# Mental Health Prediction Among Unemployed Graduates

# Using Machine Learning Approach: BD Perspective


Prepared By-

**Name: Nilima Islam Bristy**

**ID: 2018-2-55-010**

Supervised By-

**Dr. Mohammad Arifuzzaman**

**Faculty of sciences and Engineering**

**Chairperson & Associate Professor**

**Department of Electronics and**

**Communication Engineering**


A Thesis Paper Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Electronics and Telecommunications Engineering

DEPARTMENT OF ELECTRONICS & COMMUNICATIONS ENGINEERING

**January 2023**

## APPROVAL

The thesis paper "Mental Health Prediction Among Unemployed Graduates Using Machine Learning Approach: BD Perspective" that Nilima Islam Bristy (2018-2-55-010) submitted to the Department of Electronics and Communications Engineering at East West University, Dhaka, Bangladesh, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Electronics & Telecommunications Engineering. Its formatting has also been approved.

**Approved By**

--------------------------------------

**(Supervisor)**

**Dr. Mohammad Arifuzzaman**

Faculty of sciences and Engineering

Chairperson & Associate Professor

Department of Electronics and Communications Engineering

## DECLARATION

We certify that neither this University nor any other University has previously accepted our work for the purpose of awarding a degree. To the best of my knowledge and belief, this thesis contains no material that has already been published or authored by another person, with the exception of passages where appropriate citation has been made. We hereby state that the research we conducted under the guidance of Dr. Mohammad Arifuzzaman, Chairperson, Associate Professor in the Department of Electronics and Communications Engineering at East West University in Dhaka, Bangladesh, is the source of the work given in this thesis.

**Countersigned**

**--------------------------------**

**(Supervisor)**

**Dr. Mohammad Arifuzzaman**

**Signature**

**--------------------------------**

**Name: Nilima Islam Bristy**

**ID: 2018-2-55-010**

# DEDICATION

This thesis is dedicated

To

Our beloved parents and honorable teachers

For their endless love, support and

encouragement

# ACKNOWLEDGEMENT

First and foremost, I would want to express my gratitude to Almighty Allah for bestowing upon me the courage and knowledge necessary to finish my thesis. To my honorable supervisor Dr. Md. Arifuzzaman Assistant Professor, Department of ECE, East West University, Aftabnagar, Dhaka, Bangladesh, I would like to express my sincere gratitude and appreciation for his guidance, encouragement, kind assistance, knowledge sharing, and constant inspiration throughout this thesis work. My profound gratitude to our honorable teachers for their welcoming demeanor and passionate assistance throughout the past four years. We would like to extend our profound gratitude to my parents, instructors, family, and friends for their guidance, financial assistance, unending encouragement, constant inspiration, and excellent contributions from the beginning to the end. Without any of them, this report's final output would not be feasible.

## ABSTRACT

According to emerging advancements, university graduates looking for work are more likely to experience common psychological illnesses including obsessive thinking, panic attacks, nervousness, or distress. Nevertheless, Bangladesh has still not looked into the difficulties with mental health faced by unemployed graduates. We have developed a model that employs a typical psychological screening using machine learning algorithms to identify the various stages of certain psychological illnesses among unemployed Bangladeshi graduates in order to identify these problems at a young age. In order to effectively assist unemployed graduates, our focus was to obtain accurate and reliable sets for identifying trends such as age, gender, the causation of sadness, the data of behavioral change, as well as several other things. The questionnaires that have been obtained throughout the duration of this study has been of considerable assistance in terms of forecasting depression and giving counsel. On the real datasets for predicting depression, we applied seven different types of ML algorithms in our suggested model: Random Forest Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Logistic Regression, Support Vector Machine, Gradient Boosting Classifier, AdaBoost. Our model's outcomes show that Random Forest, Logistic Regression, Gradient Boosting Classifier performed admirably overall with the accuracy of 80.6%. Remaining algorithms outperformed the others we used by a significant margin. We have a clear focus to add to this conversation with the recommendation model that we developed as part of this study. Using the assistance of Machine Learning by monitoring and analyzing those behaviors, we believe that we will be capable of assisting them in the near future with website or mobile applications as they can aware themselves.

**Keywords:** Mental Health; Unemployed Graduates; Depression; Prediction; Machine Learning.

# Table of Contents

**List Of Tables**



**List Of Figures**

# Chapter 1

## Overview

Extensive research has been conducted on the relationship connecting physical and psychological symptoms brought on by poor mental health as well as unemployment. This chapter provides a detailed summary of the study that was carried out to eliminate the problem, together with the study's goal, objectives, and method of problem-solving.

## 1.1: Introduction

Worldwide, rates of mental disease are increasing at epidemic levels, and according to a WHO prediction, one in four people may have mental or neuropsychological illnesses at some stage in life. The process of diagnosing mental illness is complex and requires numerous phases. Actually, a properly structured questionnaire with questions about the indications and medical background, as well as occasionally completing a physical examination, is where the diagnosis begins. Numerous psychological tests are indeed carried out to confirm that the signs are solely brought on by issues with mental health and therefore not associated with any other issues.

The causes of unemployment have been the subject of a great deal of investigation. There seem to be numerous causes for all of this, and every person deals with unique issues. Yet, mental health issues rank among the major factors contributing to unemployment. The influence of poor mental health and therefore any extra physical or mental effects that could very well develop over time for older unemployed individuals are the main subjects of this study. The accompanying physical and mental symptoms and potential repercussions of poor mental health are stated in the study area based: exhaustion, compulsive tendencies, severe anxiety, tearfulness, excessive thinking, unhappiness, anxiousness, and inability to concentrate.

Therefore, earning a college degree is seen as a crucial element of achievement in both the professional and society spheres. Due to the strong demand, there has been a noticeable growth in the number of university graduates during the previous 50 years, as well as a notable growth of universities as well as their campuses. It is projected that more university graduates will then enter the workforce, however because the employment market is not growing quickly enough already to accommodate the growing number of recent graduates, heightened competitiveness and job insecurity have formed. Conversely, as more graduates join the workforce and administrations and the private enterprise struggle to find suitable inclusive employment, the proportion of unemployed and underemployed university graduates must have constantly climbed.

## 1.2: Aim of Study

Our primary goal is to investigate the connection between mental illness and unemployment. Our research, which confirms that individuals suffering from mental illness have a high level of unemployment. So, in order to assist the psychiatrist and counselor at the beginning of the patient's therapy, we will create an approach to estimate the intensity of the patients' high levels of anxiety. This should enable them to choose a plan of care that is appropriate for the individuals. The best training set for our current proposal to identify the seriousness of the problems will then be determined by doing direct comparison between various performance metrics of the machine learning techniques we will be utilizing throughout this research.

## 1.3: Problem Statement

Researching mental health will teach you a great deal about how mental illness is treated as well as the systems in place to handle crises and offer assistance. You will, however, also develop your own unique viewpoint on mental health and wellbeing. Nowadays, mental illness is a problem that affects people of all ages and from all over the world. The diagnosis of mental illness is typically based on the patient's self-report, which calls for the use of questionnaires created to identify particular emotional or social patterns of communication. Many individuals suffering from mental illnesses or emotional disorders should be able to heal with the right care and therapy. The primary implications of these results are that worry and poor focus significantly influence the likelihood that someone may experience mental and physical health issues that will cause them to lose their jobs. This research seeks to identify plausible causes of mental disorder, predict its occurrence, and explore its correlation with unemployment.

## 1.4: Research Methodology

Our method for examining the literature was influenced by the work with a significant focus on providing a depiction of the state-of-the-art of fundamental progress, which typically offers a much more critical particular perspective on the issues being discussed. This investigation has been conducted over a considerable amount of time. Nevertheless, we will carry on with this investigation since we want to broaden it and gather more data. We then started reading articles that were similar to our ideas. The formation of our idea was still in its beginnings at the time. At the same time, with the assistance of our supervisor, we were able to finish our hypothesis and launch our inquiry to find pertinent justifications and components that are connected to this idea. We discovered how to properly conduct a survey by being very cautious when we were designing our questionnaire forms. Following formal planning and implementing all required procedures, we then started our survey. Prior to using machine learning techniques on the data, it was crucial to clean as well as pre-process it after the data collection procedure was complete. We started preparing chapters 1, 2, and 3 of our thesis paper on September 1. After that, we wrote chapters 4, 5, and 6 for around a month. We finally finished drafting the final chapter of our thesis paper. After carefully going through our work, we spent roughly 14 days using Microsoft Word to organize it according to the East West University's requirements. Our approach is all shown in Figure **1.1.**

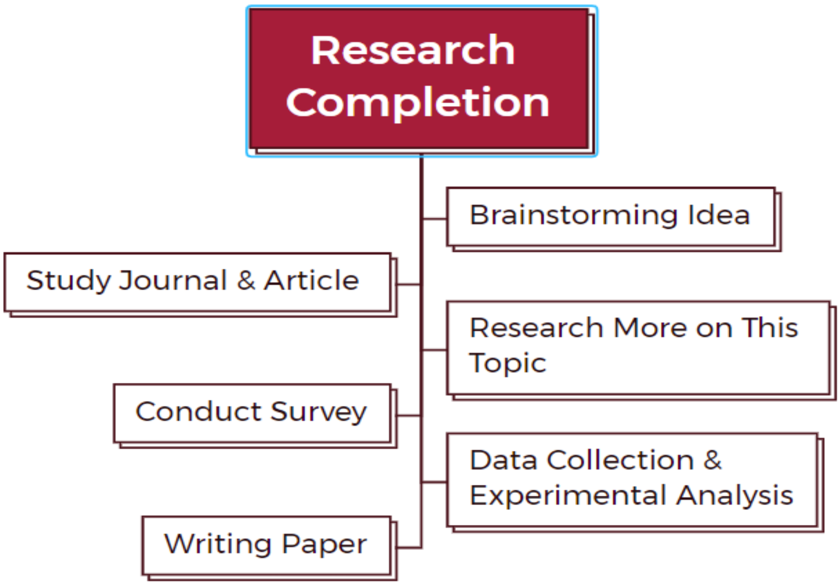



**Figure 1.1: Research Completion Process**

## 1.5: Thesis Outline

The below demonstrates how the thesis is put next to each other.

- In Chapter 2, background data on relevant and important studies that have been conducted in this field of study are given.
- In Chapter 3, the underlying idea behind the applied prediction algorithms is explained.
- In Chapter 4, it describes the detailed procedure. The dataset, the preparation technique used, the application of the predictive analysis, and the functional traits are all described in this chapter.
- In Chapter 5, the results of the algorithms are discussed, with tables and diagrams serving as illustrations. The limitations of the examination and the techniques are also presented, along with the results pertaining to the research subject.
- Finally, Chapter 6 explains the research point as well as suggests additional study

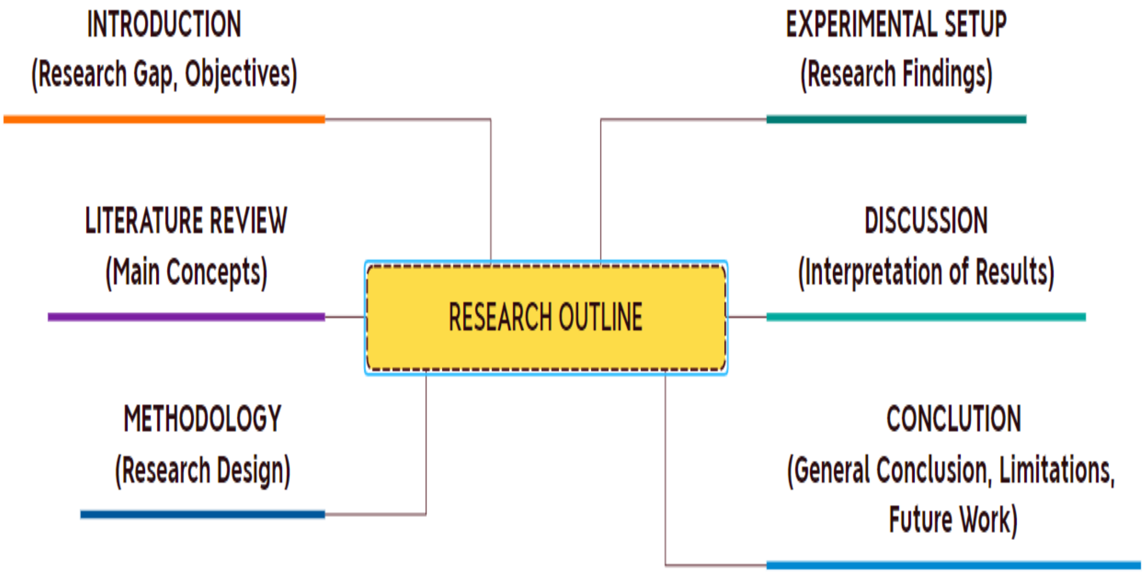



**Figure 1.2: Research Outline**

# Chapter 2

## Background Study

This study sought to determine whether emotional and cognitive issues brought on by poor mental health could lead to unemployment. This chapter continues by elaborating on the issue from the perspective of the people who responded we chose, the modification techniques we used, and evaluations of earlier studies.

## 2.1: Context of Research

The dataset from [1] served as the foundation for this study. The goal of this study was to demonstrate the link connecting mental disorder and unemployment. Two papers that were recently released on massive unemployment as well as psychological problems served as the impetus for research. So many of those analyses on physical and mentally issues brought on by poor mental health and significant unemployment is the NAMI study. The NAMI organization operates to improve the lives of those with mental diseases. In 2012, every state's demographic in America was analyzed by NAMI. And according to research, employment helps people improve their mental health, yet only 2% of people succeed in finding employment. The research article, which served as the inspiration for this paper's [1] current study, includes work by this paper [2]. Throughout 2009 and 2010, those researchers analyzed the job role of those Americans with mental diseases. The results of the poll, which already had 77,326 participants, showed that people with multiple mental diseases have difficulty finding jobs. Additionally, it was discovered that individuals above the age of 49 with serious psychological diseases had lower employment rates than those with no history of depression or only minor psychiatric disorder.

Gonçalves et al. [3] investigated if psychiatric diseases and various correlations with unemployment may be anticipated using Corley's [1] complete dataset. They employed the Gradient Boosted Trees, Random Forest, Decision Tree, and K-Nearest Neighbor computational methods categorization models. Random Forest outperformed them all with a prediction performance of 94.14% [3].

A mental disorders screening assessment system was proposed by Masri R.Y. and Jani H.M. in [4] to help mental health professionals identify and cure their people with mental illnesses. The diagnosis and recommendation of therapeutic approaches were done using three artificial intelligence methods: Fuzzy Logic, Rule-Based Reasoning, and Fuzzy-Genetic Algorithm. David D. Luxton [5] examined the application of machine intelligence in psychiatric treatment. Discussions have been had concerning the present trend, its capabilities, and consequences.

A decision support system for such identification of schizophrenic disorder was designed by Razzouk D., et al. [6] and has produced good results of 66-82%. To grade depressive symptomatology, Acharya et al. [7] created an optimal solution method.

## 2.2: Mental Health

The idea of psychological health was preferred over the idea of mental disease since it is more appropriate for the study considering the variables employed, and it is a core concept throughout this research work. Since much of the studies have been conducted on individual mental health, there are multiple interpretations for mental wellbeing. Henceforth, the World Health Organization's useful understanding of health defines it as "a condition of well-being that occurs when the individual possesses his or her strengths, can cope with the everyday challenges of life, can collaborate effectively and fruitfully, and can make a significant contribution to his or her community"—is used during this whole thesis (WHO, 2001). The understanding of health sickness, apparently, is focused on mental diseases, whereas the World Health Organization's definition is focused on contributing to community health issues in the workplace tolerance in their daily lives.
Healthcare is important because it promotes psychosocial development [8]. However, 25% of people worldwide experience mental health issues [9]. Traumatic childhood experiences can harm mental wellbeing, but mental diseases can also develop suddenly [10]. The consequences of poor mental health is emphasized by the Global Economic Forum [11]. This company released a Worldwide Risks Report upon the effects of persistent psychological and physical ailments brought on by poor mental health. Individuals who already live here have a higher risk of having lower endurance and a lower ability to handle stress and frustration and their money.

Poor mental health can lead to physical and psychological symptoms in a variety of contexts [12]. Many strategies have been developed to help persons with psychological issues. For instance, there are numerous courses available for parents to attend in order to comprehend, educate, and relate to their youngsters [13]. Corresponding to this, schools provide lessons that emphasize helpful elements for kids to promote excellent mental health [14].
The Netherlands invests 22 billion euros yearly [15] on workers; others become unwell and are less profitable or lose their jobs as a result of poor mental health, as was noted in the opening paragraph [16]. As a result, managers are taught in organizations how to identify certain aspects and deal with it accordingly [17]. Additionally, businesses launch initiatives to help spread the word and start a dialogue about the issue [18]. These steps are made to enhance the workforce's psychological state and bring down the cost associated with it.

## 2.3: Negative Effects of Psychiatric Condition on Unemployment

People's personal employment position and mental health are linked in a number of different ways. According to this [19] findings, the others who lose their jobs have a higher risk of being mentally ill than someone who has jobs. This is due to the mounting financial strain that results in a decline in their self-esteem and assurance, which causes self-doubt [20]. Individuals with poor mental health consequently believe they are much less capable of changing their employment position [21]. Some other study [22] revealed that those with poor mental health and unemployment display decreased full time position behavior. Work consequently significantly influences mental state, but finding work also greatly influences psychological health.

The preceding paragraph made the connection between psychology and professional health, and this section concentrates on potential fixes. The issues associated with poor mental health and employment have already been addressed in a realistic way. This approach entails taking action to improve mental health and, when needed, intervening to avoid psychological disorders [23]. According to studies, these strategies and initiatives can be used with kids under the age of 10 in home and in the workplace to be successful later on [24]. Additionally, unfavorable working circumstances, a large schedule, or the caliber of the task can have an impact on one's mental well-being [25]. Because of this, persons with weak mental health are supported through schooling and full-time position training. In this regard, analysis demonstrates that individuals can obtain employment more quickly, are much less likely to get jobless, and can better combine the amount and quality of the finished product.

## 2.4: Recent Findings on This Subject

There has recently been extensive research on psychological health and job stability. Numerous scholars have conducted investigations using diverse datasets. By offering such people a time-restricted budget, Rössler et al. [36] investigated whether persons who experience mental complaints because of compromised mental health find employment more quickly or stay indefinitely at their existing job. In order to forecast when individuals who suffer from mental illnesses will lose their jobs, machine learning techniques were not used in this study. Instead, they looked at whether giving people with poor psychological health compensation could help reduce their concerns.

However, the primary objective of a recent review concentrated on the connection among both mental health issues and joblessness. According to findings by Butterworth et al. [37], those who have indications of mental chronic conditions are more inclined to lose their jobs. When it comes to the mental health problems which people have experienced after losing their jobs, this research differs compared to most others. To conduct this study, an analytical approach was used as the methodology. The study made use of a database of roughly 5,600 Australian participants.

Nevertheless, a large number of studies on individual personal wellbeing have made use of algorithms for machine learning. The aforementioned models were employed by Srividya et al. (2018) to forecast people's choices and problems with mental health which are Logistic regression, Support Vector Machine, Decision Trees, Naive Bayes Classifier, K-Nearest Neighbor Classifier. According to this study, the effectiveness of Support Vector Machine, K-Nearest Neighbors, and Random Decision Trees was nearly identical. The addition of classification models then helped to even further increase the prognostications. As a consequence, the algorithms' performance is improved to 90%. The authors of the study goal was to add depth to current studies on individuals suffering from mental illnesses.

Going to follow up somewhat on earlier research, a study was undertaken on three significant disorders of mental health: exhaustion, nervousness, and depressed mood. In an attempt to catch up on past studies, an investigation on the major conditions of mental health of tiredness, anxiety, and depression was conducted. This study's [38] key finding though is that the collection lacked consistency. They determined the Highest accuracy for the optimal threshold choice to handle this issue. The Naive Bayes algorithm was shown to perform with the best classification afterwards when. The Random Decision Tree generated the greatest proposed model.

The most recent study by Cho et al. [39] examined which machine learning approaches anticipate mental health concerns the most accurately. According to this research, among the most commonly employed models in research on mental health issues are the Support Vector Machine, Random Decision Tree, K-Nearest Neighbors, Naive Bayes, and Gradient Boosting Machine. Each of these approaches has advantages and disadvantages. Even several researchers emphasize their decisions, according to recent studies. The Support Vector Machine, however, has the best performance in terms of mental health, according to Cho et al.'s [39] work. Another explanation for this may be that, despite the need for exceptionally precise preparation of complicated data, the Support Vector Machine performs reasonably effectively.

The Random Decision Tree and Gradient Boosting Machine are employed associated with SVM for feature reduction or for a categorization that doesn't need feature engineering. These techniques' benefits include their reasonable practicability. When dealing with ensemble learning, the techniques immediately overcome any prior deficiencies. Both two algorithms do, nevertheless, both have certain drawbacks. For instance, training learners in Gradient Boosting Machine models takes more time than training them in those other algorithms. Furthermore, it is not easy to interpret the relationship between variables and predictors with the Random Forest.

The extensive literature review reveals that, on the one hand, numerous research projects are underway to computerize the diagnosis of psychological health diseases. And on the other hand, attempts are underway to accurately detect mental health issues utilizing machine learning approaches. To boost the precision of screening with a condensed collection of information from individual profiles, hybrid systems machine learning algorithms are being used. The goal of this study is to analyze specific machine learning methods for predicting the likelihood of significant mental health issues like attention deficit disorder, forms of anxiety, discipline failure, and mood disorders among unemployed graduates. The standard of living for unemployed graduates is improved when these problems are identified early and treated. Comparisons with fewer qualities have also been done.

# Chapter 3

## Research Methodology

We must organize our research into multiple distinct divisions because our work involves a lengthy schedule, systematic procedures, auxiliary jobs, and a predetermined order. There are many objectives and sub - functions in each segment. Though one year since our study project required both online and physical meetings for organization, data collection, and academic and individual consultation. Additionally, in order to prepare and evaluate the data set online, I had to collaborate across many technologies research programs.

## 3.1: Proposed Methodology Recruitment and Procedure

In order to accomplish our objective, we shall describe our study strategy and procedures underneath this paragraph. Five primary categories and just a few subdivisions make up my work schedule. Additionally, there were moments when we simultaneously collected, analyzed, and applied that information. There are five elements to the likely work schedule. Below is the strategy that we suggest (see Fig. 3.1)

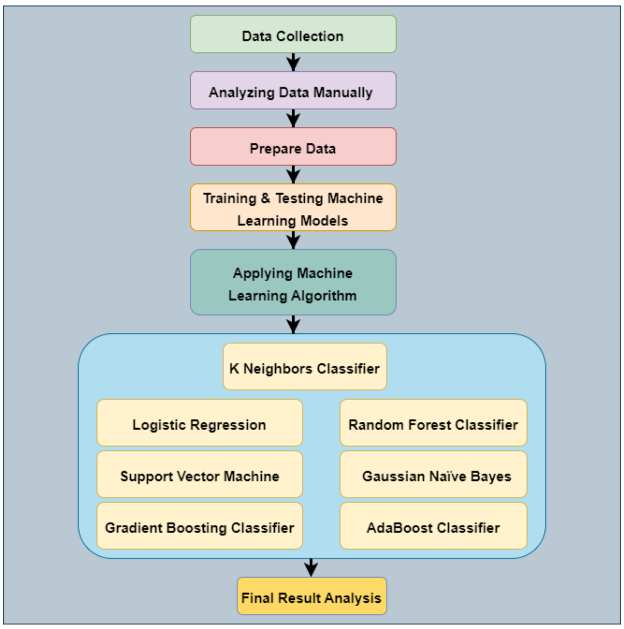



**Figure 3.1: Proposed Research Methodology**

## 3.2: Machine Learning:

Machine learning has applications in deferent area like Natural Language processing[47][48][49][50], medical[51], cyber security[52][53]. The approach of comprehending data, acquiring knowledge from it, and then making a decision or prognostication about any aspect of the environment based on that knowledge is machine learning at its most basic level. The computer is programmed to use vast amounts of data and algorithms to execute the work, rather than manually developing software processes with specific guidelines. Machine learning techniques have been used in several studies to study Computer Vision, AI-driven initiatives, and a range of other subjects. There are numerous methods intended to guide machine learning, and different algorithms are being developed to support a more comprehensive decision-making pathway.

## 3.3: Supervised Learning

In our investigation, the supervised learning methodology was used. It is feasible, for instance, to work with a data collection that already contains the inputs and the outputs. It denotes that the structure is aware of the solution and thus has made the decision based on it. Based on the most recent [26], supervised learning, often widely recognized as supervised machine learning or supervised ML, is a subcategory of artificial intelligence. Its distinctive feature is the use of annotated data sets to develop algorithms that accurately classify information or forecast events. As a core component of the cross-testing procedure, the model alters its parameters as input data is provided into it, continuing until the system is reassuringly fit. Employing supervised learning, a large number of actual problems are addressed.

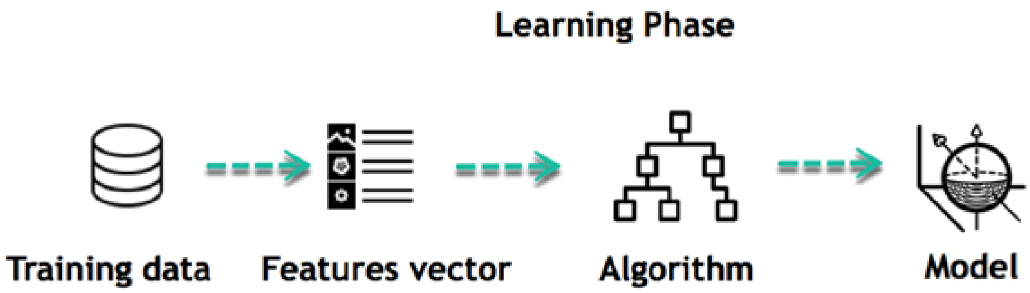



**Figure 3.2: Supervised Learning Workflow**

## 3.4: Considering Implemented Algorithms

To try and make sure that our thesis had much more precise findings than we could obtain using other methods, we applied a number of techniques. We employed a total of 7 algorithms.

### 3.4.1: Support Vector Machine:

One of the most well-liked algorithms among data scientists is this one. With decreased conversion efficiency, it can deliver exactness that is noticeable. Finding a subset of features in an N-dimensional region that can characterize the input variables in a distinctive way is the goal of the support vector machine approach [27]. To create a higher dimensional space, the Support Vector Machine approach chooses the sets of data that have the highest intense [28]. A classification model, often referred to as a higher dimensional space, is shown in the accompanying graphic and it is utilized to categorize two distinct classifications.

This approach, which uses supervised learning, is employed to detect changes and forecast outcomes in large datasets. The SVM makes an effort to justify the smallest performance of the classifier [40]. Also every result in the training examples is categorized into a certain category. Those relevant factors in the specified class are additionally distinguished by type. 4 distinct kernels—the longitudinal kernel, radial kernel, quadratic polynomial kernel, and activation functions used to categorize the identifiers [41]. To every type, these kernels offer a separate modification mechanism. The SVM is capable of doing synthetic data in addition to the kernels specified below. When the information cannot be mathematically differentiated in 2-D, this method is utilized. The characteristics are then transformed into 2-D with multiple classes [42].

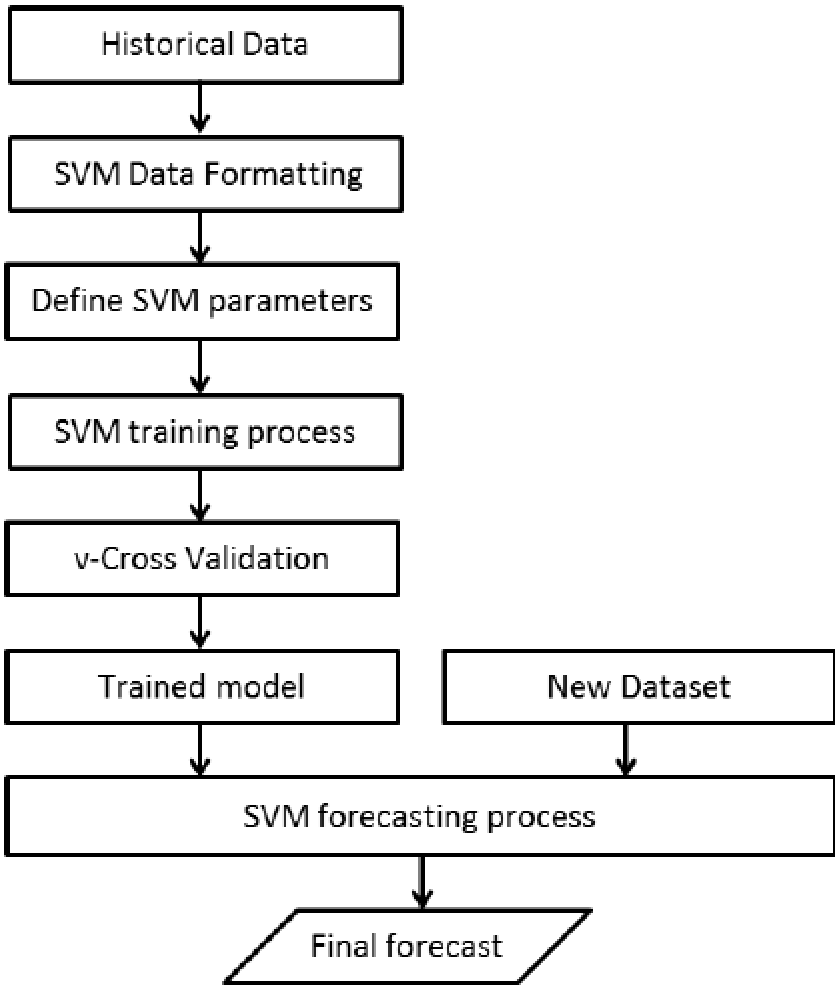



**Figure 3.3: Support Vector Machine Workflow**

## 3.4.2: Random Forest Classifier

One ensemble approach is the Random Forest classification algorithm. It incorporates constructing many tree structures at once using scaling, followed by aggregate, also known as bagging. With decision trees, we might encounter the overfitting problem, which raises uncertainty about the reliability of the information. To obtain the required data, Random Forest classifiers are utilized even though they guarantee that each clustering algorithm will be combined and generalized. Because of this, its efficiency is vastly enhanced [29].

Many Decision Tree algorithms make up an RF model. The technique has a solid track record in the regression analysis and classification sectors. A RF model is additionally utilized for image segmentation in addition to classification and regression analyses. This methodology assigns a characteristic a ranking based on how crucial it is for categorization. The rankings of each attribute can be determined by two approaches.

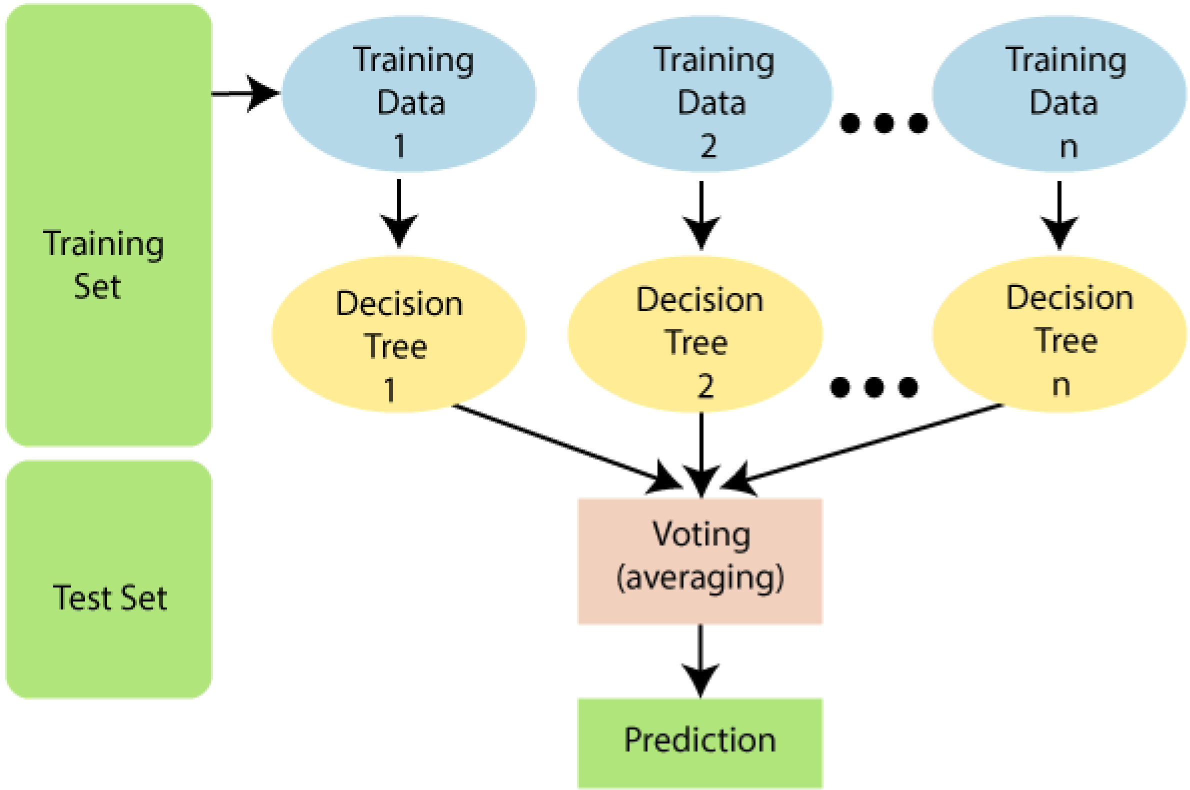



**Figure 3.4: Random Forest Classifier Workflow**

### 3.4.3: K-Neighbors Classifier

K-Nearest Neighbor determines the distance between several testing data and everything inside an attempt to identify which category the data getting examined belongs to, of the training corpus ought to be given. The next step is to select the K number of points which are most directly connected to the testing dataset. The K-Nearest Neighbor algorithm functions by calculating the probability that the test results will fall into any of the K groups of the practice data. The category with the maximum probability is then selected. It accomplishes that by using a set of statistics to determine the Distance metric among each set of binary points. Afterwards, the information that is most likely to be accurate is that which is to it with this neighboring cohort [30].

Given that it is the underlying and easiest procedure to employ at the beginning of such a dataset, this approach has been used first [43]. Depending on the closest point inside the training instance, K-NN identifies items [44]. An essential parameter is K. The neighboring size indicates how many training sample points need to be closest to an unknowable source of data in order to classify it [45]. If too K is chosen, the K-NN 13 structure would not perform as expected. This is due to how challenging the task will be for the algorithm to make the proper categorization. And besides, the data could include distortion or incorrectly tagged points. Since there are anomalies in the dataset, categorization will be less successful if K is set to a high value. To choose the right K for this investigation, two strategies have been used. The very first approach was choosing a randomized K and then seeing how the system reacts to it. In furthermore, an error margin and K value graph was created, and the K value with both the lowest mistake percentage was selected.

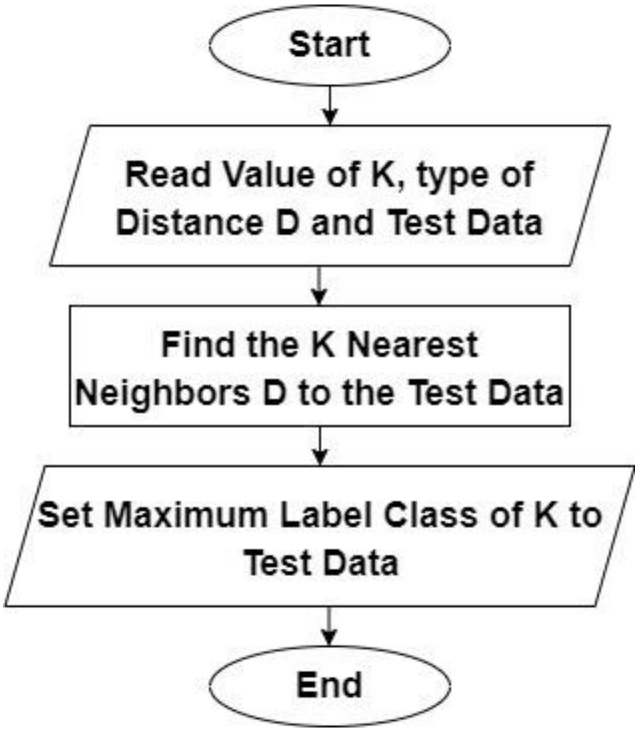



**Figure 3.5: K-Neighbors Classifier Workflow**

### 3.4.4: Naive Bayes Algorithm

One of its relatively simple Supervised Learning algorithms is the statistical classification model known as Naive Bayes. It is founded on the mathematics principle known as that of the Bayes Theorem. The Naive Bayes classifier is a method that is not always quick and moreover efficient and dependable, especially for large amounts of data [31].

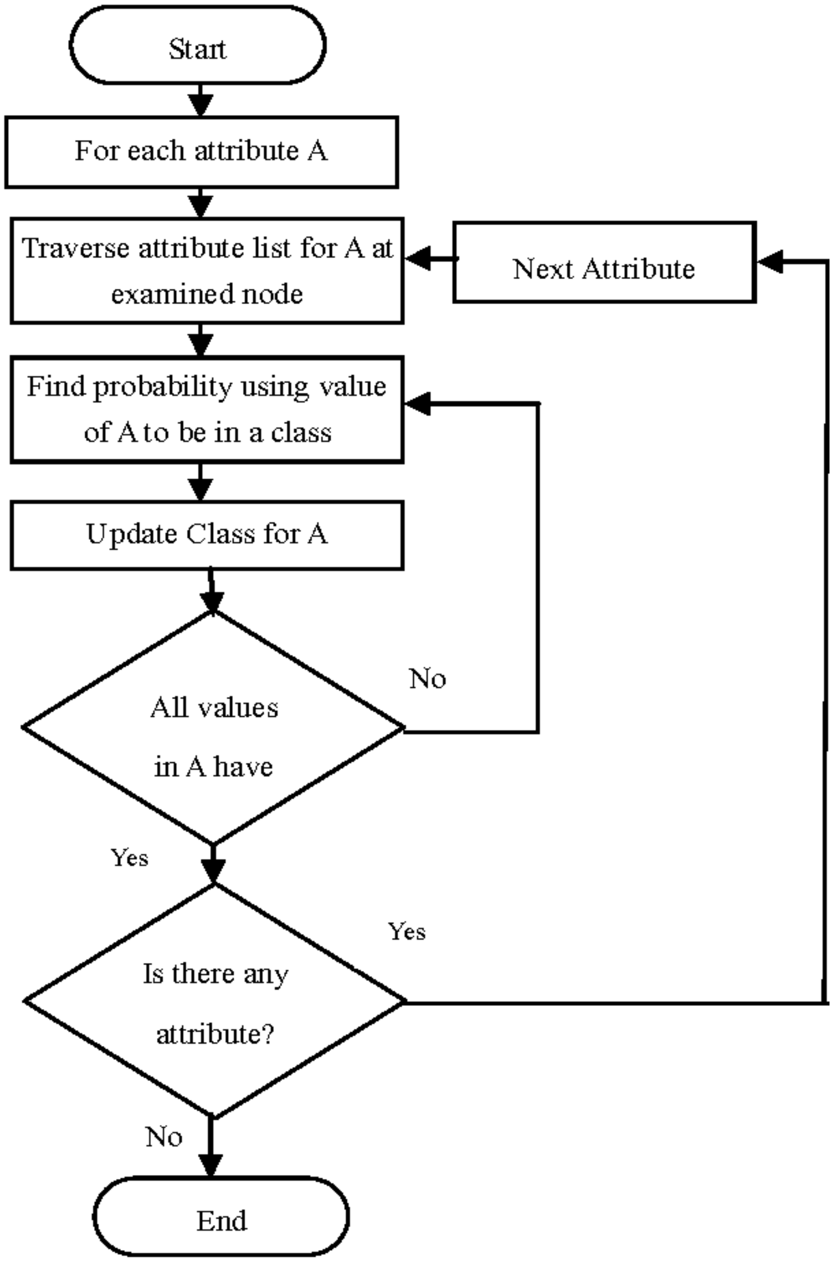



**Figure 3.6: Naive Bayes Algorithm Workflow**

### 3.4.5: Logistic Regression

A whole other statistical method that machine learning has adopted is logistic regression. Throughout this study, the standard approach will be based on logistic regression. The typical methodology used in this study will indeed be dependent on logistic regression. This is due to the employment of computational learning techniques in machine learning using advanced algorithms to search the information for connections and description [33]. However, it is useless to draw any conclusions from the material ahead, whichever version will produce better results. For instance, Lynam et al.'s [34] study indicates which Content consistency is greater with the LoR modeling than the selected machine, as evidenced by the ROC algorithmic learning.Even though a LoR approaches a simple linear regression approach for only binary regression models, it was selected for this study [46]. The ability of each model to handle mathematical expressions is thus a need for this work. Throughout this study, the foundation approach will indeed be based on multiple linear regression. This must have been selected due to the fact anticipated that machine-learning algorithms will outperform mathematical tools in this situation. This seems to be due to the fact that computational complication algorithms are used by machine learning techniques to search for correlations and recurring themes in the data. However, it is difficult to predict in advance whichever paradigm will produce the greatest results based on the available research.

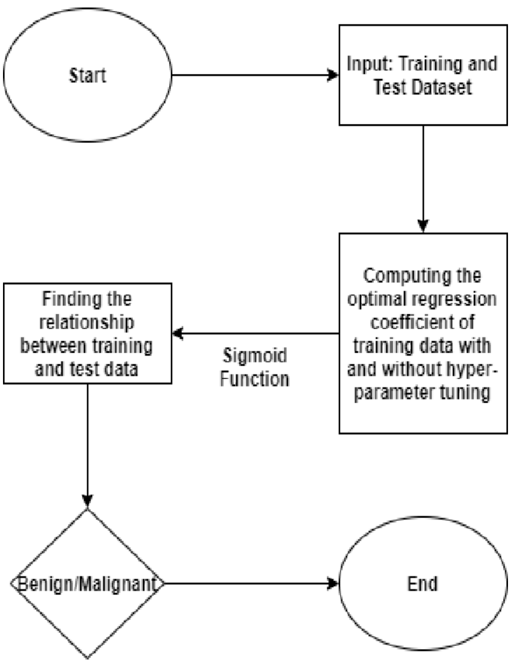



**Figure 3.7: Logistic Regression Workflow**

### 3.4.6: Gradient Boosting Classifier

A highly efficient machine learning approach for categorization is gradient boosting and issues with coefficient of determination. Continually upgrading itself based on the preceding facts from relatively weak assumptions. Following a given number of simultaneous permutations, it produces accurate predictions. In essence, it looks for the prior error and deletes it if it happens afterwards [32].

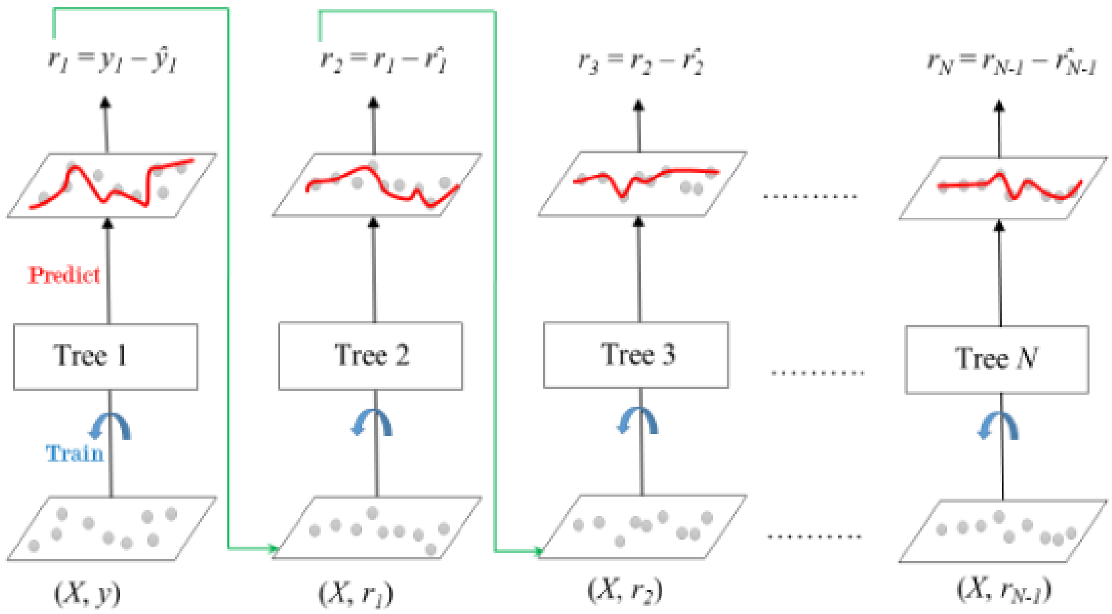



**Figure 3.8: Gradient Boosting Classifier**

### 3.4.7: AdaBoost

AdaBoost, also known as Adaptive Boosting [35], is a machine learning methodology applied as an optimization learning. Decision trees including one stage, or Decision trees with just a single split, are the most commonly used algorithms for AdaBoost. This algorithm creates a model while assigning each data piece an appropriate amount. Then, it was an information framework that incorrectly categorized equal weight. The very next algorithm now lends weight to all the elements with higher weights. If no lower error is encountered, it will continue to train the classifiers.

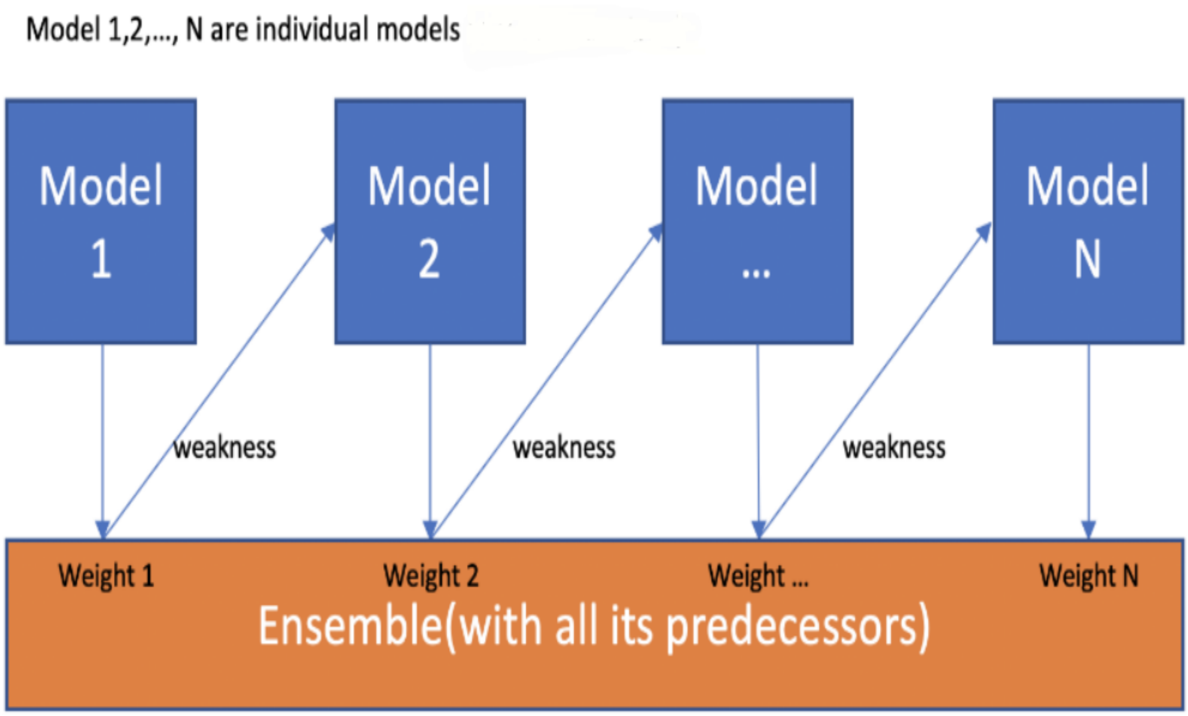



**Figure 3.9 AdaBoost Algorithm**

# Chapter 04

## Experimental Evaluation

The datasets and even the procedures used to analyze the information are described in this subsection. The development of the classifications, as well as data preprocessing algorithms is also explained in great depth. In-depth explanations are provided for the creation of the classifications, and data pretreatment procedures. We have selected 17 attributes to assess the accuracy of our predictions approach and the models. The 17 attributes we selected are relatively common, and the bulk among them were suggested by the supervisor and in addition to what we learned via in-depth research readings.

### 4.1: Data Collection and Preparation

The hardest and longest portion of our thesis proposal was gathering the data. So, we would just have to create our entire dataset. We have gathered different standardized scaled questionnaires for the predominant psychiatric disorder, obsessive thinking, including panic attacks due to this particular reason. The real valid questionnaires were gathered from several study papers, which were discovered through careful research studies. The depressive scale's compatibility to Bangladesh's environment and society is one of its primary selling points. Each of these survey questions include several questions about living that are relevant to the manner of life experienced by the individuals in this society.

### 4.1.1: Survey Form

A structured questionnaire form was created by us to get information from the graduates. We reasoned that collecting data straight from the vast majority of unemployed graduates would enable us to develop a comprehensive picture of the various psychological phenomena of Bangladeshi recent graduates. The research problem noted that we were required to design the questionnaire survey so that:

1. The inquiries are simple to comprehend.
2. They aren't too intimate to respond to
3. There ought to be sufficient inquiries so that together we can trace the actual statistics.

4. The answers shouldn't contain the doubts.

5. Lastly, the inquiries must strike a balance between someone being lengthy and brief, slightly personal in relation to depression, while also not being rude or offensive.

It can be inferred from the aforementioned criteria that developing the inquiries was not simple. We had to refer to earlier research in pertinent fields as a consequence. The questions were designed in a way that let us collect detailed information from unemployed graduates about the root of their symptoms, the techniques they used to attempt to feel more confident, and the prospective techniques for overcoming fear if they succeeded. It must have been highlighted that no personally identifiable information about the graduates will be gathered, guaranteeing their anonymity and their willingness to help us. Given that it is impossible to track someone's Google form response while acknowledging their name, email address, or other important details, we adopted additional security measures to preserve the contributors' anonymity.

### 4.2: Analyzing Data Manually

Data collection from survey forms was generally simple entirely due to Google Form. With the Online Questionnaires Survey, we employed integrated Google Sheet Response. It is simpler to extract information and examine through using Google Sheets. We would just have to trace every participant, thus we had to use a strategy that allowed us to constantly analyze, improve, and alter the database while keeping the ongoing process intact. Implementing Google Cloud Platform represents one of Google Sheets' enabling technologies. There were no alternative viable solutions other than using embedded Google Sheet Response because we had to export the datasets to Python in a sequence of steps. To protect the participants' confidentiality, we haven't shared and won't share any of the data we receive through the form with anybody else.

### 4.3: Machine Learning Tool and Extract Survey Data

We have been using the Google Sheet to gather form submissions in order to export data on live. Then, we obtained the responses from the Google Sheet using the Python environment provided by Jupyter Library. The data gathered from the answer has been examined by the most well-known machine learning library. The main factors—basic competency, data integrity, data pre-processing, task accomplishment for data mining, implementation of machine learning algorithms, model interpretations, assessment, and visualization—are the foundation of a strong machine learning system.

- Python Language
- Python Libraries (NumPy, Pandas, Matplotlib, Scikit-Learn etc.)
- Jupyter Note
- Google Colab
- Microsoft Excel

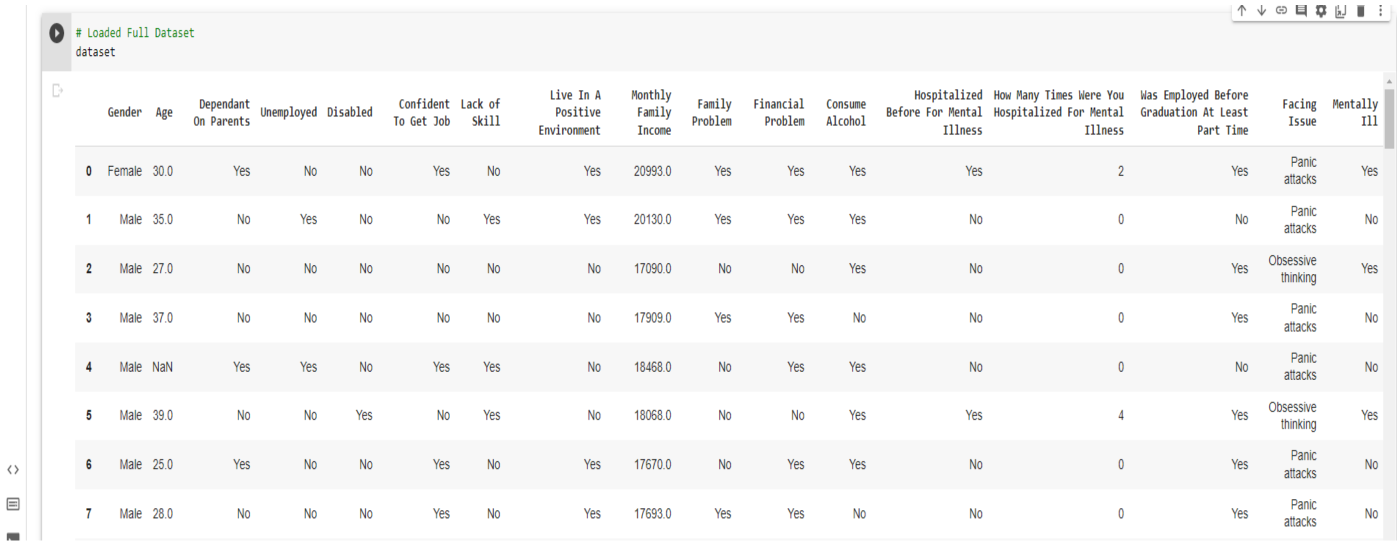
## 4.4: Dataset Description

We used legitimate questions to evaluate their mental wellbeing as part of this assessment. We had already asked them about 17 different things, starting with their age, gender, skills and so on. Furthermore, we pose more current and recent concerns about depression. The questions were constructed using the techniques previously discussed, but we borrowed the questions from research and added them to our own questionnaire. A participant was asked multiple choice questions, the subjects of which were the highest likely causes of their depression.

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 335 entries, 0 to 334
Data columns (total 17 columns):
 #   Column                                         Non-Null Count  Dtype
---  ------                                         --------------  -----
 0   Gender                                         335 non-null    object
 1   Age                                            327 non-null    float64
 2   Dependant On Parents                           335 non-null    object
 3   Unemployed                                     335 non-null    object
 4   Disabled                                       335 non-null    object
 5   Confident To Get Job                           335 non-null    object
 6   Lack of Skill                                  335 non-null    object
 7   Live In A Positive Environment                 219 non-null    object
 8   Monthly Family Income                          308 non-null    float64
 9   Family Problem                                 335 non-null    object
 10  Financial Problem                              335 non-null    object
 11  Consume Alcohol                                335 non-null    object
 12  Hospitalized Before For Mental Illness         335 non-null    object
 13  How Many Times Were You Hospitalized For Mental Illness  335 non-null    int64
 14  Was Employed Before Graduation At Least Part Time  335 non-null    object
 15  Facing Issue                                   335 non-null    object
 16  Mentally Ill                                   335 non-null    object
dtypes: float64(2), int64(1), object(14)
memory usage: 44.6+ KB
```

**Figure 4.1: Dataset Attribute Information**



```
# Loaded Full Dataset
dataset
```

| | Gender | Age | Dependant On Parents | Unemployed | Disabled | Confident To Get Job | Lack of Skill | Live In A Positive Environment | Monthly Family Income | Family Problem | Financial Problem | Consume Alcohol | Hospitalized Before For Mental Illness | How Many Times Were You Hospitalized For Mental Illness | Was Employed Before Graduation At Least Part Time | Facing Issue | Mentally Ill |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Female | 30.0 | Yes | No | No | Yes | No | Yes | 20993.0 | Yes | Yes | Yes | Yes | 2 | Yes | Panic attacks | Yes |
| 1 | Male | 35.0 | No | Yes | No | No | Yes | Yes | 20130.0 | Yes | Yes | Yes | No | 0 | No | Panic attacks | No |
| 2 | Male | 27.0 | No | No | No | No | No | No | 17090.0 | No | No | Yes | No | 0 | Yes | Obsessive thinking | Yes |
| 3 | Male | 37.0 | No | No | No | No | No | No | 17909.0 | Yes | Yes | No | No | 0 | Yes | Panic attacks | No |
| 4 | Male | NaN | Yes | Yes | No | Yes | Yes | No | 18468.0 | No | Yes | Yes | No | 0 | No | Panic attacks | No |
| 5 | Male | 39.0 | No | No | Yes | No | Yes | No | 18068.0 | No | No | Yes | Yes | 4 | Yes | Obsessive thinking | Yes |
| 6 | Male | 25.0 | Yes | No | No | Yes | No | Yes | 17670.0 | No | Yes | Yes | No | 0 | Yes | Panic attacks | No |
| 7 | Male | 28.0 | No | No | No | Yes | No | Yes | 17693.0 | Yes | Yes | No | No | 0 | Yes | Panic attacks | No |

**Figure 4.2: Dataset**

335 people who were at least 18 years old responded to the questionnaire in total. 18 questionnaires, containing ones about the two conditions such as depression brought on by panic attack, obsessive thinking, were instructed by the respondents. The characteristic "unemployment" was added to the list of few for health-related mental and physical concerns, and it was utilized to make predictions. The results revealed that, out of the 335 respondents, 80 had received a formal diagnosis of a mental condition. The characteristics that were picked are all binary types of data. For instance, the unemployment attribute contains the information about the respondent's job or joblessness. The strings "yes" and "no" were used to convey this. In this instance, "yes" denotes employment and "no" denotes unemployment. The dataset's remaining predictive factors also got the same approach.

## 4.5: Data Cleaning and Preprocessing

One dataset, used in this investigation. The primary step was to comprehend and filter the dataset because there was no requirement to integrate disparate datasets. To load the Excel sheets into Python, this was done. Understanding the statistics was the preliminary step. This involved examining the overall number of people who were unemployed, the proportion of participants who indicated experiencing mental or physical health issues, and the age categories that are reflected in the dataset.

After a long period of time, we have eventually compiled a total of 335 responses from people surveyed who are graduates. Keeping this in consideration, we subsequently move the dataset to the research and development stage. Whenever you implement the techniques to judge the accuracy of the data, we went through the entire data, evaluated for any potential abnormalities, and then eliminated them from our data set. With this same data, we choose to move on to the pre-processing stage.

### 4.5.1: Filling Null Value

From the dataset, a total of 151 values were missing. By first identifying the frequency of the responses and then putting the most common one, the imputation approach we utilized fills in the missing data. Here in the Figure, we can see the missing data and cleaned data in those figures.

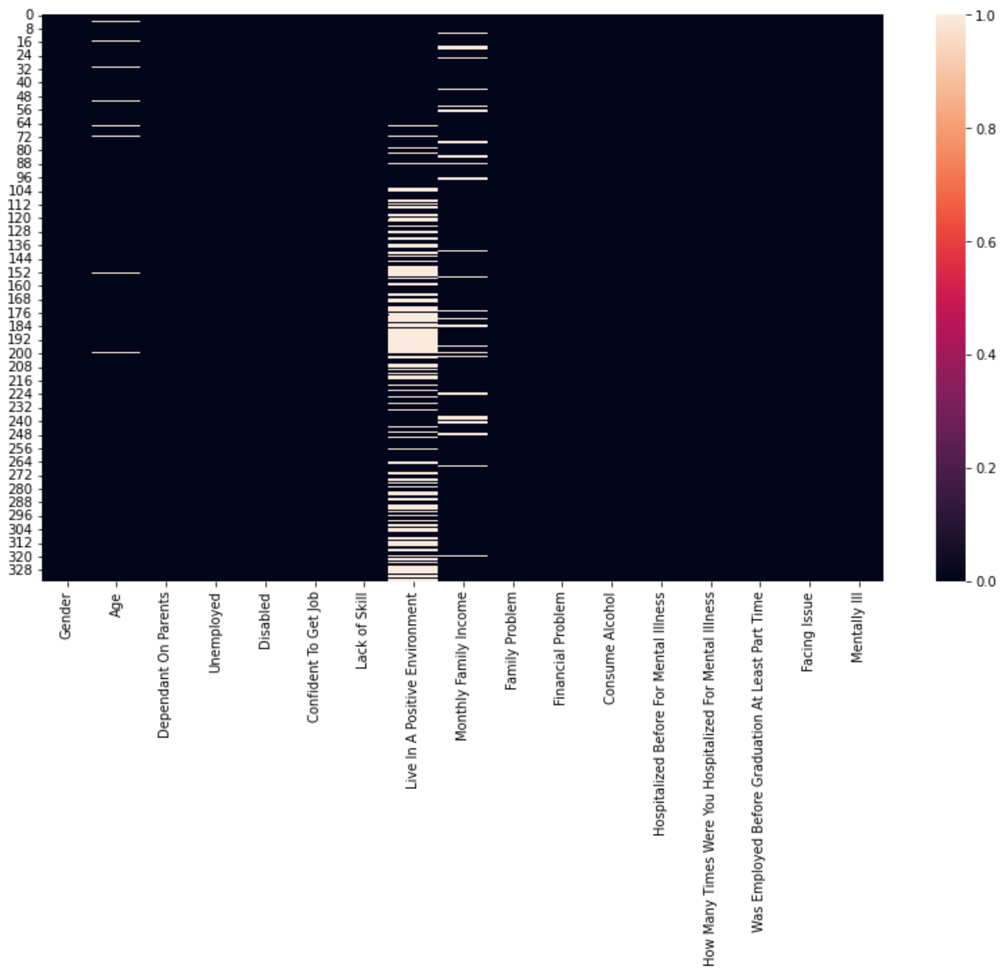



**Figure 4.3: Unprocessed Dataset**

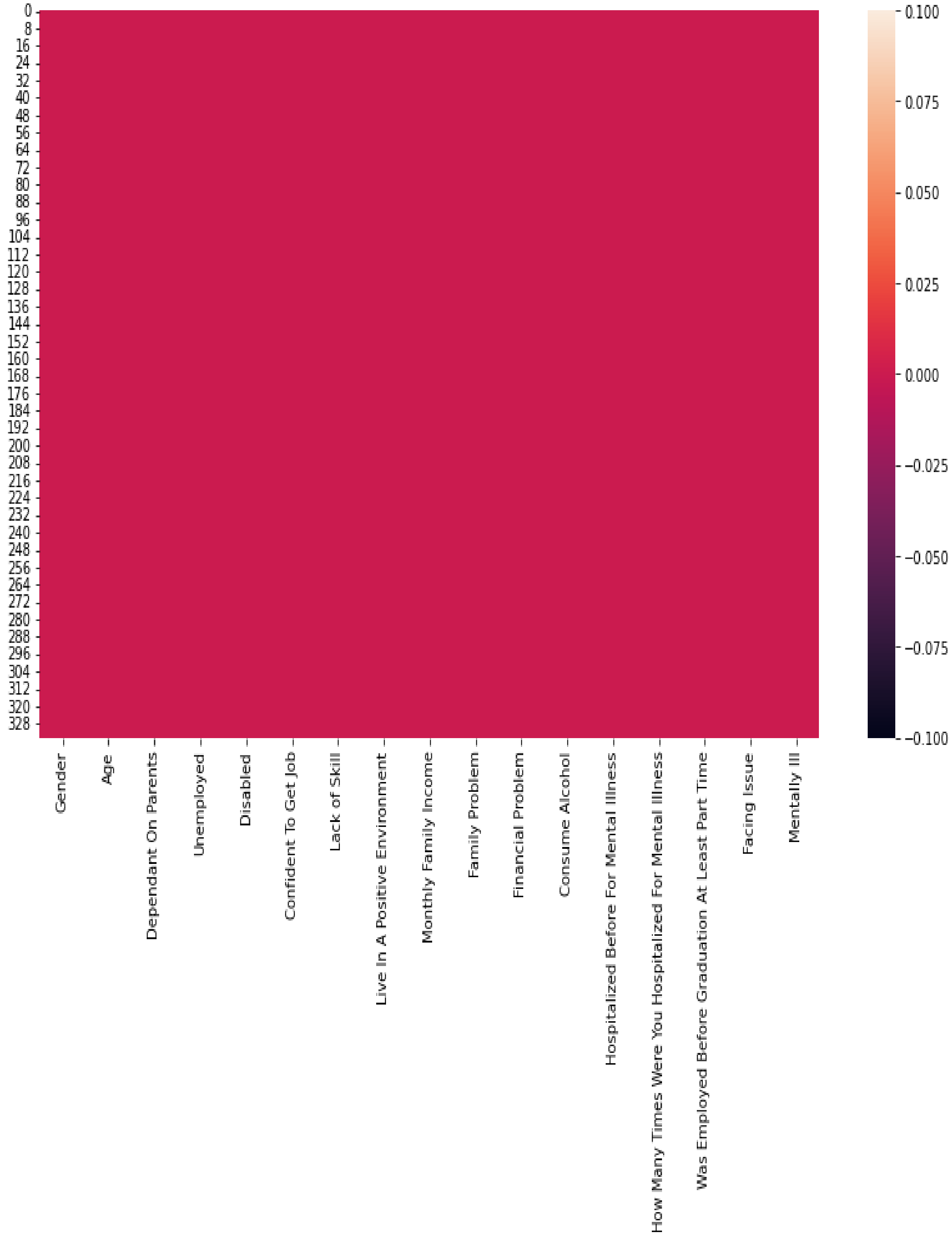



**Figure 4.4: Cleaned Dataset**

### 4.5.2: One Hot Encoding

The category variables were transformed into binary numeric variables using the one-hot encoding technique. The dataset under this study includes "yes" and "No" descriptions including all variables that are applicable. For "Yes" and "No," these initially needed to be encoded to 1s and 0s. This is required so that algorithms for selecting and classifying features can employ the data.

### 4.5.3: Feature Scaling

We balance these out to make sure our algorithms for machine learning perform properly and because our dataset contains certain properties with greater values. There are multiple machine learning techniques that require feature scaling to be done. It is performed in order to comply with extremely changeable degrees, values, or units.

## 4.6: Detailed Data Visualization

In the interest of comprehending the data, visualization is required. Because we have now used up the required volume of content, this section is labeled "Detailed Data Visualization." In good enough condition to see greater variety, we plan to gather additional information. Furthermore, we have found some connections between most of the information we have currently collected.

On the following chart (Fig. 4.5), depression status (Are you dealing with mental health-related concerns?) is displayed for participants to respond to. From where we can observe that among 52 percent of participant facing ‘Obsessive Thinking” and 48 percent participants are dealing with ‘Panic Attacks”.

Due to a gap among both skills gained (supply) and skills needed, a shortage of suitable skills leads to underemployment. The motivation for these students to pursue higher education is typically provided by their families, who think that doing so reduced the chance that they will become unemployed. But after completing graduation, many graduates do not manage a job because lack of skills which is clearly visible from Fig. 4.6.

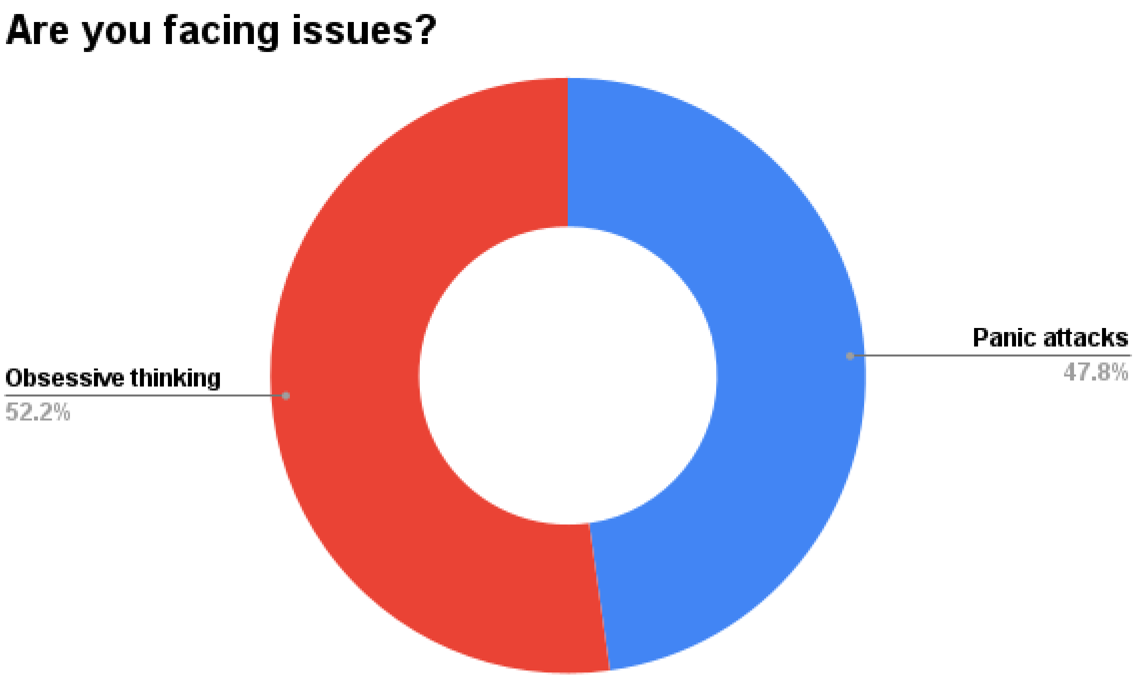



**Figure 4.5: Depression and Facing Issue**

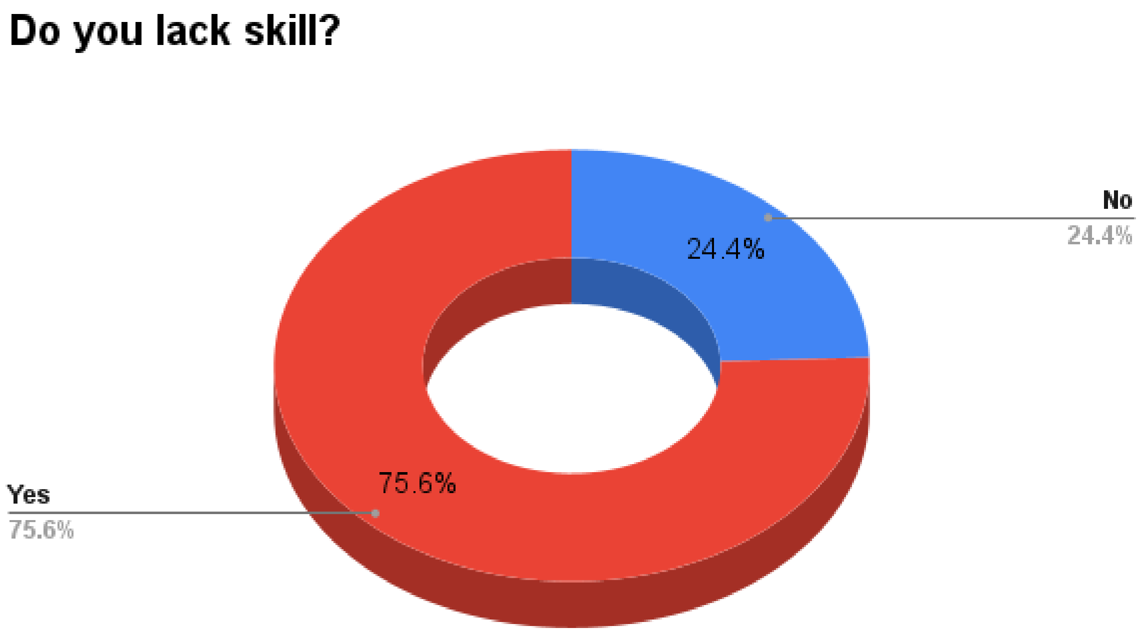



**Figure 4.6: Depression and Lack of Skill**

We regret to inform you that graduates have been severely impacted by the depressingly high unemployment rate in this generation. From Fig. 4.7, it is shown that out of 50 percent are depressed because of they are unemployed.

From our survey, we observed from Fig. 4.8 that, females are mentally sound that males. On the other hand, males are less depressed than females where as the reason can be society, family pressure and so on. Also from Fig. 4.9, we can see that graduates are finding work in some capacity, and competition is equal for men and women.

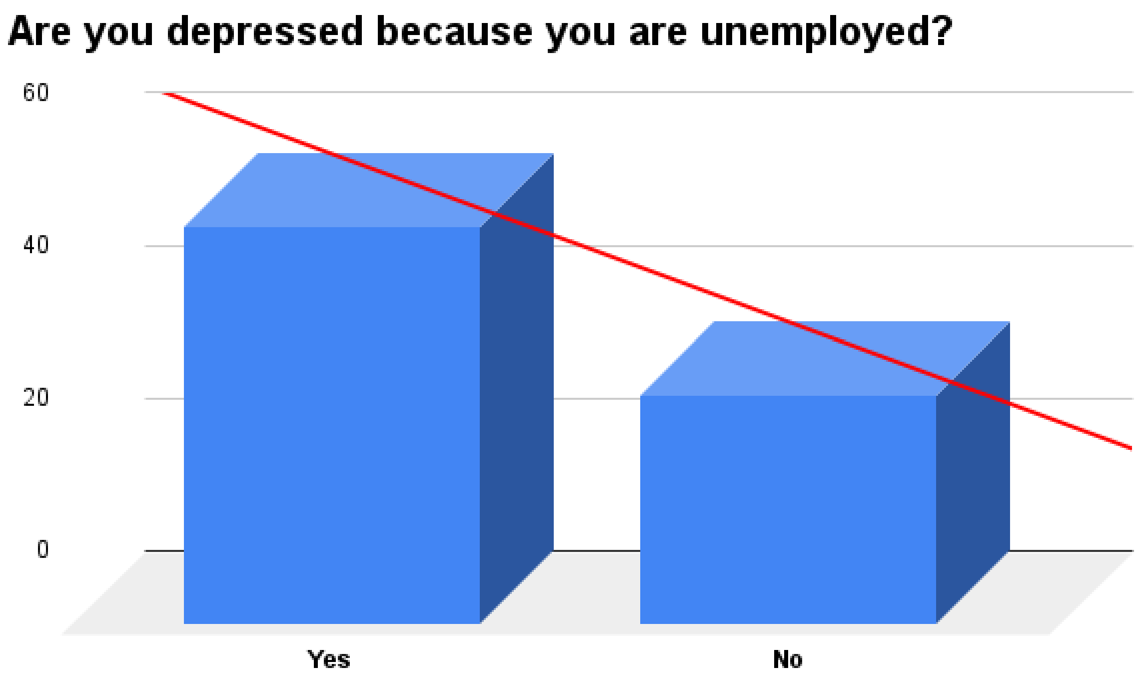



**Figure 4.7: Depression and Unemployment**

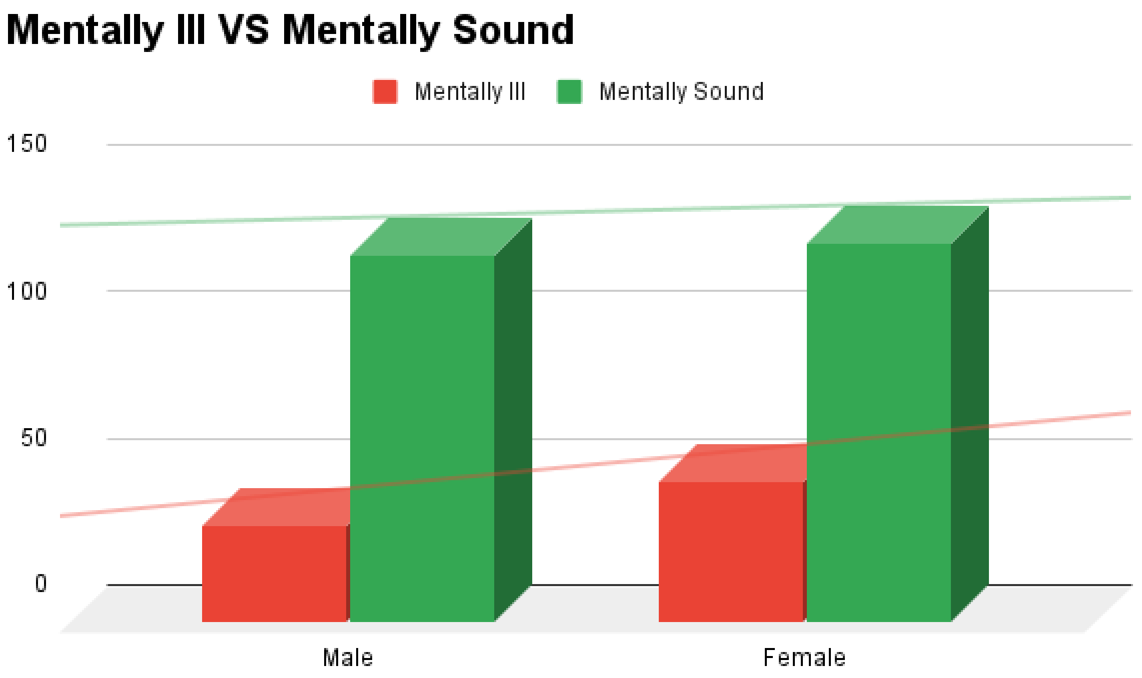



**Figure 4.8: Mentally Ill VS Mentally Sound**

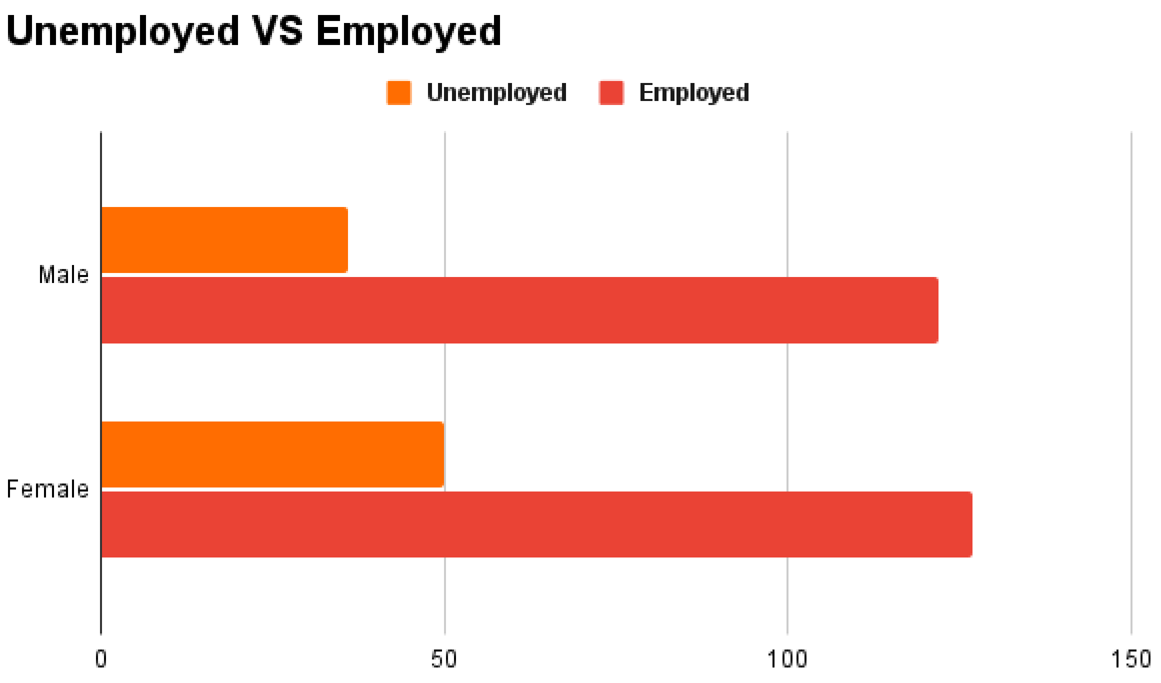



**Figure 4.9: Unemployed VS Employed**

On the following chart (Fig. 4.10), depression status (Mentally ill Vs dependency on parents) is displayed. From where we can observe that among 20 percent of participant facing mental depression and they are dependent on parents and 80 percent of the participants are mentally ill and not dependent on parents. The reason can be after the graduation; they are still unemployed and there might be some pressure on them from family and society. The participants are dealing with mental illness but not dependent on parents is higher than that of who are dependent on parents because they are more worried and overthinking about their future.

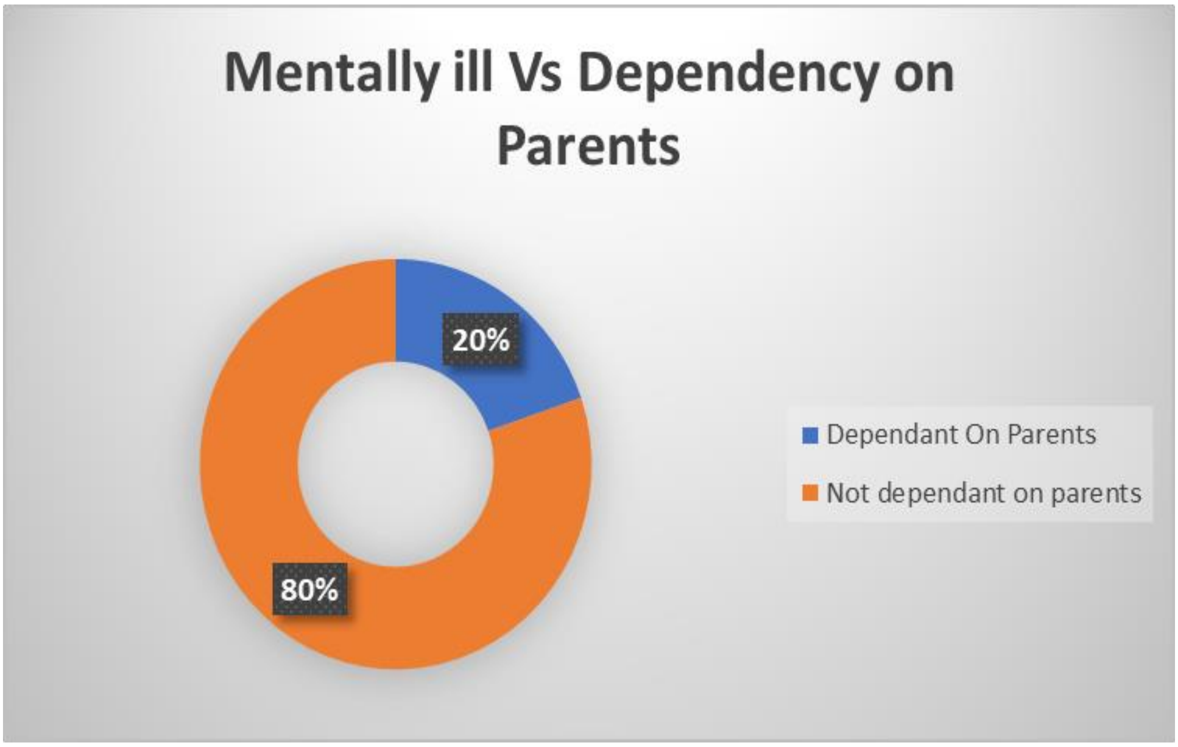



**Figure 4.10: Mentally ill Vs Dependency on Parents**

From our survey, we observed from Fig. 4.11 that, only 30 percent of the participants suffering from mental illness has hospitalized before for mental illness. On the other hand, most of the participants almost 70% have not hospitalized before for mental illness. From Fig. 4.10, we can say that most of the people are unaware or not getting proper treatment for their mental illness.

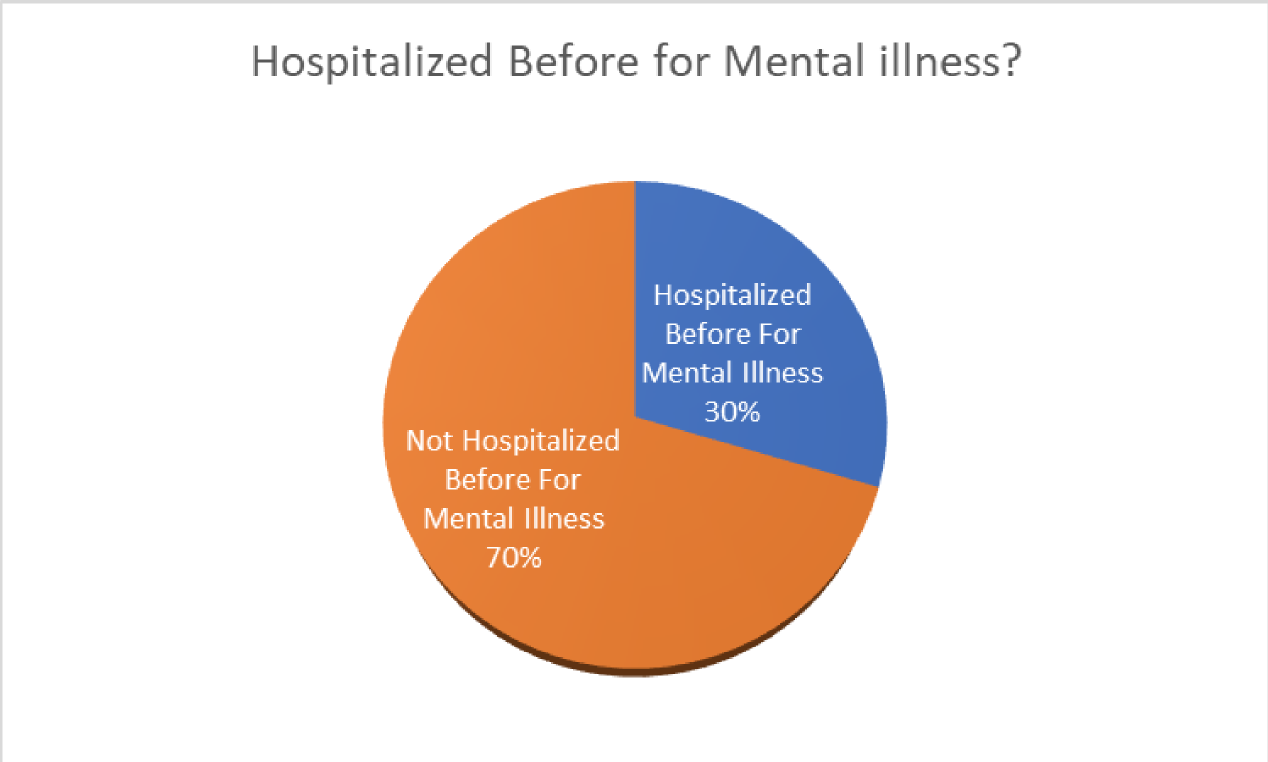



**Figure 4.11: Hospitalized and Mental illness**

From figure 4.12 we can see 22% of the participants are having mental illness because of their disability but other 78% are not disabled and still unemployed and facing mental issues. So, we can say that, physical disability is not the main reason for their mental illness but the unemployment, depression and family reasons are the key factor of mental illness.

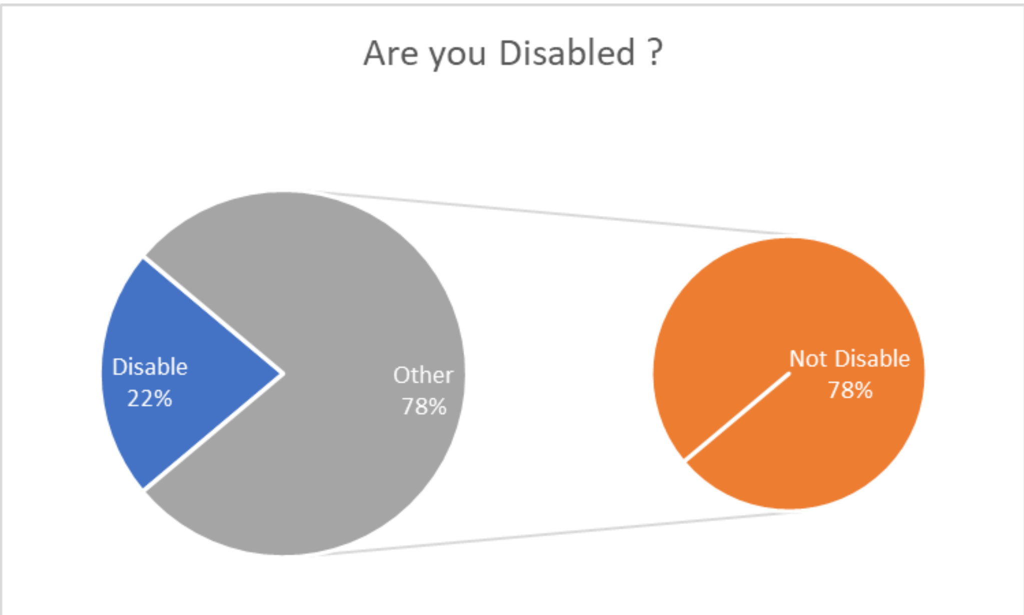



**Figure 4.12: Disability and Mental health**

# Chapter 5

## Result Analysis

Right now, our primary concern is predicting depression among BD graduates. We used the method that was explained to accomplish a high level of accuracy because the dataset we have there is multi-classified. We successfully examined the accuracy, precision, recall, and f-measures of the modeling techniques that we have proposed using seven distinct algorithms. Because the effectiveness of the model may depend on how precise the predictions are and how excellent their f-measures are, these criteria are crucial.

## 5.1: Performance Metrics

- **Accuracy**

Accuracy (A) in multiclass classification refers to the overall frequency of conclusions reached by the system out of all possible guesses.

A = (TP + TN) / (TP + FP + FN + TN)

Where

TP = true positive value,

TN = true negative value,

FP = false positive value and

FN = false negative value.

The individual's forecasts are divided into two categories: correct forecasts (TP and TN), and all other forecasts (TP AND FP)

- **Precision**

Precision (P) is the measuring indicator that counts both TP and FP strategies in line that are successfully predicted from all forecasts in the positive class.

P = (TP) / (TP + FP)

The value of 1 denotes the greatest precision grade, while the value of 1 indicates the very worst precision marks.

- **Recall**

Recall (R) is the parameter that evaluates the percentage of positive patterns that are successfully categorized.

R = (TP) / (TP + FN)

Whenever recall is 1, we can claim that perhaps the outcome has the highest precision; yet, when recall is 0, we might state that the outcome is the least effective.

- **F1-Score**

The effectiveness of a model on a dataset is evaluated by the F-score, also known as the F1-score. It's employed to assess binary categorization schemes that label examples as "positive" or "negative."

F = 2 x [(PRECISION x RECALL) / (PRECISION + RECALL)]

- **Confusion Matrix**

In order to figure out both the number of TP examples and the number of FP cases, we created a confusion matrix. We employ a classification scheme with various classes. There have been two possible results in our scenario, one for each possible situation. Whichever among these two answers—"yes," or "no”.

Therefore, we employ a two-by-two grid as even the confusion matrix for every approach. For present example, the confusion matrix is shown in Figure 5.1.

| | | **Predicted** | |
|---|---|---|---|
| | | Negative **(N)**<br>**-** | Positive **(P)**<br>**+** |
| **Actual** | Negative<br>**-** | True Negative **(TN)** | **False Positive (FP)**<br>**Type I Error** |
| | Positive<br>**+** | **False Negative (FN)**<br>**Type II Error** | True Positive **(TP)** |

**Figure 5.1 Confusion Matrix for 2 Classes**

## 5.2: Results Obtained

The analysis of our dataset involved seven distinct classification methods. We choose to utilize the metrics that have been the most generally used and accepted in order to examine categorizing systems. Check out Table 5.1 for the accuracy scores of various algorithms.

| Algorithm | Accuracy |
|---|---|
| Random Forest Classifier | 80.6% |
| Gaussian Naive Bayes | 73.13% |
| K-Neighbors Classifier | 77.61% |
| Logistic Regression | 80.6% |
| Support Vector Machine | 79.1% |
| Gradient Boosting Classifier | 80.6% |
| AdaBoost | 76.12% |

**Table 5.1: Accuracy Score of Various Algorithms**

### 5.2.1: Random Forest Classifier

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.83. Here, we discovered that 52 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 11. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were 2 false positives

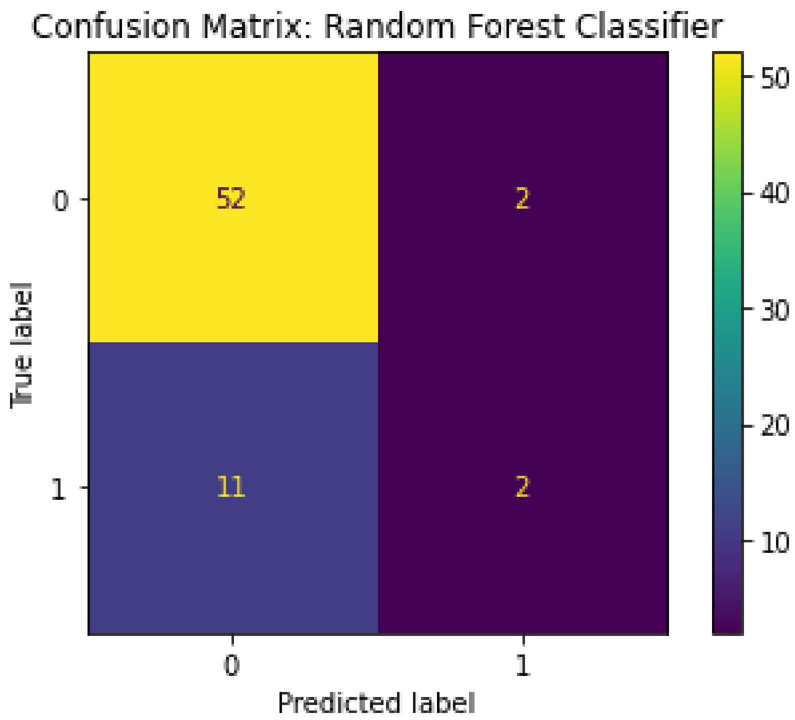



**Figure 5.2: Confusion Matrix: Random Forest**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.83 | 0.96 | 0.89 |
| 1 | 0.50 | 0.15 | 0.24 |
| Macro Avg. | 0.66 | 0.56 | 0.56 |

**Table 5.2: Classification Report: Random Forest**

### 5.2.2: Gaussian Naive Bayes

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.80. Here, we discovered that 48 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was

12. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were a total of 6 false positives.

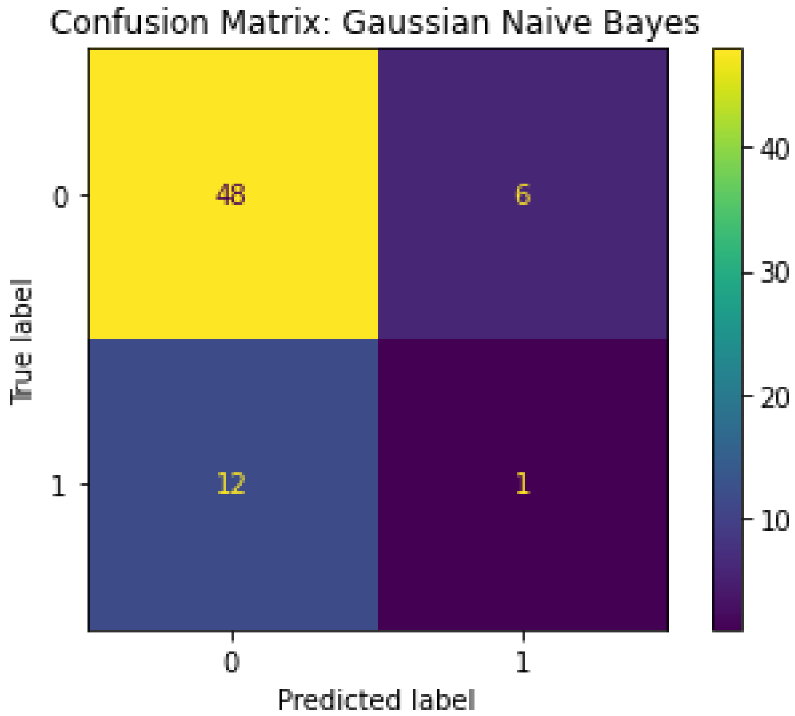



**Figure 5.3: Confusion Matrix: Gaussian Naive Bayes**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.80 | 0.89 | 0.84 |
| 1 | 0.14 | 0.08 | 0.10 |
| Macro Avg. | 0.47 | 0.48 | 0.47 |

**Table 5.3: Classification Report: Gaussian Naive Bayes**

### 5.2.3: K-Neighbors Classifier

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.82. Here, we discovered that 50 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 11. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were a total of 4 false positives.

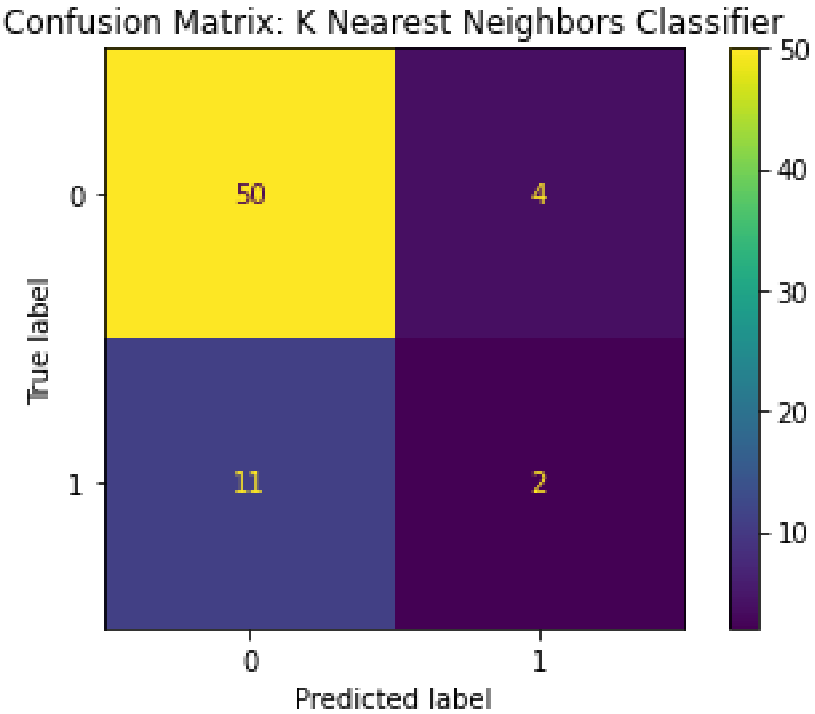



**Figure 5.4: Confusion Matrix: K-Neighbors Classifier**

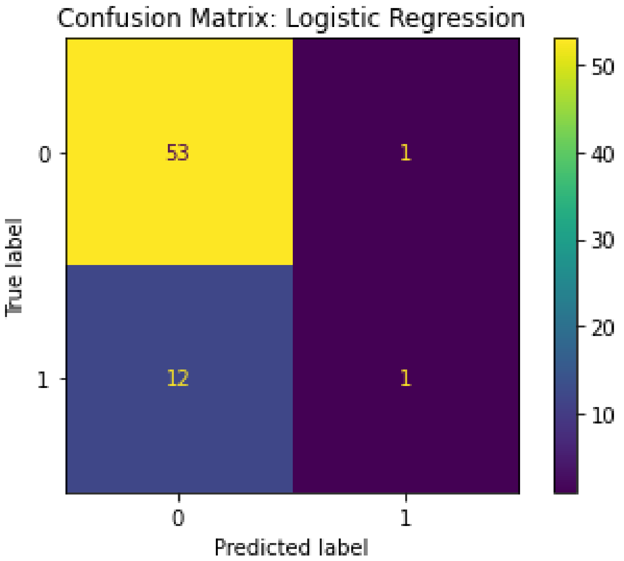
Classification Report of KNN:

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.82 | 0.93 | 0.87 |
| 1 | 0.33 | 0.15 | 0.21 |
| Macro Avg. | 0.58 | 0.54 | 0.54 |

**Table 5.4: Classification Report: K-Neighbors Classifier**

### 5.2.4: Logistic Regression

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.82. Here, we discovered that 53 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 12. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were a total of 1 false positive.




**5.5: Confusion Matrix: Logistic Regression**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.82 | 0.93 | 0.87 |
| 1 | 0.33 | 0.15 | 0.21 |
| Macro Avg. | 0.58 | 0.54 | 0.54 |

**5.5: Classification Report: Logistic Regression**

## 5.2.5: Support Vector Machine

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.82. Here, we discovered that 52 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 12. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were a total of 2 false positives.

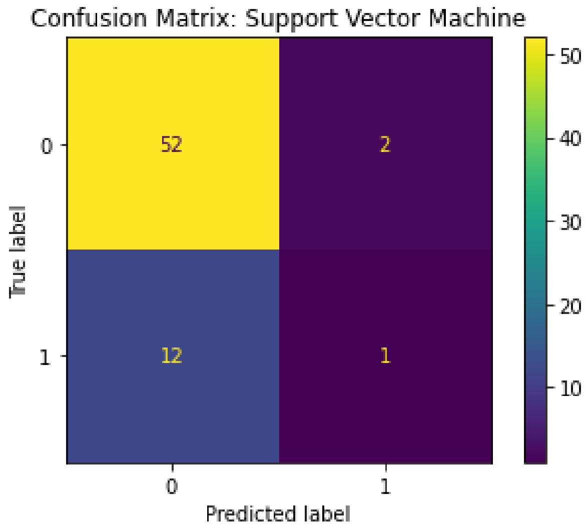



**Figure 5.6: Confusion Matrix: Support Vector Machine**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.82 | 0.93 | 0.87 |
| 1 | 0.33 | 0.15 | 0.21 |
| Macro Avg. | 0.58 | 0.54 | 0.54 |

**Table 5.6: Classification Report: Support Vector Machine**

## 5.2.6: Gradient Boosting Classifier

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.81. Here, we discovered that 54 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 13. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were 0 false positives.

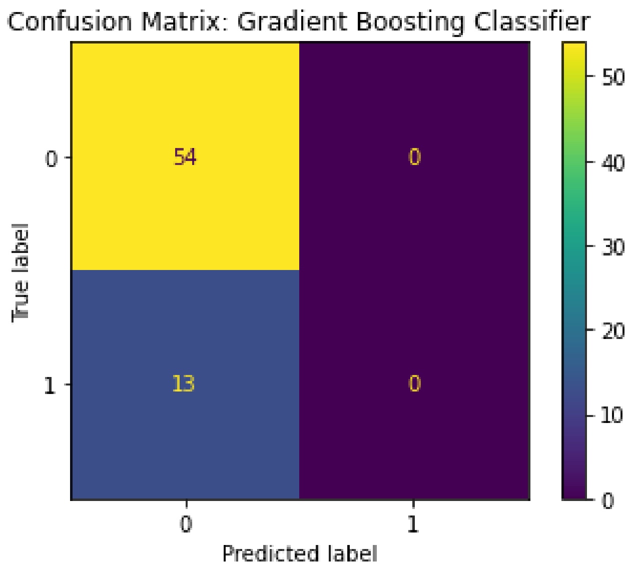



**Figure 5.7: Confusion Matrix: Gradient Boosting Classifier**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.81 | 1.00 | 0.89 |
| 1 | 0.00 | 0.00 | 0.00 |
| Macro Avg. | 0.40 | 0.50 | 0.45 |

**Table 5.7: Classification Report: Gradient Boosting Classifier**

## 5.2.7: Ada Boost

The confusion matrix as well as classification report provide us a clear understanding. By dividing the total number of TP by the product of the overall number of TP and the overall number of FP, we have been capable of determining the precision to be 0.82. Here, we discovered that 49 graduates in all had genuinely excellent outcomes. that individuals actually do experience depression. The overall number of false negatives we discovered was 11. The recall was then determined by dividing the total number of true positives by the sum of the true positive and false negative values. In this instance, there were a total of 5 false positives.

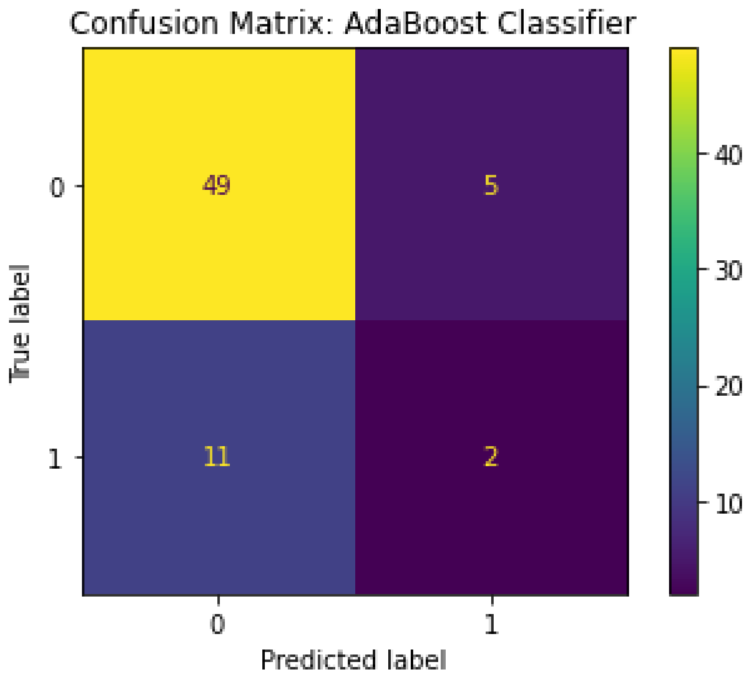



**Figure 5.8: Confusion Matrix: AdaBoost**

| | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 0.82 | 0.91 | 0.86 |
| 1 | 0.29 | 0.15 | 0.20 |
| Macro Avg. | 0.55 | 0.53 | 0.53 |

**Table 5.8: Classification Report: AdaBoost**

## 5.3: Experimental Discussion:

After reviewing the information gathered through all of the algorithms and models used across 17 features, we found that three of the methods are showing a remarkable similarity with each other in figure 5.9. The highest degree of 80.6% accuracy was shown by Random Forest, Logistic Regression, Gradient Boosting Classifier. In this case, Support Vector Machine delivered outcomes that were 79.1% more accurate than K-Neighbors Classifier, Ada Boost, Gaussian Naive Bayes which delivered results that were 77.1%, 76.12%, 73.13%, of accuracy. We are fully conscious that a higher f-measure implies a system's better accuracy as well as its general performance.

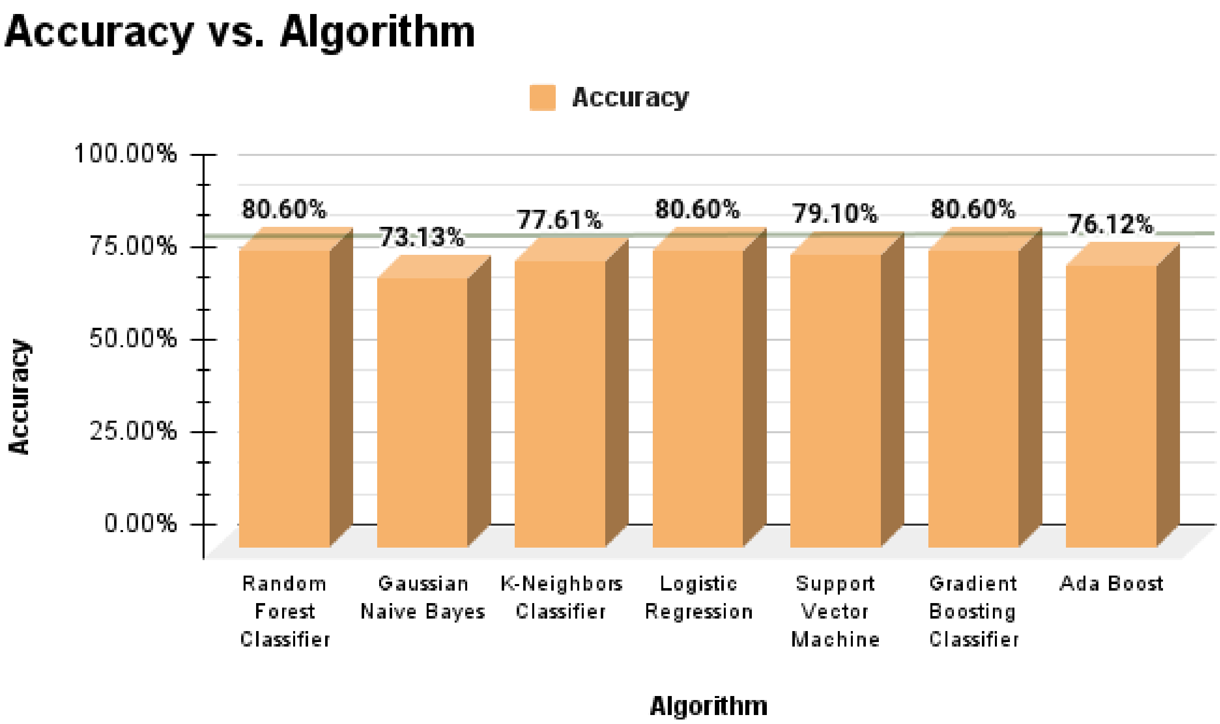



**Figure 5.9: Comparison: Accuracy VS Algorithm**

Additionally, we can see that False Positive and False Negative are nicely balanced. Moreover, we may infer from either the accuracy, recall, and f-measure in Fig. 5.10 indicating that in both cases, Random Forest, Logistic Regression, and Gradient Boosting Classifier provide the highest outcomes for our model. In concluding, based on our evaluation of accuracy, precision, recall, and the f-measure value, we might decide that using the suggested 16 characteristics is practical for our model. By doing this, this same dataset is not only kept from ever being overfit but also includes more measure the parameters dimensionality. The comparison of results of their accuracy shows that they demonstrated an even greater level of competency.

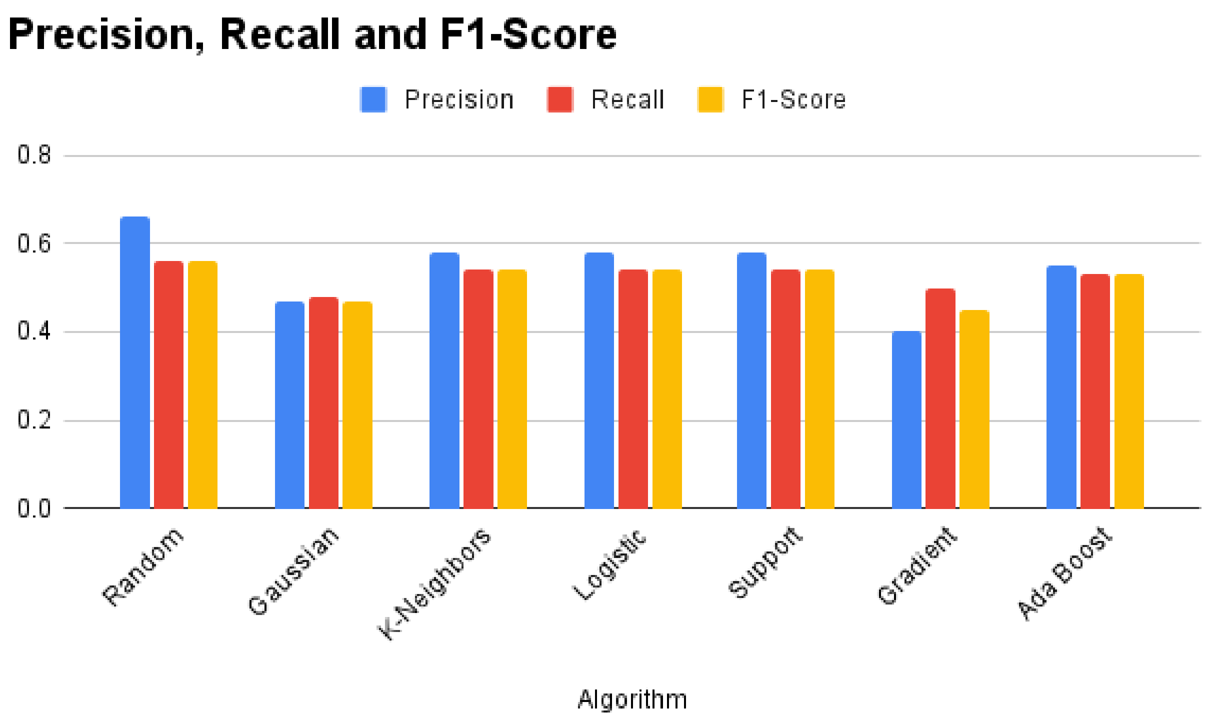



**Figure 5.10: Comparison: Precision, Recall, F1-Score**

# Chapter 06

## Final Remarks

The findings related to the study objective and its sub-questions are presented in this part analyzed as stated in the introduction. The findings of the study goal are reiterated, and the outcomes are those that are outlined in Sections 4 and 5. Significant discoveries are then addressed and paying very close attention to the dataset's significance as well as its imbalance. In order to demonstrate the additional value of this study to this field of study, this part concludes by contrasting the research findings to recent literature.

## 6.1: Research Challenges

While conducting a qualitative experiment, there were a few obstacles that we had to overcome, even though it was really difficult to do this research because it has been stigmatized in our nation.

1. It must have been challenging to get several responses for an accurate run of the survey formulas. Simply stated, the machine will function better if you only have more relevant information. Though gathering feedback online or the option of visiting online groups was not practical. Nevertheless, we made an effort to push all that we could.
2. It appears conceivable that the information we received from the survey didn't also include details that could help us quantify the depth of the depressive episodes. As a result, it is reducing our model's capacity for forecasting. Certainly, even though we do not now have an accurate magnitude matrix to experiment with the device.

## 6.2: Limitations of Research

More analysis of the outcomes, vital part of this process, and performance was evaluated lead to a crucial discovery. Despite the fact that our model accurately forecasts both psychological distress, there were some constraints as well. The survey seemed to be undoubtedly the most difficult part. We have endured immense hardship during the survey part since we frequently observe graduates demonstrating an absence of interest when completing the survey. The size of the statement whether the questions content or the arrangements are at fault. Despite our assurances to the participants that we would maintain the confidentiality of their information, it's conceivable several of them were still uneasy.

## 6.3: Future Scope

Our long-term goal for the system is that it must operate accurately, consistently, and be able to successfully identify psychiatric disorders at a preliminary phase. Mental health problems were the target factor in this research, which ignored an unbalanced dataset. By removing the unimportant variables and utilizing the most pertinent features, we intend to increase the accuracy of our system. This will assist in limiting the options and in order to improve the accuracy. In the future, we are aiming to provide this approach as a post-release monitoring tool for professional therapeutic advancements, with the capacity to serve as a platform for analysis of the sufferers. It will make the procedure of post-psychotherapy follow-up easier. When this system's trustworthiness can be relied upon by psychiatrists, we will start putting the infrastructure upgrade to work. Additionally, the system can be improved. Dependable and precise based on the reactions of the system's potential users built specific, acceptable methods to address cross-referencing across illnesses to calculate a person's mental steadiness. a set of surveys created by professional psychologists will indeed be available and employed to gather more information that would aid in connecting depressive symptoms with more convincing causes. As stated by the current results of our efforts, and how promising this system's future appears feasible. The aim of our building an environment where people feel at ease speaking more candidly about depression such that it is no longer considered as a socially undesirable disease in the upcoming years.

## 6.4: Conclusion

Today, a variety of knowledge - based systems are used in the medical field to properly anticipate diseases and illnesses so that efficient and productive treatments choices can be taken. Young people, particularly recent graduates, are still dealing with these psychological illnesses. Instead of getting aid for psychiatric treatment, they are attempting to manage with it that also, even in this day of globalization, is still controversial. The individual going experience these conditions constantly increase the likelihood of hypertensive, heart attack, cerebral abnormality and so on. WHO (World Health Organization) reports that individuals are lately increasingly susceptible to mental illnesses. The majority of patients with mental disorders don't come out and ask for assistance, and the cost of therapy is rather high and pricey for the general public. We have suggested a system to end these phenomena. It uses supervised machine learning to estimate the degree of sadness and anxiety algorithmic learning. The dataset used to train the system included seven several kinds of algorithms. These algorithms' outputs were compared, and Random Forest, Logistic Regression, Gradient Boosting Classifier therefore shown to be the most accurate method for identifying both depression among Bangladeshi graduates with the accuracy of 80.6%. This research has demonstrated to be able to forecast unemployment with accuracy using emotional and physical discomfort. Using our approach, we can empower the psychotherapists and counselors to help with clinical follow-ups, providing their management easier and much more effectively.

## References


1. Luciano, A., & Meara, E. (2014). Employment status of people with mental illness: national survey data from 2009 and 2010. *Psychiatric Services*, *65*(10), 1201-1209.
2. Luciano, A., & Meara, E. (2014). Employment status of people with mental illness: national survey data from 2009 and 2010. *Psychiatric Services*, *65*(10), 1201-1209.
3. Gonçalves, C., Ferreira, D., Neto, C., Abelha, A., & Machado, J. (2020). Prediction of mental illness associated with unemployment using data mining. *Procedia Computer Science*, *177*, 556-561.
4. Masri, R. Y., & Jani, H. M. (2012, June). Employing artificial intelligence techniques in the mental health diagnostic expert system. In *2012 International Conference on Computer & Information Science (ICCIS)* (Vol. 1, pp. 495-499). IEEE.
5. Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice*, *45*(5), 332.
6. Razzouk, D., Mari, J. D. J., Shirakawa, I., Wainer, J., & Sigulem, D. (2006). Decision support system for the diagnosis of schizophrenia disorders. *Brazilian journal of medical and biological research*, *39*, 119-128.
7. Chattopadhyay, S., Kaur, P., Rabhi, F., & Acharya, R. (2011, February). An automated system to diagnose the severity of adult depression. In *2011 Second International Conference on Emerging Applications of Information Technology* (pp. 121-124). IEEE.
8. European Commission. (2005). *Green Paper: Improving the Mental Health of the Population: Towards a Strategy on Mental Health in the European Union*. OOPEC.
9. World Health Organization. The World Health Report 2001: Mental Disorders Affect One in Four People. 2001.
10. Jané-Llopis, E., Anderson, P., Stewart-Brown, S., Weare, K., Wahlbeck, K., McDaid, D., ... & Litchfield, P. (2011). Reducing the silent burden of impaired mental health. *Journal of health communication*, *16*(sup2), 59-74.

11. van Nieuwenhuizen, R. F. H. (2021). *Prediction of Unemployment based on Physical and Mental Complaints due to Impaired Mental Health* (Doctoral dissertation, Tilburg University).
12. Higgins, A., Kilkku, N., & Kristofersson, G. K. (2022). Into the Future: Challenges and Opportunities for the APMHN Role. *Advanced Practice in Mental Health Nursing*, 503-519.
13. Stewart-Brown, S. L., & Schrader-Mcmillan, A. (2011). Parenting for mental health: what does the evidence say we need to do? Report of Workpackage 2 of the DataPrev project. *Health promotion international*, *26*(suppl_1), i10-i28.
14. Weare, K., & Nind, M. (2011). Promoting mental health of children and adolescents through schools and school based interventions: Evidence outcomes: School based interventions. *Health Promotion International*, *26*(Suppl. 1), i29-i69.
15. Digitalisering in de GGZ. ING Website. https://www.ing.nl/zakelijk/kennis-over-de-economie/themas-van-de-toekomst/nieuwefase-voor-de-GGZ-4.html
16. Palmer, S., & Dryden, W. (1994). Stress management: Approaches and interventions. *British Journal of Guidance & Counselling*, *22*(1), 5-12.
17. Robertson, I. T., Cooper, C. L., & Johnson, S. (2011). *Well-being: Productivity and happiness at work* (Vol. 3). Basingstoke: Palgrave Macmillan.
18. Czabała, C., Charzyńska, K., & Mroziak, B. (2011). Psychosocial interventions in workplace mental health promotion: an overview. *Health promotion international*, *26*(suppl_1), i70-i84.
19. Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational behavior*, *74*(3), 264-282.
20. Kulik, L. (2000). Jobless men and women: A comparative analysis of job search intensity, attitudes toward unemployment, and related responses. *Journal of Occupational and Organizational Psychology*, *73*(4), 487-500.
21. Šverko, B. (2004). Psychological Impact of Unemployment.
22. Taris, T. W. (2002). Unemployment and mental health: a longitudinal perspective. *International Journal of Stress Management*, *9*(1), 43-57.
23. Blank, L., Peters, J., Pickvance, S., Wilford, J., & MacDonald, E. (2008). A systematic review of the factors which predict return to work for people suffering episodes of poor mental health. *Journal of occupational rehabilitation*, *18*(1), 27-34.

24. DeSocio, J., Stember, L., & Schrinsky, J. (2006). Teaching children about mental health and illness: a school nurse health education program. *The Journal of school nursing*, *22*(2), 81-86.
25. Hatcher, S., Coupe, N., Wikiriwhi, K., Durie, S. M., & Pillai, A. (2016). Te Ira Tangata: a Zelen randomised controlled trial of a culturally informed treatment compared to treatment as usual in Māori who present to hospital after self-harm. *Social psychiatry and psychiatric epidemiology*, *51*(6), 885-894.
26. I. Education, Deep learning, sv, Online]. Available: May 1, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/deep-learning..
27. Ghandi, R. (2018). Support Vector Machine—Introduction to Machine Learning Algorithms (SVM Model by Scratch). *Towards Data Science, Jul, 7*.
28. Support vector machine algorithm, pt, [Online]. Available: [Online]. Available: https://www.javatpoint. com/machine -learning - support - vector -machine - algorithm.
29. Misra, S., Li, H., & He, J. (2019). *Machine learning for subsurface characterization*. Gulf Professional Publishing.
30. A. Christopher, K-nearest neighbor, en, [Online]. Available: Feb. 2, 2021. [Online]. Available: https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4..
31. Alam, B. (2022). Implementing Naive Bayes Classification using Python. *Recuperado el*, *20*.
32. P. Grover, Gradient boosting from scratch, en, Online]. Available: Dec. 9, 2017. [Online]. Available:https://blog.mlreview.com/gradient-boosting-from-scratch-1e317ae4587d
33. Chen, L. P. (2019). Mehryar mohri, afshin rostamizadeh, and ameet talwalkar: Foundations of machine learning.
34. Lynam, A. L., Dennis, J. M., Owen, K. R., Oram, R. A., Jones, A. G., Shields, B. M., & Ferrat, L. A. (2020). Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnostic and prognostic research*, *4*(1), 1-10.
35. Website: https://towardsdatascience.com/understanding-adaboost-2f94f22d5bfe
36. Rössler, W., Ujeyl, M., Kawohl, W., Nordt, C., Lasalvia, A., Haker, H., & Hengartner, M. P. (2019). Predictors of employment for people with mental illness:

results of a multicenter randomized trial on the effectiveness of placement budgets for supported employment. *Frontiers in Psychiatry*, *10*, 518.

37. Hatcher, S., Coupe, N., Wikiriwhi, K., Durie, S. M., & Pillai, A. (2016). Te Ira Tangata: a Zelen randomised controlled trial of a culturally informed treatment compared to treatment as usual in Māori who present to hospital after self-harm. *Social psychiatry and psychiatric epidemiology*, *51*(6), 885-894.
38. Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, 1258-1267.
39. Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S. H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation*, *16*(4), 262.
40. Sadiq, M. H., & Ahmed, N. S. (2019). Classifying and predicting students' performance using improved decision tree C4. 5 in higher education institutes. *Journal of Computer Science*, *15*(9), 1291-1306.
41. Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.
42. Lee, M. H. (2020). Robust random forest based non-fullerene organic solar cells efficiency prediction. *Organic Electronics*, *76*, 105465.
43. Devroye, L., & Wagner, T. J. (1982). 8 Nearest neighbor methods in discrimination. *Handbook of Statistics*, *2*, 193-197.
44. Edelstein, H. A. (1998). *Introduction to data mining and knowledge discovery*. Two Crows.
45. Liang, Y., Liu, C., Luan, X. Z., Leung, K. S., Chan, T. M., Xu, Z. B., & Zhang, H. (2013). Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification. *BMC bioinformatics*, *14*(1), 1-12.
46. Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica*, *24*(1), 12-18.
47. Hasan, M. R., Maliha, M., & Arifuzzaman, M. (2019, July). Sentiment analysis with NLP on Twitter data. In 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.
48. Bhowmik, N. R., Arifuzzaman, M., Mondal, M. R. H., & Islam, M. S. (2021). Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary. Natural Language Processing Research, 1(3-4), 34-45.

49. Bhowmik, N. R., Arifuzzaman, M., & Mondal, M. R. H. (2022). Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms. Array, 13, 100123.
50. Gope, J. C., Tabassum, T., Mabrur, M. M., Yu, K., & Arifuzzaman, M. (2022, February). Sentiment Analysis of Amazon Product Reviews Using Machine Learning and Deep Learning Models. In 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) (pp. 1-6). IEEE.
51. Basunia, M. R., Pervin, I. A., Al Mahmud, M., Saha, S., & Arifuzzaman, M. (2020, June). On predicting and analyzing breast cancer using data mining approach. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1257-1260). IEEE.
52. Ripa, S. P., Islam, F., & Arifuzzaman, M. (2021, July). The emergence threat of phishing attack and the detection techniques using machine learning models. In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) (pp. 1-6). IEEE.
53. Toma, T., Hassan, S., & Arifuzzaman, M. (2021, July). An Analysis of Supervised Machine Learning Algorithms for Spam Email Detection. In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) (pp. 1-5). IEEE.