DEPRESSION PREDICTION AMONG STUDENT BASED ON THEIR DAILY ACTIVITIES

BY

MISKATUN AHMED PRANTY

ID: 201-15-13713

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

Supervised By

DR. TOUHID BHUIYAN
Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MR. ABDUS SATTAR
Assistant Professor & Coordinator M.Sc.
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

APPROVAL

This Research titled **"Depression Prediction Among Student Based On Their Daily Activities"**, submitted by **Miskatun Ahmed Pranty**, ID No: **201-15-13713** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *25th January, 2024*

BOARD OF EXAMINERS

Narayan Ranjan Chakraborty (NRC)
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Saiful Islam(SI)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Shayla Sharmin (SS)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Sazzadur Rahman (MSR)
Professor
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that this research has been done by us under the supervision of **Dr. Touhid Bhuiyan, Professor, Department of Computer Science and Engineering and** co-supervision of **Mr. Abdus Sattar, Assistant Professor & Coordinator M.Sc. Department of Computer Science and Engineering** Faculty of Science and Information Technology, Daffodil International University. We also declare that neither this research nor any part of this research has been submitted elsewhere for the award of any degree.

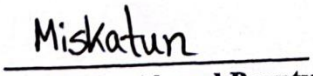
Supervised by:

Dr. Touhid Bhuiyan
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc.
Department of CSE
Daffodil International University

Submitted by:



Miskatun Ahmed Pranty
ID: 201-15-13713
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At First we would like to express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year thesis successfully.

We really grateful and wish our profound our indebtedness to **Supervisor Dr. Touhid Bhuiyan, Professor**, Department of CSE, Faculty of Science and Information Technology, Daffodil International University, Dhaka. His Deep Knowledge & keen interest with supportive instructions helped us in the field of machine learning based research, finally we completed our work on "**Depression Prediction Among Student Based On Their Daily Activities**". Their endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Head,** Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish our project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again we want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Depression affects most people in modern life. Inadequate treatment even leads to many people taking their own lives. Early detection and treatment of depression in patients are very easy to achieve. We are unable to make the finest decision at the right moment since we are unaware of the severity of the depression. The foundation of a nation is its pupils. Students educate and better their nation, representing it to the outside world. A number of things, including the difficulties Bangladeshi teenagers face in their schooling, contribute to depression. Our study's goals are to ascertain the frequency of depressed symptoms, the factors that contribute to them, and methods for lowering depression among college students. In this study, an online student depression dataset have been used for predicting the depressed or not. Two class have been consisting this dataset. Multiple algorithms have been run on this data. and have reached the maximum level of precision. This initiative will assist us in determining depression levels. To determine their degree of despair, we employ a form of algorithm. Five algorithms have been selected for this study. XGBoost classifier, Random Forest algorithm, SVM, Naive Bayes, and Decision Tree classification are among them. The forecast made by the XGBoost classifier performs the best overall. The algorithm that yields the greatest results for this research is 98.70%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

The word "depression" describes a psychiatric disorder that is rather common in today's world. This is a sign of an adverse psychological state; depression is commonly caused by a range of common variables, such as frustration, stress, and social anxiety in addition to family, health, and educational difficulties. Depression frequently has a detrimental effect on a person's ability to operate on a daily basis; a person's psychological or mental condition is detrimental to their ordinary way of living. This usually hinders a person's ability to progress both personally and professionally. Their mental state makes it impossible for them to focus or behave correctly. Depression and other psychological issues are common among students. Many people feel hopeless as a result of bad communication or arguments with family, friends, or other people. These days, student depression is a major problem that impacts students everywhere. A unique set of individuals known as university students are undergoing a period of transition as they work to integrate into their freshly established academic and social contexts. Melancholy strikes certain pupils more often during this period. A significant number of them have had anxiety in the past, and others have additional psychological issues such communication difficulties, which is a more often identified disorder that makes most of the students feel hopeless. Research on sadness in students is scarce in our nation.

Sadness is typical when an individual accepts verbal or physical humiliation. Depression is often more likely to occur after the death of a close relative. Depression can occasionally result from a deficiency of specific biological substances in the brain. Sadly, a lot of individuals either don't want to admit they have depression or don't understand it. Not to be confused with a passing mood swing or "the blues," depression is a serious, frequently protracted illness that has to be treated. If left untreated, it can cause more significant health issues and impair with a person's everyday life, career, and relationships. Although there are numerous distinct causes of depression, the condition is frequently brought on by a confluence of psychological, environmental, and hereditary variables. Some people may be especially susceptible to depression because of a physical condition, a history of abuse or trauma, or a family history of

the illness. Seeking assistance from a mental health expert is crucial if you suspect you may be suffering from depression. A person may experience depression in one of the many forms. The following lists a few prevalent forms of depression:

It is noteworthy that depression may manifest in a multitude of ways and exhibit a wide range of symptoms. If you think you may be depressed and would need assistance in determining the sort of depression you might be suffering from as well as creating a treatment plan, it's crucial that you get in touch with a mental health specialist. Depression can have a wide range of possible symptoms, and each individual may experience them differently. The following are some typical signs that a person suffering from depression may encounter:

- Constantly feeling depressed, hopeless, or empty;
- Losing interest in once-enjoyable activities;
- Difficulty focusing or making decisions;
- Improvements in appetite and weight;
- Sleep trends, such as trouble falling or remaining asleep, or laying down too much;
- Fatigue and decreased energy;
- Feelings of guilt or worthlessness;
- Problems with in relationships;
- Suicidal thoughts

It's important to keep in mind that everyone experiences these stages from time to time and that it's normal to feel unhappy after a setback or loss. However, these feelings may be a sign of depression if they persist and interfere with daily tasks. You ought to seek professional mental health help if these symptoms are significantly impacting your daily life or causing you significant distress. They can help you pinpoint the cause of your issues and recommend appropriate next steps.

1.2 Objectives

The modern period is characterized by rapid technical advancement. Technology has the potential to solve any issue. Analyzing depression entails looking at a range of topics, including its causes, effects, and symptoms. The context—clinical studies, public safety, or personal well-

being—determines the goals of depression analysis. Our primary objective is to accurately diagnose depression and anticipate the depression label. There are two categories: those who are not depressed and those who are. Our goal using machine learning techniques is to classify depression or not.

1.3 Motivation

Students are the foundation of a nation. Students educate and build the country, representing it to the outside world. But if these young people quit college every year, it would be dangerous for all countries. Depression is partly influenced by the challenges Bangladeshi adolescents face in their home life, work, relationships, education, and drug usage. Thus, the purpose of our research is to determine the prevalence of depressed symptoms, the variables that contribute to them, and strategies for reducing depression among college students. A person might experience sadness in a variety of ways, depending on his life. Depression is a common problem for people. The main problem is that not enough people in our country are aware of mental illness concerns. The difficulty for individuals to understand their despair. Neither a doctor nor a psychologist sees them. Since there hasn't been much more research in this topic, we examine the literature. There should be a lot more research done in this area. In light of this, we are using the popular and extensively applied machine learning approach to complete this depression detection assignment. Therefore, it must assess an individual's level of sadness based on their routine activities and way of life. We thus contemplate creating a system or process where machine learning may be highly successful in assessing an individual's mental health.

1.4 Rationale of the study

There have been a lot of research done on the incidence of depression among students. In contrast, Bangladeshi students' research opportunities in depression are severely restricted. Research on university students has been done a lot. Research on college and high school students is scarce. However, the rate of depression among all these impressionable youth is rising at the moment for a variety of reasons, which is hazardous for the nation as a whole. In order to conduct research and use machine learning techniques to predict depression, we gathered students depressed data for our study.

1.5 Research Questions

This study has been completed with a great deal of enthusiasm and effort. Completing this task was difficult for us as well. Coming up with a plan that is equitable, workable, and accurate faces a number of challenges. The purpose of the following questions is to aid researchers in understanding these concepts and provide a solution:

- Can raw data be obtained for scientific research?
- Is it possible to preprocess the raw data using a machine strategy?
- Do these methods and approaches have an opportunity to improve student or medical outcomes?
- How can these methods and efforts help students?

1.6 Expected Outcome

The goal of our effort is to make depression easily identifiable from other types of behavior. Identifying when we are sad constitutes one of our biggest problems. We at Daffodil International University 3 make it easy for people to consider their mental health. Consequently, a significant segment of the population will no longer have mental health difficulties. For this basis we used student stress depressed dataset.

In society, we have to communicate with a lot of people. Frequently, we notice that several of our pals have wrapped themselves. They aren't participating in any social activities anymore. There are more of these people in society than ever before—not fewer. This article will help identify people's mental health conditions.

1.7 Layout of the Report

The first chapter provided an example of how to make the endeavor more transparent by stressing the project's determination, drive, difficulties encountered during the study, and expected research outcomes. The next section of the report goes into more detail on the article's overall structure.

The work that is currently being done in this field is covered in Chapter 2. The last part that makes up the next section shows how thorough it is because of the emphasis on this topic. A brief discussion is given of the study's primary limitations or problems. The section that follows

includes sections on pertinent works, an overview of the study, and problems that need to be resolved in order to complete the project.

Chapter 3 provides an explanation of this research effort's theoretical analysis. This chapter offers more information on the statistical techniques used in order to address the theoretical element of the investigation. Additionally, procedural methods to machine learning techniques are demonstrated in this chapter. The procedure for gathering datasets and the system for preparing data are described in the following chapter. The final part of this subdivision also includes confusion matrix analysis to evaluate the model and provide the classifier's accuracy tag. Incorporating implementation analysis is vital to guarantee authentic correctness when employing machine learning methodologies. This section addresses a variety of subjects that have come up during the project's development, such as the study topic and instruments, workflow, information gathering protocol, data processing, suggested methodology, mode training, and implementation needs. Every algorithm comes with a detailed description of every machine learning technique and classification used in this study.

An outline of the study, a summary of the findings, and a discussion of the results are provided in Chapter 4. This section includes some experimental material and labeled data sets to help bring the project to life. A description of the components of the ml techniques and an evaluation of their results follows this section.

Chapters 5 and 6 contained the study summary, the follow-up activities, and the conclusion. The purpose of this part is to make clear whether or not every project report item complies with the criteria. implications for sustainability, society, and the environment. The limitations of our efforts were covered at the conclusion of this phase, and they may have an effect on anyone hoping to work in similar subjects in the future decades.

CHAPTER 2

Background

2.1 Introduction

The modern age is well familiar with the word "depression." However, depressed individuals face many difficult situations. This section will cover the scope of the problem, the challenges we faced, and relevant research that has previously been conducted by other researchers. It will also include a summary of their findings. A few research, similar works, and their applicable methods, classification techniques, and levels of accuracy that are pertinent to our investigation are included in the part that focuses on comparable work. In the study summary section, we compile a summary of every work and display it as a table to facilitate understanding. Regarding the scope of the issue, we discuss how we may support or forward this work. The final part, "Challenges," describes the many problems and barriers we ran into during our research endeavor and how we resolved them. Consequently, it is critical to identify and treat depression. Numerous research has been done in recent years to define depression. Several foreign scholars have been active on their own platforms and in their own languages. Depression has not gotten as much attention as sentiment analysis. Not much study has been done on Bengali language distress identification among them. At this point, we've tried to draw attention to research that has been done by various academics on the identification of melancholy in social media posts or comments. The use of social media comments and status updates by researchers to detect depression is being studied. Data mining has gained popularity as a technique for researching scientific subjects in the context of academic study in recent decades. In this chapter, we'll talk about data mining in connection to data classifier research hypotheses. The first section will detail earlier relevant work; the second will highlight the nature of the problems with our research-related work; and the third piece will address the difficulties we ran across while conducting our study.

This section presents the research challenges, pertinent literature, and a summary of the findings. We will analyze research publications by other writers and discuss how their methods and accuracy relate to our own in the area under "Related Works." In the section on related studies, we will discuss the methods, findings, and applicability of other research articles that are linked to our study. Our related studies will be summed up in the research summary section. We discuss

how we increased the reliability of the layer in the difficulties section and address every issue encountered while doing this investigation. Every topic has been covered.

2.2 Related Works

Scientists from looked at college freshmen. They used the (GAIN-SS) to evaluate mental health issues. The relationships between 12-month psychological issues were computed using generalized linear models. At last, they discovered that there were even fewer distinct kinds of mental health issues: internalizing issues made up 14.2% (SE=0.56) of the total, externalizing problems accounted for 8.6% (SE=0.46), drug use 1.7% (SE=0.21), and antisocial disorders 0%[1].includes research on the mental health problems that Bangladeshi students encounter. Out of the total data collected, only 590 were utilized in the final analysis. They employ eleven different kinds of questions. For this study, they employed the DASS-21 in Bangla. Cronbach's alpha for the downward trend was 0.77. Their research's Feature Set is seven in size. With a 95% accuracy rate, logistic regression is their most accurate method [2]. The writers of concentrated on depression in students. They created an internet-based survey system that resembles a traditional psychometric instrument. They could extract 121 features. Their technique was used to the evaluation of 466 students' data. It was shown that 25.32 percent of students suffer from depression [3]. The focus of the authors was first-year students. They gathered 400 bits of data for their inquiry. Despite utilizing several classifiers, logistic regression proved to be the most successful. In their surveys, they classified the 14 criteria they used into four categories. 52.3% of applicants are men and 47.8% are women, per their data [4]. After looking at four courses, they observed that fifty-two percent of the students had significant depression. Out of the 523 individuals that the writers questioned, only 468 got their data examined. The dataset employed in this study contained several psychometric tests as well as sociodemographic inquiries, albeit it is not stated. According to the final results, 27.1 percent of individuals reported feeling nervous, and 44% reported feeling depressed [5]. The investigators looked at alexithymia. 14 features and a sample size of 492 were present. The data analysis was done using SPSS version 21. They used many methods to assess sadness, anxiety, and stress. They found that alexithymia affected 24.6% of the participants. Depression, dread, and tension were prevalent in 28.5, 38.4, and 22.6 % of the population, respectively [6]. Artificial learning-based analysis of sentiment was studied by Khan, M. R. H. Daffodil International University 6 et al. using the Bengali depression dataset.

Data is collected from books, poetry, quotations from many notable individuals, and posts on social media. They employ a number of methods, including as support vector machines, random forests, k-nearest neighbors, decision trees, and xg boost. Additionally, they obtained the greatest accuracy of 86.67% using Multinomial Naive Bayes. Shukla, D. M. et al. [7] employed statistical indicators and strength analysis. proposed a method for utilizing voice or speech cues to detect sadness in an individual. They conclude that one may tell if someone is depressed or not simply listening to their voice cues. This makes use of the Ryerson RAVDESS audiovisual dataset, which has moving music and voice. They employ the Perceptron Multi-Layer technique. The average accuracy for both training and cross-validation sets was 81.56%. Deshpande, M. et al. have conducted research on the use of emotional artificial intelligence in depression diagnosis [8]. Observing Twitter messages, they could recognize melancholy.10,000 tweets were obtained from Twitter using the Twitter API in order to generate the initial training and test data. Depression was identified using the SVM and Naive Bayes methods of classification [9]. Support vector machines yield a 79% increase in accuracy. However, the Multinomial Naive Bayes method achieves the maximum level of accuracy. Multinomial Naive Bayes achieved the maximum accuracy of 83%. In order to identify depression, Asad, N. A. et al. examined user posts on social media. Information is gathered from Facebook and Twitter. By using delicious soup to extract data from Twitter, 150 people's Facebook information was personally retrieved from Facebook with their consent. This paper uses SVM, a Bayes classification technique. They use the well-known BDI-II questionnaire technique to gauge how severe their depression is. It is classified as non-depressed, normal, borderline, and moderate. If a person's depression score ranges from 1 to 55%, they are considered depressed. Naive Bayes achieves an outstanding degree of accuracy and precision of 74% and 100%, respectively [10].

2.3 Research summary

The primary emphasis of our project endeavor is the variety of methods that community at large has to offer. We have used five different approaches in all, and we have applied additional algorithms to our dataset. The main source of data in this case has been our internet-compiled dataset. As previously said, our dataset consists of both newly gathered data and data that has previously been used. We will be able to assess the accuracy of the five methods we used, as

well as things like the impact of further data we contributed from the same source. It signifies that since the same kinds and classes of labeling exist. The extraction of features and data preparation processes, which included ML approaches and methods of classification to predict depression, were mostly conducted using Python.

2.4 Scope Of The Problem

Essentially, our research entails utilizing machine learning techniques to assess the given data and create a model. Our model could be able to diagnose depression. The people in this society will be greatly impacted by this initiative. In today's world, there are many people who battle mental health issues. They do not realize that they are depressed, though. They withdraw inside themselves and don't know what to do. As a result, their first assessment was flawed. He requires a technique that will help him recognize his problem and decide whether or not he is truly depressed.

2.5 Challenges

Considering how hard it was to analyze this one massive data file, combining and evaluating each of these data sets seemed to be the main problem of this research. We employed a number of tools and methods to clean up and normalize the dataset. Considering the size of the data sets and the several layers they contained from various eras, it took some time to reach the desired outcome. Further data sets are located in this field. We didn't finish the research, so I must dedicate a lot of effort into every assignment that is linked to it, which makes it challenging for me to think of the best responses right quickly. Managing this enormous amount of data and creating an environment for categorizing the sadness were issues that arose again when defining the topic of how to identify student depression using machine learning models.

CHAPTER 3

Research Methodology

3.1 Introduction

This research section aims to give a summary of the entire procedure and the techniques applied to diagnose depression in students. We're going to go through the entire study methodology in this section. Various strategies may be employed to solve each analysis. First, we collect data from online and analyzing the data, we need to eliminate some empty values and categories that are unnecessary for our investigation. After that, a ML algorithm is chosen. We have said that we are employing five distinct machine learning approaches, therefore before we can create the framework and fit the algorithm, we need to create a data collection. After that, the model is trained using this set of information. Feature selection operates on this foundation. After then, the data was divided into training and testing sets. The training dataset and the testing dataset are other names for it. After the data has been fitted into several machine learning approach models and trained using a training dataset, we then only obtain a large fraction of the data required to evaluate our model. The accuracy of the model is then evaluated. We've been giving an overview using our standard process flow chart, but we'll go over some of the techniques using equations and diagrams to better understand how they operate. The process for the full research endeavor is displayed below, along with a synopsis of the complete study effort. Some essential elements are the data collection, analysis, and suggested model, which is further clarified by the appropriate formula, graph, appearance, and explanation. This work uses online databases for data collection and employs the most precise ML classification models for prediction.

3.2 Subject of the Research and Equipment

A research project's topic is usually one that other people are presently looking into and evaluating, which means that it has the potential to be used for model construction, data collecting, value perseverance, and model distribution. We talk about the tools and methods we employ for instrumentation. We utilized several more libraries, such as NumPy, SkLearn, OpenCV, etc., when coding in Python. I used the Co Lab web page application for all of my monitoring and investigation. The Google Lab at Google leverages Py to develop web apps that classify labels like "Depressed or Not" using statistical techniques connected to machine learning.

Libraris used:

- **Matplotlib:** The Pyplot calculating plotting charting set of techniques is one of Matplotlib's visualization tools. This could be used, for example, to indicate corners in a tale or to draw the limits of a plot when creating forms.
- **NumPy**: The Python NumPy utility has made vector processing easier. In particular, matrix computations, the indices transform, and the Fourier transform are covered in the topic material. A multitude of tools and techniques for interacting with matrices of various types are available in the NumPy packages for Python. NumPy allows for the rational and realistic design of devices. To put it simply, NumPy is a Python addon designed primarily for quantitative assessment. This comment can also be known as "computation. Py."
- **Sklearn:** Sklearn is an effective and user-friendly tool for predicted data analysis. Its construction made use of three Python tools: NumPy, SciPy, and Matplotlib, which may be acquired for free and tailored to each user's needs.
- **Seaborn:** You may use matplotlib with the next version of this popular Python information visualization application. This is an easy-to-use tool for creatively showing data.
- **H5py:** Users can access binary HDF5 code by using the Python scripting language's h5py module. Large amounts of data, particularly integers, are managed with NumPy and subsequently stored in HDF5.
- **TensorFlow:** This AI toolbox appears to be free. It offers a wide array of tools for creating and implementing different machine learning simulations, such as NN. TensorFlow is a great option for a variety of applications involving AI due to its adaptability, effectiveness, and ease of use.
- **Pandas:** This shareware toolbox for information analysis and manipulation specific to a language is offered by Pandas. Fundamental types of data and techniques for statistical analysis are offered to make structured data administration, particularly the management of summary information, easier.
- **OS:** Because of the various features of this Python OS module, programmers may interact with the language of programming that these people use.

3.3 Workflow

It's possible that this study employed several working methods or approaches to ascertain the evidence. In order to conduct this investigation, a workflow was selected, data was processed, collected, word clouds were created, stop words were eliminated, text was cleaned, and efficiency was assessed using the output of the XGBoost model.

Phase 1: Data Collection: Since there aren't existing archives created just for each type of fact, I obtained all of my data online. The following categories were recently used to define information.

Phase 2: Processing Data: Each data element was assessed independently once all realistic methods of gathering data had been used. There are several instances of improper and stop words around. I really manage the last stage of the selected dataset before putting it into practice.

Phase 3: Preparing Dataset: Getting the dataset ready Preprocessing and data extension are still done in accordance with content and label. We must arrange the data, remove stop words, and visualize it for training purposes. We just finished the very minimum of tasks required to prepare the data for splitting.

Phase 4: Model selection: To increase dependability, we choose a prediction method, train it, and then evaluate it using my data. Machine learning makes use of several filters. Even though different models were employed to increase efficiency with the device's architecture and to facilitate the ML model's categorization of false reports, only one device was finally chosen to assess the correctness of the data.

Phase 5: Performance Assessment: All implications are covered in this section. These methods provided us with a limited degree of dependability for every one of the two instructive label classes—Depressed, Not Depressed—after training and testing. Accuracy statistics and f1 measures were produced in addition to the perplexity matrices. Use machine learning methods to determine which students are deficient.

Phase 6:Conclusion and Future Work: This section will provide a synopsis and a development roadmap.

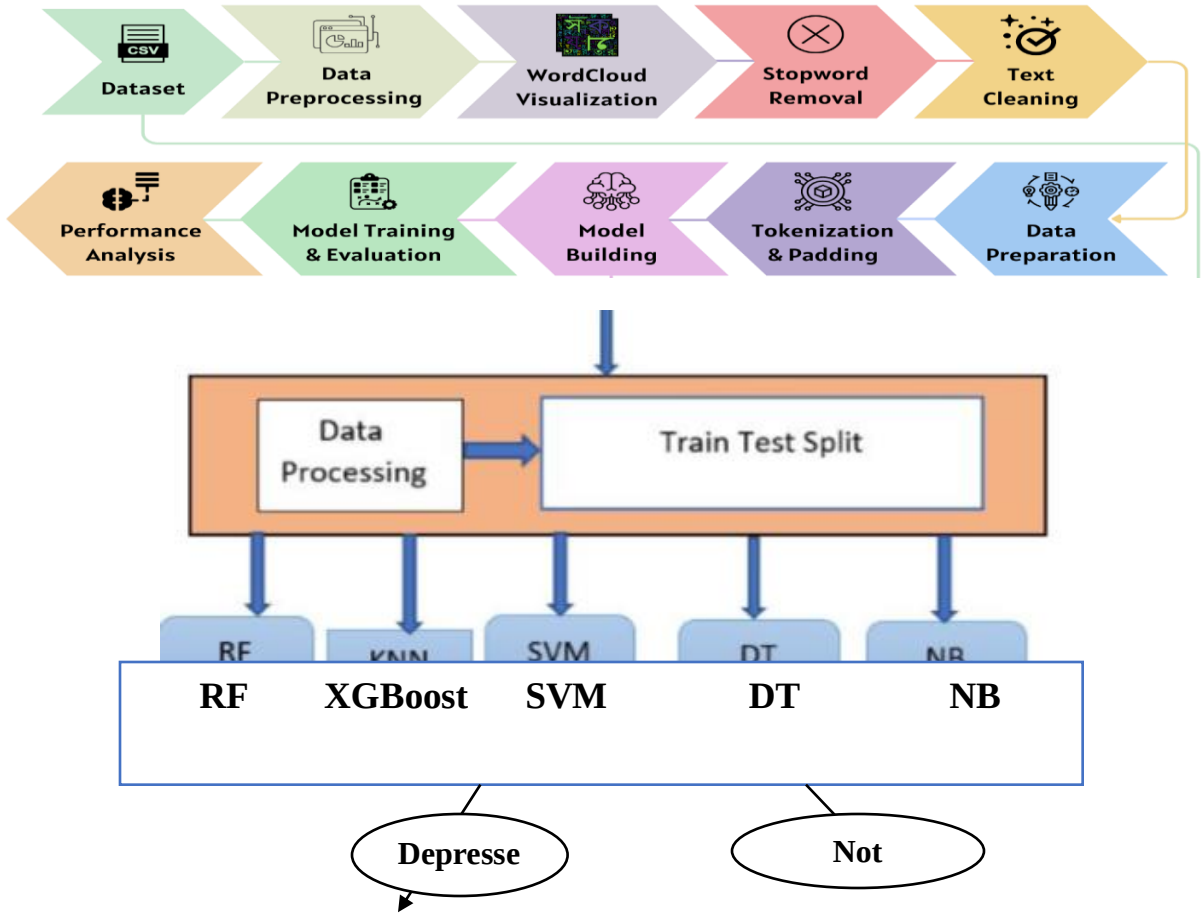


Fig 3.1: Research Proposed workflow.

The procedures in our study method that might assist students with anxiety and depression are shown in Fig. 3.1. Our sources of web data were diverse. We gathered information from a Kaggle dataset that is made available to the public. To make certain that each data set would only contain correct and pertinent information, we have gone over this data collection, eliminated any superfluous phrases, and cleaned up the terminology. Word clouds are used as graphics to make the often used phrases more understandable. We investigate the topic of ML techniques by creating and improving models utilizing pre-existing data. Additionally, we address the problem of class heterogeneity with permutation techniques to guarantee equitable representation and enhance the overall performance of the model. In addition to providing an overview, our strategy looks for methods to decrease the number of false negatives and false positives. Through the implementation of a holistic strategy that combines language knowledge with machine learning techniques, we can effectively address deception in student sadness.

3.4 Data Collection

The data required for this investigation was sourced from the open-source Kaggle database. A created dataset of 6,980 material data points is made up of 2 columns like: “label” & “text”. The dataset is divided into two parts: Train and Test. Train includes 5,576 and label data, compared to 1,394 for Test after delete duplicating data into two classes: “Depressed” and “Not Depressed”.

	text	label
0	oh my gosh	Depressed
1	trouble sleeping, confused mind, restless hear...	Depressed
2	All wrong, back off dear, forward doubt. Stay ...	Depressed
3	I've shifted my focus to something else but I'...	Depressed
4	I'm restless and restless, it's been a month n...	Depressed
5	every break, you must be nervous, like somethi...	Depressed
6	I feel scared, anxious, what can I do? And may...	Depressed
7	Have you ever felt nervous but didn't know why?	Depressed
8	I haven't slept well for 2 days, it's like I'm...	Depressed
9	I'm really worried, I want to cry.	Depressed

Fig 3.2: Sample data of depression.

Table 3.1, which includes a list of each and every field in each one of these files, is broken down as follows:

Table 3.1: Describe the Column in the Dataset.

Column Name	Description of the Column
text	Description of depression text
label	1 as “Depressed”, 0 as “Not Depressed”

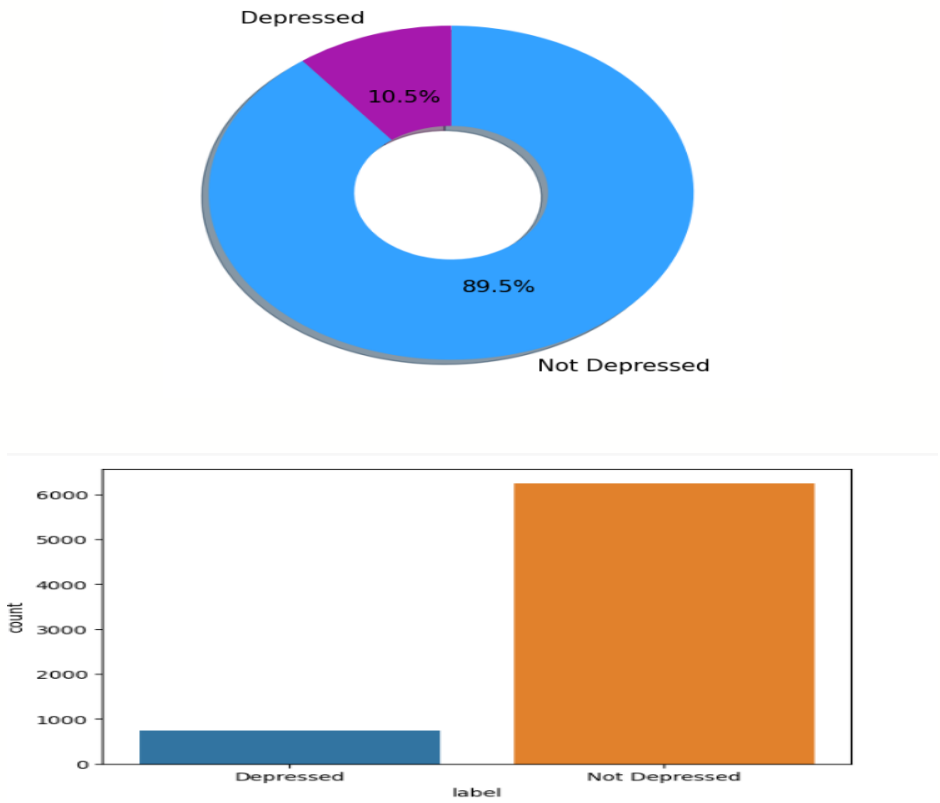


Fig 3.3: Depression Class Distribution

Encoder for labels such as:

ML experts often work with datasets that include several features, regardless of the quantity of columns in the dataset. These distinctive IDs might be a mix of characters and numbers. Words are often employed in training to classify data in an attempt to enhance user convenience and comprehension.

One method of converting signals into text that can be read by computers is to encode the labels. As a component of the previously mentioned procedure, identifiers need to be translated in order to construct a structure that transmits numerical values. The programmers who create ML systems ultimately decide on how to utilize such designations. Considering any development that is being tracked, initial processing of the standardized data is now required.

3.5Statistical Analysis

3.5.1 Analyzing Data

To transform the unprocessed written content into a framework appropriate for evaluation and model training is our aim for data pre-processing. First, from datasets containing depression detection, the required "text" and "label" variables must be extracted. The collected datasets are then combined to provide a collection of data for training and testing. We went ahead and fixed

errors by eliminating unnecessary letters and punctuation after the database was set up. Enhancing the appropriateness of word choice in the text is the aim of these exercises. The majority of word repetitions are graphically represented using word cloud visualization, which helps with language structure comprehension. In machine learning models, tokenization is the act of converting words into mathematical vectors so that they may be concatenated with ease. Our computational models are more capable to recognize depression or Not by employing this thorough data processing approach, which further improves our capacity for in-depth analysis.

3.5.2 Word Cloud Visualization

As Figures 3.5-3.6 demonstrate, highly visualizing phrases play a significant role in our analysis as well as in our training and testing datasets and another for the two classes "Depressed" and "Not Depressed." This approach helps identify trends and patterns in the language utilized by providing a visual depiction of the key terms used in the content of the report. It is possible to learn crucial information on implicit and distinctive characteristics. The picture is a useful tool that directs the generation of our class-wise data. The following makes it feasible and provides a realistic image of how English language has developed. This straightforward method helps us make the connections between unprocessed data and insightful analysis, which is helpful for making informed decisions while developing research development is false-positive.

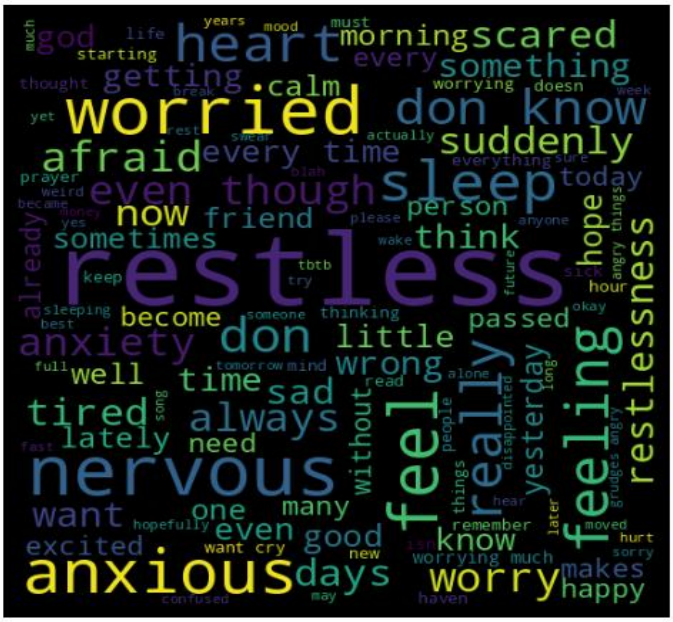


Fig 3.4: Word visualization with Depressed Class.

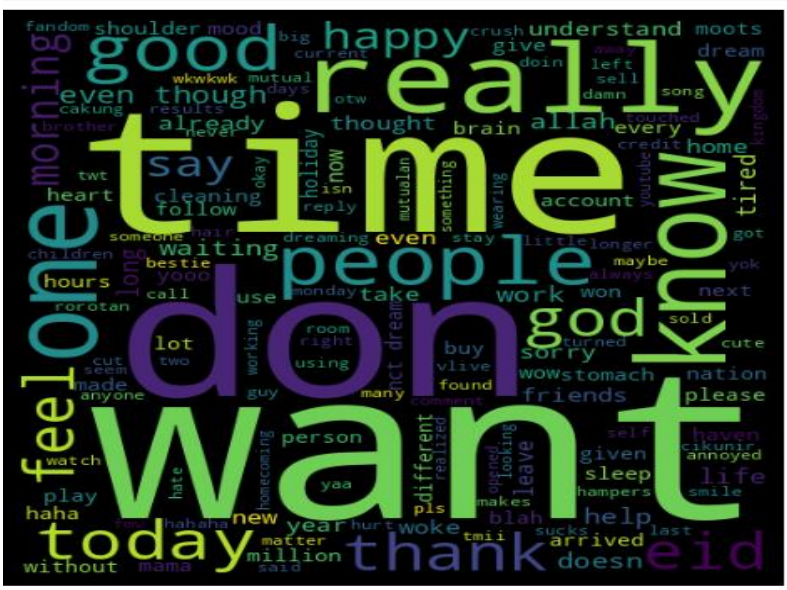


Fig 3.5: Word visualization with Not Depressed Class

3.5.3 Stop word Removal

We appreciate the language that was utilized as we worked to make the content better. Despite being widely used, the aforementioned terms often have no meaning within the categorization system and no language context. We employ terms that stop by using a bespoke module titled "Stop words" which we acquired from GitHub in order to get around this problem. The vast collection of Bengali numerals in the library has been painstakingly conserved to follow the syntactic framework of our language. We can reduce the number of unnecessary characters and improve the speed at which we analyze data by integrating these libraries. We frequently omit punctuation to improve the logic of our text categorization attempts, as the aforementioned examples demonstrate:

"it", "ve", "!","re", "me", "the", "oh", "we"etc.

```
# Define a function to remove specified words (stop words) from a sentence
def remove_stop_words_and_single_alphabets(sentence):
    words = sentence.split()
    stop_words = ["it", "ve", "!","re", "me", "the", "oh", "we", "you","so", "he", "is", "." ,"she", "lik
    "world", "back", "this", "if", "let", "mkr", "been", "thing", "should", "anything",
    "did", "its", "day", "still", "first", "too", "cant", "And", "had", "going", "make",
    "these", "only", "see", "has", "go", "why", "were", "there", "will", "because", "how",
    "the", "then", "an", "he", "if", "its", "Im", "no", "by", "at", "what", "u", "do",
    "amp", "i", "or", "so", "have", "be", "my", "who", "was", "are", "I", "to", "a", "the"]
    words = [word for word in words if len(word) > 1 and word.lower() not in stop_words]  # Remove single
    return ' '.join(words)
```

Fig. 3.6. Stop words of English.

3.5.4 Text Cleaning

We do text modification, which is a crucial step in the augmentation of the dataset. This approach consists of two primary techniques intended to improve the caliber and pertinence of textual material. First, news is delivered via a sorting mechanism that only keeps stories above a preset word limit of 100. Filtering techniques guarantee our dedication to providing high-quality, legally compliant, and educational content. Additionally, a text correction approach is used to remove extraneous characters like tabs, stop stages, special markings, etc. in order to systematically adjust the content. Standard symbols, line breaks, and some English characters have been removed from the text itself. We guarantee that the material we collect will be prepared for in-depth examination by employing a comprehensive preservation procedure.

3.5.5 Tokenization and Padding

Padding and tokenization are crucial components of contemporary data use. Tokenization is the process of turning words into sequences of numbers so that our model can comprehend language. Assigning a distinct numerical value to every word establishes the link between written speech and symbolic numbers. During pattern training, padding increases the likelihood of agreement by guaranteeing that every sequence has a fixed length. Our algorithms can assess the various linguistic intricacies of text since they translate phrases into mathematical forms and verify that sequence length varies so frequently. This stage is crucial for our computers to correctly comprehend the data and discern between depressed and non-depressed data.

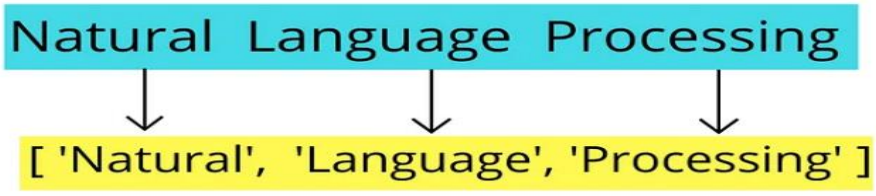


Fig 3.7: Example of Tokenization.

3.5.6 Data Preparation

We did not split the data for model training and testing at random during the data preparation phase. We extract the most significant "text" and "label" columns, comprising 6980 data, from the depression dataset.. Train and Test are the two sections of the dataset. Train has 5,576 and

labeling data, whereas Test has 1,394 after duplicate data was removed and split into two groups, "Depressed" and "Not Depressed."

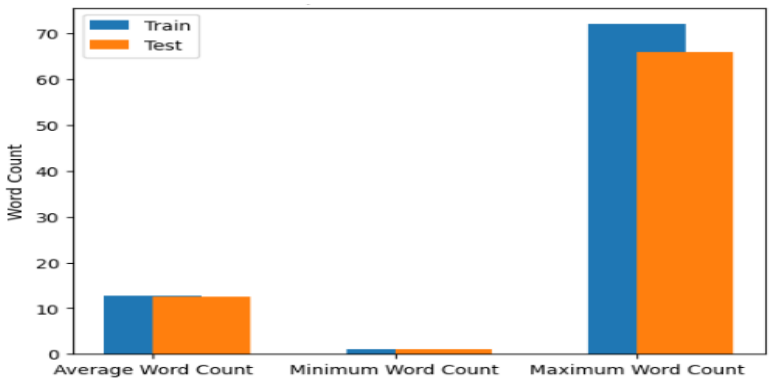


Fig 3.8: Word Count comparison between Train and Test Data.

3.5.6 Models

Our study uses a variety of techniques to understand depression reliability. We looked at the Random Forest method, SVM, Naive Bayes, XGBoost classifier, and the decision tree ML models. To take advantage of their capacity to describe background, patterns, and linguistic variances, we employ several models. This thorough investigation helps us achieve our objective of developing a trustworthy method for detecting depression in students.

1. **Decision Tree:** One technique for supervised machine learning is the decision tree. This approach can be used to tackle problems with regression and classification. It is a tree-structured classifier, where each leaf node represents the result, internal nodes indicate the dataset's attributes, and branching represent the classification rules. The Decision Node and the Leaf Node are the two nodes that make up a decision tree. The result of a decision made using a choice node, which has several branches, is represented by a Leaf node, which has no more branches. Based on the properties of the provided dataset, tests are run or choices are made. Decision trees are typically simple to understand since they show how people think while making decisions. The decision tree's logic is easy to understand since it presents a structure similar to a tree.

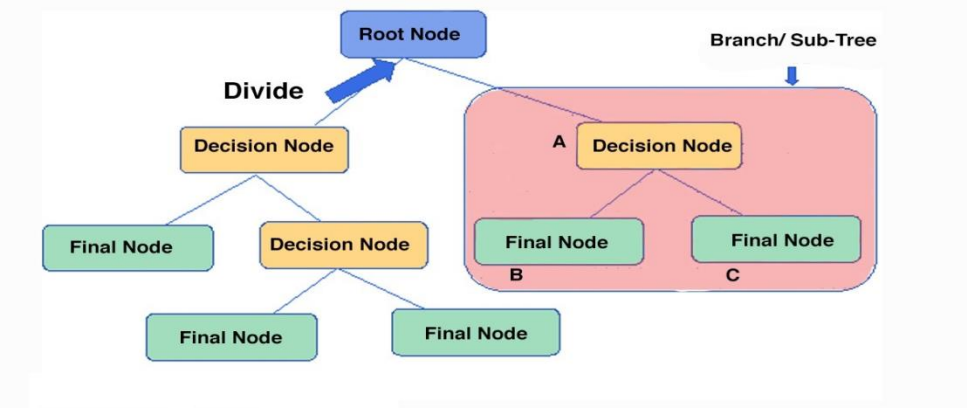


Fig 3.9: Decision Tree classifier workflow.

2. **Random Forest:** Use the RF Classifier's tree-based approach for regression as well as classification. In machine learning, it is an algorithm that generates a hierarchy based on trees. AI uses this technique to create "decision trees" that are arranged hierarchically. The Random Forest Classifier use an ensemble approach to generate a huge number of decision trees, which are then averaged. This clarifies the problems with overfitting. Because machine learning can be applied in every situation where there is a large amount of data, it is one of the most popular topics in business today. Because of the several advantages that RF Classifier has over other methods, machine learning with it is quite popular. The method was first published in 1997 and was designed to handle large datasets.

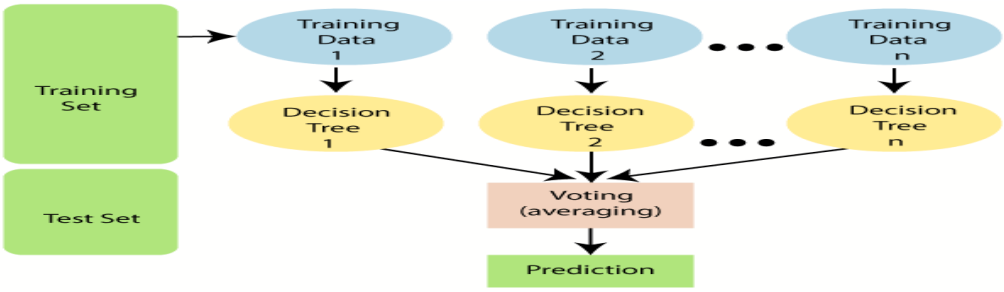


Fig 3.10: Random Forest Classifier workflow.

3. **SVM:** It's known as SVM, or Support Vector Classifier. Since SVC does not assume any information about the fundamental makeup of the data, including the number of clusters and their relative sizes, it is by definition a nonparametric clustering technique. Because of this, if your data has a large dimensionality, you will usually need to do some

preliminary processing like principal component analysis. It works well with low-dimensional data, as previous studies have demonstrated. The original technique has been modified in several ways, and these changes provide specific ways to compute the clusters by analyzing only a subset of the boundaries in the matrices of adjacency. Numerous modifications have been released.

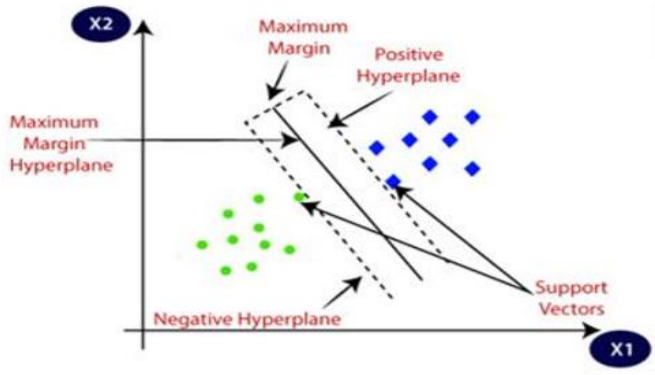


Fig 3.11: Figure of SVM.

4. **Naïve Bayes:** Since the Naive Bayes machine learning model can handle vast volumes of data, it is suggested even when dealing with data that comprises millions of records. It is quite good at NLP tasks like emotional analysis. Classifying is a simple, fast process. The Bayes theorem must be understood before we can comprehend the naive Bayes classifier. To begin, let us examine the Bayes Theorem. This theorem is predicated on conditional probability. The likelihood that something will happen in light of the occurrence of another event is known as the conditional probability. The possibility of an occurrence may be determined by using conditional probability and our past knowledge. Algorithms based on naive Bayes are extensively employed in several domains such as recommendation systems, sentiment analysis, and spam filtering. Their primary drawback is the need for independent predictors, despite their speed and ease of usage. As is the case in most real-world situations, the classifier performs poorly when prediction factors are dependent.

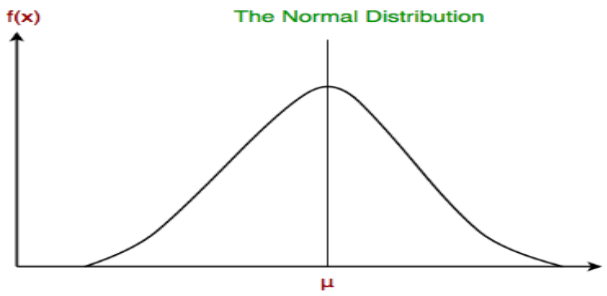


Fig 3.12: How The naive Bayes Functions

5. **XGBoost:** eXtreme Gradient Boosting, or XGBoost, is a well-liked and potent machine learning technique that may be applied to regression and classification problems. It is a fast and efficient way to implement gradient enhanced decision trees. As an ensemble learning technique, XGBoost builds a better model by aggregating the predictions from several weak learners, usually decision trees. The weak learners receive instruction one after the other, with each new student fixing the mistakes of the previous one. Different normalization penalties are included in XGBoost to prevent overfitting. Penalties regularizations result in effective training, enabling the model to achieve sufficient generalization. Non-linearity: Non-linear data structure may be recognized and learned from via XGBoost. Cross-checking: comes pre-assembled and ready to use.

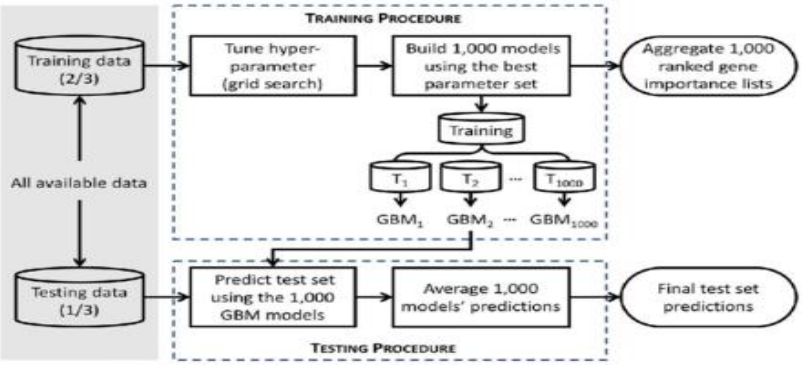


Fig 3.13: XGBoost model structure.

3.6 Implementation

After completing the previous procedures, we must gather the data in order to check correctness. For the necessary setup, our work was divided into nine sections. Every one of those tasks must be finished in order to reach our objective.

- Gathering data.
- Data pre-processing.
- Word clouds are visualized.
- Avoid Stop words.
- Text cleaning.
- Preparing Data
- The tokenization and padding.
- Make all five algorithms' models.
- Discuss the outcome and accuracy.

We had to get started on the method's code as the initial stage in implementing it. We evaluated the accuracy of five different methods. We assessed the algorithm's accuracy when it was completed. We evaluated the accuracy and came to the conclusion that one would be better for our purpose. which, in regards to the student despair, has been quite predictable. Following a thorough examination of the relevant theoretical and numerical concepts and techniques, a list of prerequisites has been developed that are required for any attempt at depressed classification. The following are potential required effects:

1.Equipment and Software specifications

- Operating System (Windows 7 or above)
- Hard Disk (minimum 1 TB)
- Ram (Minimum 4 GB)

2. Creating Tools

- Python Environment
- PyCharm.
- Google Colab.
- Visual Studio code.

CHAPTER 4
Experiment Results and Discussion

4.1 Introduction

This section covers the method used to categorize anxiety and sadness in students. There were just a few phases involved in the entire process of constructing a model: choosing a model, gathering and evaluating data, displaying word clouds, eliminating stop words, cleansing the text, and evaluating performance using output associated with depression detection. We offer the experiment results, which are covered in the part that follows.

4.2 Experimental Result

We are aware that no technology can produce perfect results. In a similar spirit, we may modify our model's parameters during training to increase accuracy. Despite this, the accuracy we discover using a variety of techniques is rather high. Below is a brief summary of our study efforts through photos. These images display the following: accuracy, precision, recall, f1 score, support, and heatmap.

Here, we identify the depression groupings that are most prevalent using our data.

4.3 Classification and Descriptive Analysis using ML Best Model

The different results we got depended on the method we used. Five machine learning techniques helped us predict the student depression label with accuracy. These approaches were employed to determine the relative effectiveness of every part of the overall framework, and several verification techniques were then utilized to arrive at the forecast's conclusion. Once the entire data set was selected, each model used a single file including both data that was freely accessible from online data sources and knowledge from our own experience. We tested the algorithms' performance using MATLAB and its already provided libraries after finishing the dataset approach. Using a comparable dataset, the next stage is to determine whether the content will be classified as depressed or not depressed in student emotions. Here, we used important performance parameters to undertake a thorough analysis of several models. Metrics including as accuracy, precision, recall, and overall F1 score provide a comprehensive picture of the models' performance.

Table 4.1: Accuracy Table

Classifier	Accuracy Score (AUC)
Decision Tree	98.35%
Random Forest	95.26%
XG Boost	98.70%
SVM (classifier)	98.42%
Naïve Bayes	89.89%

Several classifiers' performances are shown in this section. Throughout the entire process, two open-source tools were used: PyCharm and CoLab. In total, five classifiers were employed: XGBoost, Random Forest, Decision Tree, Support Vector Classifier, and Decision Tree.

	precision	recall	f1-score	support
Depressed	0.94	0.94	0.94	159
Not Depressed	0.99	0.99	0.99	1235
accuracy			0.99	1394
macro avg	0.97	0.97	0.97	1394
weighted avg	0.99	0.99	0.99	1394

Fig 4.1: XGBoost classification reports.

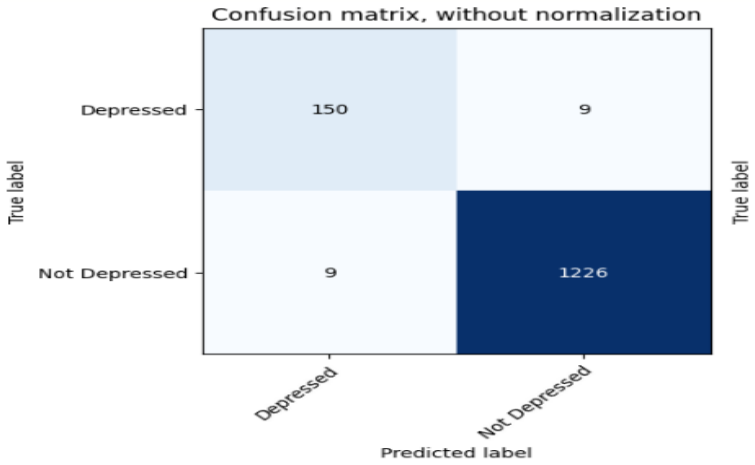


Fig 4.2: Confusion matrix of XGBoost.

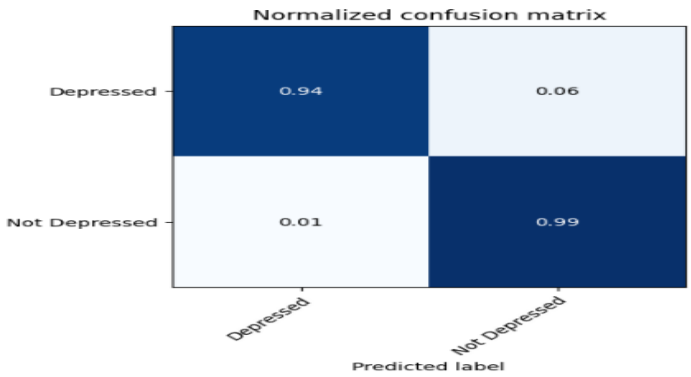


Fig 4.3: Normalization Confusion matrix of XGBoost.

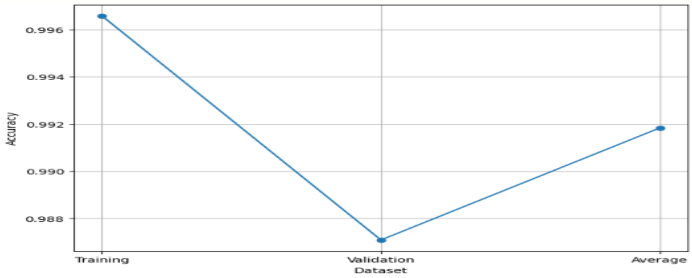


Fig 4.4: Training and Validation Accuracy for XGBoost.

To get the best accuracy, we are simply displaying the XGBoost classifying all report and normalizing matrix.
Describe the process of depression identification using the ML XGBoost classifier and the projected english student depression components in the label of the resultant data, as shown in Figs. 4.5 below. The two categories of false information that can be classified as depressed or not depressed, based on models, are listed below. Compared to detection or prediction, which is typically the case, the use of machine learning techniques is one of the most effective methods of learning without supervision. Compared to earlier approaches, it performed rather well.

```
# Preprocess new text data (replace 'new_text_data' with your actual new text data)
new_text_data = ["This is why my heart is beating so fast. Very restless. But I don't know what I'm worried about. It's a great feeling.....",
                 "if fate allows us to be together again for the second time. I promise I will tell you all the things that I went through when I wanted to forgive you. and how cry
                 "sometimes what is needed when there is a problem is to laugh until you forget that there is a problem, when you remember it, you feel restless like that well, it
                 "If the guy is sick, the girl actually feels sick indirectly. He must be feeling restless and restless. :')",
                 "When you are married you are asked when you are pregnant â . When you are married, you are asked whether you want to have children or not â... Humans are increasi

new_text_data_processed = [clean_text(text, CONTRACTION_MAPPING) for text in new_text_data]
new_text_data_sequences = tokenizer.texts_to_sequences(new_text_data_processed)
new_text_data_padded = pad_sequences(new_text_data_sequences, maxlen=X_SEQ_LEN)

# Use the trained model for predictions
new_text_predictions = text_clf_XGBoost.predict(tokenizer.sequences_to_texts_generator(new_text_data_padded))

# Decode the predicted labels to class names
predicted_class_names = encoder.inverse_transform(new_text_predictions)

# Check the predicted class for each text
for text, predicted_class in zip(new_text_data, predicted_class_names):
    print(f"Text: '{text}' is '{predicted_class}'.")
```

```
s a great feeling.....' is 'Depressed'.
 that I went through when I wanted to forgive you. and how cry' yes I was that day. :')' is 'Not Depressed'.
m, when you remember it, you feel restless like that well, it turns out that I still have a sad burden blah blah blah' is 'Depressed'.
s. :')' is 'Depressed'.
ther you want to have children or not â... Humans are increasingly wanting to love God, aren't they? The question is just like asking when are you going to die?:' is 'Not Depressed'.
```

Fig 4.5: Students depression detection.

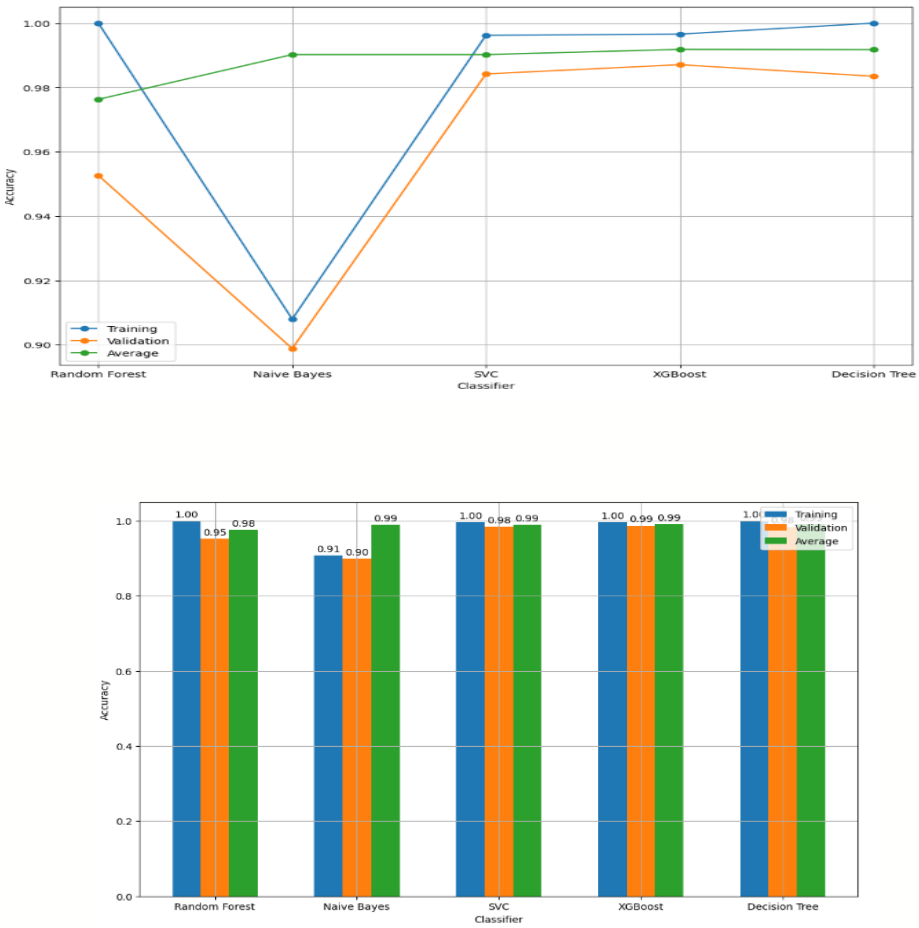


Fig 4.6: Comparison of Train, validation and average accuracy of five classifiers

4.4 Discussion

We will use machine learning (ML) models in our study to predict student depression. Every phrase should be really vital to the categorization in any area research. From the beginning, our research's objective has been to identify the depression text. The data is one of the most important parts of every investigation. The results of an identical experiment may vary greatly depending on the data that are provided. Since we used a combination of datasets, we were aware that the results that somebody else would get using one of the experiment's datasets that was previously available would differ. We were able to reach our goal by utilizing a variety of machine learning algorithms for the average and accuracy score. For this project, we used Five different algorithms in all. We had to look for a number of items before starting our assignment. After choosing the algorithm, we did get to work on it. Next, we discovered each

algorithm's accuracy.

Utilizing approaches, we were able to achieve the highest accuracy of 98.70% in XGBoost, the classifier for the following five models.

Precision: Precision is the capacity of an indicator to discern true positives from all anticipated positives. The following formula is accessible:

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

Recall: The model's recall refers to how well it can distinguish between each genuine positive and the forecasted positives. The structure of the equation is explained as follows:

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

F1-Score: Recall and accuracy together allow the F1 score to give a meaningful evaluation of the effectiveness of the model. It may be computed with the following formula:

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Accuracy: The percentage of all reported instances and accurately predicted occurrences in the data set is known as accuracy. computed in this manner:

$$Accuracy = \frac{TP + TN}{Total\ Instances}$$

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The most significant component of society are the students. Public health issues, such as recurrent mental health issues, may lead pupils to participate in risky behaviors, such as suicide behavior in extreme conditions. Students' mental health problems have a detrimental influence on society. A student who has an unforeseen psychological problem finds it challenging to focus on his academic work. This horrible sickness has an impact on their day-to-day existence. The pupils become unconscious of all that is around them. A student who is experiencing psychological problems tends to isolate himself from friends, family, and the larger community. As a result, a student cannot successfully carry out his civic duty. A socially acceptable mental health problem affecting a large number of students will have an impact on society as a whole and may lead to an unexpected occurrence such as suicide.

5.2 Impact on Environment

Compared to others who are not depressed, someone with depression has a decrease in their sense of wellbeing. He damages the environment in addition to himself. A depressed person is more prone than a normal person to get addicted to various substances, which puts him and the society at large in danger. Additionally, a depressed person is not particularly good at protecting the environment. Depression thus has detrimental impacts on the body, the surroundings, as well as the psyche. Our research aims to detect depression, which will help a person resume a healthy lifestyle. We anticipate that this will also be advantageous to the environment.

5.3 Ethical Aspects

The depression detection equipment is not unethical nor does it violate human rights in any way. Since the model does not collect any personal information, such as name, identification, etc., there won't be a privacy concern. This tactic raises awareness as opposed to diminishing someone's capacity to understand or make use of something. As a result, the depression detection model could be managed with ease by applying machine learning modern technology.

5.4 Sustainability Plan

A sustainable strategy for machine learning (ML)-based depression diagnosis must carefully take data security, privacy, and ethics into account. Work together with mental health providers and organizations to have access to ethically and anonymized data.

Give user confidentiality and privacy a priority. Make sure that the usage of data is made clear and that all data protection laws are followed. To deal with outliers and missing numbers, clean up and preprocess the data. To prevent bias in the model, make sure the dataset is representative of a range of demographics and is balanced. Try out several machine learning models, such neural networks, decision trees, and support vector machines. Assess models according to pertinent metrics such as sensitivity and specificity to maintain a balance between true positives and erroneous negatives.

CHAPTER 6

Conclusion and Future Research

6.1 Summary of the Study

This investigation has taught us a great deal about this topic. Depression predictors remain a controversial topic. This has a major influence on the annual drop in patient use of student's emotion anxiety depression detection. As a result, we used machine learning to classify two classes of people: Depressed (1) and Not Depressed (0), which represent a serious risk. The ML model has also been used to determine student sadness based on looks.

As we already mentioned, the goal of the research is to learn as much as feasible about this topic. We did this by combining data from kaggle.com, an internet resource that has two labels: depressed and non-depressed. We were capable of to examine and train our five strategies on the exceptional precision of the classifier developed by XGBoost in detecting the precise "text" of the "label" with the aid of the labels that were extracted from the data sets. The model prediction approach makes it easier to determine if someone is sad or not. At initially, a few problems were fixed. We succeeded in achieving the intended result. Different strategies provide different outcomes for different students. More of this was explored in the next section.

6.2 Conclusion

Our goals were to find out how common stress, anxiety, and depression were among Bangladeshi students, what risk factors were linked to them, and how we might help them get well by offering them a treatment plan. Nowadays, depression is a pretty common occurrence among people. Among Bangladeshi students, it has now become an unbearable and unexpected mental health issue. The purpose of this study was to ascertain how common anxiety is among Bangladeshi students. By producing, processing, and evaluating vast volumes of data and utilizing state-of-the-art machine learning techniques, we hope to enhance depression detection. Our models' high recall, precision, accuracy, and F1 scores demonstrate their efficacy in detecting false positives. However, the journey is far from finished. After applying five classifier XGBoost get highest accuracy 98.70% for prediction student depression.

6.3 Possible impacts

We contend that the results of our work vary with respect to the observation site. As a result, we worked harder to complete our tasks precisely. We take security measures to ensure that the job we perform is legitimate. We have chosen to utilize an online dataset because of this. ML models have the capacity to detect depression patterns and indications sooner than conventional diagnostic techniques. Timely intervention and therapy following early discovery may improve overall results. More people may obtain mental health examinations thanks to the deployment of ML-based technologies in a variety of contexts, such as web platforms and mobile applications. By becoming more accessible, we may be able to assist those who would not have otherwise sought assistance.

6.4 Implications of Further Study

Recent developments in artificial intelligence, data science, and contemporary technology have streamlined, expedited, and improved all facets of human existence. In the future, we hope to develop our method into an Android or web application. We will work to improve our models' accuracy in the future. We'll expand the repository to hold a vast amount of personal data and improve the datasets by adding more layers of categorization. Furthermore, the website or mobile application created for the device may be made available to everyone, including medical experts, by providing an immersive, user-friendly graphical user interface. The efficacy and usefulness of the model can be substantially increased by including more features, parameters, and algorithms. A comprehensive dataset may be created in the future by collecting information from a wider range of people based on their activities, age, and location. Furthermore, the Department of Mental Health could support the concept's expansion.

Reference

[1] R. Bruffaerts, P. Mortier, G. Kiekens, R. P. Auerbach, P. Cuijpers, K. Demyttenaere, J. G. Green, M. K. Nock, and R. C. Kessler, "Mental health problems in college freshmen: Prevalence and academic functioning," Journal of AffectiveDisorders,vol.225,pp.97–103,2018.

[2] D. Y. Mahmood, "Classification trees with logistic regression functions for network-based Intrusion Detection System", IOSR Journal of Computer Engineering, 19(03), 48–52, (2017).

[3] "Social consequences of psychiatric disorders, I: Educational attainment" (1995) American Journal of Psychiatry, 152(7), pp. 1026–1032. Available at:

[4] M. A. Mamun, M. S. Hossain & M. D. Griffiths, "Mental health problems and associated predictors among Bangladeshi students" International Journal of Mental Health and Addiction, 20(2),657–671, 2019.

[5] M. A. Bhuiyan, M. D. Griffiths & M. A. Mamun, "Depression literacy among Bangladeshi pre-university students: Differences based on gender, educational attainment, depression, and anxiety". Asian Journal of Psychiatry, 50, 101944, 2019.

[6] G. Mei, W. Xu, L. Li, Z. Zhao, H. Li, W. Liu & Y. Jiao, "The role of campus data in representing depression among college students: Exploratory research. JMIR Mental Health", 2020. 7(1).

[7] A. K. Masum, S. Abujar, S. Akter, Ria, N. J., & Hossain, S. A. (2021). Transformer based Bengali chatbot using general knowledge dataset. 2021 20th IEEE International Conference on Machine Learningand Applications (ICMLA).

[8] D. M. Shukla, K. Sharma & S. Gupta , "Identifying depression in a person using speech signalsbyextracting energy and statistical features. 2020IEEEInternationalStudents' Conference on Electrical,Electronics and Computer Science (SCEECS), 2020.

[9] M. Deshpande & V. Rao, "Depression Detection Using Emotion Artificial Intelligence.", 2017 International Conference on Intelligent Sustainable Systems (ICISS), 2017.

[10] N. A. Asad, M. A. Mahmud Pranto, S. Afreen & M. M. Islam, "Depression detection by analyzing social media posts of user." 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), 2019.

[11] A.D. Furlan et al. ,"2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group," Spine, 34(18), pp. 1929–1941, 2009. Available at:

[12] Y. Ding, X. Chen, Q. Fu & S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm.", IEEE Access, 8, 75616–75629, 2020.

[13] D. Dahiwade, G. Patle& E. Meshram, "Designing disease prediction model using machine learning approach", 2019 ,3rd International Conference on Computing Methodologies and Communication(ICCMC).

[14] N. Ignatiev, I. Smirnov & M Stankevich, "Predicting depression with text, image, and profile data from Social Media." Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods, 2020.

[15] S. H. Hamaideh, "Alexithymia among Jordanian University students: Its prevalence and correlates with depression, anxiety, stress, and Demographics. Perspectives in Psychiatric Care," 54(2),274–280, 2017.

[16] T. Richter, B. Fishbain, A. Markus, G. Richter-Levin and H. Okon-Singer, "Using machine learning-based analysis for behavioral differentiation between anxiety and depression",2020, *Scientific reports*, *10*(1), p.16381.

[17] A. Ahmed, R. Sultana, M.T.R. Ullas, M. Begom, M.M.I. Rahi and M.A. Alam, "A machine learning approach to detect depression and anxiety using supervised learning", In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-6), 2020, December.. IEEE.

[18] P. Kumar, S. Garg and A. Garg, "Assessment of anxiety, depression and stress using machine learning models", *Procedia Computer Science*, *171*, pp.1989-1998, 2020.

Depression prediction among student based on their daily activities by Miskatun Ahmed Pranty (ID: 201-15-13713)

ORIGINALITYREPORT

12%	5%	3%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARYSOURCES

1	dspace.daffodilvarsity.edu.bd:8080 InternetSource	5%
2	SubmittedtoDaffodilInternationalUniversity Student Paper	2%
3	Submittedto University of Southern Mississippi Student Paper	2%
4	Submittedto University of NorthCarolina, Greensboro Student Paper	1%
5	ijisae.org InternetSource	1%
6	SubmittedtoJacksonvilleUniversity Student Paper	<1%
7	Barnali Adhikari, Syeda Ismatara Era, MohamedDahirMohamed,Jueal Mia. "Depression Level Prediction For Students UsingMachine LearningInThe Context of Bangladesh", 2021 12th International	<1%