PREDICTING DEPRESSION AMONG THE UNIVERSITY STUDENTS IN THE COVID-19 PANDEMIC SITUATION USING MACHINE LEARNING ALGORITHMS

BY

HRITHIK MOJUMDAR
ID: 181-15-1814

AFROZA RAHMAN ASHA
ID: 181-15-1977

AND

MOHAMMAD ABDUR RAHMAN
ID: 181-15-2063

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

Supervised By

MD. SABAB ZULFIKER
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By

MD. MAHFUJUR RAHMAN
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project/internship titled "**PREDICTING DEPRESSION AMONG THE UNIVERSITY STUDENTS IN THE COVID-19 PANDEMIC SITUATION USING MACHINE LEARNING ALGORITHMS**", submitted by **Hrithik Mojumdar,** ID No: 181-15-1814, **Afroza Rahman Asha,** ID No: 181-15-1977 and **Mohammad Abdur Rahman,** ID No: 181-15-2063 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 January 13, 2022.

<u>BOARD OF EXAMINERS</u>

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

Shah Md. Tanvir Siddiquee   **Internal Examiner**
Assistant Professor
Department of Computer Science & Engineering
Daffodil International University

Md. Reduanul Haque   **Internal Examiner**
Assistant Professor
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

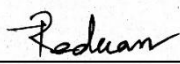**Dr. Dewan Md. Farid**   **External Examiner**
Associate Professor
Department of Computer Science & Engineering
United International University, Bangladesh

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Sabab Zulfiker, Lecturer (Senior Scale), Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md. Sabab Zulfiker
Lecturer (Senior Scale)
Department of Computer Science & Engineering
Daffodil International University

Co-Supervised by:

Md. Mahfujur Rahman
Lecturer (Senior Scale)
Department of Computer Science & Engineering
Daffodil International University

Submitted by:

Hrithik Mojumdar
ID: -181-15-1814
Department of Computer Science & Engineering
Daffodil International University

Afroza Rahman Asha
ID: -181-15-1977
Department of Computer Science & Engineering
Daffodil International University

Mohammad Abdur Rahman
ID: -181-15-2063
Department of Computer Science & Engineering
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Sabab Zulfiker, Lecturer (Senior Scale),** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of **"Predicting Depression Among the University Students in the Covid-19 Pandemic Situation Using Machine Learning Algorithms",** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Professor **Dr. Touhid Bhuiyan**, Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

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

ABSTRACT

Depression is a very common word term. But the problem of depression is not normal. People think it is a disease but it is not a disease. This is a mental problem. People suffer from depression for various reasons. At present, the amount of depression among the people seems to have increased a lot. We have tried to credit depression among all the university students of Bangladesh during the Covid-19 pandemic period. When someone is depressed we can find out what they want to do or what they have done to get rid of depression. Doctors say depression is a mental disorder. The person who is suffering from this problem is basically worried about his own wrong doing or any failure in life or there is no one close to him anymore and he cannot accept it. Depression can also be caused by family strife or financial problems. Depressed people don't want to talk to anyone else. They always try to hide their depression. In our research, we have used some scales to understand how a depressed student thinks about himself or what he wants to do and how he can free himself from depression. The survey will give us a clear idea about these issues. In this research Logistic Regression, Random Forest, Multi-Layer Perceptron, Multinomial Naïve Bayes, Gradient Boosting, K-Nearest Neighbours, AdaBoost, Support Vector Machine have been used. Among these algorithms, Random Forest Algorithm has shown the best result. Random Forest has an accuracy of 85.57% which is approximately 86%.

TABLE OF CONTENTS

CONTENTS   **PAGE**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Depression is a mental condition marked by continuous sorrow for at least two weeks. It results in difficulty to conduct everyday tasks, and depressed people lose their self confidence, joy, and positivity in doing the things they normally love [1] (World Health Organization, 2021). Approximately 280 million people [1] are suffering from depression due to the Covid-19 pandemic situation worldwide. Every day the world is getting more and more modern. To keep pace with this growing modernity in the world we are facing different kinds of problems. Such as physical, mental, social, religious, economic, etc. Because of these problems we get depressed anytime, anywhere. When these problems are in the initial step, we don't pay attention to these problems or try to solve them. This depression makes us emotionally and physically damaged, isolated from family and society and even leads to suicidal tendencies.

Nowadays the maximum young generation is facing this depression. The causes of this depression in the younger generation are somewhat different such as study problems, career problems, social life, body shaming, cyberbullying, sexual harassment, etc. These problems become even more frightening when the younger generation does not share these problems with anyone or find a way to solve them, which leads them to loneliness. Our study is to find out the cause of this kind of problem in the university students of Bangladesh in the Covid-19 pandemic situation and find out a way to solve it.

For this study, we surveyed more than 413 university students. In this survey, we used two Psychometric Scales which are PHQ-9 (Patient Health Questionnaire-9) and AIS (Athens Insomnia Scale), and a few other questionnaires suggested by experts. PHQ-9 is a psychometric scale where a doctor asks a patient 9 questions related to depression. Every question has four options and every option has a score of 0 to 3. Similarly, AIS is also a Psychometric scale to find a score to define a person have insomnia or not. It has 8 questions. Using these two medical science scales we will get a number/point. These numbers/points indicate a stage or level of depression and insomnia that is depressed or not depressed and insomnia and not insomnia. According to the PHQ-9 scale if the number/point is greater than 4 then that person will be depressed and in AIS if the number/point is greater or equal to 6 then that person have insomnia.

In our research, we used 8 classifier algorithms which are Logistic regression, Multi-layer perceptron (MLP), Random Forest, Support Vector Machine (SVM), Multinomial Naive Bayes, Gaussian Naive Bayes, Gradient Boosting, AdaBoost, etc. Finally, the efficiency of the classifiers was differentiated by us.

1.1 Background:

1.1.1 What exactly is depression?

Depression is a significant but serious mental disorder that affects your emotions, thoughts, and actions. We often think of sadness and depression as one. Sadness is a temporary upset that goes away after a while. No treatment is required for this. Depression, on the other hand, is a long-term problem. Appropriate treatment and advice are needed to get rid of it.

1.1.2 What causes depression?

Depression is a complex disease. It is not possible to say exactly why this disease occurs. However, in many cases there are some common reasons for the origin of the disease.

- Humiliation: Many suffer from depression if they admit to being humiliated emotionally or physically.
- Insecurity or loneliness: Many suffer from depression due to social and family insecurity. In addition, many suffer from depression due to a lack of relationships or disagreements with parents, friends, or other close people.
- Mortality: The death of a loved one increases the risk of depression in many cases.
- Hereditary effects: If someone in the family has depression, it can affect others.
- Major lifestyle changes: Many people suffer from depression when there is a big change in life. Many suffer from depression when they lose their jobs, retire, lose income, change places, get divorced, or even remarry.
- Side effects of a major disease: If there is a major disease, the patient may suffer from depression.
- Effects of Medications: Some people also suffer from depression as a result of taking certain medications. For example, taking isotretinoin or the antiviral "interferon-alpha" used to treat acne can lead to depression.

There are also many other reasons why people suffer from depression. Differences can be seen due to depression in individuals.

1.1.3 Who suffers more from depression?

About 10 to 20 percent of the population in any country suffers from depression. Girls are twice as prone to this disease as men. Unemployed and destitute people (old, abandoned wife) also suffer more from this depression. The rate of depression is also higher among those who use different types of drugs. Statistics show that depression is more common between the ages of 18 and 45. There was an idea before, the depression is usually in the elderly. But this idea does not seem to last now. The pressures of modern civic life are turning all calculations upside down.

1.2 The Research's Motivation:

According to the media report [2], more than 150 students committed suicide in the covid-19 pandemic situation. This report mentions that most of the students feel lonely, they do not share their stress with anyone, they don't get adequate emotional support from family/society, etc. In pandemic time, students try to spend their time where They felt comfortable spending time doing what they loved like gaming, cooking, using social media, etc. but some of their families forcefully rejected their choices.

The motivation for doing this research is to know the specific few reasons for depression or stress. Also, we can make people aware about depression and its bad side. In this research, Machine Learning algorithms prognosticate depression among Bangladeshi University Students.

1.3 The Research's Objective:

The harmful effects of Covid-19 are not limited to ordinary people. It has had a terrible effect on the lives of students. The main purpose of this research is to highlight how much the students of our country are suffering from depression because of Covid-19. All kinds of educational institutions have been closed for a long time. Besides educational institutions, people's working areas are also closed and as a result, many people have become unemployed. So there have been financial problems. Many families have become economically weak. Many parents are not able to provide food for their families. In this situation, it is very difficult to continue studying. Many students are dropping out of their studies in search of work to provide food for their families. Some are getting jobs, some are not. Whoever finds the job may not be able to support their families properly. As a result, they are suffering from financial and mental problems. The rate of depression is higher among university students because they are in a situation where they can't get a good job even if they want to and can't drop out of their studies. That's why they feel like a burden to their family.

1.4 Expected Outcome:

- We can know the causes of depression in a specific way.
- We can find out the type of depression by gender and age.

1.5 Statement of the Issue:

At present the rate of depression is increasing day by day. This rate of depression is more noticeable among students. The rate of depression among students is increasing due to spending time alone at home in this corona epidemic. Since most students are between the ages of 18 and 26, they do not want to share their stress with anyone else. Due to which the amount of their depression, mental and physical damage increases more.

Through this research we will be able to determine the amount of depression, mental and physical condition based on the experiences of the students already and now rather than taking the personal information of the students.

1.6 Social Impact:

Depression may have a negative impact on every aspect of a person's life. Depression's side effects can be evident at work, school, and home, as well as in the patient's personal relationships. Depression has physical consequences on the brain, heart, and other areas of the body. Many studies demonstrate that depression has a detrimental impact on the brain.

Our research will aware people in the society about depression. People will learn about the main causes of depression and those who suffer from depression will have a comprehensive idea of how to recover from depression.

1.7 Research Scope:

Depression might not be implemented by machine learning Classifier algorithms. Using the machine learning technique, we got good perfection. As human minds, we will work hard to learn in order to make judgments in the future.

1.8 Report Layout:

The following is a list of the contents of this research paper:

✔ The first chapter discusses the research's objective, justification for the study, research questions, and predicted outcomes.

✔ The second chapter discusses related works, a study summary, the problem's scope, and problems.

✔ The workflow of this study, the data gathering technique, statistical analysis, and feature implementations are all covered in Chapter 3.

✔ Chapter four experimental evaluation and other related debates, as well as the numerical and graphical results of study.

✔ The significance of this research on society is discussed in Chapter 5.

✔ The sixth chapter offers a summary of this study project, as well as a list of references.

CHAPTER 2

LITERATURE REVIEW

2.1 Related works

Several studies have been performed around the world to predict depression. As previously reported, they have proposed many additional systems. Several explanations and reviews are provided below to demonstrate why our work is comparable with their method.

Mamun et al. [3] worked with problematic internet use. They mention that they worked with problematic internet use. The mention that the internet can be the cause of depression if the user is too much addicted to it. According to them, the majority of students use the internet daily. Also, the participant has sleep issues. They did not talk about spending time with family. Participants use the internet badly because they can't spend time with their family properly.

Van Eeden et al. [4] studies about mood and anxiety disorders. They compare 96 models then compare them with 3 methods in 04 sets of predictor variables over two outcome sets on four follow-up waves. They use multinomial Logistic regression, naive Bayes classifier, and auto-sklearn. We are researching on Depression. Covid-19 affected all over the world. Students are affected by this. They stopped going to school or college because of this pandemic. So we research on student's depression. And we see that they are really depressed. Gradually they stop playing outside. They are seriously attacked by mobile phones, computers, and laptops for online gaming. So we research on student's depression. And we see that they are really depressed. Gradually they stop playing outside. They were seriously attacked by mobile phones, computers, and laptops for online gaming.

Hasan et al. [5] worked with only medical students and their research was in 2013's data. They did not mention any specific reason for depression and there is no suggestion or solution for depressed people. Most of the students were suffering from depression because of their academic pressure. From their study they find out that the tendency of suicide was very high.

Akhtar et al. [6] were working about depression in the covid-19 pandemic period. They do not use other methods or subjects with this. Their questionnaires are made with a mental health state, understanding quarantine, etc. They worked with less than 300 pieces of data. Their study mainly focused on people's knowledge about Covid-19 related safety issues or

lockdowns. Because of the lockdown people were facing some mental health issues because their financial status was going poor day by day.

Choudhury et al. [7] worked on only undergraduate students for predicting depression. For gaining the accuracy of their model, they applied k fold cross-validation. In their study they predict whether they will be depressed or not. They did not discuss the reason behind this depression. Their research finds out the reason for depression at an early age and finds the solution of recovery also.

Islam et al. [8] mentioned a piece of good information about how covid-19 Impacts our mental health. According to them when they used PHQ-9 scale to find Persistent Symptoms on covid-19 infected they found severe depression 48.2% and minimal 30.4%.

Mashreky et al. [9] were working on suicide case for mental pressure. People commit suicide due to some mental pressure or any physical pressure. In this paper, they write the majority of suicide cases happened because of poor. But they did not mention whether those victims were depressed or not.

Rasheduzzaman et al. [10] were working on students depressed by campus bullying, family problems, career and study problems, etc. but they did not say about student's financial problems, sleep induction, indecision, etc. Also, they did not use any verified methods to calculate depression levels.

Priya et al. [11] worked on depression, stress in modern life of people. They used five different Severity levels and classification of Depression and stress. They determine five anxieties, stress and depression. From their study they get that It's a psychological disorder.

Biilah et al. [12] worked on people's depression and they used the association rule of mining technique. They identified eighteen significant features by association rule mining. They collected data from students of several universities. They use apriority algorithm to manage database records. Then they collected 23 attributes and utilized it.

Li et al. [13] used virtual reality to cognitive function for moderate depression subjects. It explores the patient of depression.

Cvetkovi☐ et al. [14] worked on depression among Breast cancer people. 30-78 years old people are participating in this research. Here they use three types of algorithm for showing best performance. They divided the patients into two phases. They measure the depression range by BDI test. They research on 30-78 years old people only.

Na et al. [15] worked on future depression on socio-familial relationships. It indicates the measure of happiness or satisfaction on socio-familial and health. Here they use a Random Forest classifier. This research accuracy is 86.20%. They research relationships.

Richter et al. [16] studied a diagnostic support system for finding the differences among clinical anxiety and depression disorders. To complete this research, they perform various tasks. But they didn't find the reason for anxiety and depression.

Moon et al. [17] studied about depression in the job sector. Here they show the happiness or depression in people's workplace. Who is happy or who is not happy and why they don't like their job and the effect on their family is explained in the paper. They applied RF Classifier, RF Regression, Naive Bayes and K Neighbors Classifier for accuracy.

Govindasamy et al. studied on detecting depression using the machine learning techniques on Twitter data is the topic of the paper [18]. In social media, people share their emotions, happiness, sadness. Sometimes people share their thoughts about their mental health or illness. Connecting with social media people start to share everything. Now people like to share their words via social media. Now they have no need for privacy. Collecting data from people's social media they analyzed them and figured out their level of mental health.

Zhou et al. [19] studies clinical remission in depressed patients. In their research, almost 400 patients enrolled. After 8 weeks of treatment, 67.8% achieved clinical remission. They use 78 features and machine learning algorithms, GBDT, SVM, RF, LR, etc.

Bailey et al. [20] studies EEG. EEG is electroencephalography. It means the change of brain activity or brain disorder. It has two parts. one is responder and another is non-responder. They measure it by theta and alpha connectivity. They comprised their dataset with 42 participants.

Zhang et al. studied about PPD. PPD is a very serious problem among pregnant women mentioned in paper [21]. PPD means Postpartum depression. It usually seems among pregnant women. It is a mental health condition. After childbirth women face this kind of problem. Because of poor marital relationship, stressful life or low socioeconomic status the problem of PPD increases. They use EHR data. Their 80% data is from WCM as a training set and 20% data come from their test set.

Depression and post-traumatic stress disorder in sexual abuse research on paper [22] study by Gokten et al. Sexual abuse mostly happens in childhood. Girls are facing this problem too much. Neurological changes occur in the brain of the victim. Using a Random forest classifier, they analyze the problem of depression and stress in sexual abuse.

de Souza Filho et al. [23] studied screening tools for depression in primary care. By screening, they find out the undiagnosed patients and reduce them. They used AI and by this they have several successful implementations. In our research, we collect data from

students to see their depression level and we use machine learning. Then we get the accuracy of depression level. We find out the problem of depression.

Poletti et al. [24] studies on the variance between bipolar and unipolar depression. It is a mood disorder. There are two types of mood disorders. One is a Major disorder and the other is Bipolar disorder. In the world, 33% of depressed people didn't take any medication. In this research, they compare the patients of MDD and BD. And they find out that BD patients have a higher load of antipsychotics, frequency.

Wang et al. [25] mentions specific neural circuits in their research. Multi-model MRI and machine learning was used. They generate two major findings. One of them was neuropsychological impairment and the second one was neuroanatomical substrates. Neuropsychological impairment is unclear whether deficits in the domain represent cognitive deficits. Neuroanatomical substrates are the underlying key factor to a deficit of rLOD.

Jacobson et al. [26] studies anxiety and depression. They used python to learn. They used many models. The model has two layers. One was the Base layer and the other was the Averaging layer. Using this layer, they get the result.

Su, D., Zhang et al. to depression in elderly in china paper [27]. As per their report they said that the prevalence of depression between the ages 55-74 is too high. Because of their different physiological and social relation their depression level is different. In China 4.2% of people are depressed and 22.7% of people of age 60 are highly depressed.

Rahman et al. [28] author worked on the covid-19 effect on mental health and the uses of the internet in lockdown. They did not use any kind of medical science scale of depression. But they apply data visualization algorithms for simulating psychosocial health and find out the percentage of depression.

Wardenaar et al. [29] studies about nine year depression and course-trajectories on specific determinants. Through their research paper they expressed anxiety and depression as chronic-intermittent course-trajectories. This is a nine years' cohort study. To determine the severity, they compare pure depression, anxiety, and psychological distress trajectories. But research papers are not compared with anything. It is based on Covid-19 pandemic effect on student's mental health.

Narayanrao and Kumari [30] mentioned an important thing that 4.4% of the globe are suffering from depression and symptoms of depression between males and females are the same. They also said that depression causes mental disorders. They performed predictions using predictive analytics. There they just predict depression by machine learning.

2.2 Analysis and Summary

The research is about university students due to the Covid-19 pandemic situation. To get accurate results we use two scales. The scales are PHQ-9 (Patient Health Questionnaire-9) and AIS (Athens Insomnia Scale). On those scales, there are some Depression related and insomnia related questions. Like- less sleeping, worried about the future or not, facing sexual harassment, losing interest in work, suffering indecision, thinking about suicide, less sleeping, health issues, etc. The research is mainly for university students. Because of the Covid-19 pandemic situation, all kinds of educational institutes are closed. Students are becoming alone because they cannot meet with their friends, cannot go out, or study also. Some families lose their father or mother; some are facing financial problems. Because of those problems students are losing hope in their life or family. Here we find out the depression level of university students due to the Covid-19 pandemic. Sometimes students are trying to suicide because of depression, some are losing health or cannot sleep properly. We reviewed others' papers about depression but did not find any papers about university students' depression. In other papers, they study depression because of job pressure, breast cancer, clinical remission, Covid-19 effect on mental health, bipolar and unipolar depression, neural circuits, screening tools, problematic internet use, mood and anxiety disorders, depression of medical students, undergraduate student's depression, suicidal case, depression for campus bullying, depression and stress in modern life, etc. Covid-19 pandemic situation badly affected people's regular life. It changed everything. No one can think that students can also suffer from depression because of this pandemic. Students of 18-25 years old are much more depressed and from them, females are more depressed than males. We gave visual data of our research. We work with two medical science scales and other points which are suggested by experts including the covid-19 effect in our both mental and physical health.

CHAPTER 3

RESEARCH METHODOLOGY

This survey has been applied among 413 university students in Bangladesh in the pandemic situation. Through this study, we have started working with the data of 413 university students of Bangladesh. The system will decide with this data using some machine learning algorithms. This decision is about whether these students are suffering from depression or not. The following subsections explain the complete methodology of this investigation.

3.1 Data Collection:

First, we analyzed the requirements of this research. One of the most important requirements here is data. The first step of starting our research was to collect data. Then we think about how to collect new qualitative data without using pre-prepared data. For this, we have to create some questionnaires. But in this case, we can't create any kind of questions. So we create questions with the help of experts. And we take the help of our advisors as the questions are easier and less time consuming for our participants.

We have included 36 questions in this survey form. Of these, 16 are general questions that are suggested by experts, and the remaining 20 questions are psychometric scales which are PHQ-9 and AIS questionnaires. To find the depression level we used here PHQ-9 and for finding insomnia we used here AIS.

To make our questionnaires easier to avoid grammatical errors and to find other mistakes of questionnaires we made a pilot survey. In this survey 10-15 people participated. With their valuable review and suggestion, we have corrected our questionnaires and also have made them more accurate for the actual survey.

Variables for Predicting Depression:

This table is created based on dataset attributes. Here Variable name refers to all column names of the dataset. In our dataset, we divide the attribute into two parts. That is the predictor and predictor target. Variable descriptions have short descriptions about the record of those attributes. In possible value refers to an example of attributes data.

Variable Name	Variable Type	Variable Description	Possible Values
AGE	Predictor	Participant's age in year	Number (18-30)
GENDER	Predictor	Whether Participant's are Male or Female	Male, Female
PMCOVID-19	Predictor	Whether the participant suffer from any physical or mental health problems due to Covid-19 pandemic situation	Yes, No
FFP	Predictor	Financial Problem of the Participant's or Participants family in the pandemic	No Financial Problem, Low Financial Problem, Serious Financial Problem
FPPS	Predictor	Participant's family problem or personal separations in the pandemic time	Yes, No
UPSET	Predictor	Upset for some reason	Yes, No
HISSUE	Predictor	Big health issue of the participant or someone in his / her family during the pandemic	Yes, No
ALCDRUINB	Predictor	Whether the participant's take drugs/alcohol or not	Yes, No
ANGER	Predictor	Whether the participant's get angry in a small matter or not	Yes, No

SL6-8	Predictor	Whether the participant's Sleep less than 6 to 8 hours or not	Yes, No
SUICIDAL	Predictor	Does participant's tried to Commit suicide/Suicidal thought or not	Yes, No
TISM	Predictor	How many hours the participant's spend in social media	Less than 1 hour, 2-4 hour, More than 5 hour
TWF	Predictor	How many hours the participant's spend with family	Less than 1 hour, 2-4 hour, More than 5 hour
WORRIEDF	Predictor	Does participant's worried about life/future career or not	Yes, No
HARASS	Predictor	Whether the participant's recently faced physical/ mental/ sexual harassment or not	Yes, No
LOIW	Predictor	Whether the participant's loss interest in work recently or not	Yes, No
INDES	Predictor	Whether the participant's suffering in the decision/bad decision or not	Yes, No
RIDICULED	Predictor	Does participant's get ridiculed in public/among friends or not	Yes, No
NIGHTSCREENTIME	Predictor	How much time the participant's use mobile or computer regularly at night	Less than 1 hour, 2-4 hour, More than 5 hour
WORRIEDP	Predictor	Does participant's worried about his/her past or not	Yes, No
DEPRESSED	Predictor Target	Depressed is a variable that indicates whether or not a participant is depressed	Yes, No

Table 1. Variables for predicting the severity of Depression.

This study has targeted Bangladeshi university students. Students answered our questions without giving their personal information and we are just collect their problems data. We have created the form in such a way that it is possible to answer the question in 5-8 minutes. More than 413 university students participated in our survey. We have made all the questions mandatory so that no data is missed. Figure 1 has shown the percentage of total participants in terms of gender.
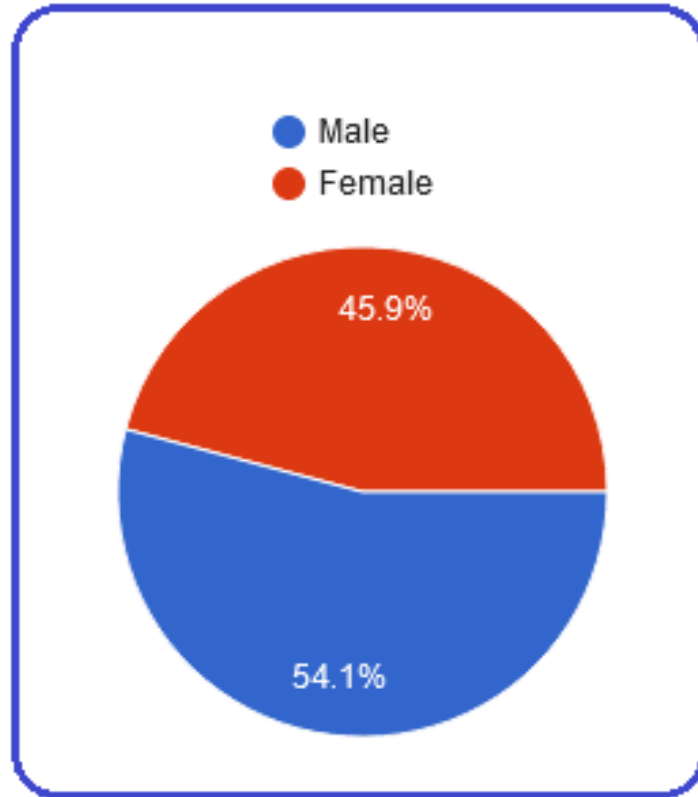

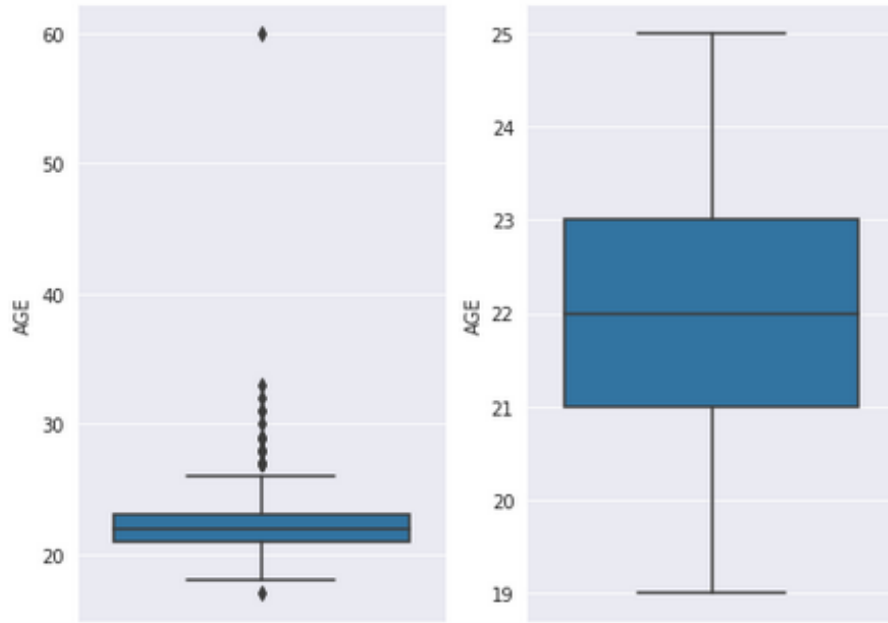Figure 1. Graphical visualization of Participant's responses by Gender

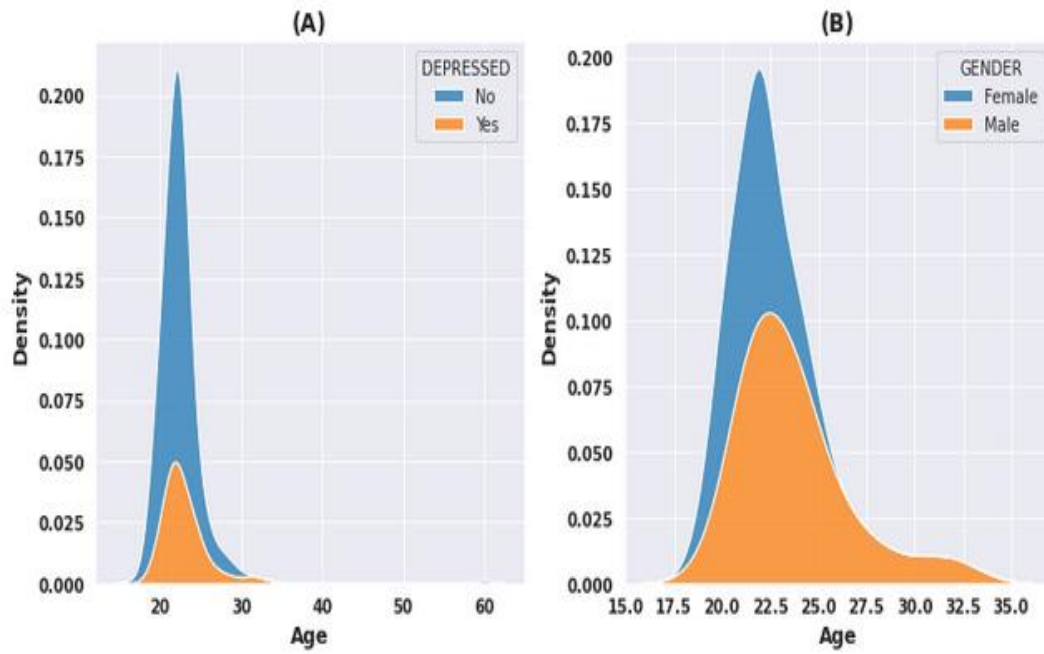Figure 2. Outliers and Without Outliers in Age

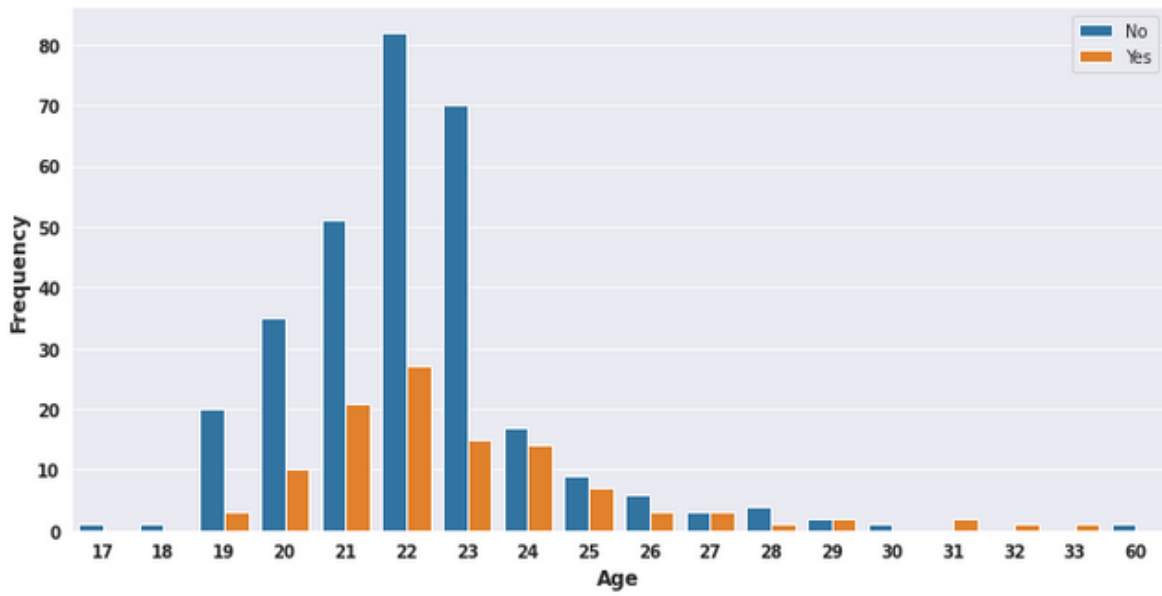Figure 3. Total Depressed Density and Gender Wise Depressed Density
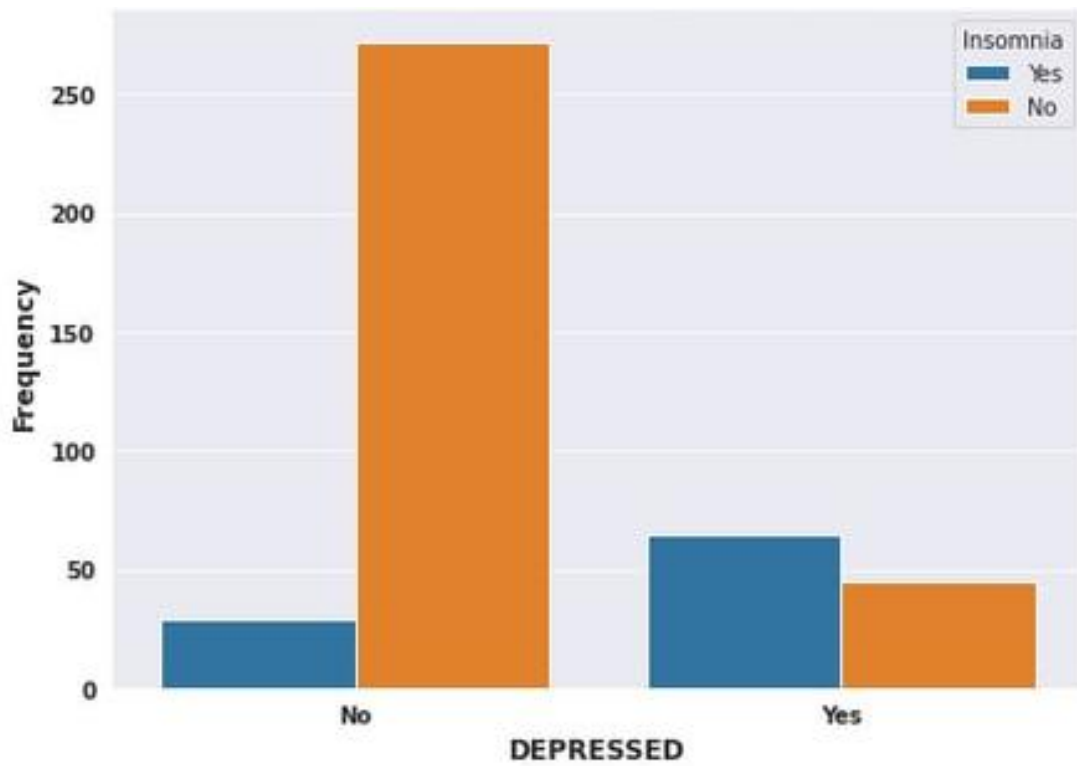
Figure 4. Specific Depressed Age

Figure 5. Insomnia in Depressed Students

In Figure 2 we see that there are various values in the age column in the dataset. Most of the values range from 20-25 and others are less than 20 or greater than 25. Because of some outliers, the algorithm cannot perform well. For that reason, we make the outliers range and remove the outliers. In the dataset, we have a total of 412 records. After removing outliers, the new dataset has 381 records. These 381 records are more accurate than previous data. Here we used the Outliers technique in the age column because it was the user input value and the rest of them was categorical value. For the use of these data, the accuracy of our algorithms has improved 0 to 5% more.

We have collected the dataset of 413 participants. 104 participants were found depressed. The dataset was collected in a pandemic situation. So, we found an extra number of people who were depressed which we did not expect. That is all for the Covid-19 pandemic situation. In Figure 3A. we can see that 25% of total participants are depressed. Among these 25% depressed people, female participants are clearly much more depressed than men which are shown in Figure 3B. In these total participants, Figure 4 Shows that 20 to 25 age participants are much more depressed than others. According to our research, we can say that specifically 21,22,23 and 24 are more depressed. In Figure 5 we can see that few students have insomnia who are not depressed. but in depressed students this ratio is high. In depressed students, this number of insomnias is more than 50. In not depressed students this number is less than 30.

3.2 Data Pre-processing:

All the data was raw when we created our dataset. Algorithms perform better with numerical values. So there is a need to clean this data into structured data. For cleaning the dataset first we read the dataset into our workstation using pandas. Pandas is a python library that is used in machine learning and data analysis. Before using pandas functionality we import pandas using [import pandas as pd]. After reading the dataset we store it into a variable called [depression]. Now we can see the dataset using [depression.head()]. In our dataset column names are too much longer and have space in the column names at the initial state. For this reason, we faced some issues that's why we removed all space from the column name. Even then column names are too big. So we renamed the column name into two or three words. These two or three words are meaningful with column data. Numerical values show better accuracy than other values. So all the data that is in yes or no we have taken 0 and 1. Here 0 for NO and 1 for YES. We have converted all the data that has multiple options into 0,1,2,3 in phases. For converting data strings to numeric we have created some functions. These functions create similar columns like we have too many columns which record have yes or no. Using one function we convert all these similar columns in this function, we create a variable mapping and it has properties like yes:1, no:0. Then we perform a loop and map into that column. After creating the function, we just passed the column name as a parameter into the function. Now all of the data is converted into numeric values.

Now we check if there is any missing value in this dataset or not. Although we have made all the questions mandatory. But due to technical reasons, some value may be missing. Luckily there is no missing value in our dataset.

3.3 Used Algorithms:

In this research, we have used the classifier algorithm because we will diagnose a person's depression through YES and NO. Classification algorithms are a technique of supervised learning. These classification algorithms can identify the new observation category based on training data. When we give training and testing values in our system, these classification algorithms learn from training values and apply the learning in testing values. In our systems, 8 classifier algorithms give better accuracy than other algorithms. These 8 algorithms are given below.

3.3.1. Logistic Regression (LR): LR is a process where it describes the supervision on the possibility of changing aim to a target variable.

3.3.2. Multi-layer Perceptron (MLP): MLP is a feed forward AI neural network augmentation. Input layer provides the processed signal as input. The output layer predicts and also used to categorize. Here from input layer to output layer signal passes.

3.3.3. Random Forest: Random forests is the study of classifying works for huge decision trees. According to the priority of trees, the result is selected for classification.

3.3.4. K-nearest neighbor (KNN): KNN gives solutions for classification and regression issues. Distance measurements may be used to assess the similarity between the occurrences. K value is the main indicator.

3.3.5. Gradient Boosting (GB): Gradient Boosting (GB) creates a better and latest model than the ancient one. The loss function is computed by GB using the gradient descent method. To minimize overfitting issues, boosting should be halted as soon as possible utilizing stopping criteria.

3.3.6. Multinomial Naive Bayes: In Natural Language Processing, the Multinomial Naive Bayes method is a Promising learning technique. Bayes theorem is used for guessing text tagging.

3.3.7. Adaptive boosting (AdaBoost): AdaBoost algorithm is used as an ensemble method. Adaptive Boosting refers to the process of re-assigning weights to each instance.

3.3.8. SVM: SVM finalizes classification and regression problems. The SVM classifier is comparatively better as frontier than other classes.

3.4 Algorithm Implementation:

Machine learning classifier algorithm is used for understanding a value yes or no. in our case it's depressed or not depressed. To define a person depressed or not depressed we used more than 15 classifiers algorithms. For applying the classifier algorithm in our dataset we clean our dataset. Because in the initial stage our dataset was long column name, null value, string, etc. which is not a good dataset for applying algorithms. That's why we rename all columns in a single word, delete the unnecessary columns, remove all null values and convert all string values into numeric values 0 and 1. After cleaning the dataset we divide our dataset into two parts which are predictor column (x) and target column (y). Then we train our algorithms by giving 75% data and test our algorithm by 25% data. When data training and testing are complete we used our classifiers algorithm to check the accuracy of prediction. At first, we use a logistic regression algorithm. For implementing this algorithm, we import the library of logistic regression using (from sklearn. linear_model import LogisticRegression). Then we store the logistic regression function into a variable. Using this variable we fit the predictor training data (x_train) and target training data (y_train). Also, we use the pandas predict function for x_test data using this variable. Getting the accuracy of prediction, we use (metrics.accuracy_score) function and we passed two parameters which y_test and that variable which we created before.

3.5 Implementation Procedure:

The implementation of this study has been completed step by step. There are seven steps in implementation procedures which are given below in figure [6] and descriptions of these are given.

This Figure [6] shows a complete overview of our research. There are five subsections in this section.
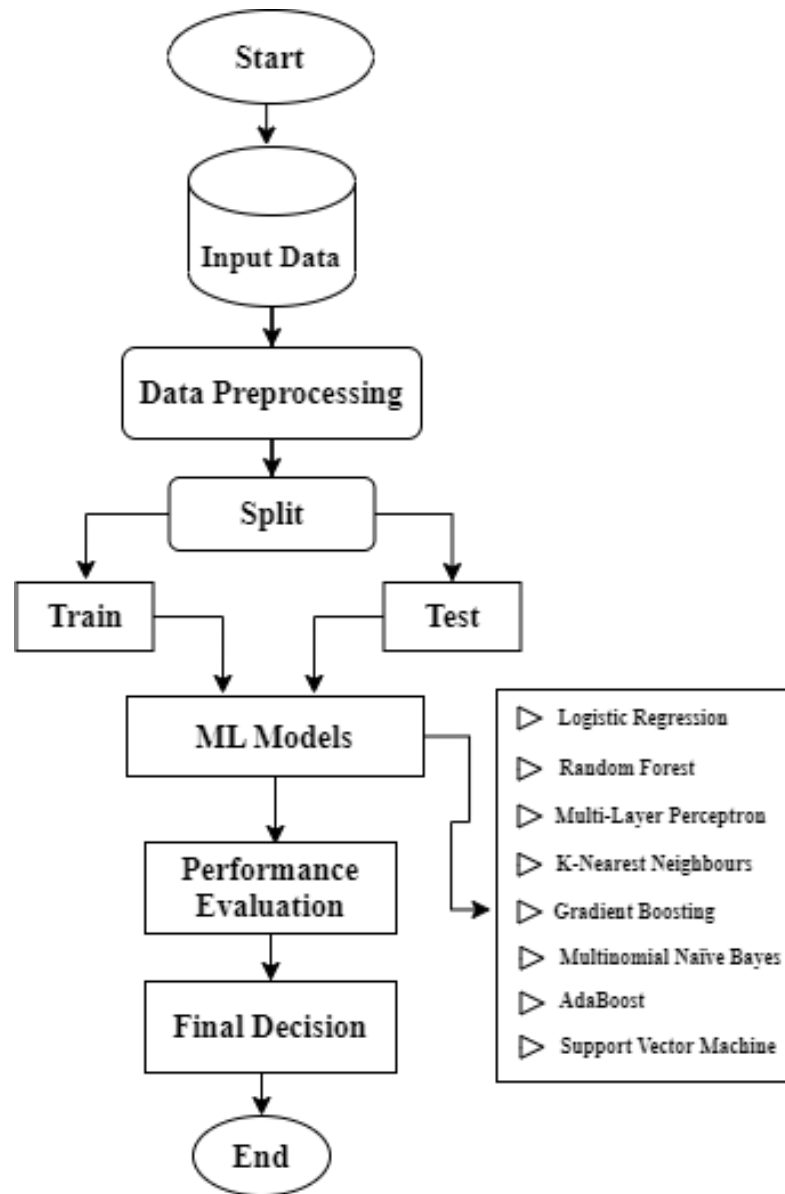

Figure 6. Methodology Diagram

3.5.1. Input Data: For this research, the surveyed dataset was used as input data in our system. This dataset has some questionnaires of PHQ-9 and AIS scale. Also, have a few other questions which were suggested by experts. This dataset has 413 records of Bangladeshi university students.

3.5.2. Data Preprocessing: In this dataset, mapping is used for converting unstructured data into structured data. Machine learning algorithms perform better with numerical data. We could use label encoding in unstructured data but in our system mapping gives high accuracy rather than label encoding.

3.5.3. Data Splitting: Since we used the classification algorithm in our system, we have to divide the dataset into predictor and target predictor attributes. This dataset has 21 attributes where "DEPRESSED" is the Predictor Target attribute and the rest of them are predictor attributes.

3.5.4. Train and Test: Before performing classification algorithms, the system needs to learn about the dataset. For that reason, we give training and testing value to the system. The system learns from training data and applies its learning to testing data. In our system, we give 75% data for training and 25% for testing.

3.5.5. ML Models: In this research, 8 classifiers algorithms were used which are LR, MLP, RF, SVM, Multinomial Naive Bayes, Gaussian Naive Bayes, Gradient Boosting, AdaBoost.

3.5.6. Performance Evaluation: The performance of the 8 algorithms used in this system is good enough. accuracy of these algorithms between 84% to 86%. Also, the precision, sensitivity, and specificity of these algorithms are sufficient.

3.5.7. Final Decision: Among these 8 classifier algorithms, Random forest has better performance than other algorithms. it has 86% accuracy which is the highest in our selected algorithms.

3.6 Training and testing for predicting depression:

The training datasets are used to train the classifiers, which include KNN, Decision Tree, SVM, AdaBoost, Random Forest etc. Following training, each of these classifiers was used to predict depression in test dataset individuals.

3.7 Algorithm Comparison:

In our research, we implement more than 15 classifier algorithms. Among them, 8 algorithms have high accuracy which is 84% to 86%. These 8 algorithms are Logistic Regression, Random forest, Multi-layer Perceptron (MLP), Multinomial Naive Bayes, Gradient Boosting, K-Nearest Neighbors (KNN), AdaBoost, Support Vector Machine. For comparing these 8 algorithms we calculate the value of sensitivity, specificity, and f1-measure. Also, we compare its accuracy, precision, and recall.

After comparing the 8 classifier algorithms, we came to the conclusion that RF Algorithm shows comparatively better accuracy and also the best efficiency than other algorithms in our system.

CHAPTER 4

RESULTS & DISCUSSION

Result:

In this project the accuracy of various classifiers algorithms was calculated. besides the precision, sensitivity, specificity, recall, f1 score and also confusion matrix of these algorithms were measured. True positive rate here is considered as Sensitivity. The ratio of accurately diagnosed positive tuples to the number of positive tuples are defined as sensitivity.

$$\text{Sensitivity} = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100\% \ \text{----------------} \ (1)$$

The true negative rate is also known as specificity is the ratio of correctly diagnosed negative tuples to the total number of negative tuples.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \times 100\% \text{----------------} (2)$$

Measuring exactness is also referred to as Precision which is the Quotient among genuine positive values with the expected positive value.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100\% \ \text{----------------} \ (3)$$

With the help of f1-score, recall's harmonic mean and also precision is calculated.

$$F1\ Score = \frac{2 \times True\ Positive}{2 \times True\ Positive + False\ Positive + False\ Negative} \times 100\% \ \text{----------------} \ (4)$$

The Confusion matrix measures ML classifications. It will run the classification models in the opposite to the test data and find out the true negative, true positive, false negative, false positive and values in a tabular manner.

All the methods of the Confusion Matrix is applied in the model which is described in Table 2. and also in that table every classifier of the model evaluation is described.

Algorithms	Confusion Matrix				Algorithms	Confusion Matrix			
LR	True class		No	Yes	MLP	True class		No	Yes
		No	68	6			No	67	7
		Yes	11	19			Yes	9	21
		Predicted Class					Predicted Class		
RF	True class		No	Yes	SVM	True class		No	Yes
		No	67	7			No	68	6
		Yes	8	22			Yes	11	19
		Predicted Class					Predicted Class		
MNB	True class		No	Yes	Ada Boosting	True class		No	Yes
		No	65	9			No	68	6
		Yes	7	23			Yes	11	19
		Predicted Class					Predicted Class		
GB	True class		No	Yes	KNN	True class		No	Yes
		No	68	6			No	71	3
		Yes	10	20			Yes	14	16
		Predicted Class					Predicted Class		

Table 2. All classifier's confusion matrix.

Performance of the different classifiers Algorithms:

In table 3. we have shown the results of the performance of different classifier algorithms. This performance data is collected from the classifier algorithm that we used in our system.

No.	Classifier Name	Accuracy (%)	Precision (%)	Recall (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
01	LR	83%	86%	92%	63%	91%	89%
02	MLP	84%	88%	91%	70%	90%	89%
03	RF	85%	89%	91%	73%	90%	90%
04	SVM	83%	86%	92%	63%	91%	89%
05	MNB	84%	90%	88%	76%	87%	89%
06	GB	84%	87%	92%	66%	89%	89%
07	AdaBoost	83%	86%	92%	63%	91%	89%
08	KNN	83%	84%	96%	53%	95%	89%

Table 3. Classifier performance evaluation

In our system we have seen that LR, AdaBoost, KNN and SVM give 84% accuracy. Compared to the other algorithms these four algorithms give the lowest accuracy. Because of Multi-layer Perceptron (MLP), Multinomial Naive Bayes, Gradient Boosting has 85% accuracy. Which is the mid-level accuracy of our selected algorithms. And Finally, Random forest algorithm gives approximately 86% accuracy. Which is high accuracy among our selected algorithms. Accuracy difference is almost 1%.

Logistic regression has a precision of 86% and recall of 92%. If we compare with random forest, we see that random forest has a precision of 89% and recall of 91%. That means Random forest and logistic regression have not too much difference in recall but there is some difference in precision between these two algorithms. On the other hand, Multi-layer Perceptron (MLP) has a precision of 90% and recall of 88%. We can compare this with K-Nearest Neighbours (KNN). In K-Nearest Neighbours (KNN) has a precision of 84% and recall 96%. There is too much difference in Precision and recall. In that case Multi-Layer Perceptron (MLP) is better for our system. Multinomial Naive Bayes is more accurate when we compare with Gradient Boosting. Because Multinomial Naive Bayes has a precision of 90%, recall of 88% and Gradient boosting has a precision of 87%, recall 92%. When comparing AdaBoost and support vector machines, AdaBoost gives The precision of 86%, recall 92% and Support vector machine gives the precision of 86%, recall 92%. In these comparisons, both algorithms perform the same.

Discussion:

About 17% of participants have suffered physical or mental problems due to covid-19 and Almost 87% of participants have faced financial problems during this pandemic situation. In these participants they said 16% are faced with serious financial problems, 71% have low financial problems and the rest of them haven't had any financial problems during covid-19. Family or personal separation is the cause of 18% of people who are depressed and 14% of them upset with themselves or upset with someone in his/her family.

In total participants, most of them like to spend time on social media, not with family and they use mobile or computer regularly at night. 9% of the participants take drugs or alcohol and 15% of participants have short anger. Few participants said that they worried about their life or future career. less than 5% of participants faced mental, physical, or sexual harassment and 21 of the total participants tried to commit suicide.

CHAPTER 5

RECOMMENDATIONS & CONCLUSIONS

5.1 Conclusion:

In this research by using machine learning algorithms, 02 distinct levels of anxiety and depression intensity have been assessed. We collected data by using a perfect questionnaire that evaluated the most frequent symptoms of anxiety and sadness.

Although Random Forest was determined as the best model, the accuracy of naive Bayes was found to be the greatest. Because this problem yielded imbalanced classes, the best-model selection was based on the f1 score, which is utilized in instances of uneven partitioning.

We have analyzed the data survey and found that people suffer from problems like depression for various reasons. We have collected about 413 data so far. From this it is understood that people do not suffer from depression for any specific reason. Moreover, the amount of depression varies between men and women. In the case of women, they rarely break down. Within a few days, they became very weak physically and mentally. However, if they or someone close to them can diagnose the cause of the breakdown, it is possible to get out of this depression.

In this study we were able to measure the amount of depression in a human being using machine learning algorithms. We also get an idea of exactly what causes a person to suffer from depression and what the depressed person wants to do or feel at this time.

5.2 Future work:

We have now collected 413 Raw Data and completed our research. From these 413 data we have been able to determine who is more depressed, what age people are comparatively depressed. We have future plans for this research. We are in the future

- We will collect more data for our research.
- Our focus is to increase the accuracy of the algorithms.
- We will show the data analysis visually.

REFERENCES

01. Depression, *Who.int* (2020), https://www.who.int/news-room/fact-sheets/detail/depression. Accessed: 01 Aug 2020

02. https://www.newagebd.net/article/141238/frustration-and-suicide-rising-among-youths

03. Mamun, M.A., Hossain, M.S., Siddique, A.B., Sikder, M.T., Kuss, D.J. and Griffiths, M.D., 2019. Problematic internet use in Bangladeshi students: The role of socio-demographic factors, depression, anxiety, and stress. *Asian journal of psychiatry*, *44*, pp.48-54.

04. van Eeden, W.A., Luo, C., van Hemert, A.M., Carlier, I.V., Penninx, B.W., Wardenaar, K.J., Hoos, H. and Giltay, E.J., 2021. Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. *Psychiatry Research*, *299*, p.113823.

05. Hasan, M.T., Hossain, S., Gupta, R.D., Podder, V., Mowri, N.A., Ghosh, A., Mahmood, H.R., Ahmmed, F., Khatun, M.H., Nodi, R.N. and Koly, K.N., 2020. Depression, sleeping pattern, and suicidal ideation among medical students in Bangladesh: a cross-sectional pilot study. *Journal of Public Health*, pp.1-9.

06. Akhtar, S., Ahmed, R., Chaity, A.S. and Billah, M., 2021. Association between knowledge and depression at rising time of COVID-19 in Bangladesh. *International Journal of Public Health*, *10*(2), pp.387-393.

07. Choudhury, A.A., Khan, M.R.H., Nahim, N.Z., Tulon, S.R., Islam, S. and Chakrabarty, A., 2019, June. Predicting Depression in Bangladeshi Undergraduates using Machine Learning. In *2019 IEEE Region 10 Symposium (TENSYMP)* (pp. 789-794). IEEE.

08. Islam, M., Islam, U.S., Mosaddek, A.S.M., Potenza, M.N. and Pardhan, S., 2021. Treatment, persistent symptoms, and depression in people infected with COVID-19 in Bangladesh. *International journal of environmental research and public health*, *18*(4), p.1453.

09. Mashreky, S.R., Rahman, F. and Rahman, A., 2013. Suicide kills more than 10,000 people every year in Bangladesh. *Archives of Suicide Research*, *17*(4), pp.387-396.

10. Rasheduzzaman, M., Al Mamun, F., Faruk, M.O., Hosen, I. and Mamun, M.A., 2021. Depression in Bangladeshi university students: The role of sociodemographic, personal, and familial psychopathological factors. *Perspectives in Psychiatric Care*.

11. Priya, A., Garg, S. and Tigga, N.P., 2020. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, pp.1258-1267.

12. Biilah, M., Raihan, M., Akter, T., Alvi, N., Bristy, N.J. and Rehana, H., 2022. Human Depression Prediction Using Association Rule Mining Technique. In *International Conference on Innovative Computing and Communications* (pp. 223-237). Springer, Singapore. [Google Scholar](#)

13. Li, F. and Ding, Y., 2019, August. Data mining in Cognitive function training of depression patients applications. In *2019 10th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 98-101). IEEE. Google Scholar

14. Cvetkovi□, J., 2017. Breast cancer patients' depression prediction by machine learning approach. *Cancer investigation*, *35*(8), pp.569-572.

15. Na, K.S., Cho, S.E., Geem, Z.W. and Kim, Y.K., 2020. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. *Neuroscience letters*, *721*, p.134804.

16. Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G. and Okon-Singer, H., 2021. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, *141*, pp.199-205.

17. Moon, N.N., Mariam, A., Sharmin, S., Islam, M.M., Nur, F.N. and Debnath, N., 2021. Machine Learning Approach to Predict the Depression in Job Sectors in Bangladesh. *Current Research in Behavioral Sciences*, p.100058.

18. Govindasamy, K.A. and Palanichamy, N., 2021, May. Depression Detection Using Machine Learning Techniques on Twitter Data. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 960-966). IEEE.

19. Zhou, S., Ma, Q., Lou, Y., Lv, X., Tian, H., Wei, J., Zhang, K., Zhu, G., Chen, Q., Si, T. and Wang, G., 2021. Machine learning to predict clinical remission in depressed patients after acute phase selective serotonin reuptake inhibitor treatment. *Journal of Affective Disorders*, *287*, pp.372-379.

20. Bailey, N.W., Krepel, N., van Dijk, H., Leuchter, A.F., Vila-Rodriguez, F., Blumberger, D.M., Downar, J., Wilson, A., Daskalakis, Z.J., Carpenter, L.L. and Corlier, J., 2021. Resting EEG theta connectivity and alpha power to predict repetitive transcranial magnetic stimulation response in depression: a non-replication from the ICON-DB consortium. *Clinical Neurophysiology*, *132*(2), pp.650-659.

21. Zhang, Y., Wang, S., Hermann, A., Joly, R. and Pathak, J., 2021. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *Journal of Affective Disorders*, *279*, pp.1-8.

22. Gokten, E.S. and Uyulan, C., 2021. Prediction of the development of depression and post-traumatic stress disorder in sexually abused children using a random forest classifier. *Journal of Affective Disorders*, *279*, pp.256-265.

23. de Souza Filho, E.M., Rey, H.C.V., Frajtag, R.M., Cook, D.M.A., de Carvalho, L.N.D., Ribeiro, A.L.P. and Amaral, J., 2021. Can machine learning be useful as a screening tool for depression in primary care. *Journal of Psychiatric Research*, *132*, pp.1-6.

24. Poletti, S., Vai, B., Mazza, M.G., Zanardi, R., Lorenzi, C., Calesella, F., Cazzetta, S., Branchi, I., Colombo, C., Furlan, R. and Benedetti, F., 2021. A peripheral inflammatory signature discriminates bipolar from unipolar depression: A machine learning approach. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *105*, p.110136.

25. Wang, Z., Yuan, Y., Jiang, Y., You, J. and Zhang, Z., 2021. Identification of specific neural circuit underlying the key cognitive deficit of remitted late-onset depression: A multi-modal MRI and machine learning study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *108*, p.110192.

26. Jacobson, N.C. and Nemesure, M.D., 2021. Using Artificial Intelligence to Predict Change in Depression and Anxiety Symptoms in a Digital Intervention: Evidence from a Transdiagnostic Randomized Controlled Trial. *Psychiatry Research*, *295*, p.113618.

27. Su, D., Zhang, X., He, K. and Chen, Y., 2021. Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. *Journal of Affective Disorders*, *282*, pp.289-298.

28. Rahman, M.M., Saifuzzaman, M., Ahmed, A., Mahin, M.F. and Shetu, S.F., 2021. Impact of COVID-19 on mental health: A quantitative analysis of anxiety and depression based on regular life and internet use. *Current Research in Behavioral Sciences*, *2*, p.100037.

29. Wardenaar, K.J., Riese, H., Giltay, E.J., Eikelenboom, M., van Hemert, A.J., Beekman, A.F., Penninx, B.W. and Schoevers, R.A., 2021. Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders*, *293*, pp.295-304.

30. Narayanrao, P.V. and Kumari, P.L.S., 2020, March. Analysis of machine learning algorithms for predicting depression. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1-4). IEEE.

31. https://adaa.org/understanding-anxiety/depression/symptoms .

32. https://www.webmd.com/balance/stress-management/stress-symptoms-effects_of-stress-on-the-body#1.

33. Chi-Square Test (Definition, Formula, Properties, Table & Example), *BYJUS* (2020), https://byjus.com/maths/chi-square-test/. Accessed 02 Sep 2020

APPENDIX

Abbreviation:

AdaBoost = Adaptive Boosting

RF = Random Forest

SVM = Support-Vector Machines

LR = Logistic Regression

KNN = K-Nearest Neighbors

MNB = Multinomial Naive Bayes

MLP= Multilayer perception

DTC = Decision Tree Classifier

GB = Gradient Boosting

QDA = Quadratic Discriminant Analysis

PMCOVID-19 = Physical or mental problems due to the Covid-19

FFP = Family facing any Financial Problem.

FPPS = Family problems or personal separations.

UPSET = Upset with yourself or someone in your family.

HISSUE = Health issues.

ALCDRUINB = Taking Alcohol or Drugs.

ANGER = Angry in a small matter

SL6-8 = Sleep less than 6 to 8 hours.

SUICIDAL = Suicidal thoughts.

TISM = Time in social media.

TWF = Time with family.

WORRIEDF = worried about future career.

HARASS = Physical / Mental / Sexual Harassment.

LOIW = Lose interest in work.

INDES = Indecision or bad decision.

RIDICULED = Ridiculed in public or among friends.

NIGHTSCREENTIME = Mobile or computer uses at night.

WORRIEDP = Worried about something from past.

Appendix: Research Reflections

During the time when we were completing our project, it was very difficult for finding the problems and conditions.

Firstly, we determined the perfect algorithms among all of the others for getting better results and perfection. Besides, with the help of machine learning and python, all had to take a deep idea about that. In the corona period, gathering and collecting a huge dataset wasn't as easy as we expected.

In spite of all critical and a vast challenging period, we were finally able to finish that successfully.

PLAGIARISM REPORT

PREDICTING DEPRESSION AMONG THE UNIVERSITY STUDENTS IN THE COVID-19 PANDEMIC SITUATION USING MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

20% SIMILARITY INDEX	17% INTERNET SOURCES	13% PUBLICATIONS	16% STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	2%
3	Md. Sabab Zulfiker, Nasrin Kabir, Al Amin Biswas, Tahmina Nazneen, Mohammad Shorif Uddin. "An In-Depth Analysis of Machine Learning Approaches to Predict Depression", Current Research in Behavioral Sciences, 2021 Publication	1%
4	Submitted to University of East London Student Paper	1%
5	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2022 Publication	1%
6	Submitted to AUT University Student Paper	1%

7 pure.amc.nl
Internet Source
1%

8 pure.rug.nl
Internet Source
1%

9 Thalia Richter, Barak Fishbain, Eyal Fruchter, Gal Richter-Levin, Hadas Okon-Singer. "Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders", Journal of Psychiatric Research, 2021
Publication
<1%

10 Submitted to Leiden University
Student Paper
<1%

11 irep.ntu.ac.uk
Internet Source
<1%

12 Submitted to Southern New Hampshire University - Continuing Education
Student Paper
<1%

13 Submitted to University of Bedfordshire
Student Paper
<1%

14 Tabassum Ferdous Mumu, Ishrat Jahan Munni, Amit Kumar Das. "Depressed People Detection from Bangla Social Media Status using LSTM and CNN Approach", Journal of Engineering Advancements, 2021
Publication
<1%