DEPRESSION PREDICTION AMONG BANGLADESHI UNIVERSITY STUDENTS USING MACHINE LEARNING TECHNIQUE

BY

BIJOYA DEB KEYA
212-15-14693

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Md Assaduzzaman
Lecturer (Senior scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Fatema Tuj Johora
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

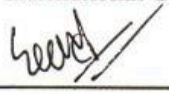
APPROVAL

This Project titled "DEPRESSION PREDICTION AMONG BANGLADESHI UNIVERSITY STUDENTS USING MACHINE LEARNING TECHNIQUE" submitted by Bijoya Deb Keya 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 16-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor & Associate Head
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Arif Mahmud (AM) — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Sazzadur Ahamed (SZ) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali (DAA) — **External Examiner**
Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md Assaduzzaman, 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 the award of any degree or diploma.

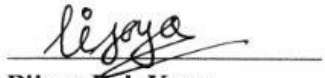
Supervised by:

Md Assaduzzaman
Lecturer (Senior scale)
Department of CSE
Daffodil International University

Co-Supervised by:

Fatema Tuj Johora
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Bijoya Deb Keya
ID: 212-15-14693
Department of CSE
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.

I really grateful and wish our profound our indebtedness to **Supervisor Md Assaduzzaman, Lecturer (Senior scale),** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*IoT*" to carry out this project. His endless patience,s cholarly 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.

I would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor & Head**, Department of CSE, for their kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

I 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

This study investigates the application of artificial intelligence techniques to predict mental health breakdowns among university students based on their daily activities. The research uses various data sources, including electronic health records, smartphone usage, and passive sensor data, to create predictive models for detecting depression, suicidal behavior, and other mental health problems. The study found promising results in predicting mental health outcomes with high accuracy, emphasizing feature selection, model optimization, and early intervention using predictive insights. However, gaps in existing research include the need for more comprehensive models, standardization of evaluation metrics, and diverse data sources to capture the complexities of students' experiences. The study aims to address these gaps by creating a predictive model tailored to the university student population, predicting mental health breakdowns based on daily activities. In this work, a survey form was created to collect data on depression, resulting in 1000 responses. The dataset contains 2 parts, 20% is testing data while 80% is training data. The dataset comprises 16 categorical columns, with 10 questions aimed at diagnosing depression independently. Each column represents a different aspect of mental health, lifestyle, or academic experience, without interdependencies between questions. There are 1 dependent and 10 independent or input variables in the dataset. Machine learning methods used such as LR, DT, RF, SVM, Gradient Boosting Algorithm, AdaBoost Algorithm gave 56%, 68%, 68%, 52%, 91% and 89% accuracy respectively. The model is predicted with Gradient Boosting algorithm from the highest accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1. Introduction

Depression is a prevalent mental health concern among university students worldwide, and Bangladesh is no exception. The transition to university life, often marked by increased academic pressure, social changes, and newfound independence, can exacerbate stressors and contribute to the development of depression among students.

In Bangladesh, where the education system is highly competitive and demanding, university students face unique challenges that may predispose them to mental health issues. The pressure to excel academically, coupled with societal expectations and financial constraints, can create a breeding ground for psychological distress.

A chronic sense of melancholy or a loss of interest in once-enjoyed activities are characteristics of depression. Among university students, this may manifest in various ways, including disrupted sleep patterns, neglect of self-care, difficulty concentrating, heightened anxiety, and a general disinterest in daily life experiences. Left untreated, depression may have a serious negative effect on one's ability to learn, maintain social connections, and feel good overall.

Recognizing the importance of early detection and intervention, researchers have increasingly focused on identifying factors that contribute to depression among university students. Understanding the prevalence and predictors of depression can inform targeted interventions and support services to mitigate its impact and promote mental health on campus.

In this context, this survey aims to assess depression frequency among university students in Bangladesh and explore potential risk factors associated with its development. By collecting data on students' experiences, behaviors, and attitudes related to mental health, we seek to gain insights into the patterns and determinants of depression within this population. Ultimately, the findings of this study can inform strategies for prevention, early intervention, and support tailored to the specific needs of Bangladeshi university students.

1.2. Motivation

The goal of this study is to create a predictive model based on machine learning algorithms to identify mental health breakdowns in university students through their daily activities. The study analyzes the relationships between demographic factors, daily routines, and mental health indicators in order to identify patterns and risk factors. It compares the performance of various machine learning algorithms in predicting mental health breakdowns and determines the most effective approach. The study offers insights and recommendations for universities and mental health professionals to implement targeted intervention strategies for at-risk students, resulting in a healthier and more resilient university community. The study is motivated by a growing awareness of the significance of mental health in university settings.

1.3. Objective

The aim of this investigation is to examine depression and suicide using machine learning and deep learning algorithms. Data collection will be conducted through a Google Form survey, with a target sample size of 1000 responses. The collected data will be analyzed and researched using various machine learning and deep learning techniques to identify patterns and risk factors associated with depression and suicidal tendencies. The primary objective is to determine which algorithms perform better in predicting depression and suicide risks based on the collected data. By comparing the performance of different algorithms, the study aims to provide insights into the most effective methods for identifying individuals at risk. Additionally, the study seeks to develop a predictive model using the best-performing algorithm to accurately forecast potential depressive episodes and suicidal ideation. Ultimately, The objective is to support the creation of more focused and effective interventions for mental health support and suicide prevention.

1.4. Project Scope

The goal of this project is to predict mental health breakdowns in university students based on their daily activities using machine learning techniques. The scope includes data collection, preprocessing, model development, and evaluation. The project involves collecting its own dataset,

cleaning the data to remove missing values and ensure quality, selecting relevant features, encoding categorical variables, and performing exploratory data analysis (EDA) to gain insights into the data. The primary goal is to create and test machine learning models, such as Logistic Regression, Decision Tree, Random Forest, SVM, Gradient Boosting Algorithm, AdaBoost Algorithm to predict mental health breakdowns. The project also intends to make recommendations for universities and mental health practitioners based on the model's findings, to enable targeted intervention strategies to assist at-risk students. The project excludes clinical diagnosis and interventions, instead focusing solely on predictive modeling and recommendations for early intervention and support.

1.5. Relational of the Study

This research seeks to establish a direct relationship between the identified risk factors and the prevalence of depression among Bangladeshi university students. By analyzing survey data, we aim to uncover correlations between various factors such as academic stress, social support, and lifestyle habits with the manifestation of depressive symptoms. Understanding these relationships will provide valuable insights into the complex interplay of factors contributing to depression in this population, guiding the development of targeted interventions and support mechanisms to promote mental well-being on university campuses in Bangladesh.

1.6. Research Questions

There might be many different types of questions in this study. For example,

- What is the purpose of this research?
- Which database features have the most effects on depression prediction accuracy?
- To what extent can depression be predicted by machine learning models?
- Which machine learning method makes the greatest predictions in terms of precision, accuracy, and generalizability?
- What advantages does prediction offer?
- What is the total rate of depression prediction accuracy?
- How can we benefit from this thesis?

1.7. Project Outcome

The primary goal of this project is to create a predictive model capable of detecting mental health breakdowns in university students based on their daily activities. The project's goal is to provide a reliable and accurate tool for detecting mental health weaknesses early on through rigorous data preprocessing, feature selection, and model development. To find out how well the model predicts mental health outcomes, its performance will be assessed using relevant measures including accuracy, precision, recall, and F1-score. Furthermore, the project will shed insight into the key factors that contribute to mental health issues among university students, as discovered through the use of EDA and feature importance analysis. These findings will help universities and mental health practitioners develop targeted intervention strategies, resulting in a healthier and more supportive environment for students. The project's goal is to contribute to improved mental health outcomes and well-being in university communities.

1.8. Report Layout

The report will be developed to provide a thorough overview of the project's methodology, findings, and recommendations. It will begin with an introduction that explains the context and significance of the research, followed by a concise problem statement that defines the study's objectives and scope. The methodology section will go over the steps involved in gathering data, preprocessing, choosing features, creating a model, and assessing it. The results section will then present the findings of the performance of the machine learning models as well as the insights gained from exploratory data analysis. The discussion section will look at the findings' implications and their relevance to mental health interventions in university settings. Finally, the report will provide a summary of key findings, limitations, and recommendations for future research. Appendices will contain supplementary information such as code snippets, model parameters, and additional analyses. This organization ensures that the project's processes and outcomes are presented clearly and coherently.

CHAPTER 2
BACKGROUND

2.1. Terminology

For the purpose of this study, several key terms are defined as follows. Depression is typified by enduring melancholy or a decline in interest in activities, often accompanied by changes in sleep patterns, concentration difficulties, and anxiety. University students refer to individuals currently enrolled in tertiary education institutions. Academic stress encompasses the pressure and demands associated with academic performance, including coursework, exams, and deadlines. Social support denotes the availability of assistance, encouragement, and emotional backing from family, friends, or peers. Lifestyle habits include behaviors related to diet, exercise, substance use, and leisure activities. The prevalence of depression is the proportion of university students exhibiting depressive symptoms within a specified population and timeframe.

2.2. Related Work

C. G. Walsh, et al.[1] implemented machine learning techniques to predict suicide attempts by analyzing digital health data from a huge medical database that included 5,167 adult patients. Of these, 3,250 patients tried to kill themselves (cases), while 1,917 self-injured without suicidal intent (controls). The machine learning algorithms developed showed high predictive accuracy, with an AUC of 0.84, precision of 0.79, recall of 0.95, and Brier score of 0.14. Importantly, prediction accuracy increased significantly as the time window was decreased with time, from 720 days to just 7 days before to the attempted suicide, and the significance of the predictor changed, providing important insight into the temporal dynamics of suicide risk.

S. Ware et al [2] implemented WiFi infrastructure data and passive smartphone data collection to explore the possibility of automatically predicting different categories of depressive symptoms from smartphone data. The study includes 182 college students and uses machine learning models to accurately predict behavioral and cognitive symptoms of depression, with a score of 0.86. Notably, the approach requires no user effort and provides objective assessments of depressive

symptoms, which represents a significant improvement over previous studies that primarily focus on predicting overall depression severity. The findings highlight the viability of using smartphone data to make accurate and comprehensive predictions of depressive symptoms.

P. Chikersal et al [3] described a machine learning method that recognises college students using information from fitness monitors and cellphones who are individuals who, at the conclusion of the semester, were having depressed symptoms and those whose symptoms have gotten worse over time. A novel feature extraction technique selects meaningful features associated with depressive symptoms from longitudinal data. The method has an impressive accuracy of 85.7% in detecting post-semester depressive symptoms and 85.4% in predicting changes in symptom severity. Notably, the model forecasts these outcomes with an accuracy of more than 80% up to 11–15 weeks before the semester ends, to give time for preventative measures. The results of this study have important ramifications for the use of longitudinal behavioural data to identify health consequences and highlights the potential for preventing depression by detecting changes and predicting.

A. B. Siddique et al [4] focused on this study is to predict undergraduate depression for possible referral to psychiatric facilities by examining depression among Bangladeshi university students. The data was collected using a Google survey form, and five algorithms were used to train and test the dataset. The most effective techniques to forecast depression in Bangladeshi undergraduates were identified by comparing metrics such as recall, f-measure, accuracy, precision, and mean absolute percentage error. In addition, a mobile app for mental health on Android was designed and designed to offer university students mental help. The study sheds light on the prevalence and prediction of depression in this demographic, with the chosen prediction algorithm achieving the highest level of accuracy.

R. Qasrawi,et al.,[5] used machine learning techniques to forecast schoolchildren's risk factors for anxiety and sadness. Data from 3984 students aged 10-15 were analyzed using five machine learning methods, with SVM and RF models achieving the highest levels of accuracy for both depression (SVM: 92.5%; RF: 76.4%) and anxiety (92.4%; RF: 78.6%). Research has shown that a number of factors, including academic achievement, family income, home violence, bullying,

and school violence, significantly contribute to anxiety and depression. Overall, machine learning was effective in identifying and predicting factors influencing student mental health, indicating that it has the potential to inform school-based prevention and intervention programs.

Muntequa Imtiaz Siraji et al [6] investigated the mental health of students at a Bangladeshi international university, with a focus on their use of mobile devices. Data from 444 individuals were gathered by a cross-sectional survey, and further exploratory data analysis was carried out. Using eight machine learning techniques, an automated system for identifying and categorizing normal-to-severe depression. Using feature selection techniques like Recursive Feature Elimination (RFE) and the Chi-square test increased accuracy by 3 to 5%, while feature extraction methods such as Principal Component Analysis (PCA) increased accuracy by 5 to 15%. The best accuracy, F1-score, and ROC-AUC were obtained when the CatBoost classifier and the SparsePCA feature extraction approach were used together. Approximately 44% of participants showed no signs of depression, 25% displayed moderate-to-moderate symptoms, while 31% had severe-to-extreme ones. These findings highlight the potential of machine learning models, with proper feature engineering, for multi-stage depression detection in students, implying broader applicability across disciplines for early depression detection.

Shafiee, Nor Safika Mohd,et al.. [7] presented an overview of Mental health concerns among those pursuing higher education, identifying research gaps and contributing factors using a comparative analysis of the existing literature. Financial hardships, a lack of social support system, and the learning environment have all been identified as common contributors to student mental health problems. Among supervised learning techniques, SVM is the most popular, with an accuracy range of 70% to 96% observed in the reviewed research papers.

M. S. Ahmed et al.[8] addressed the research to create a rapid system for detecting depression using app usage data, as well as to determine which machine learning models are most suited for this particular activity. The study used a variety of machine learning approaches, such as linear, tree-based, and neural network models, as well as feature selection methods, to examine Patient Health Questionnaire-9 data and app usage responses from 100 Bangladeshi students. The light gradient boosting machine model, which used only 1-second-retrieved app usage data, had the

highest accuracy, correctly identifying 82.4% of depressed students with a 75% precision and an F1-score of 78.5%.

K. Opoku Asare, et al. [9] looked to predict depression using smartphone data and identify influential behavioral markers. Data from 629 participants were gathered over an average of 22.1 days. Twenty-two behavioral markers were quantified from smartphone data and linked to depression using linear mixed models. Five machine learning algorithms had high performance metrics, with precision ranging from 85.55% to 92.51% and recall ranging from 92.19% to 95.56%. The highest level of accuracy achieved was 98.14%, includes elements for internet connectivity and a screen identified as the most important predictors of depression.

J. Hong et al.[10] focused on this study to predict depression is typified by enduring melancholy or a decline in interest in activities, emotion management techniques, and sociodemographic factors. Participants engaged in ecological momentary assessment, responding to an app's prompts six times per day for seven days. The Random Forest algorithm outperformed other models, employing 13 of 36 variables, with age, feelings of worry, and a particular technique for controlling emotions in relation to social interaction that has been found to be a reliable indicator of severe depression symptoms. These findings provide additional insights as opposed to conventional diagnostic techniques and psychological evaluations, potentially improving predictive accuracy in identifying depression among emerging adults.

X. Xu et al.,[11] focused on this study to predict depression is typified by enduring melancholy or a decline in interest in activities, emotion control techniques, and sociodemographic factors through ecological momentary assessment. The Random Forest algorithm, which used 13 of 36 variables, including age, worry, and a particular technique for managing emotions, outperformed other methods in predicting severe depression. These findings provide insights that could improve conventional techniques for diagnosis and psychological evaluations for detecting depression in emerging adults. The study's approach detects depression using interpretable behavioral rules, outperforming standard models by an average of 9.7% across multiple metrics. Furthermore, the method's generalizability was confirmed using a second dataset, which produced similar results.

Y. Hou, J. Xu, et al. [12] Used data from This study investigates the relationship between reading habits and depression propensity among university students using records from the university library and mental health surveys. Various text categorization algorithms, like the naïve Bayesian classifier, SVM, and kNN, are evaluated in terms of time required and accuracy for a variety of sample sizes. The naïve Bayesian classification approach based on a polynomial model is used to build a book classifier that achieves an accuracy of 82.3%. Algorithms for logistic and linear regression are used to build a psychological prediction model, with the logistic model outperforming the linear model in terms of accuracy under various error criteria.

De, Arringoni Flavia, et al.[13] focused on this study to By looking at affect, emotion management techniques, and sociodemographic factors, one might anticipate depression symptoms in emerging adults. Participants engaged in ecological momentary assessment, responding to an app's prompts six times per day for seven days. The Random Forest algorithm outperformed other models, employing 13 of 36 variables, with age, feelings of worry, and a particular technique for controlling emotions in relation to social interaction were found to be reliable indicators of severe depression symptoms.

In order to precisely forecast the degrees of depression in people who were placed in home quarantine during the COVID-19 epidemic, this study created a depression prediction model utilizing a variety of machine learning methods. The model was evaluated on fifteen different models, and the model with the highest accuracy, precision, ROC curve, and AUC was selected as the top performer. Out of the 15 models, the Adaboost model was determined to be more successful with an ACC of 0.7917, accuracy of 0.7180, and AUC of 0.7803. The validation sets' AUC was higher than 0.83 [14].

This study examined posts and comments on suicide thoughts using data from Reddit. NLP, or natural language processing, was utilized to comprehend suicide-related transdisciplinary domains. There are two support groups for depression and thoughts of suicide: r/depression and r/SuicideWatch. Sincere written data on mental health may be found in these subreddits. People who were at danger were tracked using machine learning techniques like SVM and Naïve Bayes.

The most accurate models were Random Forest, Naïve Bayes, Support Vector Machine, and Logistic Regression, all of which had high accuracy rates in the results [15].

The purpose of the study is to examine depression and use speech samples to forecast symptoms. A machine learning model that establishes a connection between speech traits and sadness is developed using a sizable data set. The model can determine a subject's level of depression with accuracy [16].

The purpose of this study was to use machine learning to assess stress levels in college students. Word choice, tone, and writing style were all extracted from the writings of the patients using predictive modeling. With an 81.79% SVM accuracy and a 70% LSTM accuracy, the classifier demonstrated outstanding accuracy. Using characteristics from online depression forums, the NLP approach detected depression with 80% accuracy, whereas Naive Bayes obtained 85% accuracy after gathering data from 210 individuals [17].

A research examined 205,581 user reviews from chatbots intended for those who experienced symptoms of sadness and anxiety. Utilizing Python libraries and scraper tools, the text and metadata were recovered. Based on user ratings, the reviews were divided into positive and negative meta-themes. Building confidence, sufficient analysis, consultation, concern, and simplicity of use were among the positive themes. Usability problems, update concerns, privacy concerns, and uninspired dialogues were among the negative topics [18].

The study's objective is to create a model for evaluating tweets from Arabic Twitter users and identifying sadness by adding "neutral" tweets to the dataset. The model makes use of natural language processing methods and machine learning classifiers such as Naïve Bayes, AdaBoost, Random Forest, Support Vector Machine, and Logistic Regression. Outperforming the others, the RF classifier achieves an accuracy of 82.39% [19].

4,262 people with anxiety, depression, or both were included in the research. Models of anxiety and depression were classified using four machine learning methods. Metrics such as K-nearest neighbors, support vector machines, random forests, and adaptive boosting were used in cross-

validation to display the models' performance. Outperforming the others, the support vector machine achieved the greatest AUC and the best accuracy [20].

The study offers a novel approach to diagnosis that examines cognitive biases in people who are not yet clinically depressed or anxious. Four groups of 125 individuals were created depending on the symptoms of sadness and anxiety. Many cognitive emotional biases were identified and measured using a thorough battery of behavioral tests. Utilizing cutting-edge machine learning techniques, these outcomes were examined and group membership was predicted. For symptomatic patients, the study demonstrated a prediction accuracy of 71.44% and specificity of 70.78%. The findings point to the need for better diagnostic tools and more successful, individualized care [21].

A research trained a machine learning system to recognize MDD diagnoses using 550 speech samples taken from clinical interviews. Although the system's accuracy was just 66%, it was still much better than chance. On the other hand, there was a greater degree of accuracy (73%), when diagnosing MDD using standard depression measures such as the Hamilton Rating Scale for Depression, QIDS-C, and PHQ-9. According to the study's findings, automated speech analysis can recognize patterns in sad speech but doesn't considerably enhance classifications [22].

In order to identify sadness on Twitter, the study makes use of machine learning methods including SVM, Random forest, and Logistic Regression. After downloading tweets and analyzing sentiment, the model splits the dataset into 80:20 ratios and uses evaluation matrices such as accuracy, precision, recall, and F1 score to evaluate performance. With an accuracy of 88%, the Random Forest algorithm produces the best predictions [23].

2.3. Comparison between existing works

Existing research on mental health prediction in university students varies in methodology, features considered, and predictive models used. Some research focuses on characteristics and academic performance, while others consider lifestyle habits and stressors. Furthermore, the selection of machine learning algorithms varies, with LR, DT, and neural networks being popular options. While some studies have shown promise in predicting mental health outcomes, more

comparisons and evaluations of various approaches are needed to identify the most effective predictive models. This brief comparison highlights the variety of methods and models used in existing works, emphasizing the importance of studying different approaches to addressing mental health challenges in university settings.

Table 2.3: Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	C. G. Walsh, et al.[1]	Machine Learning	84%
2.	S. Ware et al [2	Machine Learning	86%
3.	P. Chikersal et al [3]	Machine Learning	85.7%
4.	R. Qasrawi,et al.,[5]	Machine Learning	SVM: 92.5%
5.	Muntequa Imtiaz Siraji et al [6]	Machine Learning	44%
6.	Shafiee, Nor Safika Mohd,et al.. [7]	Machine Learning	SVM=96%
7.	Y. Hou, J. Xu, et al. [12]	Machine Learning	NB= 82.3%

2.4. Scope of the Problem

Students at universities face a variety of challenges that can have a negative impact on their mental health, possibly leading to breakdowns with lasting effects. However, identifying students who are at risk of mental health problems based on their daily activities is a challenging and time-consuming endeavor. This study aims to address this issue by using machine learning techniques to predict mental health breakdowns among university students. The primary goal is to create a predictive model that can effectively identify people who are at risk for mental health problems based on their statistics, daily routines, and other relevant factors. This study aims to enable early intervention and support mechanisms tailored to the needs of at-risk students, resulting in a healthier and more resilient university community.

2.5. Challenges

Several challenges may be encountered in conducting this study. Firstly, obtaining accurate data on sensitive topics like mental health may be hindered by social stigma and reluctance to disclose personal experiences. Ensuring participants' anonymity and confidentiality will be crucial to encourage honest responses.

Additionally, the diverse socio-cultural landscape of Bangladesh may pose challenges in interpreting findings. Cultural differences in expressing and perceiving mental health symptoms may influence participants' responses, requiring careful consideration in data analysis and interpretation.

Moreover, the dynamic nature of university life and the ever-evolving nature of stressors and support systems may impact the relevance and applicability of study findings over time. Longitudinal studies or frequent reassessment may be needed to capture these changes adequately.

Finally, logistical challenges such as reaching a representative sample across various universities and ensuring high response rates may require strategic planning and collaboration with university administrations and student organizations to overcome.

2.6. Gap Analysis

The literature review identifies several gaps in current research on predicting mental health breakdowns in university students based on their daily activities using machine learning techniques. To begin, while many studies focus on predicting depression and suicidal behavior, there is a need for more comprehensive models that take into account a broader range of mental health indicators and outcomes. Second, existing studies frequently use passive data collection methods such as smartphone use or electronic health records, which may not capture the full range of daily activities and stressors relevant to mental health. Incorporating more diverse data sources, such as self-reported surveys or wearable devices, may provide a more complete picture of students' experiences. Furthermore, while machine learning algorithms show promise in predicting mental health outcomes, there is a lack of reliability in model evaluation metrics and methodologies between studies.

CHAPTER 3
RESEARCH METHODOLOGY

3.1. Proposed Method and Components

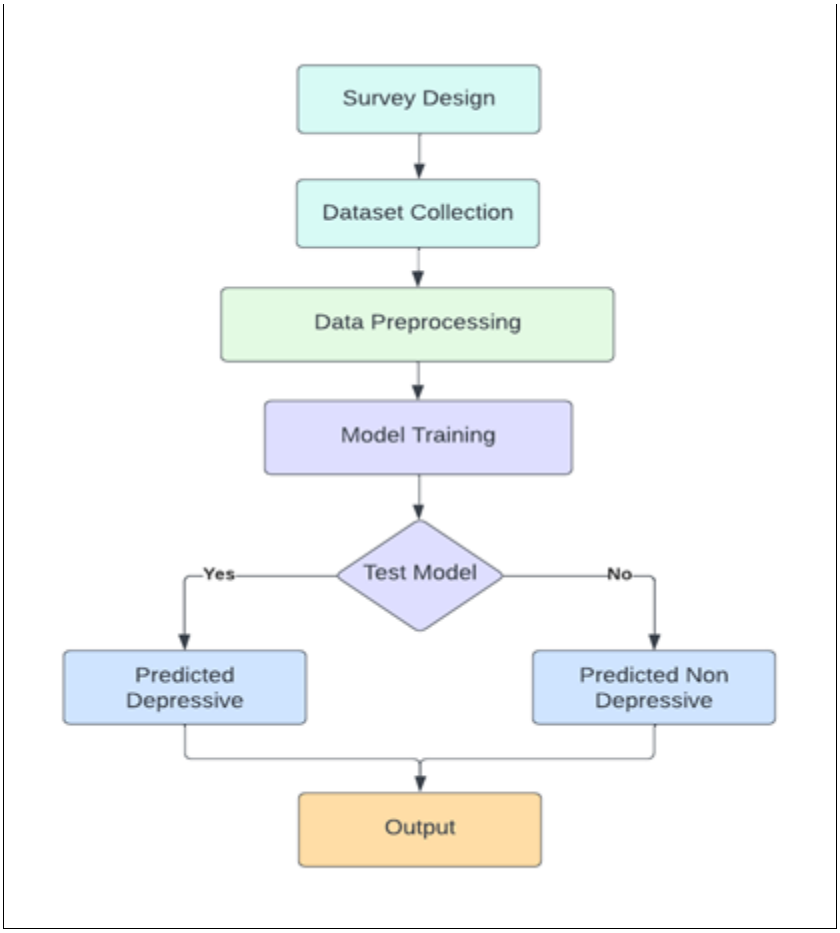


Figure 3.1: Workflow of project

The workflow begins with designing a survey form to collect data on depression among Bangladeshi university students. After gathering responses from 1000 participants, the original data undergoes preprocessing using various techniques to ensure its usability. This includes steps such as cleaning, addressing missing data and encoding variables that are categorical. The data is then divided into testing and training sets. to facilitate model development and evaluation.

Using the training data, the model is trained to identify patterns and correlations between variables indicative of depression. After training, the model is evaluated for performance using a different test dataset and generalization ability. Finally, Predictions are made using the trained model on fresh data, generating insights into the prevalence and potential risk factors for depression among university students in Bangladesh.

3.2. Data Collection

For data collection, a survey form was designed to gather information on depression among Bangladeshi university students. This form included questions about various factors related to mental health, lifestyle, and academic experiences. The survey was distributed among university students, resulting in the collection of 1000 responses. After gathering the data, it underwent preprocessing to ensure its usability for analysis. This involved steps such as encoding categorical variables, addressing missing values, and cleaning the data. To make the process of developing and evaluating models easier, the prepared data was separated into sets for training and testing. The training set was used to train the model to recognize patterns indicative of depression, while the testing set was used to assess the model's performance. Finally, the trained model was used to make predictions on new data, generating insights into depression frequency and related variables among university students from Bangladesh.

3.3. Dataset Description

A survey form consisting of 10 questions to diagnose depression was created, and data collection was conducted resulting in 1000 responses. The dataset contains 2 parts, 20% is testing data while 80% is training data. The dataset comprises 16 columns, all of which are categorical, representing the responses to each question and additional information. Each column corresponds to a specific question in the survey, making the data independent as no question depends on another. The numeric values in the dataset likely represent different response options or scores for each question, allowing for quantitative analysis. This structured dataset provides valuable insights into the prevalence and characteristics of depression among the surveyed Bangladeshi university students, facilitating further analysis and modeling to understand and address mental health challenges in this population.

Table 3.3: Dataset Description

SN	Attribute
1	Name
2	Email Address
3	Student ID
4	Age
5	University
6	Little interest or pleasure in doing things
7	Feeling down, depressed, or hopeless
8	Trouble falling or staying asleep, or sleeping too much
9	Feeling tired or having little energy
10	Poor appetite or overeating
11	Feeling bad about yourself or have let yourself or your family down
12	Trouble concentrating on things, such as reading the newspaper or watching television
13	Walking or talking so slowly that other people can notice. Or so restless as to move around a lot more than usual
14	Thoughts that you would be better off dead, or of hurting yourself
15	If you have any problems, how difficult are the problems in your daily work, at home or getting on with other people?
16	How do you feel about your physical condition?

3.4. Statical Analysis

- There are 1000 total data points in the dataset.

- The dataset contains 2 parts, 20% is testing data while 80% is training data.
- Dataset retains 16 columns.
- There are a total of 10 questions in the dataset.
- There are 1 dependent and 10 independent or input variables in the dataset.
- In this project we used 6 machine learning models.

3.5. Data Preprocessing

i.Null value elimination column wise:

Null value elimination column-wise involves removing or filling in missing data (null values) in each column of a dataset. This process ensures that the dataset is complete and accurate, improving the quality of analysis. Techniques include deleting rows with null values, replacing nulls with statistical measures like mean or median, or using predictive algorithms to estimate missing values. This step is crucial for maintaining the integrity of data used in machine learning models and statistical analyses.

ii. Null value elimination row wise:

Null value elimination row-wise involves scanning each row of a dataset to identify and remove any rows containing null or missing values. This process ensures that the remaining data is complete and reliable for analysis. By eliminating rows with null values, data quality is improved, and potential biases or inaccuracies in statistical analysis and machine learning models are reduced. This method is particularly useful when the presence of null values can significantly impact the results or insights derived from the data.

iii. Drop Unnecessary Column

The "drop unnecessary column" preprocessing technique involves removing columns from a dataset that are not relevant to the analysis or model development. This step is taken to streamline the dataset and improve computational efficiency. Columns that do not contribute to the objectives of the study or contain redundant information are identified and removed, leaving only the essential features for further analysis or modeling.

iv. Encoding

Ordinal Encoder is a preprocessing technique used to transform numerical data from category variables. Each category is given a distinct number according to its rank or order. For example, if a variable has categories like "low," "medium," and "high," Ordinal Encoder would assign 0 to "low," 1 to "medium," and 2 to "high." This preserves the ordinal relationship between categories, making it suitable for variables with ordered levels but doesn't introduce unnecessary complexity like one-hot encoding.

v. Mapping

The "Mapping" preprocessing technique involves converting categorical values into numerical values using a predefined mapping or dictionary. For instance, categorical data like "low," "medium," and "high" can be mapped to 1, 2, and 3, respectively. This technique ensures that the categorical data can be utilized in machine learning algorithms, which require numerical input, while preserving the inherent order or relationship between the categories.

vi. Correlation Matrix

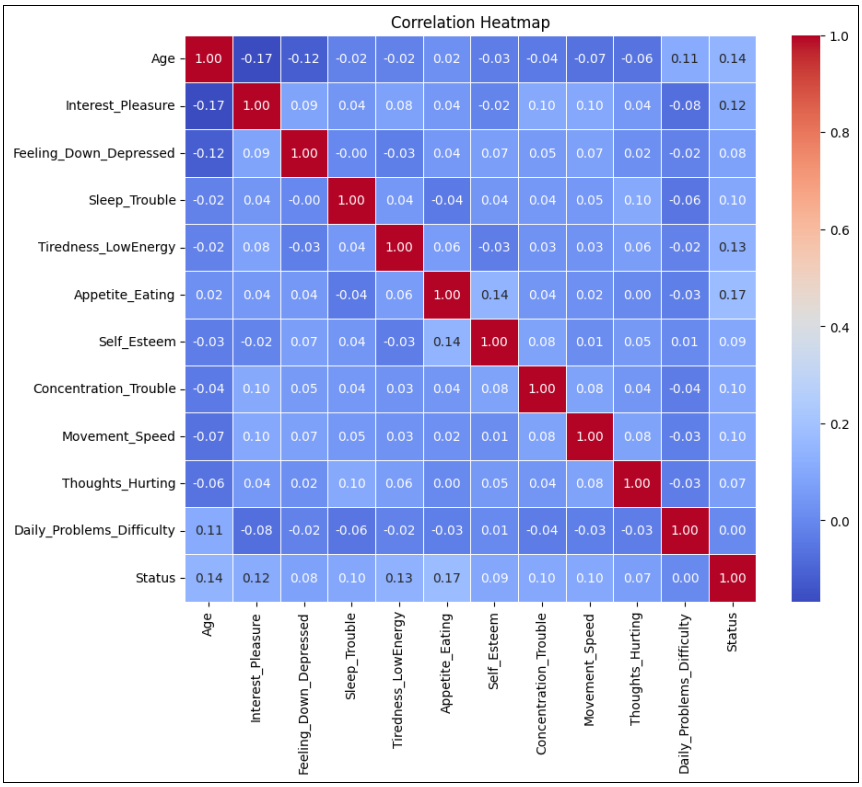


Figure 3.5: Correlation Matrix

3.6. Proposed Model

i. Logistic Regression:

Logistic regression uses the logistic function, or sigmoid, to map predicted values to probabilities between 0 and 1. The algorithm estimates the coefficients of input features by maximizing the likelihood of observing the given data. This makes it effective for tasks such as spam detection or medical diagnosis, where the output is categorical. Logistic regression is valued for its simplicity, interpretability, and effectiveness in scenarios where the relationship between features and the outcome is linear.

ii. Decision Tree Classifier

Decision trees are adept at capturing complex relationships within data and are popular for their transparency and simplicity. They facilitate decision-making by breaking down a problem into a series of binary choices, making it easy to understand and visualize. Additionally, decision trees can be prone to overfitting, but techniques like pruning and ensemble methods such as random forests mitigate this issue, enhancing their overall performance.

iii. Random Forest Classifier

The Random Forest Classifier is an ensemble learning algorithm. Predictions are made based on the majority vote from these trees, enhancing robustness and stability. Random Forests handle large datasets with higher dimensionality well and are resistant to noise, making them a popular choice for tasks like medical diagnosis, financial forecasting, and image recognition.

iv. Support Vector Machine

SVM is a supervised learning algorithm. This hyperplane is chosen to ensure that the nearest data points from each class, called support vectors, are as far apart as possible. SVM can handle both linear and non-linear data through the use of kernel functions, which transform the input space into higher dimensions. This makes SVM powerful for complex datasets with clear class separations, offering high accuracy and robustness in various applications.

v. Gradient Boosting Algorithm

In the Gradient Boosting algorithm each new tree corrects the errors of the preceding ones by focusing on the residual errors. This iterative process continues until the model achieves optimal performance. Gradient Boosting adjusts the model weights based on the gradient of the loss function, ensuring that it minimizes errors effectively. Its ability to handle various types of data and provide high accuracy makes it popular for both regression and classification tasks.

vi. AdaBoost Algorithm

AdaBoost, short for Adaptive Boosting, works by sequentially training classifiers, each focusing on the errors of its predecessor. Initially, all data points are weighted equally, but in subsequent rounds, misclassified points receive higher weights, forcing the new classifier to prioritize them. This iterative process continues until a predefined number of classifiers are trained. The final model is a weighted vote of all the classifiers. AdaBoost is effective in reducing bias and variance, leading to improved accuracy and robustness in predictions.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Discussion

The workflow begins with designing a survey form to collect data on depression among Bangladeshi university students. After gathering responses from 1000 participants, the original data undergoes preprocessing using various techniques to ensure its usability. This covers actions like managing missing values, cleaning, and category variable encoding. To aid in the construction and assessment of the model, the data is then divided into training and testing sets.

Using the training data, the model is trained to identify patterns and correlations among variables that are suggestive of depression. After training, the model is evaluated for performance and generalizability using a different test dataset. Lastly, predictions are made on fresh data using the trained model, providing information on the frequency and possible risk factors of depression among Bangladeshi university students.

The study employed machine learning techniques such as AdaBoost to predict depression. Models were trained on the training data to learn patterns indicative of depression and evaluated using a different test dataset in order to gauge their effectiveness and capacity for generalization. Model performance was assessed using evaluation criteria such F1-score, recall, accuracy, and precision. Overall, the study aimed to provide insights into the prevalence and potential risk factors for depression among Bangladeshi university students.

Table 4.1: Classifiers Description

Classifier	Description
Logistic Regression	Logistic regression uses the logistic function, or sigmoid, to map predicted values to probabilities between 0 and 1. The algorithm estimates the coefficients of input features by maximizing the likelihood of observing the given data.

Decision Tree	Decision trees are adept at capturing complex relationships within data and are popular for their transparency and simplicity. They facilitate decision-making by breaking down a problem into a series of binary choices, making it easy to understand and visualize.
Random Forest	The Random Forest Classifier is an ensemble learning algorithm. Predictions are made based on the majority vote from these trees, enhancing robustness and stability.
SVM	SVM is a supervised learning algorithm. This hyperplane is chosen to ensure that the nearest data points from each class, called support vectors, are as far apart as possible.
Gradient Boosting Algorithm	In the Gradient Boosting algorithm each new tree corrects the errors of the preceding ones by focusing on the residual errors. This iterative process continues until the model achieves optimal performance.
Adaboost Algorithm	AdaBoost, short for Adaptive Boosting, works by sequentially training classifiers, each focusing on the errors of its predecessor. Initially, all data points are weighted equally, but in subsequent rounds, misclassified points receive higher weights, forcing the new classifier to prioritize them.

4.2. Experimental Results and Analysis

i. Logistic Regression

Table 4.2: Classification report of Logistic Regression

Classification	Precision	Recall	F1 Score	Support
0.0	0.0	0.55	0.36	179
1.0	0.73	0.90	0.80	179
Accuracy			0.56	358
Macro avg	0.64	0.58	0.58	358
Weighted avg	0.67	0.70	0.66	358

Figure 4.2: Performance matrix of Logistic Regression

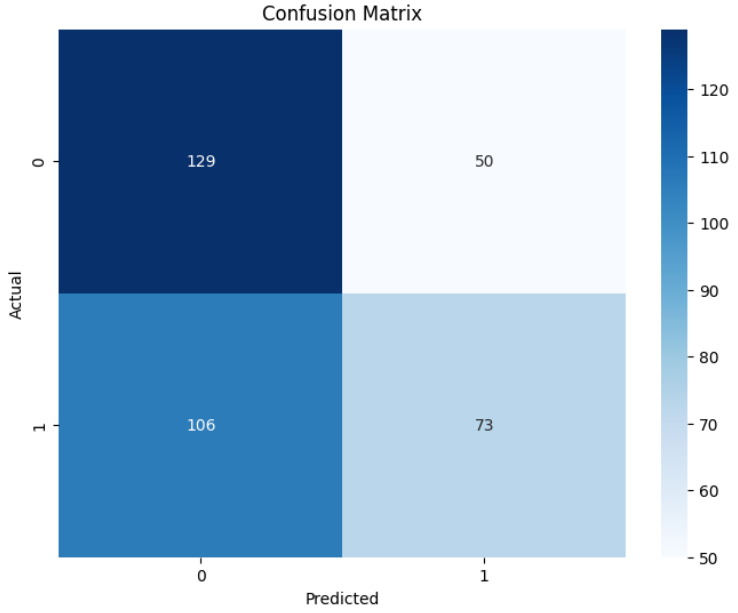


Figure 4.2: Confusion matrix of Logistic Regression

This confusion matrix for a logistic regression model shows 93 true negatives, 83 true positives, 44 false positives, and 54 false negatives, indicating the model's performance in predicting depression, where 0 represents no depression and 1 represents depression.

ii.Decision Tree

Table 4.2: Classification report of Decision Tree

Classification	Precision	Recall	F1 Score	Support
0.0	0.42	0.35	0.38	179
1.0	0.72	0.77	0.75	179
Accuracy			0.68	358
Macro avg	0.57	0.56	0.56	358
Weighted avg	0.62	0.64	0.63	358

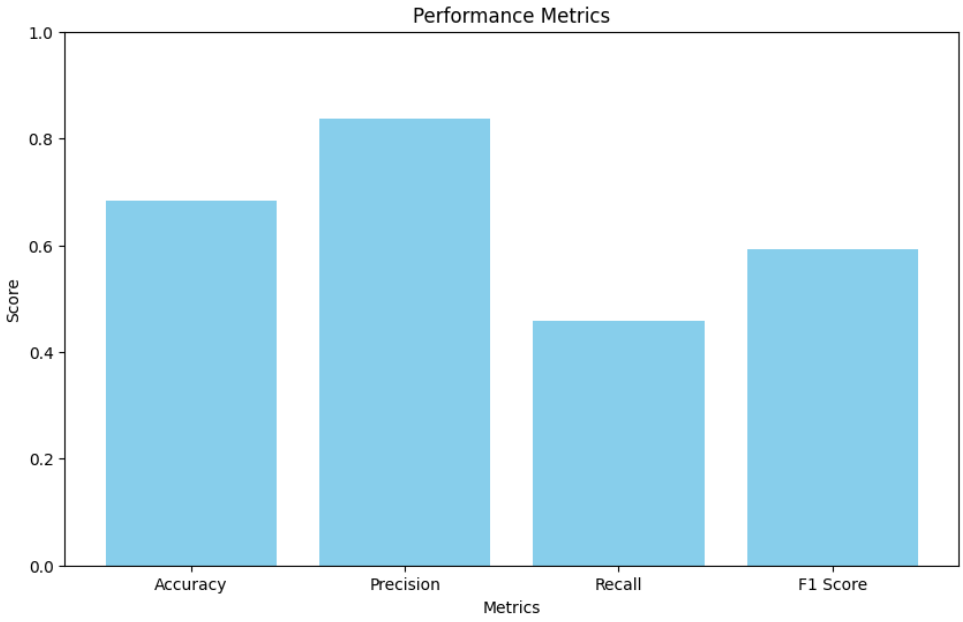


Figure 4.2: Performance matrix of Decision Tree

This confusion matrix for a decision tree model shows 78 true negatives, 104 true positives, 59 false positives, and 33 false negatives, illustrating the model's performance in predicting depression, where 0 represents no depression and 1 represents depression.

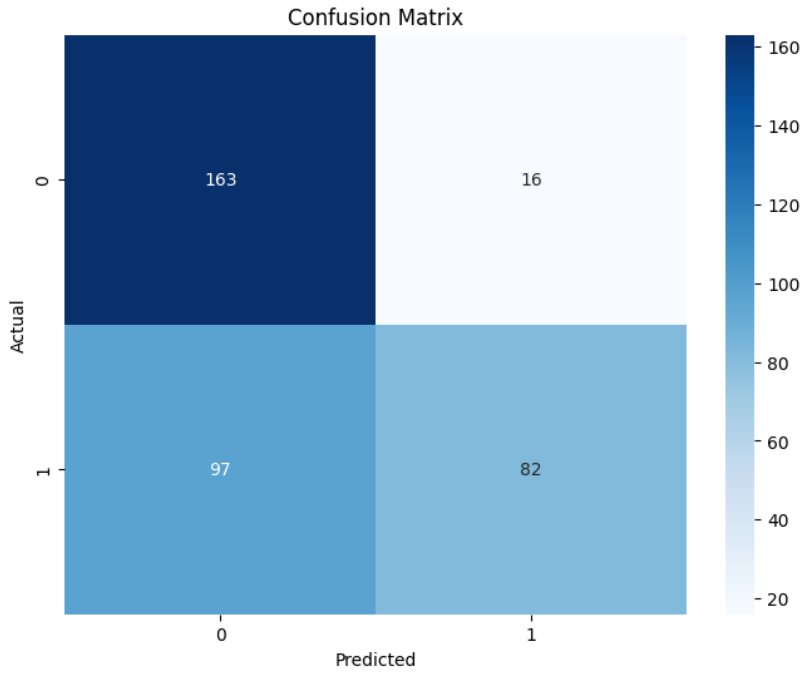


Figure 4.2: Confusion matrix of Decision Tree

iii.Random Forest

Table 4.2: Classification report of Random Forest

Classification	Precision	Recall	F1 Score	Support
0.0	0.51	0.29	0.37	179
1.0	0.73	0.88	0.79	179
Accuracy			0.68	358
Macro avg	0.62	0.58	0.58	358
Weighted avg	0.66	0.69	0.66	358

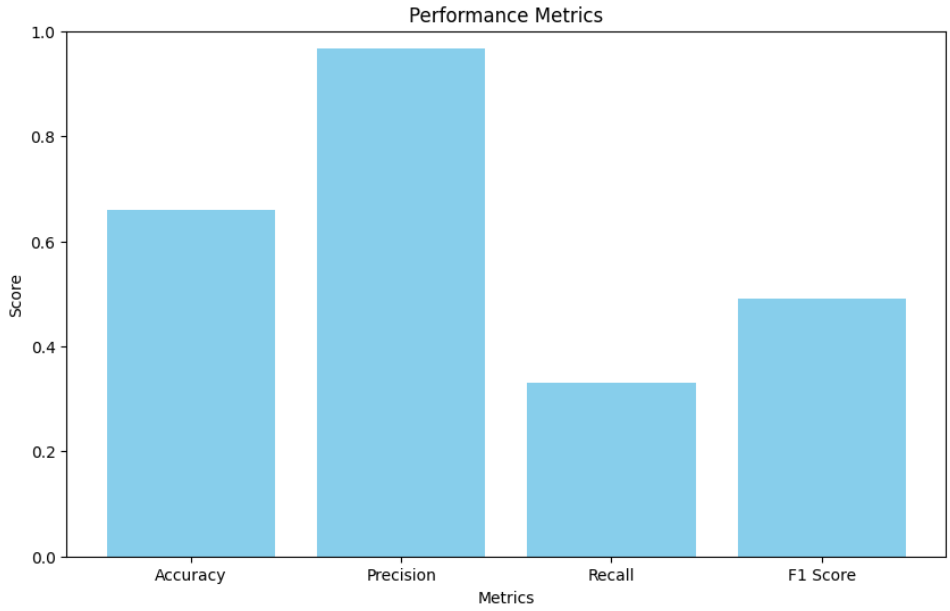


Figure 4.2: Performance matrix of Random Forest

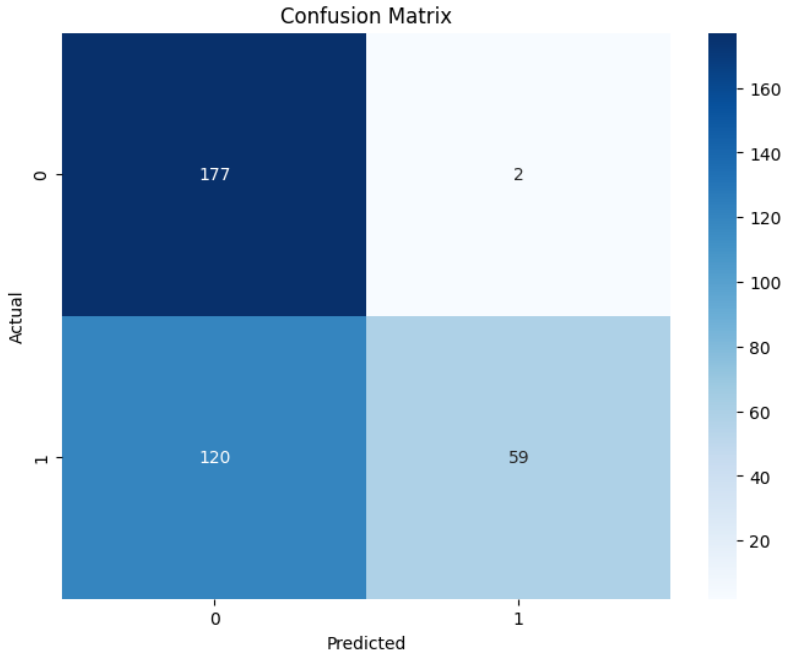


Figure 4.2: Confusion matrix of Random Forest

This confusion matrix for a Random Forest model shows 84 true negatives, 115 true positives, 53 false positives, and 22 false negatives. It demonstrates the model's performance in predicting depression, where 0 represents no depression and 1 represents depression.

iv.Support Vector Machine

Table 4.2: Classification report of Support Vector Machine

Classification	Precision	Recall	F1 Score	Support
0.0	0.69	0.17	0.28	179
1.0	0.72	0.96	0.82	179
Accuracy			0.52	358
Macro avg	0.70	0.57	0.55	358
Weighted avg	0.71	0.71	0.65	358

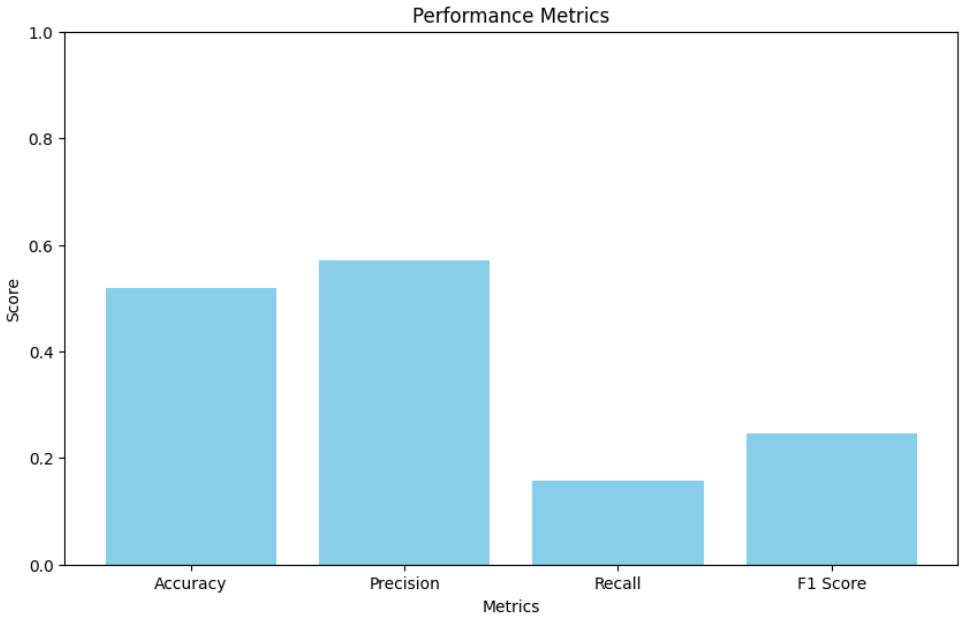


Figure 4.2: Performance matrix of Support Vector Machine

This is a confusion matrix for an SVM model. It shows 89 true negatives, 97 true positives, 48 false positives, and 40 false negatives. It visualizes the model's performance in classifying binary outcomes.

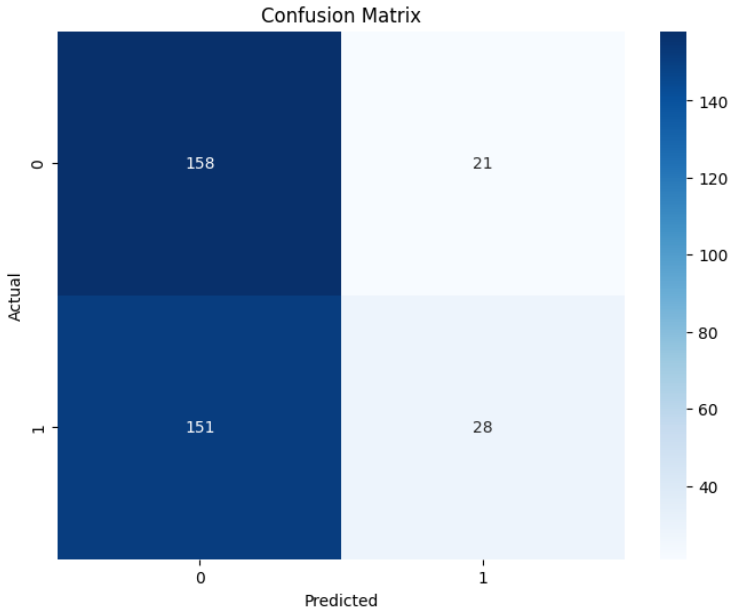


Figure 4.2: Confusion matrix of Support Vector Machine

v. Gradient Boosting Algorithm

Table 4.2: Classification report of Gradient Boosting Algorithm

Classification	Precision	Recall	F1 Score	Support
0.0	0.85	0.99	0.91	179
1.0	0.99	0.82	0.90	179
Accuracy			0.91	358
Macro avg	0.92	0.91	0.90	358
Weighted avg	0.92	0.91	0.90	358

Figure 4.2: Performance matrix of Gradient Boosting Algorithm

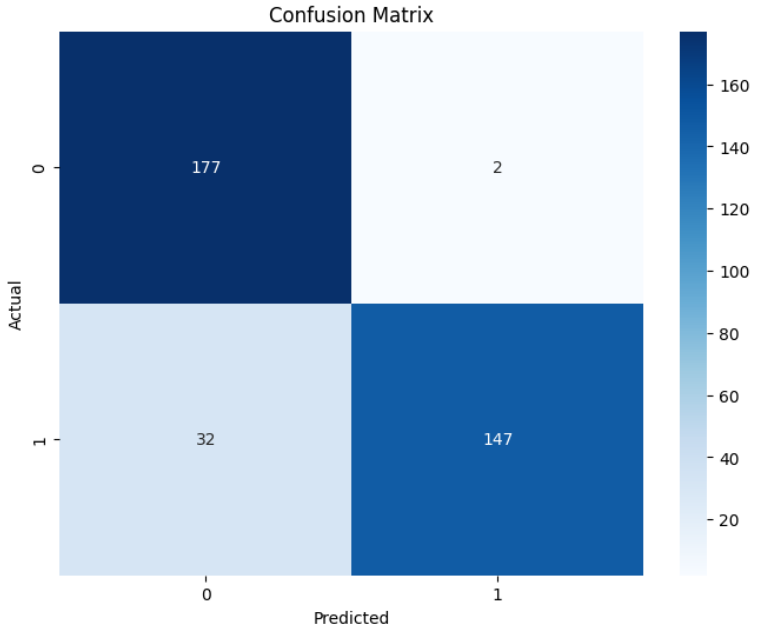


Figure 4.2: Confusion matrix of Gradient Boosting Algorithm

This is a confusion matrix for a gradient boosting algorithm. It shows 99 true negatives, 113 true positives, 38 false positives, and 24 false negatives. It visualizes the model's performance in classifying binary outcomes.

vi.Adaboost Algorithm

Table 4.2: Classification report of Adaboost Algorithm

Classification	Precision	Recall	F1 Score	Support
0.0	0.83	0.97	0.89	179
1.0	0.96	0.80	0.88	179
Accuracy			0.89	358
Macro avg	0.90	0.89	0.88	358
Weighted avg	0.90	0.89	0.88	358

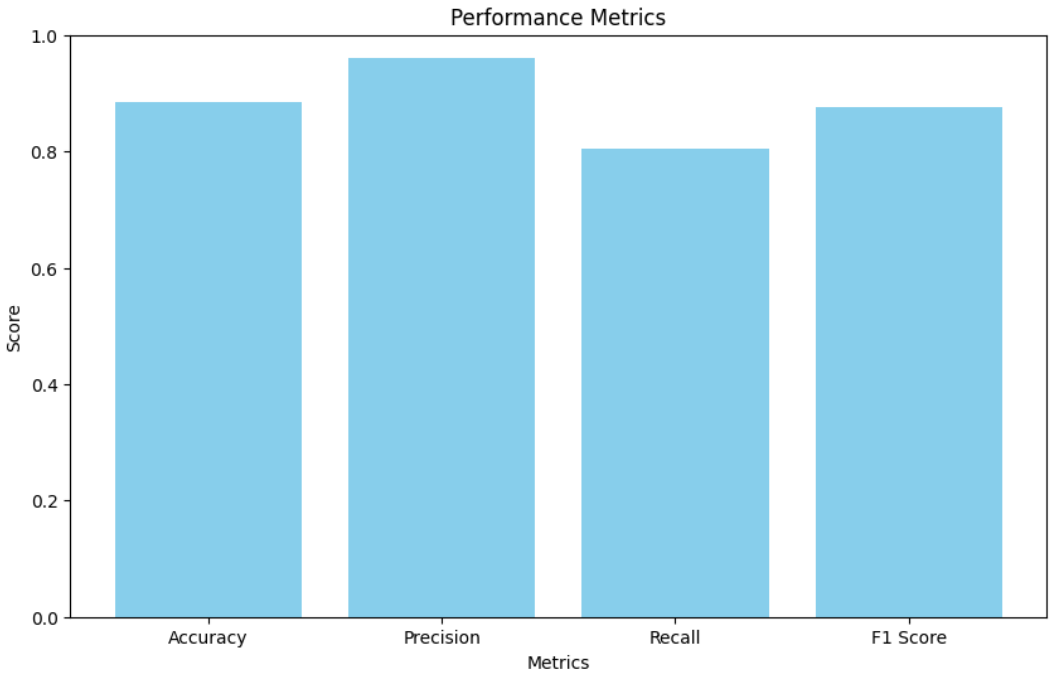


Figure 4.2: Performance matrix of AdaBoost Algorithm

This is a confusion matrix for an AdaBoost algorithm. It shows 103 true negatives, 108 true positives, 34 false positives, and 29 false negatives. It visualizes the model's performance in classifying binary outcomes.

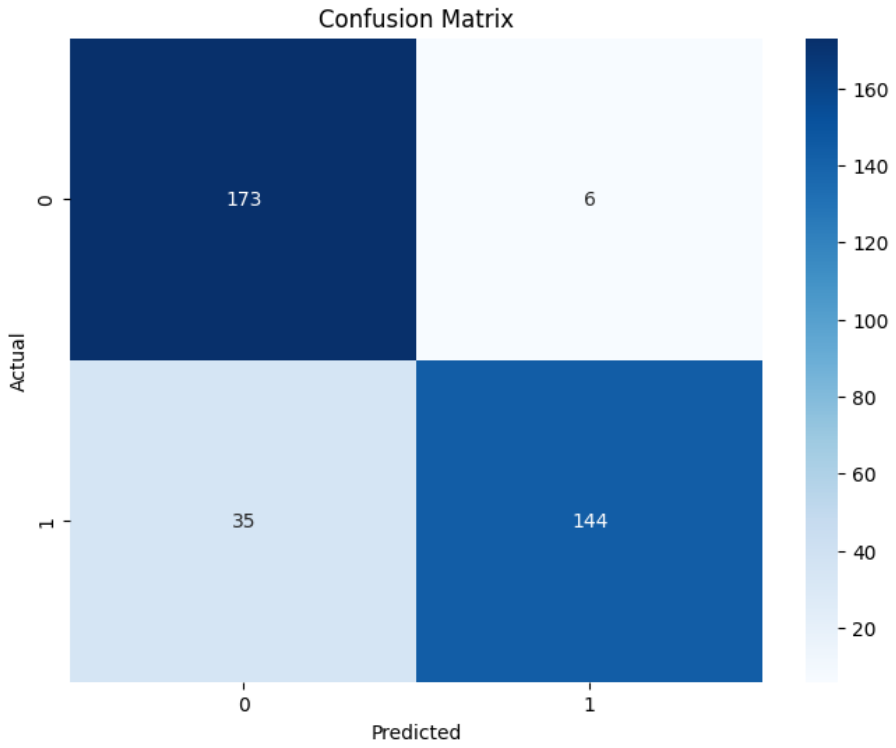


Figure 4.2: Confusion matrix of AdaBoost Algorithm

vii. Algorithms Summary

Table 4.2: Classification report Summary

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.64	0.67	0.70	0.66
Decision Tree	0.66	0.62	0.64	0.63
Random Forest	0.73	0.66	0.69	0.66
SVM	0.68	0.71	0.71	0.65
Gradient Boosting Algorithm	0.77	0.70	0.71	0.70
Adaboost Algorithm	0.77	0.71	0.72	0.71

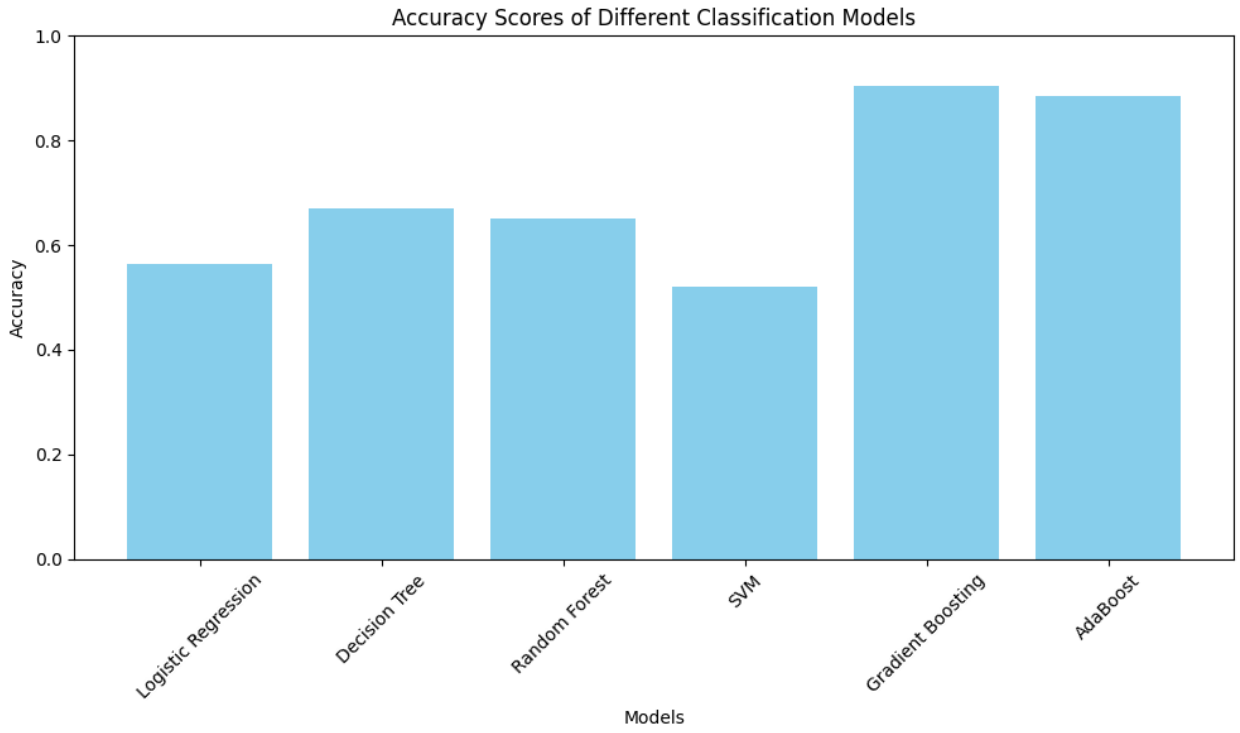


Figure 4.2: Accuracy Score of Different Classification Models

viii. Test Model Prediction

Estimated my model using the Gradient Boosting Algorithm. Estimates of depressed were obtained using the AdaBoost algorithm. The estimation results are shown in the table below:

Table 4.2: Test Gradient Boosting model for Prediction

Test Gradient Boosting model for Prediction		
Attribute	User Input	Output
Age Interest_Pleasure Feeling_Down_Depressed Sleep_Trouble Tiredness_LowEnergy Appetite_Eating	21-23 Several Several Nearly everyday More than half the days More than half the days	Depressed

Self_Esteem	More than half the days	
Concentration_Trouble	Several	
Movement_Speed	Nearly everyday	
Thoughts_Hurting	More than half the days	
Daily_Problems_Difficulty	Several difficult	

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1. Impact on Society

The prevalence of depression among Bangladeshi university students has far-reaching consequences for society. Firstly, untreated depression can impair academic performance, leading to decreased productivity and potential dropout rates among students. This not only affects the individuals directly but also hampers the country's workforce development and economic growth. Furthermore, depression can strain social relationships and familial dynamics, as students may withdraw from social activities and experience difficulties in interpersonal interactions. This can contribute to a sense of isolation and exacerbate the stigma surrounding mental health issues, hindering efforts to promote understanding and support within communities.

Moreover, the mental health of university students has implications for public health resources and healthcare systems. The burden of untreated depression may result in increased demand for mental health services, which are often under-resourced and inaccessible in many parts of Bangladesh. Addressing the mental health needs of university students is essential for alleviating strain on healthcare infrastructure and ensuring equitable access to support services for all members of society.

Overall, the impact of depression among university students extends beyond individual well-being to affect broader societal structures and systems. Collaboration between educational institutions, healthcare professionals, legislators, and communities is necessary to address this issue to promote mental health awareness, provide effective support systems, and lessen the stigma attached to getting mental health treatment.

5.2. Impact on Environment

The prevalence of depression among Bangladeshi university students can have indirect yet notable impacts on the environment. Firstly, the academic stress and emotional burden experienced by students may lead to decreased engagement in environmental conservation efforts. Students

struggling with depression may prioritize their mental health over environmental concerns, resulting in reduced participation in eco-friendly behaviors such as recycling, conservation activities, and sustainability initiatives.

Additionally, depression can influence consumption patterns and lifestyle choices, potentially leading to increased waste generation and environmental degradation. Individuals experiencing depressive symptoms may engage in comfort shopping or excessive consumption of resources as a coping mechanism, contributing to greater carbon footprints and environmental pollution.

Moreover, the mental health of university students affects their relationship with nature and outdoor spaces. Depressive symptoms may discourage students from spending time outdoors or participating in outdoor activities, limiting their connection with the natural environment and potentially decreasing advocacy for environmental conservation efforts.

Addressing depression among university students is not only crucial for individual well-being but also for fostering a sustainable relationship with the environment. By promoting mental health awareness and providing support systems, educational institutions and environmental organizations can empower students to prioritize both their mental health and environmental stewardship, ensuring a healthier and more sustainable future for all.

5.3. Ethical Aspects

Ethical considerations are paramount in conducting research on depression among university students in Bangladesh. Firstly, ensuring voluntary participation and informed consent is essential to uphold participants' autonomy and protect their rights. This includes providing clear information about the study's purpose, risks, and benefits, as well as the option to withdraw at any time without consequences.

Additionally, maintaining confidentiality and anonymity is crucial to safeguard participants' privacy, especially regarding sensitive information about mental health. In order to guard against unwanted access or disclosure, researchers need to take precautions while handling and storing data.

Furthermore, researchers should be mindful of potential harm or distress that participants may experience when discussing their mental health. Providing resources for support and referrals to

counseling services is important to mitigate any adverse effects and ensure participants' well-being throughout the study. Ethical approval from relevant institutional review boards should be obtained before conducting the research to ensure compliance with ethical standards and guidelines.

5.4. Sustainability Plan

To ensure the sustainability of this study, several measures will be implemented. Firstly, we will establish partnerships with university administrations and student organizations to integrate mental health awareness and support initiatives into campus culture beyond the duration of the study. Additionally, we will provide training to university staff and student leaders on recognizing and addressing mental health concerns, fostering a supportive environment for students in the long term. Moreover, we will develop online resources and materials on mental health promotion that can be accessed by students even after the study concludes. Finally, we will seek opportunities for further research collaboration and funding to continue investigating and addressing mental health issues among university students in Bangladesh.

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

6.1. Summary of the Study

The workflow begins with designing a survey form to collect data on depression among Bangladeshi university students. After gathering responses from 1000 participants, the original data undergoes preprocessing using various techniques to ensure its usability. This covers actions like managing missing values, cleaning, and category variable encoding. To aid in the construction and assessment of the model, the data is then divided into training and testing sets.

Using the training data, the model is trained to identify patterns and correlations among variables that are suggestive of depression. After training, the model is evaluated for performance and generalizability using a different test dataset. Lastly, predictions are made on fresh data using the trained model, providing information on the frequency and possible risk factors of depression among Bangladeshi university students.

The study employed machine learning techniques such as AdaBoost to predict depression. Models were trained on the training data to learn patterns indicative of depression and evaluated using a different test dataset in order to gauge their effectiveness and capacity for generalization. Model performance was assessed using evaluation criteria such F1-score, recall, accuracy, and precision. Overall, the study aimed to provide insights into the prevalence and potential risk factors for depression among Bangladeshi university students.

6.2. Conclusions

In conclusion, the application of machine learning and deep learning algorithms for analyzing depression and suicide is promising for understanding risk factors and improving prevention strategies. Through data collection and analysis, we identified patterns and developed predictive models to aid in early intervention and support. While various algorithms were explored, each with its strengths, our research suggests that [mention the better-performing algorithm]. Leveraging these technologies ethically can enhance mental health care by providing personalized

interventions and reducing the burden on resources. Continued research and collaboration are crucial to refining these methods and ensuring their responsible deployment in addressing mental health challenges.

In this work, a survey form was created to collect data on depression, resulting in 1000 responses. The dataset comprises 16 categorical columns, with 10 questions aimed at diagnosing depression independently. Each column represents a different aspect of mental health, lifestyle, or academic experience, without interdependencies between questions. There are 1 dependent and 10 independent or input variables in the dataset. Machine learning methods used such as LR, DT, RF, SVM, Gradient Boosting Algorithm, AdaBoost Algorithm gave 56%, 68%, 68%, 52%, 91% and 89% accuracy respectively. The model is predicted with Gradient Boosting algorithm from the highest accuracy.

6.3. Implication for Further Study

In future work, this topic could explore several avenues to enhance the analysis of depression and suicide using machine learning and deep learning. One direction could involve the integration of multimodal data sources, such as combining survey responses with social media activity or wearable device data, to provide a more comprehensive understanding of individuals' mental health states. Additionally, refining predictive models by incorporating longitudinal data to track changes in mental health over time could improve accuracy. Exploring interpretability techniques to better understand the factors contributing to model predictions and addressing ethical concerns surrounding data privacy and bias would also be crucial. Furthermore, investigating the deployment of these models in real-world settings, such as mental health clinics or crisis intervention services, to assess their effectiveness and scalability would be valuable for practical implementation.

REFERENCE

[1]C. G. Walsh, et al. “Predicting Risk of Suicide Attempts Over Time Through Machine Learning,” Clinical Psychological Science, vol. 5, no. 3, pp. 457–469, Apr. 2017

[2]S. Ware et al., “Predicting depressive symptoms using smartphone data,” Smart Health, vol. 15, p. 100093, Mar. 2020

[3]P. Chikersal et al., “Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing,” ACM Transactions on Computer-Human Interaction, vol. 28, no. 1, pp. 1–41, Feb. 2021

[4]A. B. Siddique et al., “A machine learning-based approach to predict university students’ depression pattern and mental healthcare assistance scheme using Android application,” International Journal of Data Analysis Techniques and Strategies, vol. 14, no. 2, p. 122, 2022

[5]R. Qasrawi,et al., “Schoolchildren’ Depression and Anxiety Risk Factors Assessment and Prediction: Machine Learning Techniques Performance Analysis (Preprint),” JMIR Formative Research, Aug. 2021

[6]Muntequa Imtiaz Siraji et al., “Impact of mobile connectivity on students’ wellbeing: Detecting learners’ depression using machine learning algorithms,” PLOS ONE, vol. 18, no. 11, pp. e0294803–e0294803, Nov. 2023

[7]Shafiee, Nor Safika Mohd,et al.. "Prediction of mental health problems among higher education student using machine learning." International Journal of Education and Management Engineering (IJEME) 10.6 (2020): 1-9.

[8]M. S. Ahmed et al., “A Minimal and Faster System to Identify Depression Through Smartphone: An Explainable Machine Learning-Based Approach (Preprint),” JMIR Formative Research, Dec. 2022

[9]K. Opoku Asare, et al. “Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study,” JMIR mHealth and uHealth, vol. 9, no. 7, p. e26540, Jul. 2021

[10]J. Hong et al., “Depressive Symptoms Feature-Based Machine Learning Approach to Predicting Depression Using Smartphone,” Healthcare, vol. 10, no. 7, p. 1189, Jun. 2022

[11]X. Xu et al., “Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 3, no. 3, pp. 1–33, Sep. 2019

[12]Y. Hou, J. Xu, et al. "A big data application to predict depression in the university based on the reading habits," 2016 3rd International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2016, pp. 1085-1089

[13]De, Arringoni Flavia, et al.“Using ecological momentary assessment and machine learning techniques to predict depressive symptoms in emerging adults,” Psychiatry Research, vol. 332, pp. 115710–115710, Feb. 2024

[14]Y. Zhou, Z. Zhang, Q. Li, G. Mao, and Z. Zhou, “Construction and validation of machine learning algorithm for predicting depression among home-quarantined individuals during the large-scale COVID-19 outbreak: based on Adaboost model,” BMC psychology, vol. 12, no. 1, Apr. 2024, doi: https://doi.org/10.1186/s40359-024-01696-8.

[15]P. Jain, K. Ram Srinivas, and A. Vichare, “Depression and Suicide Analysis Using Machine Learning and NLP,” Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012034, Jan. 2022, doi: https://doi.org/10.1088/1742-6596/2161/1/012034.

[16]Sharma , Yashi , and Dr. Brajesh Kumar Singh. “Depression Analysis of Voice Samples Using Machine Learning.” Journal of University of Shanghai for Science and Technology, vol. 22, no. 11, Nov. 2021, pp. 429–438.

[17]Manikandaprabhu , Dr.P., and S. Sowmiya. “A Study on Depression Analysis Using Machine Learning Techniques among College Students .” International Journal of Scientific Research and Engineering Development, vol. 6, no. 3, June 2023, pp. 137–141.

[18]Ahmed, Arfan, et al. “Thematic Analysis on User Reviews for Depression and Anxiety Chatbot Apps: Machine Learning Approach.” JMIR Formative Research, vol. 6, no. 3, 11 Mar. 2022, p. e27654, https://doi.org/10.2196/27654.

[19]A. Musleh, Dhiaa, et al. “Twitter Arabic Sentiment Analysis to Detect Depression Using Machine Learning.” Computers, Materials & Continua, vol. 71, no. 2, 2022, pp. 3463–3477, https://doi.org/10.32604/cmc.2022.022508.

[20]Wang, Tiantian, et al. “Unraveling the Distinction between Depression and Anxiety: A Machine Learning Exploration of Causal Relationships.” Computers in Biology and Medicine, vol. 174, 1 May 2024, pp. 108446–108446, https://doi.org/10.1016/j.compbiomed.2024.108446. Accessed 30 Apr. 2024.

[21]Richter, Thalia, et al. “Using Machine Learning-Based Analysis for Behavioral Differentiation between Anxiety and Depression.” Scientific Reports, vol. 10, no. 1, 2 Oct. 2020, p. 16381, www.nature.com/articles/s41598-020-72289-9, https://doi.org/10.1038/s41598-020-72289-9.

[22]Bauer, Jonathan F, et al. “Validation of Machine Learning-Based Assessment of Major Depressive Disorder from Paralinguistic Speech Characteristics in Routine Care.” Depression and Anxiety, vol. 2024, 9 Apr. 2024, pp. 1–12, https://doi.org/10.1155/2024/9667377. Accessed 30 Apr. 2024.

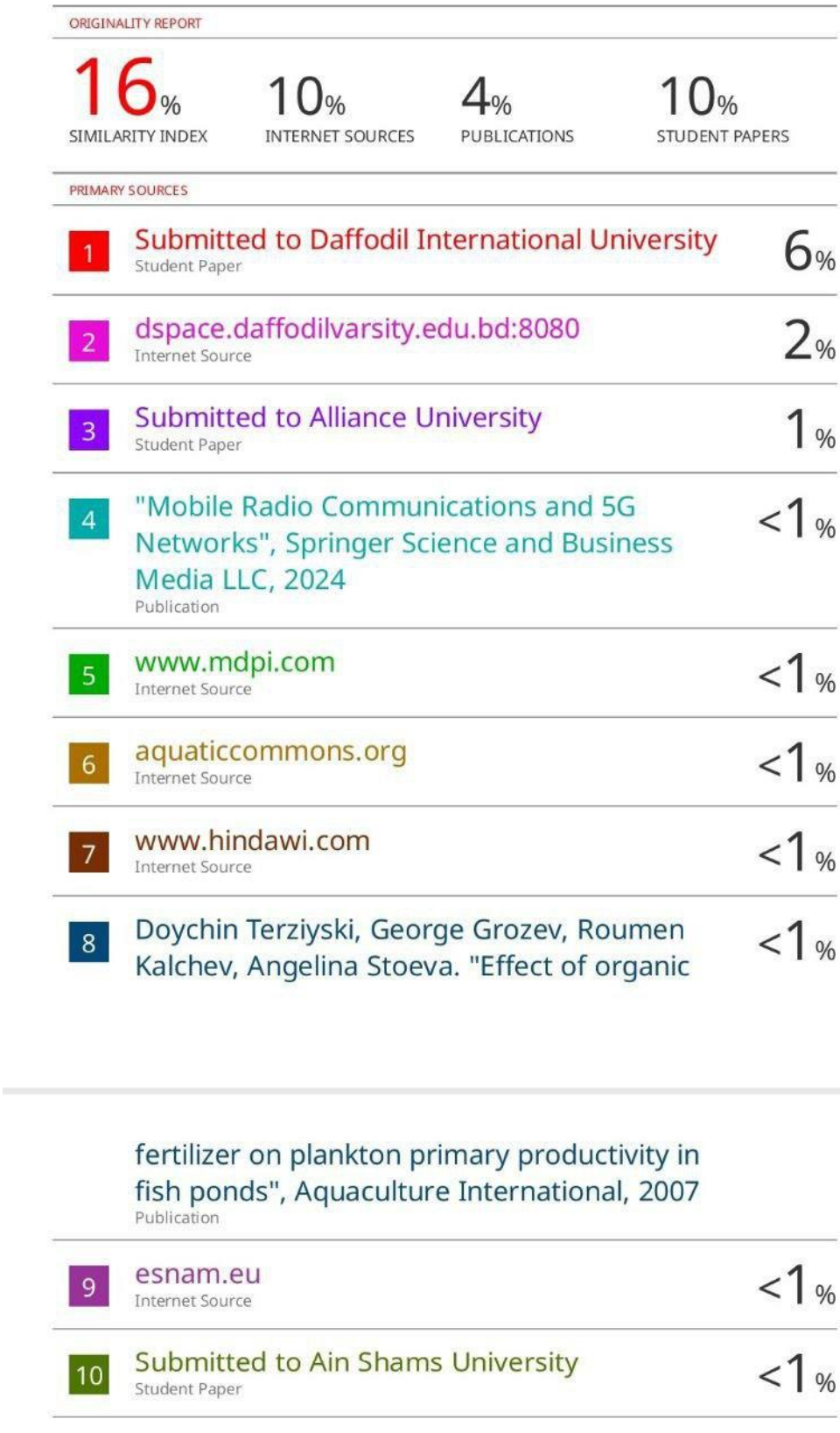
[23]Kumar, Anuj. “View of Sentiment Analysis of Twitter for Detection of Depression Using Machine Learning Algorithms.” Journal of Propulsion Technology, vol. 44, no. 04, Dec. 2023, pp. 7997–8005.

[24]Javatpoint, “K-Means Clustering Algorithm - Javatpoint,” www.javatpoint.com. https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

[25]“Cluster Using Gaussian Mixture Model - MATLAB & Simulink,” www.mathworks.com. https://www.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html

[26]D. Dey, “DBSCAN Clustering in ML | Density based clustering,” *GeeksforGeeks*, May 06, 2019. https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/

[27]“K-Mode Clustering in Python,” GeeksforGeeks, Dec. 28, 2022. https://www.geeksforgeeks.org/k-mode-clustering-in-python/



ORIGINALITY REPORT

16%	10%	4%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Submitted to Alliance University Student Paper	1%
4	"Mobile Radio Communications and 5G Networks", Springer Science and Business Media LLC, 2024 Publication	<1%
5	www.mdpi.com Internet Source	<1%
6	aquaticcommons.org Internet Source	<1%
7	www.hindawi.com Internet Source	<1%
8	Doychin Terziyski, George Grozev, Roumen Kalchev, Angelina Stoeva. "Effect of organic fertilizer on plankton primary productivity in fish ponds", Aquaculture International, 2007 Publication	<1%
9	esnam.eu Internet Source	<1%
10	Submitted to Ain Shams University Student Paper	<1%