuously updating itself from the previous data. After certain number of sequential iterations, it generates good prediction. Basically, it checks the previous mistake and omits the mistake if that occurs again [16].

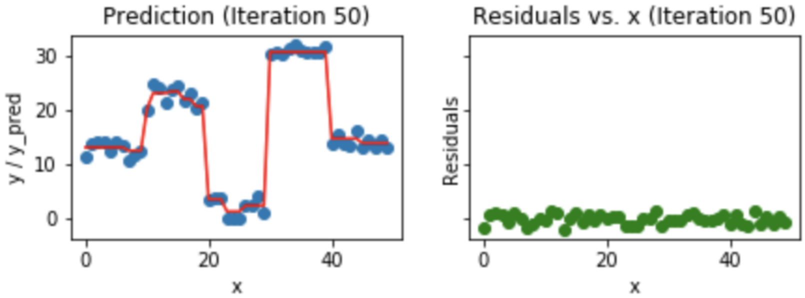

Figure 2.12: Visualization of Gradient Boosting Predictions After 50 Iteration

2.6 Related Works

We have reviewed many research papers and discovered many techniques for determining depression. The approaches include data mining combined with machine learning, scaling techniques, and others. We discovered interesting facts about this research subject by examining their research techniques.

In this research, researchers utilized machine learning classification to identify depression in the study [21]. This study included almost 935 students. To begin, researchers conducted a poll of university students and gathered data. Their survey consisted of four sections, the last two of which were Beck and Bangla scaling. They did not utilize all the data; instead, they cleaned and retained 577 records to get more precise findings. They employed a total of six methods to determine the accuracy, precision, recall, specificity, and f-measure of their original 20-feature prediction model. Using the deep learning model, they obtained a micro accuracy of 71.75%, indicating that 217 students were sad while the actual number was 132, indicated as genuine positive. The researchers then performed a Generalized Linear Model (GLM) and obtained an accuracy of 74.17%, with 49 students out of 367 receiving false positives. The accuracy rate was about 72% when using the Gradient Boosted Algorithm (GBA), with 83 out of 367 false positives. Following that, they achieved 67%, 75%, and 73% accuracy using K-NN, Random Forest, and Support Vector Machine (SVM) methods, respectively. They obtained a 72%accuracy for Deep Learning, a 75% accuracy for GLM, a 73% accuracy for GBA, a 66% accuracy for K-NN, a 73% accuracy for the random forest, and a 74% accuracy for SVM after running them with 15 optimum features. Thus, they concluded that deep learning and GBA algorithms were the best because of their high f-measure percentages. According to the research, it would benefit students, counselors, and the institution.

In a research study [7], researchers have used machine learning to predict possible depression in people in the future if they are not already suffering from it. They used synthetic data prepared with the help of a Java program to carry out the research. The training sample consisted of 600 instances whereas the test set had 400 instances of data. Out of the 600 instances, 555 were correctly classified giving an accuracy of 92.5 percent. Total of 10 out of 20 classes were predicted with a probability of 1 meaning that the prediction was absolutely correct. The smallest probabilities of predicting a YES and NO class were 0.829 and 0.778 respectively clearly illustrating that the predictions were accurate. The study mentions that the probability distributions could be helpful to physicians and could be used in other medical applications of data mining.

Another study [35] used a sizable textual dataset to predict depression symptoms. Two datasets from ung.no were used in this study consisting of the people's questions on their depressed lives and expert responses. The initial dataset yielded about 11807 messages, of which 1820 were classified as depressive text. When it came to detecting emotional states, they used an RNN-based method and had a success rate of 98 percent. When they utilized Decision Tree (DT) as a second method, they discovered that 84.2 percent of symptoms could be accurately recognized, with 1684 of 1807 depression texts properly categorized and 1730 non-depression texts correctly classified among 10000 non-depression texts. The second dataset included 21470 text samples, 1470 of which were categorized as depression text and the remainder as non-depression text. They were able to achieve a 99 percent mean accuracy using their technique. Text-based depression detection may be used by mental health services to evaluate and predict normal and severe states of mood disorders in real-time, using cutting-edge technology and intelligent surroundings.

A research [11] found that an automated method might identify depression in the elderly. The goal was to eliminate the errors that may occur with manual diagnosis and begin treating patients sooner rather than later. The Geriatric Depression Scale was used to interview a total of 60 older adults, all of whom were at least 60 years old (GDS). Classifiers such as BayesNet (BN), Logistic (LOG), Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), and Decision Table were evaluated on four metrics: Accuracy, ROC area, Precision, and Root Mean Square Error (RMSE). The data set is divided into training and test halves based on user-specified percentages. There will be another epoch after that, which will utilize test data generated from the previous 80 samples, with the 20 test samples now being used as training data. The top prediction model was SMO, which had a precision of 93.33% and an accuracy of 0.94%. SMO was followed closely by BN, which had a higher ROC Area and RMSE. In order to improve the accuracy of depression prediction, several nature-inspired algorithm-based optimization methods may be used. Additionally, the dataset may be constructed utilizing data gathered from other regions of the nation and even other countries if feasible in order to check for consistency in the findings. This study powered with data mining and machine learning might help better predict and understand depression in Bangladeshi university students.

A research [9] shows, people who are depressed have profound and repeated thoughts, which they communicate on Twitter and in real life. People diagnosed with depression were studied by looking at their tweets over one week, which served as an indicator. Suicidal thoughts, such as those related to sleep deprivation, suffering, and death, were considered. Research on depression-related rumination-using words or word combinations led researchers to seek two distinct types of rumination in the Twitter histories of all users. The depressed/study group had more Tweets regarding sleep, pain, and suicidal thoughts than the control group, according to ”Fisher's exact test,” an independent ”t-test,” and a ”Chi-square test.” They do not tell us whose universities, regions, or countries these Twitter users came from. On the other hand, this research may be helpful in future studies that we want to do.

Chapter 3

Research Methodologies

As our work has a large timeline, systematic processes, side tasks and specific sequence we have to divide our research work in several specific sections. Each section has several tasks and sub tasks. During the one year of our research project we had to work remotely as well as offline to have meetings for planning, gathering data, consulting with experts and individuals. We also had to work collaboratively in different computer research labs to train and test the data set online.

Our work plan is divided into five main divisions and a few sub-divisions. We also sometimes worked parallelly to collect, analyze and use those data at the same time.

The probable work plan consists of 5 parts. They are:

1. Data Collection
2. Analyzing Data Manually
3. Prepare Data
4. Training and Testing Machine Learning Models
5. Applying the Best Model and Distributing Suggestion
6. Further Research and Tool Development

3.1 Data Collection

3.1.1 Survey Form

We made a survey form for collecting data from the students. We thought getting data directly from the mass students would help us create an overall map for different mental states of a Bangladeshi University-going student. As mentioned in the problem statement, we had to create the survey questions in such a way that:

1. The questions are not very complex to understand
2. They are not way too personal to answer.
3. Should have enough questions so that we can map the actual data
4. The questions should not belong to the answer.
5. Finally, the questions have to maintain a balance between long and short length and somewhat personal related to depression and not too much personal.

Based on the above criteria, it can be said that creating the questions was not easy. As a result, we needed to consult prior research [21] in relevant areas. The questions are constructed in such a manner that they have assisted us in gathering comprehensive data from students about the underlying cause of their feelings, the methods they used to try to feel better, and the potential methods of overcoming depression if they did so. It was emphasized that no personal information about the students will be collected, ensuring that their confidentiality will be respected and that they will be happy to assist us.

The form is divided into various sections that identify a responder as a volunteer or contributor. A person can either share their life activities and personal information or ask for suggestions on overcoming depression, or they can share their or her experience or provide some suggestion to assist others in overcoming depression. In the first section, they were asked to provide information on their mental, physical, academic, career, and romantic status. In the second section, they were asked whether they have overcome depression or not. If they have accomplished this, they were asked to describe the actions that enabled them to do so. And, if they do not overcome depression, they were asked if they want to get suggestions or ideas to overcome it or not, which had included their volunteering for this project.

Additionally, we took certain extra precautions to protect the participants' anonymity. Since there is no way to monitor a person's Google form answer without knowing their id, name, or other identifying information, we devised a strategy based on a random code generator. Each form response was to produce a unique code at random. The volunteer and we both have access to the code. Because we will not request the code from volunteers, there will be no way to identify or detect the volunteer who filled out the form. Nonetheless, implementing this was not straightforward due to the absence of an add-on, extension, or plug-in that allows embedding a random code generator in a form. To accomplish our goal, we needed to create a GCP account, apply for access to their Google Form API, then study and write the generating program in Google's app script language.

3.1.2 Consult with Therapists, Psychologists

Instead of only relying on the information provided by the students, we have also consulted with experts in the fields of psychology and mental health to enhance our understanding of human emotions and psychology in general. Options include the BRAC University Counseling Unit, nonprofits such as ”Kan Pete Roi,” and so on. Professional counselors and therapists are at BRAC University's Counselling Unit to assist students through difficult times. Professor Mehtab Khanam, Ph.D., a world-renowned psychologist, is an advisor of BRAC University's Counseling Unit. Our team reasoned that since they are constantly dealing with issues like depression, they may have some insight, statistics, and recommendations on the subject. The online treatment program ”Kan Pete Roi” is widely used in Bangladesh. It is reasonable to assume that we have gotten comparable data from them.

3.1.3 Getting Opinions from Students

We have also interviewed some of the volunteering students to get some ideas and suggestions from them regarding this matter. We think our work would not be completed without getting some actual suggestions from this first-party itself. However, we will not disclose their identities publicly if they do not want to.

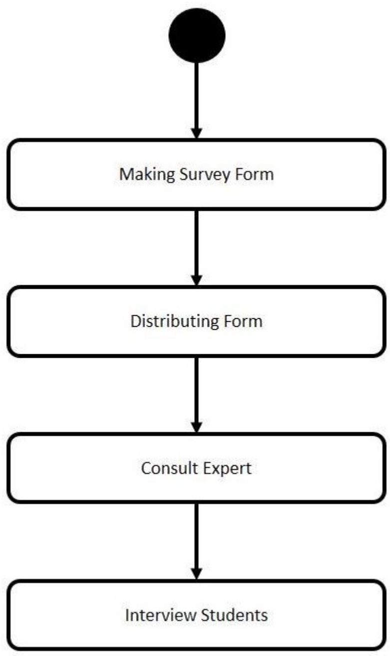

Figure 3.1: Data Collection Workflow

3.2 Analyzing Data Manually

3.2.1 Collect Survey Form Data

Thanks to Google Form, collecting survey form data was generally straightforward. We used embedded Google Sheet Response with the Google Form Survey. Using

Google sheet, it becomes easier to export data and analyze it. We had to track each of the volunteers, so we had to follow an approach where the database is live but can always be analyzed, optimized, and manipulated without destroying the live process. One of the main features of Google Sheets is, using Google Cloud Platform (GCP), the data can be exported on life using API. As we have to export the dataset to python in the following steps, there were absolutely no other valid options without using the built-in Google Sheet Response. Also, have not and will not share any data collected from the forms with anyone to maintain the privacy of the volunteers.

3.2.2 Extract Survey Data to Python and Use Data Science Tools (NumPy/Scikit-learn/Pandas, etc.) to Analyze

For exporting data on living, we used Google Sheet API to collect form responses. For this step, creating a Google Cloud Platform account was necessary to get API access permissions. Then we used Jupyter Library's python environment and extracted those responses from the Google Sheet. The most popular Data Science library has analyzed the data collected from the response.

3.2.3 Visualize Survey Data

Raw data is not always helpful for understanding the connection between different data, which is most needed in our case. Data Visualization helps to understand the connections so swiftly. Python's one of the most popular library pandas make it easy to visualize the massive amount of collected response data. This phase was included making connections between the answers to the questions and the students' depression states to get a basic understanding of what causes depression, how a certain number of students are coping with it, Etc. We have used visualization to help us make these connections. We have made several assumptions based on that understanding.

3.2.4 Examine Data Commonalities With Therapist Interviews and Student Interviews

After making the assumptions in the previous step, we have tried to find the similarities between our assumptions based on visualized form response and the interview data from students and professional psychiatrists and therapists. It has helped us to understand the data more clearly, and our assumptions are now more accurate.

3.2.5 Make Some Logical and Possible Assumptions of Root Causes, Possible Fixes and Suggestions

We have created a clear representation by basing it on the data that we have gathered from all of the sources. Before moving on to the next stage, we have made a few concluding assumptions. These have provided guidance for determining the parameters and input data frame that was used to train the machine learning models.

Figure 3.2 indicates analyzing data manually.

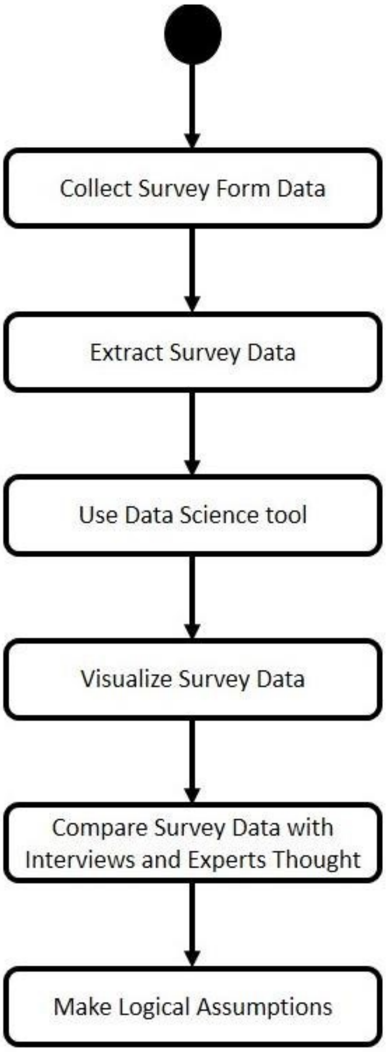

Figure 3.2: Analyze Data Manually

3.3 Prepare Data

As mentioned in the Research Objective Section. At first we have trained our machine learning models for detecting depression and giving suggestions. Then our aim was to test the models if they are accurate enough to detect and give suggestions.

In order to do that, we have separated our whole data set into two parts for training and testing. This has helped us to test our machine learning model before predicting suggestions in real life. There are 3 main steps in this section.

3.3.1 Rectify Data

We have clean up not-impactful parameters based on our previous assumptions and test it out while training the model.

3.3.2 Data Classification

As our research objective is not only to detect depression but also to give suggestions we had make our Google form data set prepared for two different training models.

1. **Detect:** This data set was used in detecting depression training models.
2. **Yes:** This data set was used for suggesting workarounds in training models. This data set is consists of the collected form responses of the students who have marked "Yes" in the "Did you overcome depression" section of the survey form.
3. **No:** This data set consists of the collected form responses of the students who have marked "No" in the "Did you overcome depression" section of the survey form. Our trained and tested models will give suggestions to these respondents as cases in the final section.

3.3.3 Data Separation for Training and Testing

Both the "Yes" and "Detect" dataset will be further separated in order to prepare them for training and testing in a separate order. We used the suffix A in the dataset's name in order to classify them. For example:

1. **DetectA**: This dataset was used for training machine learning models for detecting depression.
2. **DetectB**: This dataset was used for testing our previously trained machine learning models with DetectA dataset for detecting depression. We have got the accuracy, precision, recall and f1-score of the models here.
3. **YesSuggestA**: This dataset was used for training machine learning models for suggesting workarounds to overcome depression.
4. **YesSuggestB**: Similar to DetectB, Previously trained Machine Learning models with YesSuggestA dataset was applied to this dataset to test their accuracy, recall and precision.

5. **Optimize Data**: The data will be further optimized for machine learning training and testing if needed.

The accuracy of the model that was evaluated was significantly impacted as a result of this factor.

Prepare Data diagram showed in Figure 3.3.

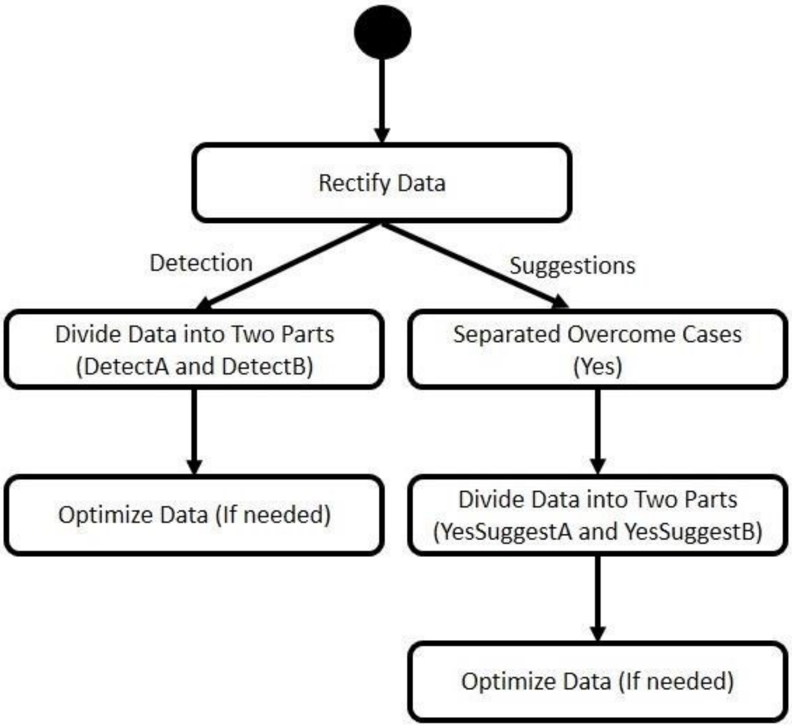

Figure 3.3: Prepare Data

3.4 Training-Testing Machine Learning and Deep Learning Models

3.4.1 Build Machine Learning Models

We programmed some most popular machine learning models for this research project. Previously implemented Python on Anaconda Server was used for programming these models. We have used the renowned scikit library of python to implement the machine learning models easily. Here are the classifiers we have used for this work.

1. Support Virtual Machine (SVM)

2. K-Nearest Neighbor (K-NN)

3. Gaussian Naive Bayes (GNB)
4. Decision Tree (DT)
5. Random Forest (RF)
6. Artificial Neural Network (ANN)
7. Gradient Boosting (GB)

3.4.2 Train Machine Learning and Deep Learning Models

We have followed these rules for testing the accuracy of our trained models:

1. DetectB and YesSuggestB data sets was used for testing.
2. We have calculated the accuracy (detection rate), recall, precision, f1-score of the models and store them for future use.
3. We have changed the model parameters and manipulate the data sets' parameters to achieve more accuracy until reaching an optimal end.
4. If the model accuracy does not meet a sufficiently high score, we have repeated the previous (Train Machine Learning Models) process and manipulate the data set and model parameters.

3.4.3 Test Machine Learning and Deep Learning Models

Strategy that we used to the testing:

1. DetectB and YesSuggestB data sets are used for testing.
2. We have calculated the accuracy (detection rate), recall, precision, f1-score of the models and store them for future use.
3. We have changed the model parameters and manipulate the datasets' parameters to achieve more accuracy until reaching an optimal end.
4. If the model accuracy does not meet a sufficiently high score, we have repeated the previous (Train Machine Learning Models) process and manipulate the data set and model parameters.

3.4.4 Determine Best Machine Learning Classifier

After calculating the accuracy, precision stats, we have determined the best classifier based on the highest accuracy. These parameters are used in suggesting ways to improve mental health to the volunteers later.

3.4.5 Enhance and Optimize Models

In order to improve our machine learning mode, we are analyzing and experiment with a wide variety of machine learning tools, models/classifiers, and deep learning algorithms, in addition to Neural Networks and Artificial Neural Networks. The Training and Testing Machine Learning Models are shown here in Figure 3.4.

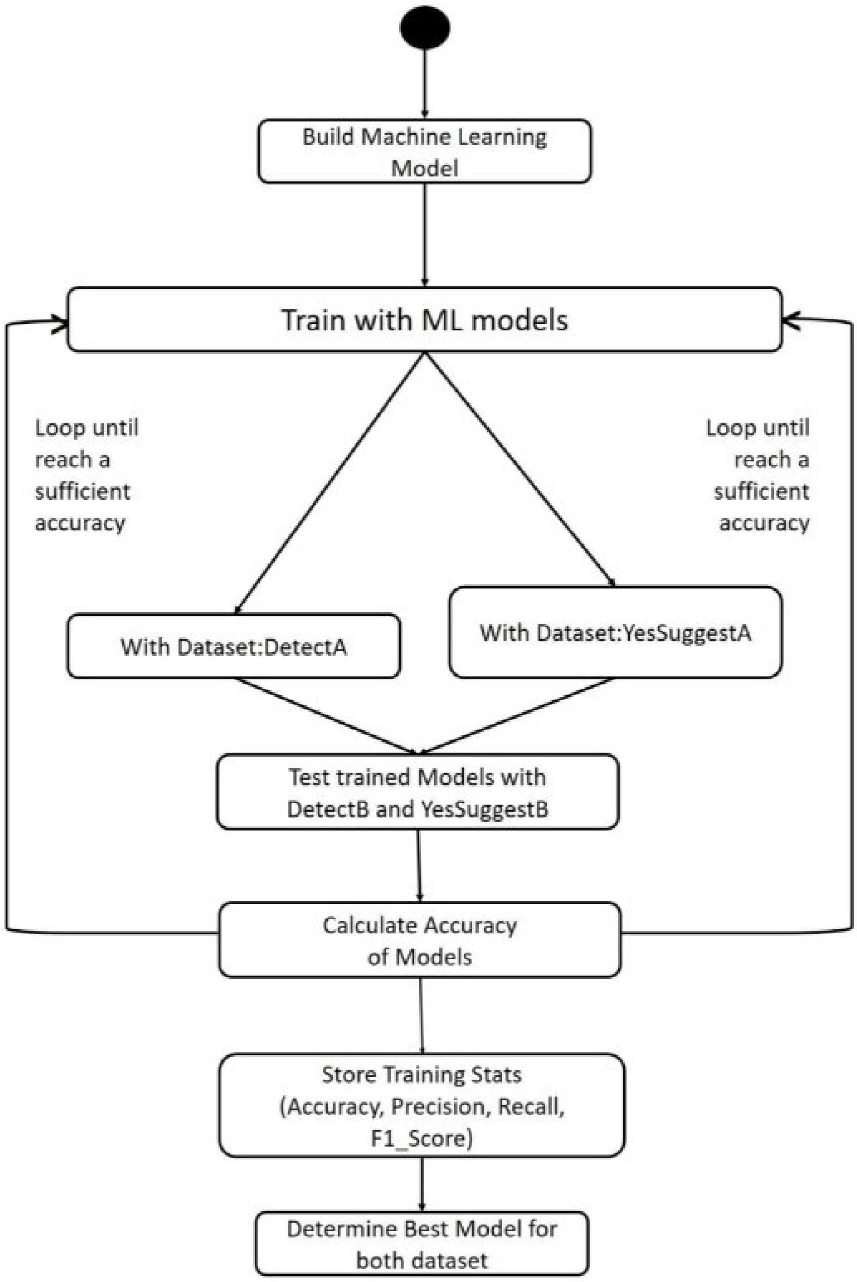

Figure 3.4: Training and Testing Models

3.5 Applying the Best Model and Distributing the Suggestions

We have divided this section into two sub-section. Where we have included the best model application to the 'No' data set and contact the volunteers and distribute the suggestions.

3.5.1 Applying the Most Accurate Model to "No" Dataset

We have applied our most accurate model to the "No" Dataset without any modifications. We have used the same and optimized parameters and data columns used in the previous testing model. It is already proven with high accuracy so changing parameters might hamper the detection rate. We have stored the prediction which is the suggestions.

3.5.2 Contact the Volunteers and Distribute the Suggestions

We will contact the volunteers who send us mail in order to give them the suggestions. We are planning to manually follow their activity frequently by contacting them in future.

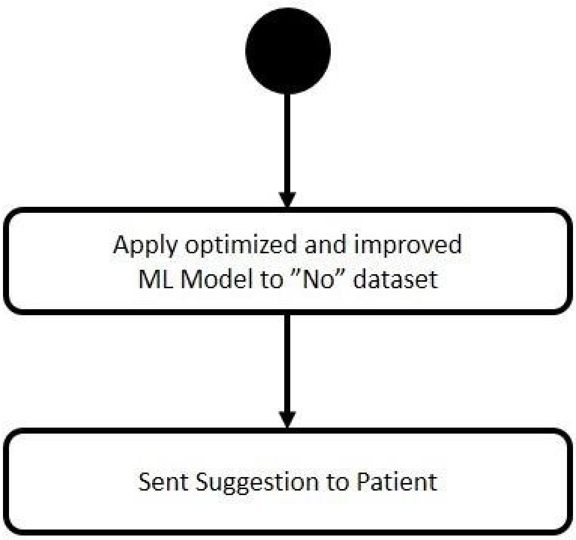

Figure 3.5: Applying Best Model and Distributing Suggestions

3.6 Further Research and Tools Development

We will distribute the feedback form to the volunteers who followed our previously distributed suggestions made by the machine learning models. In the feedback form, the volunteers will tell if they felt better following the suggestions or not. They will also give us some feedback for our project. We will use the feedback data for researching further and may build a user-accessible tool if the feedback is mostly positive.

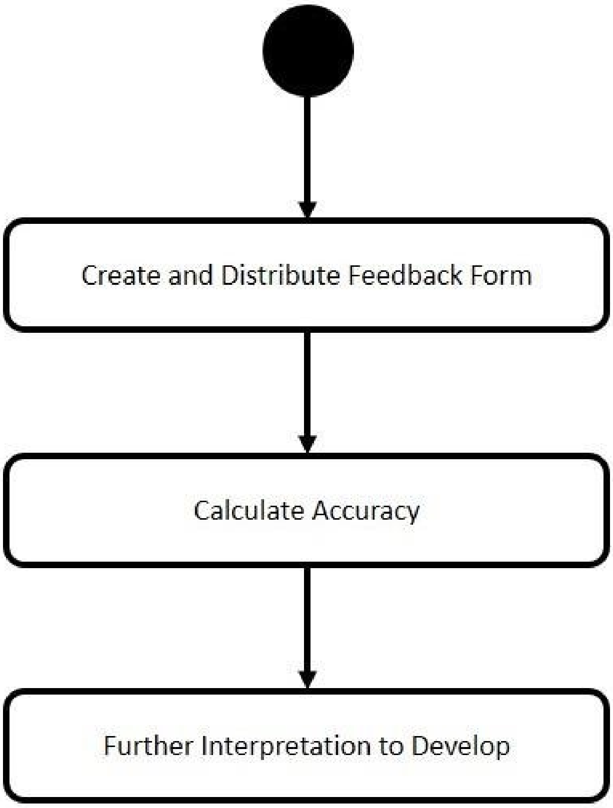

Figure 3.6: Further Research and Tools Development

Chapter 4

Data Collection and Pre-processing

This section details the data collection procedure, including how our questionnaire was constructed. Then, we'll detail the pre-processing steps we've taken with the data we've gathered so far. Please note that we marked this as primary data collection since we are still in the process.

4.1 Interview and Making Questionnaire

We read the relevant number of publications and research papers to create questionnaires for this study. Aside from that, we have spoken with a representative from "Kan Pete Roi." This organization operates over the phone and provides emotional assistance to those suffering from depressive disorders. Unfortunately, we cannot reveal the identity of the individual with whom we have spoken because of their policy. According to the counselor, they get a significant number of calls from young students from all around the nation. More on to that, we asked several questions and engaged in a thorough conversation with that individual. Following that, we began distributing our survey on the 2nd of September, 2021, and have gathered 829 responses till 18th of May, 2022.

As an additional precautionary measure, none of the students were forced to complete this survey. We made sure that their identities stayed hidden and that there was no way for us to track them down. However, we did open the possibility of individuals volunteering their time to us for additional study. As a result, we provided them with an automatically created private key and did not ask for their name. We distributed our survey form to several groups inside the BRAC University student community and club groups in the first instance. Later on, with the cooperation of our friends from different universities (both public and private), we are disseminating the form. In our Facebook post, we made a few observations, which are as follows:

1. We ask that students fill out the form carefully and truthfully.
2. We tried multiple times and observed that it takes between 10 and 20 minutes to complete the form properly.
3. It is not acceptable for the same individual to fill out the form twice.
4. They have the right to quit or opt-out of the survey at any moment if they feel uncomfortable doing so.

4.2 Describing Dataset

As part of our study, we asked them similar questions to measure their psychological well-being. There are a total of 36 aspects in which we have inquired about their age, gender, CGPA (Cumulative Grade Point Average), and the kind of varsity they attend as the first set of queries. Then we ask more pertinent questions regarding depression that are more recent. Those questions were developed in accordance with the methods described before; however, we looked to Beck's Depression Inventory for the questions and integrated those into our own. The most probable reasons for a person's depression were the subject of 18 questions that were asked to them.

4.2.1 Most Common Causes of Depression

We conducted this study on university-going students to understand that the majority are young. As a result, we sought to determine what causes depression. According to research [20], socioeconomic status has a critical impact on depression. The same research discovered that academic pressure is also a factor. They feel under strain as a result of the many exams and assignments. They have no other choice since they must maintain a certain grade level. Another research [5] discovered that familial issues, a family history of depression, and drug addiction might contribute to depression.

Besides these reasons, we get to know more reasons from 'Kan Pete Roi.' We have focused on the problems that they have brought out to us.

Student's Academic Condition

According to a study [23], if individuals cannot attain their full potential in education, they will have difficulty obtaining excellent prospects; this way of thinking will eventually lead to mental illness. Not only that, as we learn from several research studies, students are overwhelmed by academic pressure. According to a survey [8], 63.5 percent of students reported experiencing stress due to academic pressure.

Problems in Relationship

Nowadays, we find that most of the young population enters into romantic partnerships. According to an article [31], individuals might feel deceived or mistreated at times, leading to feelings of despair. Not only that, but some individuals find it difficult to maintain long-distance relationships. In addition, the dissolution of a relationship might be a contributing factor to depression. A study [14] result revealed that 60 percent of people might suffer from depression due to relationship troubles, 30 percent experiencing severe depression.

Family Violence

Family violence is a major contributor to debilitating mental health concerns. According to Ronald C. Kessler and William J. Magee, childhood exposure to family violence has a long-lasting effect on clinically significant recurrent depression in adulthood [1]. It's unsurprising that the majority of clinically depressed people had an atypical familial relationship and were frequently subjected to physical or mental violence.

Sexual Harassment or Abuse

Sexual harassment has been identified as a cause of stress that has a negative impact on the physical and mental health of individuals who are subjected to it. Despite the fact that social scientists have shown a causal link between sexual harassment and poor mental health. According to a University of Pittsburgh study, sexually assaulted victims had about four times the risk of having high blood pressure and slightly more than twice the risk of having poor sleep quality, as well as signs of worry and depression, compared to those who had not been harassed. Another study discovered that four out of every five adolescent females who are sexually abused experience post-traumatic stress disorder in the months following the assault [17].

Victim of Bullying

In terms of sadness and anxiety, bully-victims may be the most vulnerable group. The study [3] discovered that pupils who had been bullied in recent months were more likely to express depressive symptoms than non-bullied kids.

Family History of Mental Illness

According to the National Institutes of Health and Medicine in the United States [19], a family history of mental health problems is closely associated with depression in the younger generation. According to their findings, there is a genetic difference between MDD individuals with a favorable family history of depression and sporadic cases.

Not Having Friends

One of the primary reasons for depression is often a person's lack of social support, particularly friendships. Even while a lack of social support is one of the root reasons of depression, overcoming the condition may be facilitated by cultivating a few reliable and trustworthy close friendships. A significant amount of stress, which is strongly tied to mental health, may be reduced by sharing personal difficulties and thoughts with another individual.

Searching for passion

Passion plays a vital role in a mental health. Based on our learning from the related research, finding passion or not has one of the most influences on the mental health related issues. According to psychologist Vicky Bellman, passion for something creates habits, consistency and inspire us to reach a goal [10]. It thrives us to keep up with the hope. So, it can be said that, it can have effects on mental health issues in a positive way so, we included it in our survey questions set.

Drug Addict

A report from National Bureau of Economic Research shows that about 84 percent of the people who consumes cocaine has been gone through severe depression phase and needed to take medical help. Not only cocaine, similar kind of drugs and over consumed alcohol can have effects or can be the effect of depression and other mental health related issues.

Economical Condition

Having a large amount of personal and/or family debt can hamper a child's mental health. Study shows that only 10% increase in short term debt can be a cause of alarming 24% mental health. [13]

4.3 Data Cleaning

Following a significant amount of time, we have finally collected a total of 829 responses from students attending a variety of Bangladeshi educational institutions. Using that in mind, we now go to the experimentation stage with the data. We looked through all of the data, found any possible outliers, and then deleted them from our data set before you go and apply the algorithms to assess the correctness. The procedure was broken down into 2 stages. Following the completion of two stages of cleaning, we are now in possession of 750 data rows.

During the first stage of the cleaning process, we saw that a significant number of the kids were able to correctly identify themselves as being either high school or college students. Due to the fact that this study is focusing on undergraduate students in Bangladesh, we were required to delete such data as soon as possible. Next, we learned that some of the students are attending universities outside of Bangladesh for their education. We came to the conclusion that it would be best to exclude them as well given the possibility that the environment there is not the same as that of Bangladesh, which may have an effect on the reliability of the predictions. At the end of this stage, there were 750 data rows left for us. We made the decision to proceed to the pre-processing step with those data. When we started pre-processing with sklearn tool, first we generated a heatmap to determine the correlation between features. There we saw a strong a correlation between "Do you have a friend circle with whom you talk more often?" and "Do you have a best friend/ very close friend with whom you share most of the things?" So, we check the importance metric and decided to drop the the column 'Do you have a friend circle with whom you talk more often?" since we may face over fitting problems during the algorithm run. After that, we are now remained with 22 features for detecting depression among students. Since, the data were not in integer or float or binary from, we applied One Hot Encoding to and then train our data set with the help of K-fold cross validation. After that we moved forward to our algorithm applying phase.

4.4 Final Visualization of Data

It is necessary to visualize the data in order to comprehend it. This part is titled "Final Visualization of Data" since we are finished with the necessary quantity of data at this time. We anticipate collecting additional data in order to see more variety. However, we have discovered some correlations between the data we have obtained thus far. Depression status (Do you have depression or mental health-related issues?) where a participant can answer 'Yes,' 'No' or 'Maybe' is shown on the x-axis, while the number of cases is represented on the y-axis. Each question is represented by a bar that displays the options for that question. The bars in the histograms are of different color to indicate each type of answers. The color labels are annotated in the top corner of the figures. Let us analyze side by side comparison of some important features.

Question Asked About Passion

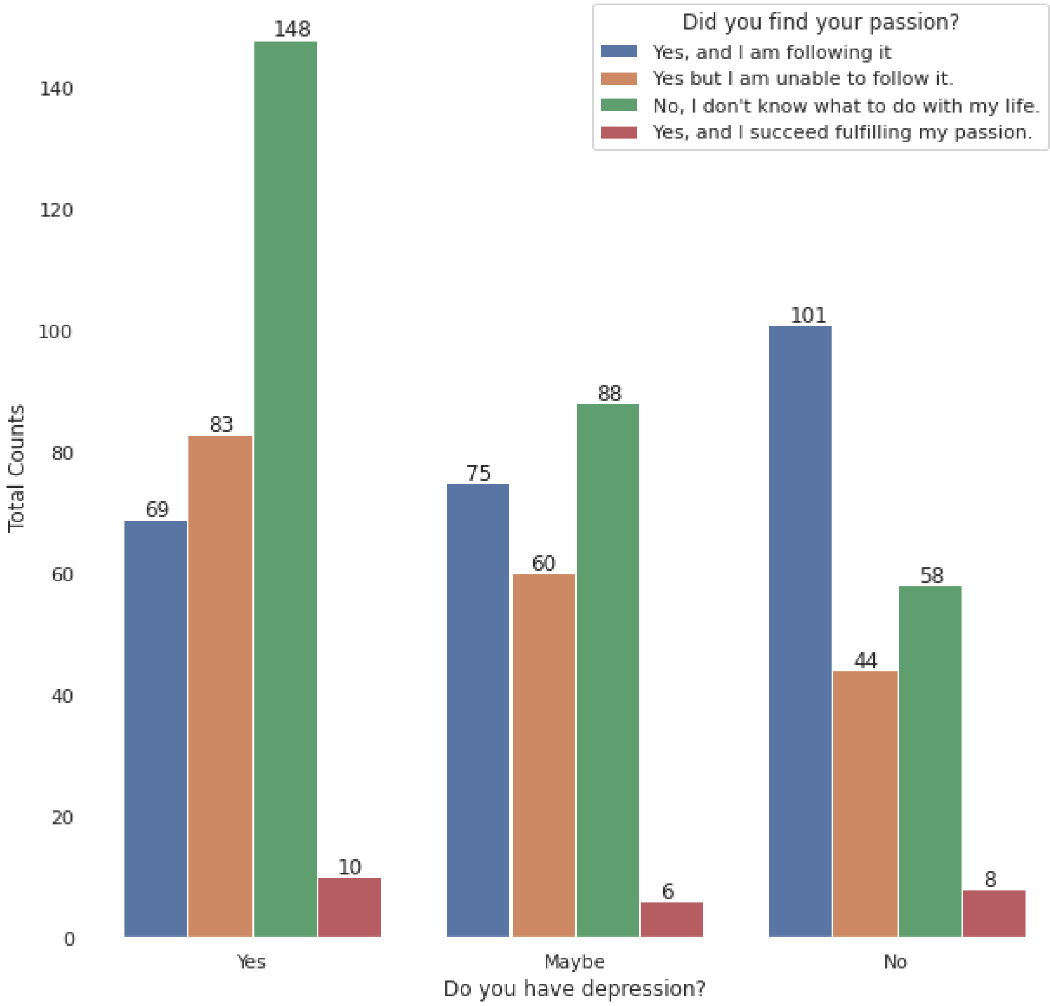

Figure 4.1: Depression and Passion Pursuing

The primary inquiry was, "Have you discovered your passion?" This question was prompted by the expert recommendations we've already stated. We hypothesized that persons who lacked a sense of passion could feel depressed as a result. Yes, they have a passion, and they are now pursuing it; yes, they have a passion, but they are unable to pursue it; yes, they have a passion, and they have succeeded in pursuing it; and no, they have no passion and have no idea what to do with their lives.

The majority of those who indicated they did not know what to do with their lives self-identified as depressed. 148 individuals out of 750 reported being sad, and 83 individuals unable to pursue their interest are also depressed. Still sad are 10 successful students and 69 students pursuing their passion. Out of 310 "Yes" replies, almost 75% of students do not discover passion or are unable to pursue passion.

There are 229 students who picked "Maybe" as a possible indication of depression. Unsurprisingly, the majority of 88 students do not know what to do with their lives, with 60 indicating they are unable to pursue their passion, 75 pursuing it, and 6 having achieved it.

221 individuals said that they are not depressed, with 101 pursuing their passion and 44 unable to do so. 58 individuals do not know what to do with their lives, whereas 8 individuals have successfully pursued their passion.

Private or Public University?

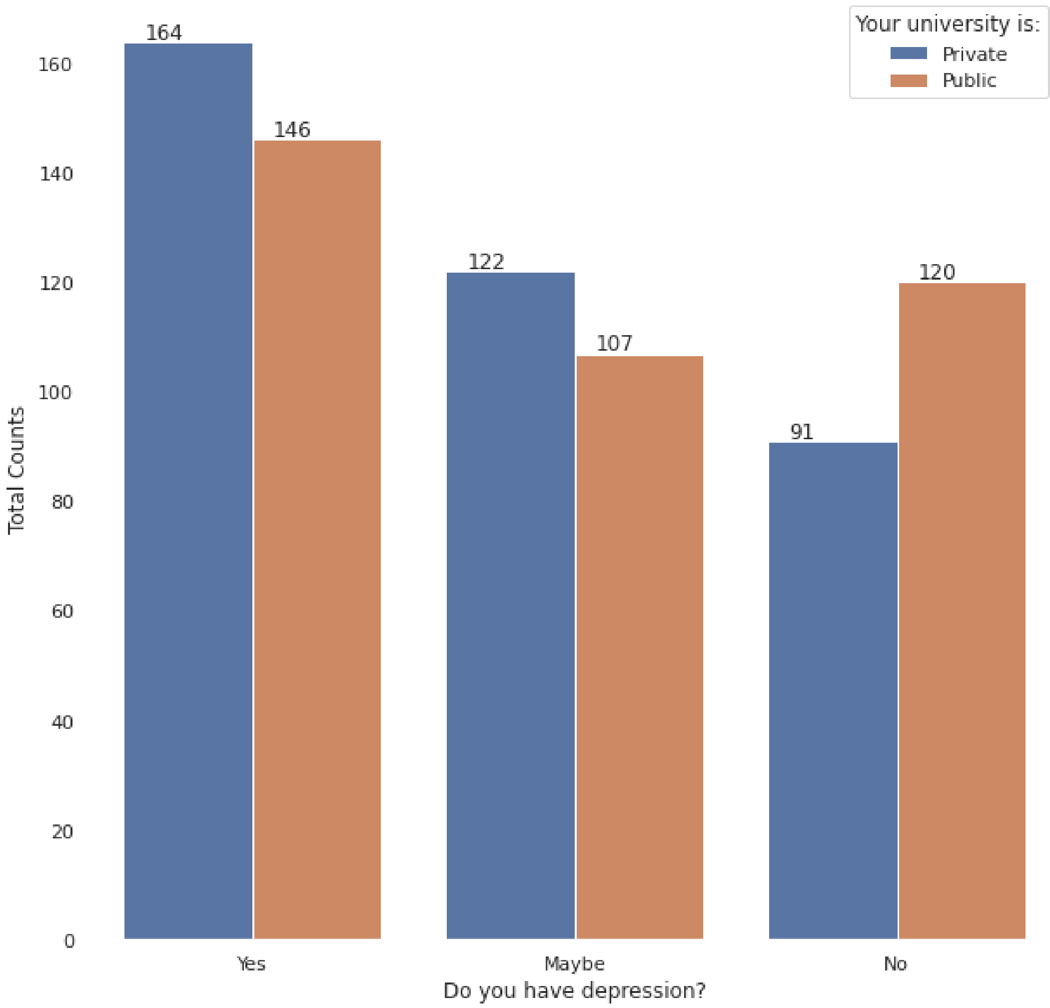

Figure 4.2: Depression and University Status

The ratio of students attending public universities to those attending private universities is about equal across all 750 data points that have been rectified. Among the 310 students who selected "Yes," we can observe that 164 students came from private universities and 146 students came from public universities. There are 229 students who are uncertain about their depression and tagged "Maybe" about it, of those 229 students, 122 are from private universities and 107 are from public universities. People who selected "No," there are 91 students from private universities and 120 students from public universities among those who selected "No." From what we can tell, the percentages are very comparable, but it seems that students attending public universities have less signs and symptoms of depression. This is a question that remains in the Google form since it is our experience that the majority of individuals in Bangladesh lean more toward attending public universities rather than private universities. For some students, attending a public institution has been a lifelong goal dating back to when they were kids. According to our research on Bangladeshi university students, this might be a contributing factor to depressive symptoms.

The Nature of Individuals

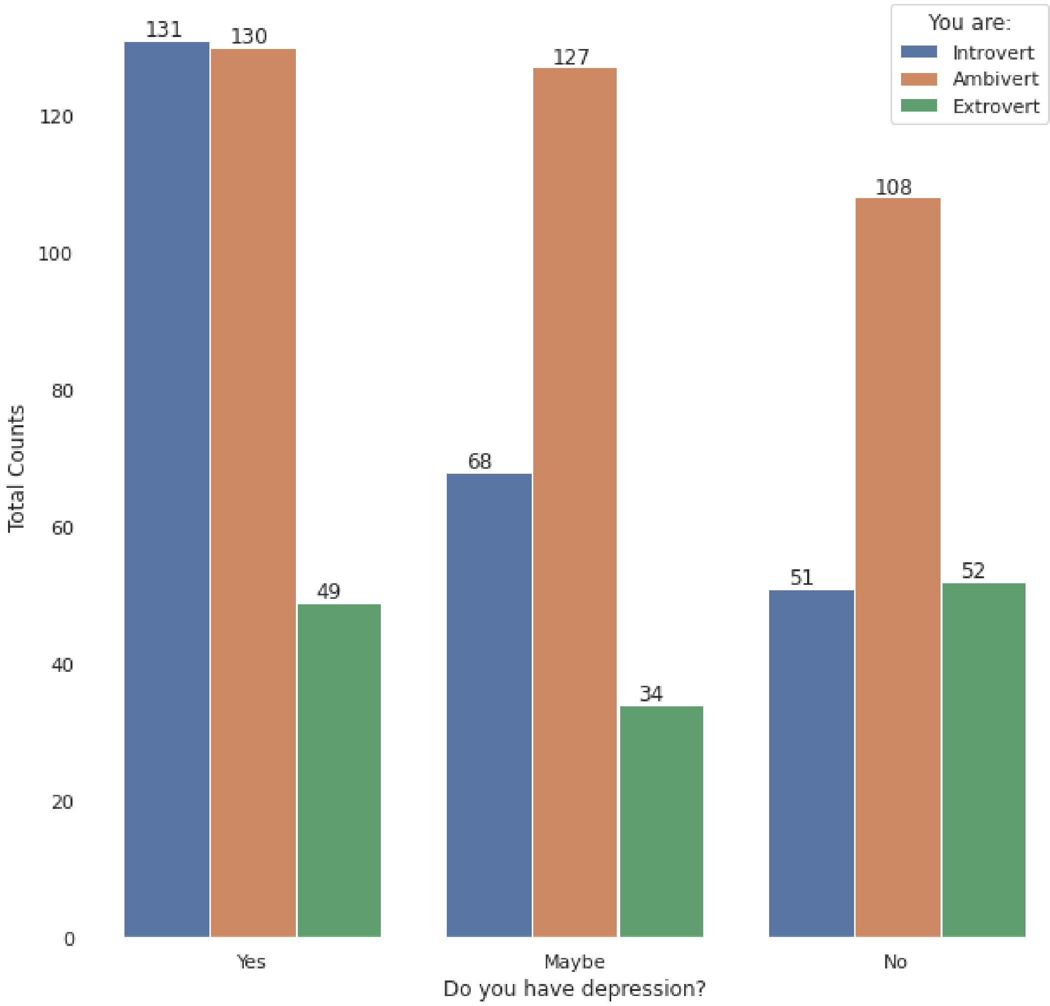

Figure 4.3: Depression and the Nature of Individuals

When designing this question, we used the assumption that persons who are more restrained in their social interactions are less likely to be open about the difficulties they face. Therefore, there is a possibility that the total will be far greater. Following the data analysis, we were able to see that the number of introvert students who are depressed is 131, while the number of ambivert students who are depressed is 130. Where only 49 out of 310 students who are extroverts suffer from depression. Students that checked ”Maybe” on the survey were broken down as follows: 68 students were introverts, 127 students were ambiverts, and 34 students were extroverts. There are 51 introverts, 108 ambiverts, and 52 extroverts among the students who selected ”No.”.

Economic Conditions

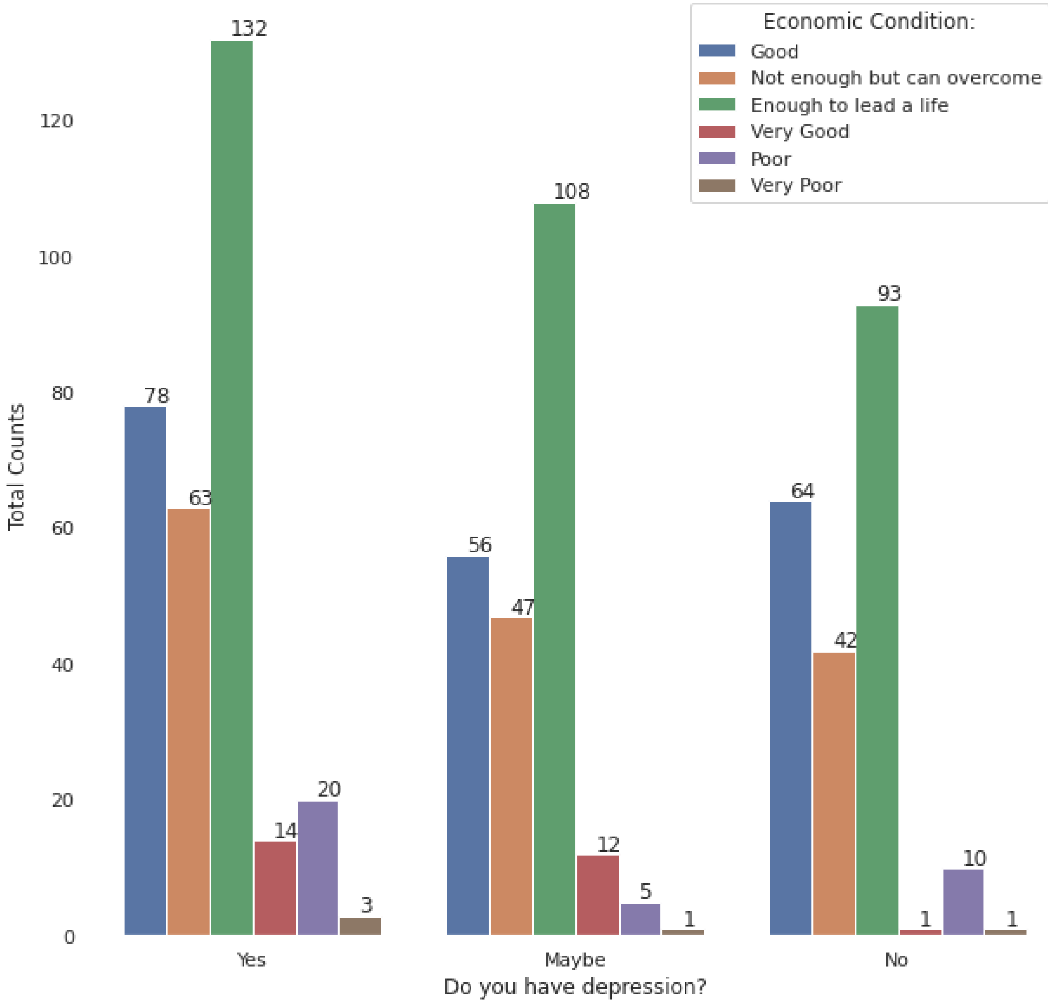

Figure 4.4: Depression and the Economic Conditions

The state of the economy has the potential to be a significant contributor to depressive symptoms. The visual representations showed that among the students who voted "Yes," 78 are leading good economic condition, 63 people are not in good economic condition but they think they can overcome that, 132 are somehow managing to lead life, 14 are in very good condition, 20 are poor, and 3 students are very poor. Among the students who voted "Yes," the majority of those who voted "Yes" are in good economic condition.

Among the students who checked "Maybe," 56 of them are in good condition, 47 of them claimed their condition was not good but they could overcome it, 108 people are getting by in some way, 12 students are in very good condition, 5 of them are in poor condition, and 1 student is in very poor condition.

Last but not least, among those who answered "No," 64 of them are in good economic condition, 42 of them are not in good economic condition but they believe they can overcome it, 93 of them are managing somehow, 1 of them is in very good economic condition, 10 of them are poor, and 1 of them is very poor.

Academic Conditions

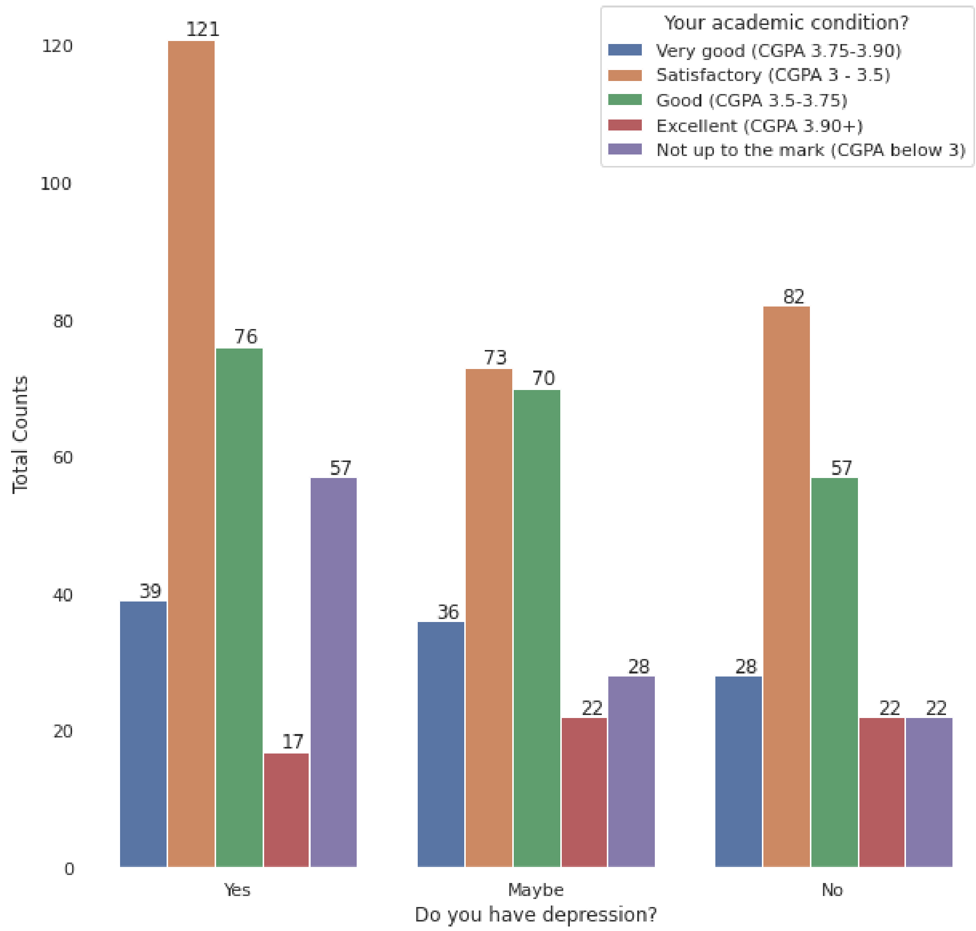

Figure 4.5: Depression and the Academic Conditions

The assumption that we made before was true in one more respect, and that was with regard to the association between the academic situations of persons and their mental health. We operated on the presumption that those who had better grades and had excellent results were statistically less likely to suffer from mental health problems. Based on our findings, we were able to determine that the figures corresponded with our anticipation in some domains; nevertheless, this does not tell the whole picture.

Around 53% of the 107 students who voted that they have a relatively low Cumulative Grade Point Average (CGPA) said that that they have depression. Around % said that they do not have depression, and 21% students was not sure about their mental health, proving that our assumption was correct.

We anticipated that students who had achieved higher grades would have a more positive mental state. The facts that we have obtained from our survey, on the other hand, suggest that the anticipation was not accurate. We have found that the majority of students who report feeling depressed have a somewhat better Cumulative Grade Point Average (CGPA) count and a good outcome. This is the case even when they report feeling depressed. Among the 276 students that were identified as having mental health issues, 121 of those students reported being happy with the result. Only 82 of the 276 students indicated that their results were adequate and that they did not have any mental health issues.

Relationship Status

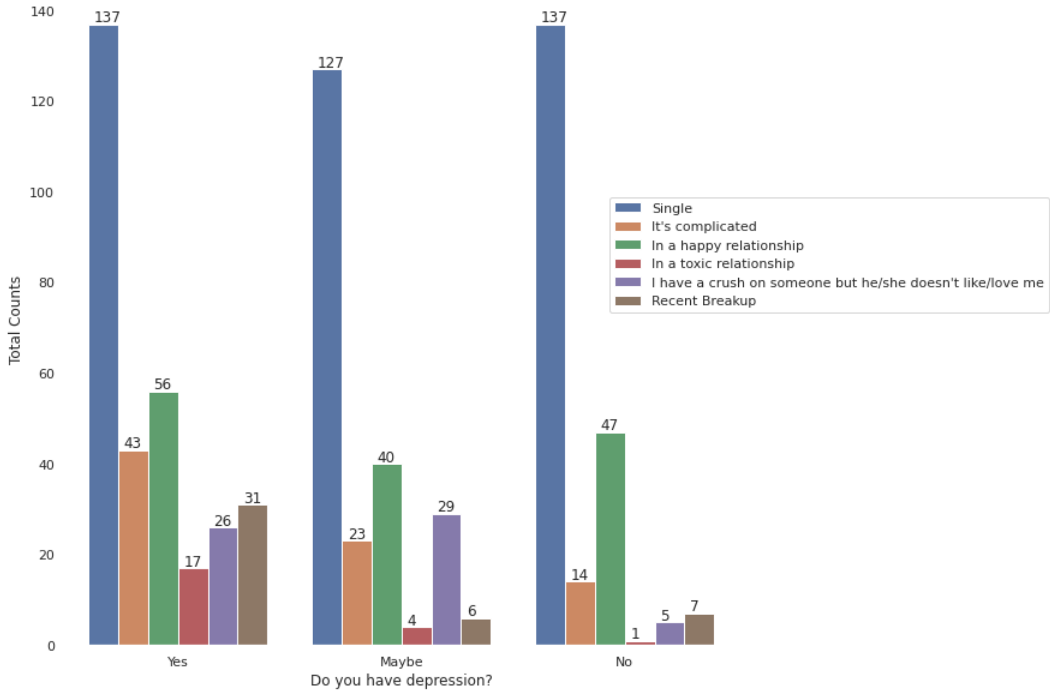

Figure 4.6: Depression and Relationship status

In terms of one's mental health, we have not yet discovered any major differences based on their relationship status. It can be shown that the majority of the students were single regardless of whether or not they suffered from depression. It is interesting to note that the number of students in the part for those with depression and the section for those who did not have depression both had precisely the same number of single students, which accounts for roughly three-quarters of the overall data.

An intriguing correlation that we have discovered is that some of the students who are content in their relationships have a greater propensity to suffer from depressive symptoms. Although, if we consider the data from verified instances of depression (56) and not confirmed cases of depression (40), they will have less mental health concerns in comparison to the other group (47).

As was to be predicted, students who had difficult relationships or relationship issues, and more especially those who were involved in unhealthy relationships, reported higher levels of depression. When compared to students who are happy in their relationships, we found that about 94–95% who are in toxic relationships tend to feel mentally sick. This is in contrast to students who are happy in their relationships.

Experience of Being Bullied

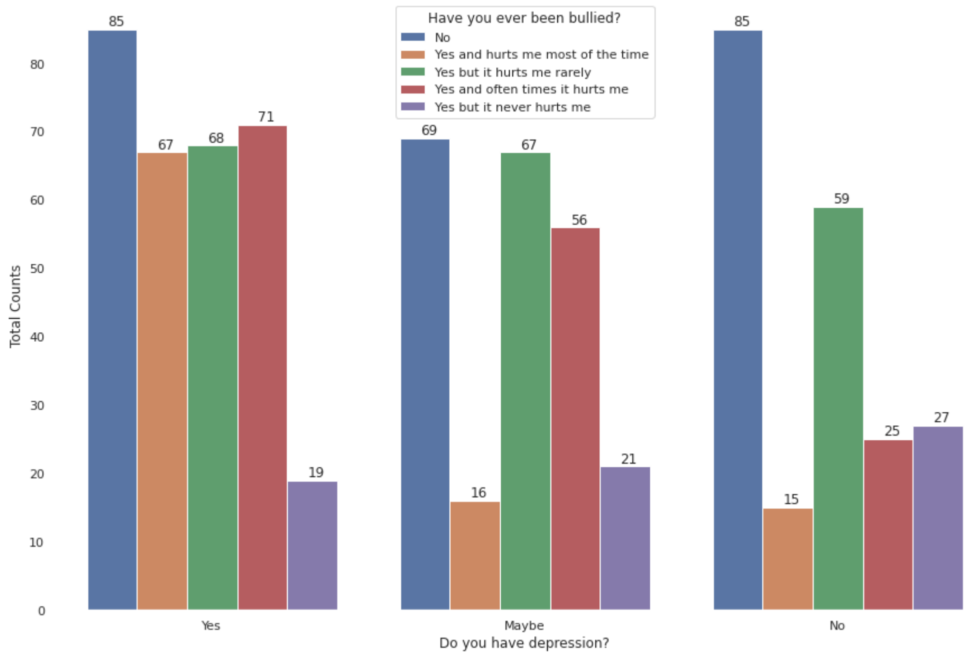

Figure 4.7: Depression and Bullying

It did not end here when we saw outcomes that were consistent with what we had anticipated happening. According to the statistics, 239 out of 750 students, which is 32%, had never been bullied before. Those who have been bullied are likely to have experienced feelings of pain often or the vast majority of the time. It is clear that recollections of having been bullied have a significant role in the development of depressive symptoms. Students who have been bullied in the past and whose memories continue to torment them usually experience increased levels of depression. A little over eighty-four percent of the pupils who can't forget the recollections of prior experiences of being bullied feel depressed, whereas just sixteen percent do not feel any depression while remembering the negative memories extremely often or even most of the time.

On the other hand, things become better if the students figure out how to cope with unpleasant memories from the past or with bullying. It has been observed that at the age of 19, those who undoubtedly experience depression are able to forget or cope extraordinarily well with bullying in the past, to the point where it no longer affects them. In contrast, the number of persons who are certain that they have a healthy mental state is 27, which is the complete opposite of the scenario that was shown before.

Sexual Abuse or Harassment

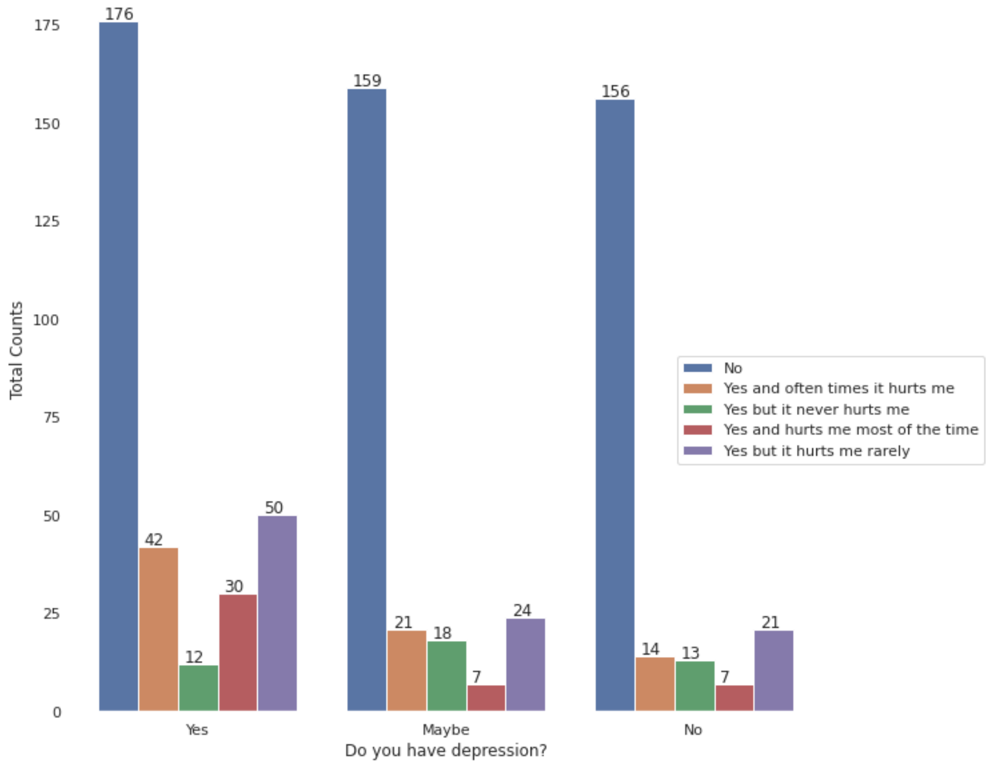

Figure 4.8: Depression and Sexual Abuse or Harassment

It is good to learn that the majority of the students who participated in this survey did not experience any kind of sexual assault or harassment. However, the number of students who had the nightmare was not much lower. Among the students who have reported being depressed, 134 have indicated that they have been subjected to sexual abuse or harassment. Whereas 21 students said that it hurts them often, 18 students stated that it never hurts them, 24 students stated that it very seldom harms them, and 7 students stated that it hurts them the most of the time. 176 students have indicated that they are depressed, yet they have never had problems with sexual harassment.

People who checked “Maybe” on the survey, which included 159 students, do not end up being sexually assaulted or harassed. There were a total of 70 individuals who were subjected to some kind of sexual abuse or harassment. In this group of 70 students, 21 of them are hurt on a regular basis by the incidences, whereas 18 of them are never affected by them. 7 of the students said that they experience pain most of the time, compared to 24 who stated they only feel it rarely.

Last but not least, of the students who responded in the negative that they do not suffer from depression, 156 of them had never been sexually assaulted or harassed. 14 students were mistreated, and this causes them pain on a regular basis. Thirteen of the participants said that it never made them sick. It is a nightmare for 7 pupils, and they are the ones who suffer the most from it. 21 of the students responded that it just sometimes hurts them.

Having a Close friend

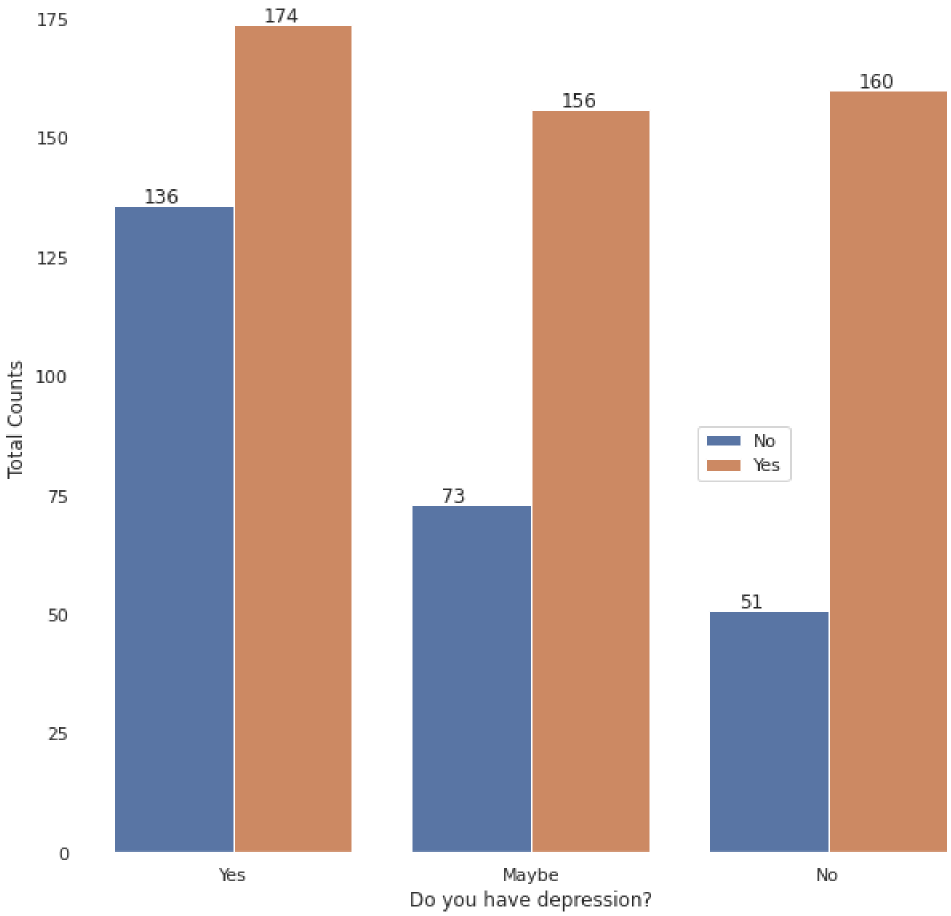

Figure 4.9: Depression and Having a Close Friend

We have asked them the question about if they have a close friend with whom they share most of things. According to the findings of our study, almost 80% of the students who do not have a best friend or a close buddy with whom they can regularly converse have symptoms of depression. The number of individuals who have been diagnosed with depression is rather large. Persons who do not have this sort of buddy are more likely to self-diagnose themselves with depression, with 136 out of 260 people falling into this category. For the same group, the number of individuals who do not suffer from depression and do not have a closest friend is 51.

Therefore, it is reasonable to suppose that having a solid relationship in which one can share their own views, beliefs, and experiences may be of great use in warding off depression.

Domestic Violence

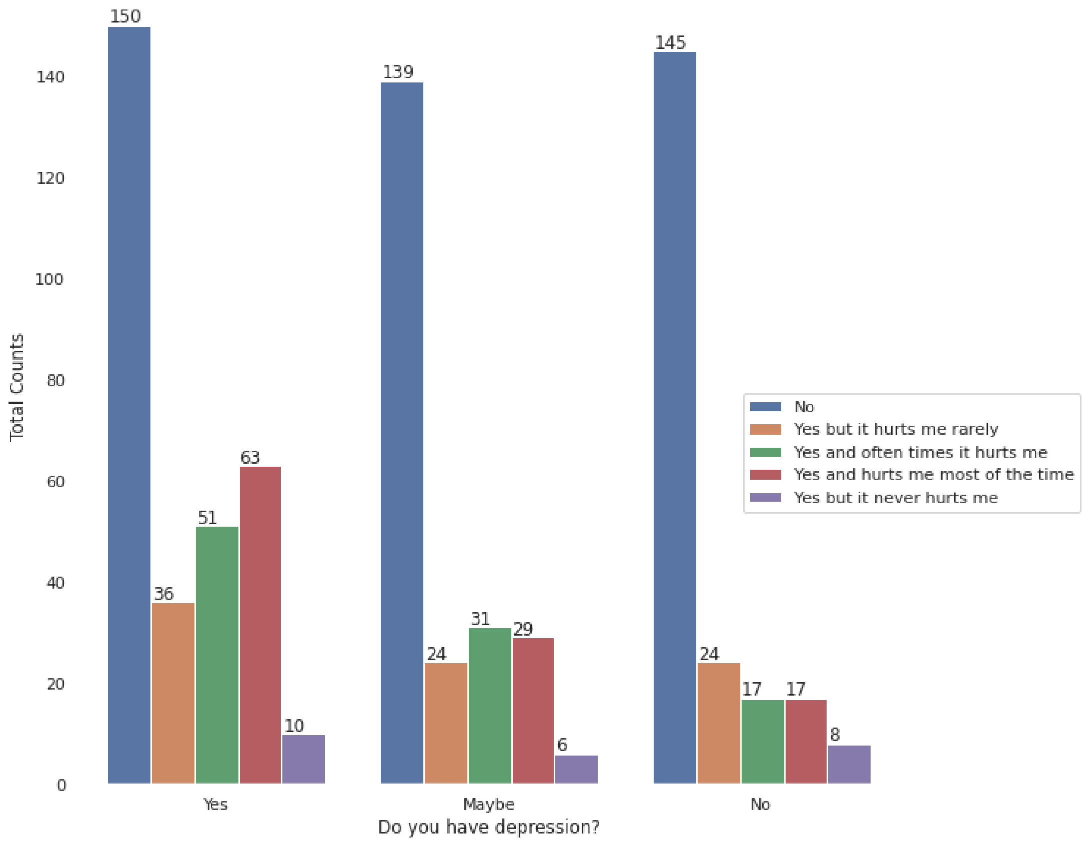

Figure 4.10: Violence in Family

We asked them about whether they faced domestic violence or not. Then we observe a lot of similarities can be seen between the chart of domestic violence and the chart of previous experiences with being bullied. The fact that as many as 42% of students have reported having seen domestic violence is certainly not a positive development. Even if the figure is higher than the percentage of students who had seen domestic violence in their own family (52%), the number ought to be as close to zero as possible. It makes little difference; in general, the majority of the students who have seen violence in their family end up being harmed and thinking about it most of the time. The number of this group of student is whopping 107, compared to the students who do not feel bad for the violence (67).

As was to be predicted, the metrics of having mental health issues are bigger on the students who are wounded and think about it most of the time about those violent events that occur in their homes. Those students who have seen regular violence in their families are around 84 percent more likely to experience depression than their peers who do not experience depression as a result of the violence. The figure is rather modest considering the possibility that the kids affected by the violence do not suffer physical or emotional harm as a result of it. (68 verified instances of depression plus 67 unconfirmed cases) vs 59 cases of depression not confirmed.

Chapter 5

Experimental Evaluation

In a previous section, we discussed our used model, the method of our data collecting, the process of data cleaning, and the pre-processing mechanism. In addition to this, we visually represent the similarities and linkages. After completing all of that, the next thing we did was begin the algorithm run for our accuracy test. In order to evaluate the reliability of our prediction method and the models, we have chosen 22 characteristics. The 22 characteristics that we have chosen are quite widespread, and the majority of them were recommended by the expert that we spoke with as well as gleaned through comprehensive material readings.

5.1 Experimental Setup

We have used Google Colab for our Data Pre-Processing, Analyzing, Model Training and Model Testing. As the collaboration process is easier in Google Colab Research as well as have the common modules to do Machine Learning work, we decided to use that instead of any local system. We used the basic Non-Pro version of Google Colab for our project which turned out to be enough for our research.

Apart from that, as we mentioned earlier, we used Google Form instead of physical papers for collecting survey data as it will be easier to import the survey data to our system and it can be anonymous and easier to track responses. We have also used Google App Engine as well as Form API to modify some of the questions, customize tracking keys etc. Then we have imported the data as CSV file and upload it in the Google Colab for further usage.

5.2 Experimental Results

Predicting depression is our top priority at the moment. Since the data set that we have is multi-classified, we employed the technique that was discussed in order to get a high level of accuracy. 7 different algorithms have been executed in order for us to evaluate the accuracy, precision, recall, and f-measures of the models that we have suggested. These variables are essential due to the fact that the strength of the model may be determined by how accurate the models are and how high their f-measures are. In the same vein, the greater the level of precision that we get, the clearer it will be to see which examples are in true positive. The recall is still another essential component of our investigation since it reveals the proportion of true positive to total yes answers.

We maintained our trust in K-fold cross validation for the purposes of training and testing. Because we have seen that problems with over fitting are common when using algorithms of this kind, we have resorted to using k-fold validation. As a result, with the help of cross validation, we are able to separate the training phase from the testing phase of this algorithm. This method divides the complete data set into K numbers of folds, each of which contains the same quantity of data. At the conclusion of each cycle, the K-1 folds are chosen to serve as the training set, while the remaining fold is used for the testing. It ensures that the value of each repetition is maintained. Therefore, at long last, we get the average of the mean squared errors in our data.

We constructed confusion matrix in order to determine the number of true positive instances as well as the number of false positive cases. We are using classification system with multiple classes. In our scenario, there are three different outcomes that may occur with each case. Which of these three is it: 'yes', 'maybe', or 'No'. As a result, the confusion matrix that we use for each method is a three-by-three grid. Figure 5.1 illustrate confusion matrix for our scenario.

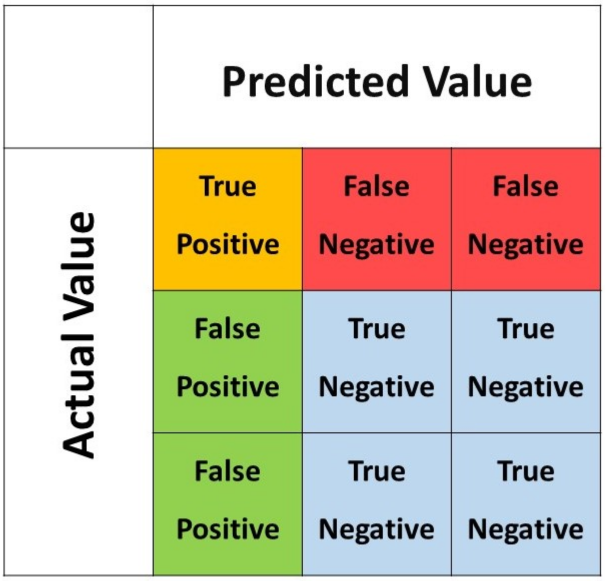

Figure 5.1: Confusion Matrix for 3 Classes

Following the various types of cleaning that we performed, we began using various machine learning algorithms with the 22 features that we had previously mentioned. Here we noted the result of 7 machine learning models.

5.2.1 Support Vector Machine (SVM) with 22 Features

Accuracy: 72.26% +/- 6.1				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	292	11	7	94.19%
True-Maybe	160	48	21	20.96%
True-No	45	78	88	41.71%
Precision	58.75%	35.03%	75.86%	

Table 5.1: Accuracy, Precision and Recall of SVM with 22 Features

Our accuracy is 72.26 %, as shown in table 5.1, which we have here (approximately 72%). We were able to determine the precision of 58.75% by doing the calculation of the total number of true positives divided by the sum of the total number of true positives and the total number of false positives. Here, we found that a total of 292 students had real positive results. That they are suffering from depression in reality. 160 and 45 were the total number of false positives that we obtained from. If we add all of them up, we get 205. We arrived to that conclusion by the use of calculations. After that, we determined the recall by dividing the total number of true positives by the aggregate of the total number of true positive and total false negative results. In this case, the total number of false negatives was 18, consisting of 11 and 7. Therefore, by dividing 292 by the sum of 292 and 18, we get a percentage of 94.19 %. In this model, there were only 18 out of 310 instances of depression that were false negatives. On the other side, there were 205 false positive cases out of 497 total, which is a very even distribution for our model.

5.2.2 Gaussian Naive Bayes (GNB) with 22 Features

Accuracy: 71.87% +/- 2.9				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	119	180	11	38.39%
True-Maybe	18	144	67	62.88%
True-No	2	57	152	72.04%
Precision	85.61%	37.37%	66.08%	

Table 5.2: Accuracy, Precision and Recall of GNB with 22 Features

Following the use of Gaussian Naive Bayes, the accuracy in table 5.2 was determined to be 71.87% (Approximately 72%). We were able to quantify the precision of our results, which came out to be 85.61%, by dividing the number of true positive instances by the total number of true positive cases and false positive cases. In this scenario, there were a total of 119 true positives out of 310, and there were 18 and 2 false positive instances correspondingly. Therefore, after doing the calculations, we found out that our model has an accuracy of 85.61%, which is not too awful. In order to compute the recall, we verified that it was 38.39% and that this figure matched the one in the database. We found the proportion of true positives to total true positives and false negatives by dividing the true positive value by this sum.

In this particular instance, the overall number of true positives was 119, but the number of false negative instances was 180, and 11 in particular. The recall was not quite as excellent as we had hoped for it to be. However, this helped us in finding a model that was more suited for our study. At conclusion, there were a substantial number of false negative instances in this algorithm, totaling 191 out of the 310. This suggests that 191 persons who self-reported that they were not depressed were, in fact, suffering from depression.

5.2.3 Artificial Neural Network (ANN) with 22 Features

Accuracy: 76.13% +/- 1.11				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	234	47	29	75.48%
True-Maybe	76	83	70	36.24%
True-No	27	51	131	62.67%
Precision	69.43%	45.86%	56.96%	

Table 5.3: Accuracy, Precision and Recall of ANN with 22 Features

The output of the Artificial Neural Network technique that was applied to our data set can be seen in Table 5.3. Using this strategy, we were able to achieve an accuracy of around 76.13% (Approximately 76%). We determined the precision and recall of this algorithm by using the same methodology that we have described for the precision and recall of previous algorithms. In all, there were 310 people, and 234 of them were predicted true positives. The number of false positives was 76 and 27. After doing this, we obtained a percentage of 69.43% in precision by dividing 234 by the total of 76, 27, and 234. Which is quite good for our model. In order to calculate recall, we have taken into account both situations of false negatives and true positives. The number of false negative were 47 and 29. The result was that we divided 234 once again using the sum of 47, 29, and 234. We were able to get a recall rate of 75.48%, which corresponds to a balanced outcome according to our model.

5.2.4 Random Forest Classifier (RFC) with 22 Features

Accuracy: 87.06% +/- 1.05				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	295	4	11	95.16%
True-Maybe	59	140	30	61.13%
True-No	23	10	178	84.36%
Precision	78.25%	90.91%	81.27%	

Table 5.4: Accuracy, Precision and Recall of Random Forest with 22 Features

According to the result that are shown in table 5.4, the random forest classifier performed with an accuracy of 87.06%, which is comparable to about 87%. Out of 310 total people, there are 295 who is considered true positives. This indicates that the algorithm correctly identified 295 individuals as being depressed who are in fact suffering from the condition. Whereas it classified 15 persons in total as not being depressed when they were, in reality, suffering with depression. We divided 295 by the total number of predicted false positive instances as well as the number of actual true positive cases. As a direct consequence of this, we obtained an precision of 78.25%. In this instance, the number of false positive cases was 59 and 23. It indicates that they did not suffer from depression but were labeled as having the condition. In order to calculate recall, we began by dividing 295 by the total number of instances which were false negatives and true positives. As a direct consequence of this, there has been a 95.16% recall. In this scenario, instances 4 and 11 both qualify as false negatives. As a consequence, we are able to observe that it provided a positive outcome for our system by displaying a good performance of recall and precision.

5.2.5 K-Nearest Neighbors (K-NN) with 22 Features

Accuracy: 77.20% +/- 2.8				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	218	72	20	70.32%
True-Maybe	64	96	69	41.92%
True-No	15	52	144	68.25%
Precision	73.40%	43.63%	61.80%	

Table 5.5: Accuracy, Precision and Recall of K-NN with 22 Features

The K-Nearest Neighbors method displayed an accuracy of 77.20%, which may be deemed to be the same as 77% (It is shown in table 5.5). Out of a total of 310 students, there were a total of 218 students who exhibited signs of depression. We were able to achieve this level of precision by dividing the number of true positive instances we had (218) by the combined number of false positive and true positive cases (total). In this case, there were 64 false positive cases, and 15 false negative ones. The precision achieved was 73.40%. Moving on to the recall, we have divided 218 by the total number of occurrences that included both false negatives and true positives. As a result of this, there has been a recall of 70.32% of the cases. In this particular circumstance, case number 72 and number 20 were classified as examples of false negatives. So from this result, we can say 218 students were truly depressed and detected correctly. But 92 people who are depressed but our system does not recognized them. But the overall proportion was good. Consequently, our model produced a satisfactory outcome in this regard as well.

5.2.6 Gradient Boost (GB) with 22 Features

When we look at table 5.6, we can see that our accuracy comes in at 64.27%, which is close to 64%. When we computed the precision, we found that it was 55.93%, and

Accuracy: 64.27% +/- 5.7				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	198	44	68	63.87%
True-Maybe	89	59	81	25.76%
True-No	67	33	111	52.61%
Precision	55.93%	43.38%	42.69%	

Table 5.6: Accuracy, Precision and Recall of Gradient Boost with 22 Features

we did so by dividing the total number of true positive instances by the sum of the total number of false positive cases and the total number of true positive cases. We found that the overall number of true positive predictions was 198, which indicates that people are depressed in the reality. The number of false positive instances is 89, while the number of false negative cases is 67. After that, we calculated the recall by dividing the anticipated total true positive by the total number of instances that included both false negatives and true positives. The percentage of items recalled was 63.87%. Which, considering our model, isn't all that bad. In this case, we projected that 44 individuals were sad when in reality they were, and we predicted that 68 people were not depressed when in reality we should have predicted 'Yes' for those people. However, when seen as a whole, the experiment was good.

5.2.7 Decision Tree (DT) with 22 Features

Accuracy: 60.61% +/- 7.21				
	Pred-Yes	**Pred-Maybe**	**Pred-No**	**Recall**
True-Yes	210	79	21	67.74%
True-Maybe	87	94	48	41.05%
True-No	108	41	62	29.38%
Precision	51.85%	43.92%	47.43%	

Table 5.7: Accuracy, Precision and Recall of Decision Tree with 22 Features

Table 5.7 indicates that the decision tree has an accuracy of 61.61%. This can be seen by looking at the table (Approximately 62%). In finding the correct level of precision, we divided the number of true positives by the total number of true positives and false positives. This means that we have divided the number of true positives, which was 210, by the total number of false positives, which was 87 plus 180. We measured it and found that its accuracy was 51.85%. This indicates that our prediction was accurate for 210 students who are depressed, while our prediction was incorrect for 87 and 108 individuals who were actually depressed but were informed by our prediction that they are not. In order to discover the recall, we have divided 210 once more. However, this time we decided to divide it by the sum of true positives and false negative totals, which occurred in 79 and 21 correspondingly. After that, we will got the 67.74% recall. Regarding our model, this is not quite as good as we had expected for. But despite that, we gained some new insights as a consequence of these findings.

5.3 Experimental Discussion

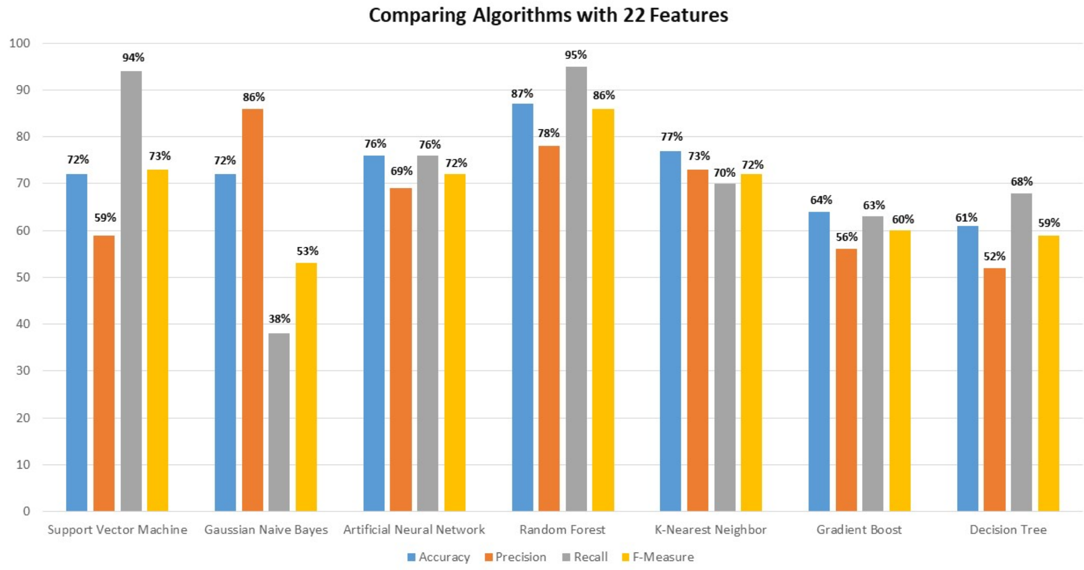

Figure 5.2: Algorithms with 20 Features Comparison

After analyzing the result collected from all of the models and algorithms over 22 features, we noticed in figure 5.2 that four of the algorithms are displaying a close resemblance to one another. Random Forest demonstrated the greatest level of accuracy, followed by Gradient Boosting and Decision Tree in descending order of accuracy. In this particular instance, Random Forest provided the most accurate results, 87%, while Decision Tree provided the least accurate results, 61%. We are aware that a greater f-measure both indicates a system's increased accuracy and serves as a measure of the system's overall quality. The Gaussian Naive Bayes model has the lowest f-measure score, which comes in at about 53%. It seems that, for the model that we are now using, it is not beneficial. The recall that we are receiving from this algorithm is quite low—only approximately 38%. This indicates that it is only capable of accurately identifying 38% of the instances. However, the precision that it indicated was 86%, which is a very high percentage. In conclusion, there was not a solution to our model that was balanced overall.

The remaining algorithms all shown satisfactory performance in terms of f-measures and accuracy. The f-measure for the Support Vector Machine is 73%, and the accuracy is%. The f-measure for the Artificial Neural Network is 72%, and the accuracy is 76%. The f-measure for the Random Forest is 86%, and the accuracy is 87%. Even in the decision tree, we have f-measures of 59 percent and accuracy of 61%. Therefore, with the exception of the Gaussian Naive Bayes method, all of the other algorithms performed well. Even in the case of accuracy, all other models, with the exception of Gaussian Naive Bayes, fared much better. Random forest was able to get extremely excellent results in every scenario. In every scenario, it demonstrated a healthy level of equilibrium. The accuracy that Random forest provided us with was 78%. Whereas SVM displayed 59%, ANN displayed 69%, K-NN displayed 73%, Gradient Boosting displayed 56%, and Decision Tree displayed

52%. In spite of the fact that it displayed the highest result of 86 percent, the Gaussian Naive Bayes algorithm was not balanced at all. According to the results of our model, Random forest did do well overall.

5.4 Suggestion Mechanism

A straightforward method was chosen for the recommendation that was made. One of the questions on the questionnaire inquires as to whether or not the individual was successful in overcoming their depression. Students who responded ”Yes” or ”Partially” to the question were directed to the next question, which was about how they overcame their despair.

Our strategy consisted of deriving the weight from the 22 characteristics that were described in the previous section, analyzing the similarities of this situation, and coming up with a suggestion based on those similarities.

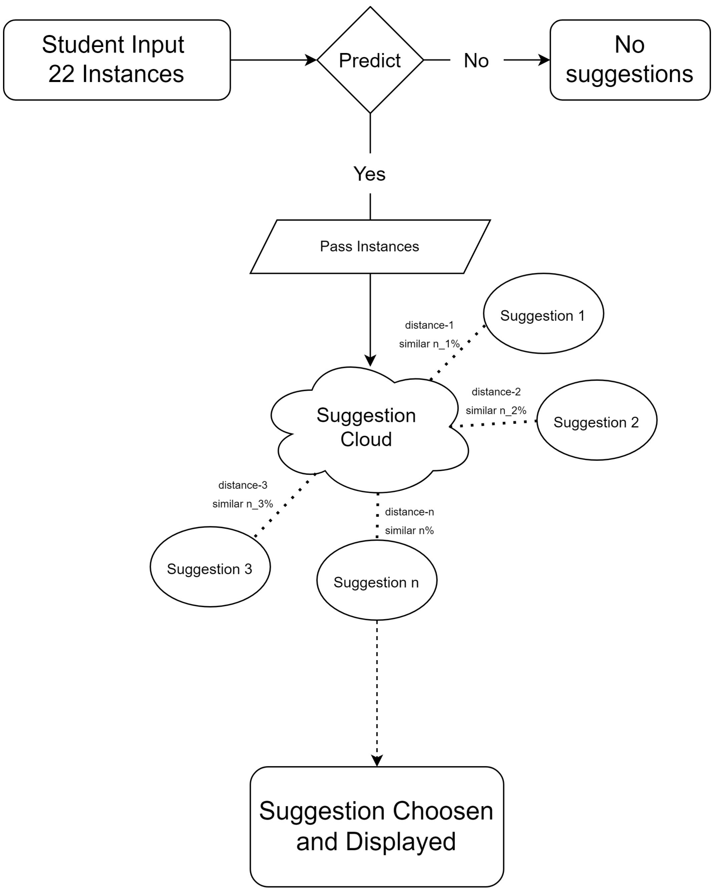

Figure 5.3: Suggestion Generate from Suggestion Cloud

The way that our algorithm generates suggestions is shown here in figure 5.3, as can be seen. First, we, the students, offer our input, and if our system anticipated that their input would be “Yes,” it continues the examples through to the next step, where it determines whether or not there is a comparable sort of instances. If it discovers a situation that is comparable, it will enter that cloud of suggestions where the previously-stored proposal is located. The 22 characteristics that are present at the beginning of the pre-stored proposals are the identical ones. Currently, it is comparing the situations to find similarities and determining the best way to act on the proposal. We have obtained 544 potential responses using the information from our dataset. Every one of the responses includes a number of recommendations. Each of the responses is being treated as its own separate batch by us. As a result of the fact that our goal is to provide the sufferer with immediate support, the ideas that we provide often consist of fundamental activities like as listening to music, watching movies, participating in games, practicing yoga or meditation, or engaging in psychotherapy, etc. However we have plans to continuously take feedback from the user to know about if the suggestion was helpful or not.

Chapter 6

Limitations and Future Work

In this section, we have made an attempt to restate the issue and briefly summarize the work that we have done so far for this study, as well as the limits of the whole work and the potential future scope of this research.

6.1 Limitations and Challenges

There were certain limitation we have faced while doing this research and it was really challenging to complete this research since it is already a stigma in our country.

1. The survey was by far the most challenging aspect. We have suffered greatly in the survey section as a result of the fact that we often see students displaying a lack of enthusiasm in filling out the survey form. It's possible that the length of the forms or the questions themselves are to blame. Even though we assured the students that we would keep their information confidential, it's possible that some of them still felt unsafe.

2. It was difficult to collect a large number of responses in order to correctly run the algorithms. Simply said, if you have more data, the machine will perform better. Because we began this study at the height of the COVID-19 pandemic, the majority of the universities were closed, and collecting responses online or visiting Facebook groups were not a viable option. However, we tried to push as many as we could.

3. It seems to be possible that the data we obtained via the questionnaire didn't include any information that may indicate how severe the depression was. Therefore, it is limiting the ability of our model to make predictions. Simply because we do not now possess a precise severity matrix with which to train the machine.

6.2 Future Works

We have a strategies in mind to improve the precision of our model in the future by adding additional data. There is a strategy in place for us to collaborate with a licensed professional psychotherapist on the enhancement of our model. Our goal is to cultivate an atmosphere in which people feel comfortable talking more openly about depression, so that the condition will no longer be seen as socially unacceptable in the years to come.

Figure 6.1: Starting Interface of App

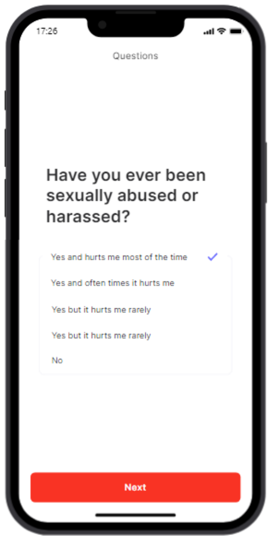

Figure 6.2: Question Interface of App

At the moment, we are working on a interface prototype version (Figure 6.1, 6.2, 6.3) of our prediction model that will be able to be loaded on any smartphone, and then students will be able to quickly access it anytime they want. Because it is impossible to predict when someone may start feeling depressed, our group works up with this concept as a solution to the problem. Therefore, the application will be quite useful. This figure displays a preview of the interface that our prototype will have.

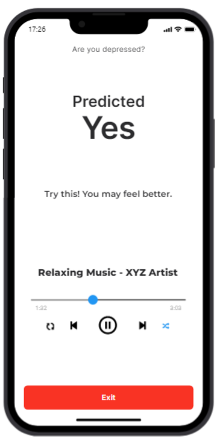

Figure 6.3: Result Interface

Chapter 7

Conclusion

Since depression is one of the most prevalent mental disorders or illnesses, there is little doubt that it has been the topic of a greater amount of research. Even though some research has been done on the issue, there is not much evidence to suggest that using technology to prevent depression is effective. In spite of this, the prospect of preventing or treating depression-related mental diseases is growing at an exponential rate as technological advancements are made. Because of this, the time has long since come for us to make use of the technical capabilities at our disposal in order to put a stop to the ever-increasing rate of depression-related suicides committed by university students. If we didn't act quickly, we wouldn't be able to notice the talent and production losses that have been generated by this massive turmoil in time. In Bangladesh, there have been numerous people who have taken their own lives as a result of being depressed. Students at reputable universities in Bangladesh are succumbing to despair and killing themselves in increasing numbers. It is indeed possible that the findings of our research won't be enough to fix the issue completely, but we believe this is an approach that has the potential to open up new avenues of research. We have observed that in our machine learning model, Random Forest was capable of predicting sadness with a high degree of precision. 6 In addition to those methods, several others were utilized, with the essential aspects under the hood. In light of this, our approach to determining the fundamental reasons and doing an analysis of the patterns of causes and actions could be a useful initial step in this ongoing research project. In addition, we believe that our research will help in the creation of a model for machine learning that may assist in the improvement of the mental health of university students.

References

[1] R. Kessler and W. Magee, *Childhood family violence and adult recurrent depression*, en, Mar. 1994.

[2] J. Andrews, "Depression today, or new maladies of the economy," en, *Social Text*, vol. 27, pp. 167–173, 2009.

[3] L. Fleming and K. Jacobsen, "Bullying and symptoms of depression in chilean middle school students," en, *Journal of School Health*, Feb. 9, 2009.

[4] B. Ahmedani, *Mental health stigma: Society, individuals, and the profession*, it, PubMed Central (PMC, 2011.

[5] G. Kumar and S. A. Jain, "Prevalence of depression and its associated factors using beck depression inventory among students of a medical college in karnataka," en, *Indian Journal of Psychiatry*, 2012.

[6] A. P. Association, *Diagnostic And Statistical Manual Of Mental Disorders*, en, 5th ed. American Psychiatric Publishing, 2013.

[7] K. Daimi and S. Banitaan, "Using data mining to predict possible future depression cases," en, *International Journal of Public Health Science (IJPHS*, Dec. 2014.

[8] S. Deb and E. S. Hansen, "Academic stress, parental pressure, anxiety and mental health among indian high school students," en, *International Journal of Psychology and Behavioral Science*, 2015.

[9] P. Nambisan, J. Luo, A. Kapoor, and T. Patrick, "Social media, big data, and public health informatics: Ruminating behavior of depression revealed through twitter," en, in *Proceedings of the Annual Hawaii International Conference on System Sciences*, Mar. 2015.

[10] V. Bellman, *Three ways in which passion can support our mental health*, en, Online]. Available: Sep. 5, 2016. [Online]. Available: concentriccounselling.com/journal/2017/5/8/passionmhaw17.

[11] I. Bhakta and A. Sau, "Prediction of depression among senior citizens using machine learning classifiers," en, *International Journal of Computer Applications*, Jun. 2016.

[12] M. COPELAND, *What's the difference between artificial intelligence, machine learning and deep learning?* en, Online]. Available: Jul. 29, 2016. [Online]. Available: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/..

[13] S. NGO, *5 signs you are under too much stress*, en, Online]. Available: Dec. 5, 2016. [Online]. Available: https://www.wallstwatchdog.com/money-career/signs-that-you-are-under-too-much-financial-stress..

[14] wwwrelateorguk, *Relationships and mental health*, en, 2016.

[15] J. Bre, *Prediction of wind pressure coefficients on building surfaces using artificial neural networks*, en, Nov. 2017.

[16] P. Grover, *Gradient boosting from scratch*, en, Online]. Available: Dec. 9, 2017. [Online]. Available: https://blog.mlreview.com/gradient-boosting-from-scratch-1e317ae4587d?gi=d259ff39adb7..

[17] N. Davis, *Sexual assault victims more likely to have anxiety and depression – study*, en, 2018. [Online]. Available: www.theguardian.com,.

[18] R. Gandhi, *Support vector machine — introduction to machine learning algorithms*, en, Jun. 7, 2018.

[19] B. Zalar, A. Blatnik, A. Maver, Z. Klemenc-Ketiš, and B. Peterlin, "Family history as an important factor for stratifying participants in genetic studies of major depression," en, *National Center for Biotechnology Information*, Jun. 2018.

[20] S. Azim and M. Baig, "Frequency and perceived causes of depression, anxiety and stress among medical students of a private medical institute in karachi: A mixed method study," en, *Journal of the Pakistan Medical Association*, Jun. 2019.

[21] A. Choudhury, M. Khan, N. Nahim, and S. Tulon, "Predicting depression in bangladeshi undergraduates using machine learning," en, in *2019 IEEE Region 10 Symposium (TENSYMP*, Jun. 7, 2019, p. 2.

[22] M. Megha, "Why do mental health patients not seek help?" en, *The Daily Star*, Apr. 18, 2019.

[23] T. Zainab and A. Zafar, "Poor academic performance as predictor of depression," en, *International Journal of Latest Research in Humanities and Social Science (IJLRHSS*, 2019.

[24] I. Education, *Deep learning*, sv, Online]. Available: May 1, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/deep-learning..

[25] "Ibm cloud education, "ibm.com," fr, *IBM*, Aug. 19, 2020, Online]. Available: [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning.

[26] K. Koly, S. Sultana, A. Iqbal, J. Dunn, G. Ryan, and A. Chowdhury, "Prevalence of depression and its correlates among public university students in bangladesh," en, *Journal of Affective Disorders*, Dec. 28, 2020.

[27] B. Merz, *Six common depression types*, fr, Online]. Available: Oct. 13, 2020. [Online]. Available: https://www.health.harvard.edu/mind-and-mood/six-common-depression-types..

[28] H. Misra, *Machine learning for subsurface characterization*, en, 2020.

[29] P. Narayanrao and P. Kumari, "Analysis of machine learning algorithms for predicting depression," en, *Computer Science, Engineering and Applications (ICCSEA*, Mar. 13, 2020.

[30] M. Torres, *Psychiatry.org*, en, Online]. Available: Oct. 2020. [Online]. Available: https://www.psychiatry.org/patients-families/depression/what-is-depression..

[31] J. Brito, *Relationship depression: How to cope*, en, 2021.

[32] A. Christopher, *K-nearest neighbor*, en, [Online]. Available: Feb. 2, 2021. [Online]. Available: https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4..

[33] U. Illinois Chicago, *Machine learning in healthcare: Examples, tips & resources for implementing into your care practice*, en, Online]. Available: Jul. 22, 2021. [Online]. Available: https://healthinformatics.uic.edu/blog/machine-learning-in-healthcare/..

[34] W. H. Organization, *Depression*, fr, [Online]. Available: Sep. 13, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression..

[35] M. Uddin, K. Dysthe, A. Følstad, and P. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," en, *Neural Computing and Applications*, Aug. 27, 2021.

[36] N. UNB, "Over 61 percent of 18-25s suffering from depression during pandemic," en, *UNB NEWS*, Jul. 10, 2021.

[37] B. Alam, *Implementing naive bayes classification using python*, et, Online]. Available: Jan. 14, 2022. [Online]. Available: hands-on.cloud/implementing-naive-bayes-classification-using-python.

[38] N. Chauhan, "Decision tree algorithm, explained," en, *www.kdnuggets.com*, Feb. 9, 2022, Online]. Available: [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html.

[39] *Support vector machine algorithm*, pt, [Online]. Available: [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm..

[40] *Uc berkeley university health services, "clinical depression*, en, [Online]. Available: [Online]. Available: https://uhs.berkeley.edu/health-topics/mental-health/clinical-depression..