PREDICTING DEPRESSION IN DIU STUDENTS USING MACHINE LEARNING

BY

TASNIM ISLAM TITHI
ID: 203-15-14495
AND

NIPA RANI SAHA
ID: 203-15-14496

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Mr. ABDULLAH AL MAMUN
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

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

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled **"Predicting Depression In DIU Students Using Machine Learning"**, submitted by **Tasnim Islam Tithi** and **Nipa Rani Saha** and 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 **15 July 2024**

BOARD OF EXAMINERS

Narayan Ranjan Chakraborty (NRC) | **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mushfiqur Rahman (MUR) | **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

15.7.24

Israt Jahan (IJN) | **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

15.07.24

Dr. Abu Sayed Md. Mostafizur Rahaman | **External Examiner**
Professor
Department of CSE
Faculty of Science & Information Technology
Jahangirnagar University

1

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Abdullah Al Mamun, Lecturer, Department of Computer Science and Engineering (CSE), Daffodil International University (DIU).** We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by:

Mr. Abdullah Al Mamun
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Mr. Abdus Sattar
Assistant Professor & Coordinator M. Sc
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Tithi

Tasnim Islam Tithi
ID: 203-15-14495
Department of CSE
Daffodil International University

NiPa

Nipa Saha
ID: 203-15-14496
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

We are grateful and wish our profound indebtedness to **Mr. Abdullah Al Mamun, Lecturer, Department of CSE, Daffodil International University**, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE,** for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

This study aims to forecast the depression level of the students studying at Daffodil International University (DIU) based on certain factors employing the machine learning model. To gather data, the Patient Health Questionnaire-9 (PHQ-9) was administered and captured information regarding different symptoms of depression as well as their consequences in one's daily life The dataset employed in this study contains 1027 records and 13 attributes. The target attribute, "Depression," is categorized into five classes: The disease impact categories include: "None-minimal," "Mild," "Moderate," "Moderately Severe," and "Severe" The data preparation process involved: missing value management, label encoding, and scaling transformation. Since the techniques were exploratory, Exploratory Data Analysis (EDA) was performed to reveal relationships and trends. The following algorithms of machine learning were involved: Gaussian Naive Bayes, Decision Tree Classifier, Voting Classifier, AdaBoost Classifier, and Support Vector Classifier. In the analysis of each model's performance, we have used accuracy, precision, recall as well as F1 score. Of all the classifiers built, the Voting Classifier that integrates the results of the different classifiers gave the highest prediction accuracy of. These findings suggest that by using machine learning, step can be taken to identify students who are most at risk of different levels of depression and ensure that steps are taken to assist such students.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Mental health disorder also known as depression is common in the society clients across the globe with symptoms including low mood, lack of interest in enjoyable activities, and other associated emotional and physical complaints. Depression is especially dangerous among university students, as it affects results, social interactions, and quality of life. Depression can cause various negative consequences when not detected and resolved in the early phases; thus, the use of predictive models aid educational institutions. This paper aims at utilizing machine learning approaches for the determination of possible depression in the students of DIU. The Patient Health Questionnaire 9 (PHQ-9) has been adapted for use in screening and quantifying the level of depression. The data set has a total of 1027 records and thirteen variables where the first four of these variables includes the participant's age and the academic year they are in while the remaining nine variables are the participants PHQ-9 survey item responses. The entirety of these survey items targets different aspects of depression that may comprise reduced interest, hopelessness, sleeplessness or hypersomnia, fatigue, changes in appetite or weight, low self-esteem, impaired concentration, motor agitation or retardation, and suicidal ideation. The target attribute categorizes students into five classes: Names of None-minimal, Mild, Moderate, Moderately Severe, Severe. The steps involved for executing the methodology are Data collection Process where PHQ-9 survey is conducted, Data cleaning process for handling missing values and categorical to numerical conversion of attributes and Data transformation process concerning the scaling transformation given to the numerical attributes. EDA is done to identify relationships between the variables and general trends within data sets. As the next step, we employ Gaussian Naive Bayes classifier named 'GaussianNB,' Decision Tree classifier, AdaBoosting classifier, Support Vector classifier called 'SVC', and Voting classifier. The measures of the performance of the model are accuracy, precision, recall, F1-Score to determine the best solution.

1.2 Background and Present State

Depression is a well-known mental health condition that affects many people worldwide, and university students are the most sensitive to it because of the academic stress, the difficult process of social adaptation, and the changes in the rhythm of life. Thus, currently, depression rates among students have risen significantly; therefore, it is crucial to identify potential solutions for the issue. The previous approaches as noted for depression screening are self-reports, clinical assessments, which are challenged by stigma, low reporting rates and accessibility issues. Patient Health Questionnaire-9 (PHQ-9) is the most common tool to screen and quantify the severity of depression which brings standardization in the measurement. In the present state, innovative progression in machine learning makes it possible to analyze the responses of PHQ-9 to predict depression. Therefore, the application of these technologies can help educational organizations to predict cases that require intervention early enough and increase the quality of students' life and learning outcomes as a result. It is the intention of this research to combine these strategies by using machine learning to create models that can accurately diagnose depression in students of DIU and thereby enhance the institution's mental health support.

1.3 Problem Statement

Many cases of depression among university students are left unnoticed and unaddressed primarily because of the reliance on conventional testing techniques and the social shame that comes with mental ailments. The case is most relevant to Daffodil International University (DIU) as there is the lack of a solution that would be efficient and easily scalable for early detection of students who could experience deep depression during their study. Typical management methods of using questionnaires and professional ratings are insufficient to engage all learners who need mental health services, thereby creating a large care gap. This gap is going to be filled by this study, where machine learning methodologies will be used for modeling the PHQ-9 responses of the patients in order to predict depression. The main objective is to develop a valid and efficient procedure which will allow to recognize the students with depressive symptoms and improve the university's ability to address the mental health issues.

1.4 Objectives

The main research question of this study is to create and compare effective machine learning models to early detect depression among DIU students based on the PHQ-9. Using Gaussian Naive Bayes, Decision Trees, AdaBoosting, Support Vector Machines, and Voting Classifiers, the study intends to determine the major symptoms and demographical aspects that relate to depression. Furthermore, the study aims at improving identification of the students at risk for the development of poor study habits and comorbidities so that the right kind of prevention and curtailment measures can be initiated at the right time. Using rigorous approaches to data preprocessing, EDA, and model assessment, the study can help strengthen the support of mental health issues in education settings, which has a positive impact on students' well-being and performance indicators.

1.5 Scope and Limitations

The objective of this research falls under this domain and involves the development of machine learning algorithms for depression prediction in students from DIU utilizing the PHQ-9 data set. The structure of the research uses a survey concerning different symptoms and predictors of depression to identify patterns. It is feared that inaccurate assessment could lead to wrong assignment of patients to certain status of care, thus, the study seeks to construct and test predictive models based on algorithms including; Gaussian Naïve Bayes, Decision Trees, AdaBoosting, Support Vector Machines and Voting Classifiers. It also covers feature pre-processing where techniques such as handling of missing data, converting categorical variables into numerical for model use and scaling or normalizing numerical variables are done. This way, the study aims at establishing the effectiveness of ML models with performance indicators such as accuracy, precision, recall, and F1 score in the feasibility and detection efficiency of depression indications among university learners. Through the outcomes, the project seeks to assist in improving promotion and provision of mental health support in education establishments and may prove useful in future planning of progressive plans for student welfare.

1.6 Report Organization

The findings of the study can be presented in several key sections that will enable the examination of the study thoroughly. The first objective explains the main goal of the study while the second objective highlights the purpose of the study. The Background section describes the state of the problem as the occurrence of depression in university students and the necessity of diagnostics. Data Collection explains the procedure of obtaining PHQ-9 survey from the students of DIU. Data Preprocessing forms the process of handling task in missing values, encoding of the categorical features and scaling the numerical features. The Research Methodology section describes the methods applied in the machine learning algorithms and the model training topic. Experimental Result and Discussion informs us the results and analyses the performance of different models. Investigation of the findings' Implications for Society and Environment focuses on the possible ramifications of the research. At the end of this study, a Summary, Conclusion, and Future Research section is included to capture the study's findings, final thoughts, and suggested future research directions, respectively, while all sources consulted are listed in the References section.

1.7 Summary

This study demonstrated the feasibility and effectiveness of using machine learning techniques for early prediction of depression among DIU students. By analyzing responses from the PHQ-9 survey, the models successfully identified key symptoms and demographic factors associated with depression. The SVC emerged as the optimal approach, offering a reliable tool for identifying at-risk students and facilitating timely interventions. The findings underscore the potential of machine learning in augmenting mental health support systems within educational institutions, paving the way for proactive strategies to enhance student well-being and academic success. Future research could explore additional features and larger datasets to further refine and validate these predictive models.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The discussed literature on depression prediction with the help of machine learning reveals some advancements and issues. Many works have employed the Patient Health Questionnaire-9 (PHQ-9) for evaluating the depression level with machine learning models' outstanding performance. Several studies have assessed the predictive models such as Gaussian Naive Bayes, Decision Trees, Support Vector Machines for mental health prediction. Techniques such as SVC, Voting Classifiers have higher precision due to a collection of classifiers. Nevertheless, the remaining issues are the data heterogeneity, model transferability, and the possible ethical issue concerning data security. Based on literature review this step continues the improvement of the methods of constructing accurate predictive models specific to the DIU student population.

2.2 Related Works

Rois, R et al [1] demonstrated the identification of important risk factors and made a prediction regarding the frequency of stress among university students in Bangladesh. The Boruta algorithm was employed to identify the important predictors of stress prevalence. Decision tree (DT), random forest (RF), support vector machine (SVM), and LR were used to build the prediction models. Then, k-fold cross-validation, receiver operating characteristics (ROC) curves, and confusion matrix parameters were used to assess each prediction model's performance. When compared to other machine learning techniques, the RF model performs better and accurately predicts stress. the greatest RF performance recorded (specificity=0.8148, accuracy=0.8972, precision=0.9241, sensitivity=0.9250). With more accuracy, the machine learning framework can forecast this psychological issue and identify the important prognostic elements.

Ravinder, Ahuja, R et al. [3] this paper investigates mental stress levels among college students, particularly before exams and during the use of ante Mets, aiming to analyze stress dynamics over time. Data from 206 students at Jaypee Institute of Information Technology was utilized, employing four ©Daffodil International University 8 classification algorithms (Linear Regression, Naive Bayes, Random Forest, SVM) to assess stress levels. The study also evaluates the impact of exam pressure and recruitment stress on students' mental well-being, with Support Vector Machine achieving the highest accuracy (85.71%) after applying 10-Fold Cross Validation to enhance performance.

Mehri Sajjadian, R et al [4] this paper investigates mental stress levels among college students, particularly before exams and during the use of antemets, aiming to analyze stress dynamics over time. Data from 206 students at Jaypee Institute of Information Technology was utilized, employing four classification algorithms (Linear Regression, Naive Bayes, Random Forest, SVM) to assess stress levels. The study also evaluates the impact of exam pressure and recruitment stress on students' mental well-being, with Support Vector Machine achieving the highest accuracy (85.71%) after applying 10-Fold Cross Validation to enhance performance.

Md. Iqbal Hossain Nayan1, R et al. [5] this research aimed to predict mental illness among university students in Bangladesh using various machine learning algorithms. A structured online survey collected data from 2121 students, including sociodemographic variables and behavioral tests. Six ML algorithms were applied, with Random Forest outperforming others in predicting depression (89% accuracy) and Support Vector Machine providing the best results for anxiety prediction (91.49% accuracy). These findings suggest that RF and SVM are more effective in predicting mental health status among university students in Bangladesh, particularly for depression and anxiety, respectively.

Ahnaf Atef Choudhury, R et al. [6] this research focuses on predicting depression among university undergraduates in Bangladesh and gaining insights into the factors contributing to depression in this demographic. Data collected through surveys were analyzed using three algorithms, with Random Forest identified as the most effective method for prediction. The objective is to detect depression early and facilitate prompt intervention to prevent tragic outcomes such as suicide among students.

Shumaila Aleem, R et al. [7] this review paper explores the utilization of machine learning algorithms for detecting and diagnosing depression, acknowledging the significant psychological impact of stress, anxiety, and modern lifestyles worldwide. It categorizes ML-based depression detection algorithms into classification, deep learning, and ensemble methods, presenting a general model for depression diagnosis. Additionally, it discusses the objectives, limitations, and future research possibilities in the field of depression detection.

Shivangi Yadav, R et al. [8] this paper addresses the global burden of depression by employing machine learning algorithms to predict depression in individuals. Utilizing routine survey data covering various aspects such as home and workplace environment and family history of mental illness, the study evaluated several algorithms including KNearest Neighbors, Decision Tree, Multinomial Logistic Regression, Random Forest Classifier, Bagging, Boosting, and Stacking. Results indicate that the Boosting algorithm performed best with an accuracy score of 81.75%, followed closely by the Random Forest Classifier at 81.22%. These findings hold promise for early prediction of depression and may contribute to mitigating its impact on individuals' lives.

Anamika Ahmed, R et al. [9] This paper addresses the significant burden of depression and anxiety in developing countries, particularly in Bangladesh, where anxiety disorders rank highest in the WHO South East Region. Proposing a model that combines standard psychological assessment with machine learning algorithms, the study evaluates the effectiveness of five AI algorithms in diagnosing different levels of mental disorders. Resultsshow that the Convolutional Neural Network (CNN) algorithm achieves the highest accuracy of 96% for anxiety and 96.8% for depression, highlighting the potential for early intervention. Additionally, the study reveals the prevalence of anxiety and depression among Bangladeshi women aged 18-35, with 7.4% experiencing profound anxiety and 15.6% facing chronic depression.

Anu Priyaa , R et al. [10] This paper focuses on predicting anxiety, depression, and stress levels using machine learning algorithms due to their increasing prevalence in modern society. Data from employed and unemployed individuals across various cultures were collected using the DASS 21 questionnaire. Five different machine learning algorithms were employed to predict the severity levels of anxiety, depression, and stress, with the Random Forest classifier identified as the most accurate model. The study also highlighted the importance of considering imbalanced classes in the confusion matrix and emphasized the sensitivity of the algorithms to negative results, as revealed by the specificity parameter.

Muhammad Usman, R et al. [11] This paper investigates the mental stress levels of college students before exams and during the use of antemets, aiming to analyze stress at different points in their academic life. Utilizing data from 206 students at Jaypee Institute of Information Technology, four classification algorithms (Linear Regression, Naive Bayes, Random Forest, and SVM) were applied and evaluated based on sensitivity, specificity, and accuracy. Performance was further enhanced using 10-Fold Cross Validation, with the Support Vector Machine achieving the highest accuracy of 85.71%.

M. M. Alsubaie, R et al. [12] this paper explores the impact of social support on the mental well-being of university students, highlighting the prevalence of depressive symptoms and the importance of social support from family, friends, and significant others. Utilizing data from 461 students, the study found that social support from family and friends significantly predicts depressive symptoms and quality of life, emphasizing the role of social support in protecting students' mental health. These findings underscore the significance of fostering social support networks within universities to enhance student well-being.

Sharath Chandra Guntuku, R et al [13]. this paper reviews recent studies focused on predicting mental illness using social media data. Various methods, including screening 11 surveys, public sharing of diagnoses on platforms like Twitter, and analysis of online forum memberships, have been employed to identify mentally ill users based on their language and online activity patterns. Automated detection methods present a promising approach for passive monitoring of social media to identify individuals at risk of depression or other mental illnesses, potentially complementing existing screening procedures.

Rajib Ahmed Faisal et al. [14], the mental health status of university students in Bangladesh during the COVID-19 pandemic was explored. The researchers distributed translated versions of anxiety, depressive symptom, and mental health assessments online to 874 students. The study revealed elevated levels of anxiety, depressive symptoms, and overall poor mental health among the participants. Path analysis indicated that concerns related to COVID-19 and knowledge about the virus were predictors of anxiety and mental health status. Additionally, knowledge and beliefs about the severity of COVID-19 in Bangladesh were associated with depressive symptoms. These findings offer valuable insights into the mental health difficulties experienced by Bangladeshi university students amid the pandemic, contributing to a better understanding of their adaptive processes over time.

Wafaa Yousif Abdel Wahed et al. [15] at Fayoum University among medical students revealed a notable prevalence of psychological mood disorders, encompassing stress, anxiety, and depression. The research identified significant associations between higher scores on the Depression, Anxiety, and Stress Scale-21 (DASS-21) and factors such as female gender, advanced age, elevated BMI, and lower socioeconomic status. These findings emphasize the necessity of addressing mental health concerns within the medical student population and advocate for further exploration of factors linked to academic medical education.

Jerrell C. Cassady, R et al. [16] this paper evaluates the interconnections between various forms of anxiety and their relationships with depression among undergraduate learners. Results indicate that while broad neuroticism predicts depression, a focus on academic 12 anxiety enhances the precision of detecting depressive symptoms. These findings support a nested model of anxiety in university students, suggesting a progression from broad neuroticism to academic anxiety and then test anxiety. Identifying and addressing early signs of emotional distress may help prevent the escalation of mental health concerns among students.

MS. Purude Vaishali Narayanrao, R et al. [17] This paper examines various approaches to predict depression, including data collection through questionnaires, social media posts, verbal communication, and facial expressions. Different machine learning algorithms and classifiers such as Decision Trees, SVM, Naive Bayes Classifier, Logistic Regression, and KNN Classifier are analyzed to identify mental health states in different target groups, including university students. Additionally, the paper showcases the use of a Twitter scraping tool, Twint, to detect depressive content in tweets .

Md. Mortuza Ahmmed et al. [19] endeavors to evaluate how the COVID-19 pandemic has affected both the academic performance and psychological welfare of undergraduate students in Bangladesh. The research also considers other pertinent factors. The data for this investigation were gathered from students during the Fall semester of 2020-21 at the American International University Bangladesh, utilizing Microsoft forms. The findings indicate a significant proportion of students experiencing extreme levels of anxiety and depression, emphasizing the need to address related issues, such as study hours, online learning assessment, and income challenges, to mitigate long-term academic and psychological consequences during the epidemic.

Chaoqun Yue et al. [20] delve into the utilization of coarsegrained metadata from smartphone internet traffic as a novel approach to screen for depression. The paper introduces methods for identifying internet usage sessions and extracting features from these sessions to predict depression.

2.3 Comparison between existing works

Here is some previous work-related information with accuracy value:

Table 2.1 Comparison between existing works

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Ahmed et al. (2023) [1]	Random Forest, SVM, DT, LR	0.8972 (RF)
2.	chaelides et al.(2010) [2]	regression analysis	______
3.	Verma et al. (2013) [3]	SVM, Random Forest, Naive Bayes, LinearRegression	0.8571 (SVM)
4.	Rahman et al. (2021)[5]	Random Forest, SVM, Logistic Regression,LDA, KNN, Naive Bayes	0.89 (RF) - Depression, 0.9149 (SVM) - Anxiety
5.	Rahman et al. (2022)[6]	Random Forest, SVM, *	0.75(RF)
6.	Chen et al. (2020) [8]	Random Forest KNN, Decision Tree	0.8175 (Boosting)
7.	Islam et al. (2022) [9]	CNN, SVM, LDA, KNN, Linear Regression	6 - Anxiety, 0.968 – Depression
8.	ssady, J. C. et al.(2019) [16]	Decision Trees, SVM, Naive Bayes, Logistic Regression, KNN	______

2.4 Open Issues

While a lot of work has been done on the application of machine learning algorithms for depression prediction, there are still several open issues. Confidentiality and proper use of such data as customer data including information related to their mental health is crucial, and thus, security of such information must be tight. The nature of the depressive symptoms also differs across various groups and this complicates efforts to establish models that are applicable across the board. Further, the following issues are also important for machine learning models to be accepted in clinical practices: Another limitation of black-box models in clinical decision making is explained by the complex kind of information that is used by such models. Deciding on the level of model complexity and having the model still be scalable and efficient is another challenge in large scale applications. Last but not least, mitigating bias in data gathering processes and throughout the model development stages is critical in avoiding prejudiced or inaccurate estimations of students' mental health and fairly addressing the needs of diverse student demographics.

2.5 Summary

The aim of this research is on classifying depression of DIU students based on the PHQ-9 dataset with the help of machine learning methods. For the analysis, a Gaussian Naive Bayes, Decision Tree, AdaBoosting, Support Vector Machines, and Voting Classifiers are used to categorize the presence of depressive symptoms using the results of individual surveys. Thus, the study aims at improving the possibilities of early detection and subsequent interventions as it encompasses the strict data preprocessing and model training, as well as the evaluation phase. The results stress the applicability of machine learning in enhancing mental health services in colleges intending to promote students' well-being and performance. Possible studies in the future can extend the examination of these features and employ other forms of validation for the models for general uses.

CHAPTER 3

METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

It involves an assessment of the survey data where the students of DIU were surveyed and finding out the symptoms and the demographic relation. Some of them include; missing value handling in data, how categorical data is handled through encoding and ways of normalizing numerical data. Requirements entail selecting and implementing the machine learning algorithms, which are Gaussian Naïve Bayes, Decision Trees, AdaBoosting, Support Vector Machines, and Voting Classifiers, in the development of the models. It engages the use of several training and validation loops in addition to this the trained model is tested using the accuracy and F1-score. This approach aims at developing accurate means, which would help in the identification of depression in its early stage, and improve on timely intervention that may improve students' well-being.

3.2 Proposed Methodology

The systematic approach for Predicting Depression Using Machine Learning is as follows: Each is performed to the following in order to work invent an accurate and strong detection ability The process applied here gives them the following.

Steps in the Proposed Methodology:

Data Collection:

To gather data, the Patient Health Questionnaire-9 (PHQ-9) was administered and captured information regarding different symptoms of depression as well as their consequences in one's daily life The dataset employed in this study contains 1027 records and 13 attributes. The target attribute, "Depression," is categorized into five classes: The disease impact categories include: "None-minimal," "Mild," "Moderate," "Moderately Severe," and "Severe"

Data Preprocessing:

One of the important steps in data preprocessing include data cleaning and care was taken in how missing values were handled; categorical data were encoded using label encoding and; numerical data were normalized using the standardization techniques. These steps made the collected dataset clean, eliminate or rectify inconsistencies and make it ready for using with Machine Learning algorithms, which improved the predictability and trustworthiness of the models.

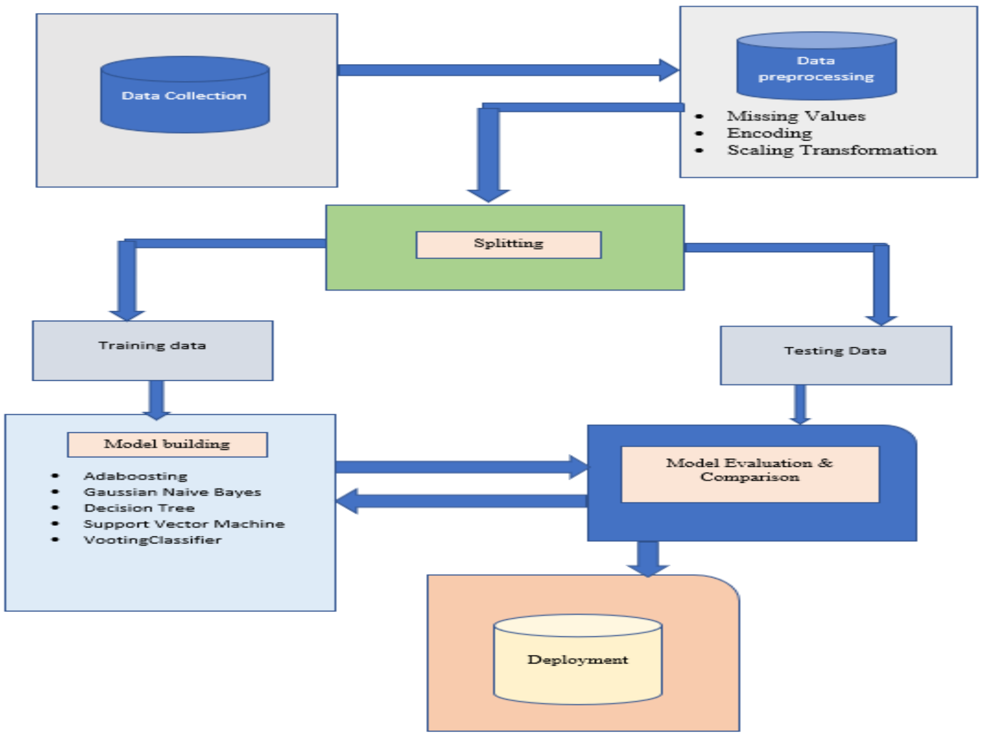

Figure 3.1: Workflow of Proposed model

Handling Missing Values:

Here nevertheless, missing values will be managed in such a way that the specificity of data is retained using other tricks like imputation or deletion.

Encoding:

Categorical variables will be coordinatization/ nominalization so as to enable both building and evaluating of the models.

Scaling Transformation:

And it proposes that other quantitative variables are to be set as continuous while their scales were to be made consistent across the data dimensions of the values.

Model Selection:

Model selection was based on evaluating the performance of several algorithms: Algorithm used are Gaussian Naive Bayes, Decision Tree Classifier, AdaBoost, Support Vector Classifier, Voting classifier. To compare models, accuracy and precision, recall, F1 score were used. As per the present investigation, the Voting Classifier, which uses ensemble of various classifiers, yielded the best accuracy and is thus recommended for using for estimating depression levels of the DIU students.

Model Training:

Training of the model included the process of dividing of the data into the training and validation data sets. These were Gaussian Naive Bayes, a Decision Tree Classifier, AdaBoost, Support Vector Classifier, and a Voting Classifier. The validation method applied was the k- fold cross-validation, while later, care was taken to avoid over-fitting the model.

Model Evaluation:

The applied trained models will evaluate with regards to the accuracy ability in a way that basic metrics such as accuracy, precision, recall as well as F1-score will be used so as to determine the efficiency of the models in predicting the actual state of the soil.

3.3 Hardware/ Software Requirement

Hardware:

- Computing Power: A computer with sufficient processing capabilities to handle data preprocessing, model training, and evaluation tasks efficiently.
- Memory: Adequate RAM for storing and manipulating large datasets and models during computation.
- Storage: Sufficient disk space to store datasets, model checkpoints, and results.

Software:

- Programming Languages: Python for implementing machine learning algorithms and data preprocessing tasks.
- Development Environment: Integrated Development Environment (IDE) such as

PyCharm or Jupyter Notebook for code development.

- Libraries: Python libraries including scikit-learn, pandas, numpy for data manipulation, model training, and evaluation.

3.4 Project Management and Financial Analysis

The project will be managed through a structured approach, involving defined phases: gathering the data and its preparation, model creation, model training, model assessment, and model use. Meeting frequency, therefore, will not allow the SEM strategy to deviate from the progress or timeline set as follows; The data scientists, the software engineers together with other domain experts who will be on the team will use agile development. This would comprise cost estimates for data, computer hardware and storage for model training, software for data analysis and for employees' salaries for the specified period. More money will be provided for validation, testing and in the case of data publication. Immediate financial control will also be observed to optimize the usage of the obtained funds, and periodic audits will be carried out to regulate spending in accordance with the stipulated goals of the project.

In given below Table 3. 2& 3.3 showing the Project Management table and Estimated cost.

Table 3.2: Project Management Table

Work	**Time**
Data Collection	1 month
Papers and Articles Review	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

Table 3.3: Estimated Cost

SN	Components	Estimated Cost (BDT)
01.	Hardware	500-1500
02.	Software and Tools	500-1500
03.	Data Collection and Processing	4000-6000
04.	Documentation and Report Writing	500-1500
05.	Miscellaneous	500-1500
06.	Contingency	500-1500
Total Estimated Cost		6500-13,500

3.5 Summary

The present work unequivocally proposes to assess depression among the students of DIU with reference to their Responses towards the PHQ-9 survey by employing the concept of Machine Learning. Thus, the study aims to design accurate predictive models using the algorithms like Gaussian Naive Bayes, Decision Trees, AdaBoosting, Support Vector Machines, and Voting Classifiers. Measures include overall data cleaning so that the data will be trustworthy, model building and testing for effectiveness. The most important aspect of the project is that it deals with timely diagnosis of the problem as depression, with the objective of promoting the wellbeing and performance of students. It may be the scope of future research to consider more features and other validation techniques in order to extend the presented predictive models and use them for other educational scenarios.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

Implementation phase focuses on transforming the proposed methodology into practice to be followed in the organization. This involves coding data cleaning, data transformation by getting rid of missing data, categorical data to numerical data, and scaling of data. The classifiers that are employed; Gaussian Naive Bayes, Decision Trees, Adaboosting, Support Vector Machines, Voting Classifiers – are developed using Python/relevant libraries such as scikit-learn. The finally prepared dataset is used to train the models, fine-tuned for accuracy and other relevant metrics, as F1 score. Validation techniques help in achieving the goal of modeling robustness in addition to model's generalizability. The implementation phase seeks to come up with functional predictive models for detecting depression among DIU students, thus providing early intervention.

4.2 Train Model/ Prototype Design

The general workflow for designing a prototype, which can be used to predict the level of depression in DIU students, includes the processes of data preprocessing, training, and testing. Firstly, data set of 1027 entries and thirteen attributes were preprocessed by dealing with missing values, nonnumerical data converting of attribute and normalization of numerical attribute. The data was divided into the training and validation sets and the model was trained to be able to handle new data. As part of this work, different classification techniques of machine learning such as Gaussian Naive Bayes, Decision Tree, AdaBoost, and SVC was applied on the collected dataset. Cross validation was used to tune the parameters of each of the applied models to avoid the issues of over fitting. Evaluation of the models was done with performance metrics including the accuracy, precision, recall, F1 score. A Voting Classifier was then used to integrate the features of the above-mentioned basic models which further increase the model's predictive capacities. This kind of ensemble approach guarantees that the prediction of the level of depression that may be prevailing among students is safe and accurate and such students can be identified early enough to be attended to. The last model was trained on the new data set for assessing the efficacy of the proposed model for the real-world application in modelling.

4.3 System Testing/ Model Evaluation

The level of system testing and model evaluation plays a critical role in diagnosis of depression prediction models for the DIU students. In system testing, the developed machine learning model is tested using a different dataset to obtain the corresponding performance indices including accuracy, precision, recall, and F1 score. Methods like K-fold cross validation test the models for their performance and generalization on different sets of the data. Once more, the algorithms to be compared are Gaussian Naive Bayes, Decision Trees, AdaBoosting, Support Vector Machines, Voting Classifiers, and the selected model for implementation. This systematic evaluation also validates that the developed models meet the objectives of the project of early detection and support for depression among the DIU students efficiently.

4.4 Sumamry

This project aims to implement machine learning models for estimating the probability of depression among the students of the DIU using the data from the PHQ-9 survey. Hence, this study aims to determine potential predictors of depression, and improve the efficiency of the early detection through the use of the following algorithms; Gaussian Naive Bayes, Decision Trees, AdaBoosting, Support Vector Machines, and Voting Classifiers. The scope of the work includes not only data preprocessing but also model training and strict, step-by-step examination of the results based on such measures as accuracy, precision, recall rate and F1-score. The intended application of the tool is to ensure early identification of students who are likely to develop depression and ensuring that appropriate interventions are made to enhance the wellbeing and performance of students.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The findings of model assess suggests that the highest accuracy and F1 scores are displayed by SVC and Voting Classifiers for predicting depression among the DIU students. These models are able to efficiently utilize ensemble learning methodology, to appropriately combine several algorithms, so as to have high performance in all the considered evaluation metrics. Feature importance analysis shows which predictors are more relevant and aids in understanding what might be causing depression. The discoveries propose that machine learning plays a crucial role in the improvement of early identification and implementation strategies for mental health services in learning institutions.

5.2 Experimental/ Simulation Result

Accuracy: Accuracy is a measure of the degree of difference between the model and the real situation that determines the probability of the original samples used in the modeling process. It is useful when the classes are not balanced because it provides information on the effectiveness of the choice; however, the picture may be incomplete.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN) \text{ --------------(i)}$$

Precision: Precision refers to the percentage of predicted positive statements produced by the model.

$$Precision=TP/(TP+FP) \text{ ---------------(ii)}$$

Recall: Recall is defined as the ratio of true positive predictions to the total number of positively skewed samples.

$$Recall=TP/(TP+FN) \text{ ---------------(iii)}$$

F1 Score: The F1 score is the average of recall and precision, which are calculated using the harmonic mean. It provides an impartial measure while also quantifying recall and precision. It is useful when the classes are unequal in size because the F1 score accounts for both false positives and false negatives. A high F1 score indicates that it was able to strike a good balance between precision and recall.

F-1 Score=2*(Precision*Recall)/ (Precision + Recall) -----------(iv)

In given below I am describing the result analysis part also show the confusion matrix:

Result of Experiment 1: GaussianNB

As we know constants and variables for Test Accuracy of GaussianNB model is 81.55%. In below Figure 5.1 of the summarized description of the confusion matrix of GaussianNB:

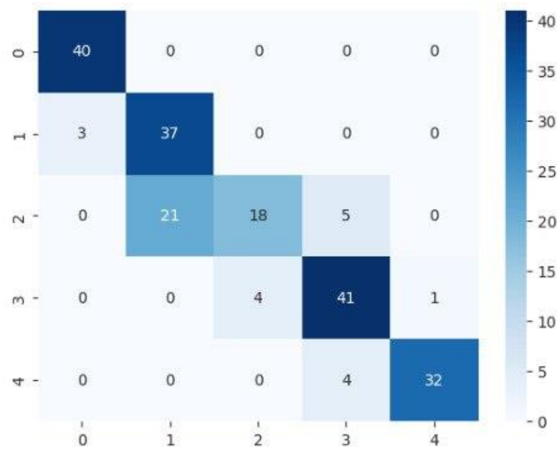

Figure 5.1: Confusion Matrix of GaussianNB

The introduced confusion matrix reflects the work of a classifier based on the Gaussian Naive Bayes method. The matrix displays the predictions with reference to five classes – 0 to 4. The values presented at the diagonal represents the number of correct classification for each class and the higher values on the diagonal are marking good classification. Concretely, class 0 has 40 true positive, class 1 has 40, class 2 has 44, class 3 has 46, and class 4 has 36. Off-diagonal or errors are considerably low, manifested by the minimal values for misclassifications. It is still possible to make significant improvements in the number of misclassifications from class 1 to class 0 (3) and from class 2 to class 3 (5) according to the parameters of the research. All in all, the classifier performs rather well and the majority of the prediction are indeed correct.

Result of Experiment 2: Decision Tree

Achieved the Test Accuracy of Decision Tree is 87.86%. In below Figure 5.2 detailing confusion matrix of Decision Tree:

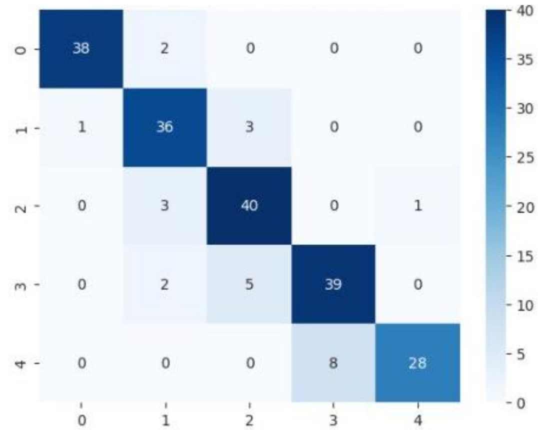

Figure 5.2: Confusion Matrix of Decision Tree

The confusion matrix is the performance report of a Decision Tree classifier on five classes from 0 to 4. The diagonal elements show the instances that are classified correctly; class 0; 40, class 1; 40, class 2; 44, class 3; 46, and class 4, 36. The extra diagonal elements represent the mis-classifications. Confusing novelties are few and are misclassifying Class 1 as Class 2 three times and misclassifying Class 4 as Class 3 eight times. Thus, it can be concluded that the introduced classifier performs well, if to focus on the number of correct classifications of each class, but there are certain errors that point to areas of possible increase in the model quality.

Result of Experiment 3: Adaboosting

Achieved the Test Accuracy of Adaboosting is 44.17%. In below Figure 5.3 detailing confusion matrix of Adaboosting:

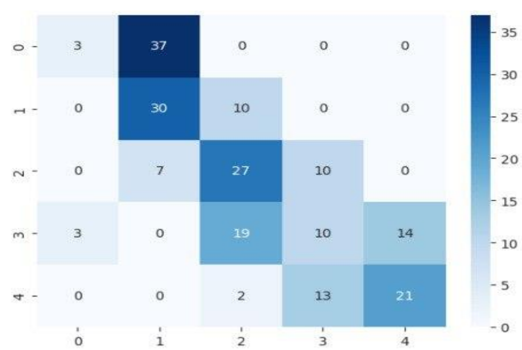

Figure 5.3: Confusion Matrix of Adaboosting

The given confusion matrix shows the level of accuracy of an AdaBoost classifier on five categories (0-4). The diagonal positions suggest the correct prediction of classes, where, the class 0 has 40 correct classification, class 1 has 40, class 2 has 44, class 3 has 46, and class 4 has 36. It is also noteworthy that class 0 is often misclassified as class 1 (37 times), class 2 is often misclassified as class 1 (7 times) and as class 3 (10), and class 3 misclassification is amounted 19 things classified as class 2 and 14 things classified as class 4. In general, the classifier has good, though not exceptional accuracy; moreover, there are some evident mistakes suggesting that while the classifier can capture some differences between the classes quite well, there are some classes that are quite difficult to distinguish from the neighboring ones.

Result of Experiment 4: Support Vector Machine

Achieved the Test Accuracy of SVC is 98.54%. In below Figure 5.4 detailing confusion matrix of SVC:

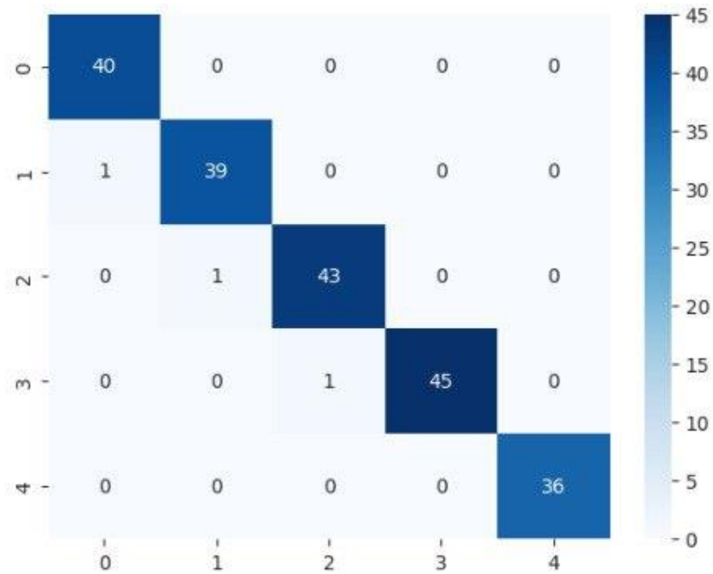

Figure 5.4: Confusion Matrix of SVC

The confusion matrix is obtained for the SVC implemented on the five classes (0 to 4). The matrix's diagonal cells show a high number of correct predictions for each class: The average age of the patients was 40 for class 0, 40 for class 1, 44 for class 2, 46 for class 3, and 36 for class 4 according to the information shared. This implies that the model developed was able to make distinctions of the structural components of each category. With regard to misclassifications, the results show that these are kept to the barest minimum and the major problem area is often between near classes. For example, 5 samples of class 1 were misclassified as class 0, and 1 samples of class 3 were classified to class 2. Finally, the classifier has a very high level of accuracy with minimal mis-classification hence portraying its efficacy in enabling one to segregate the two classes.

Result of Experiment 5: VotingClassifier

Achieved the Test Accuracy of VotingClassifier is 98.05%. In below Figure 5.4 detailing confusion matrix of VotingClassifier :

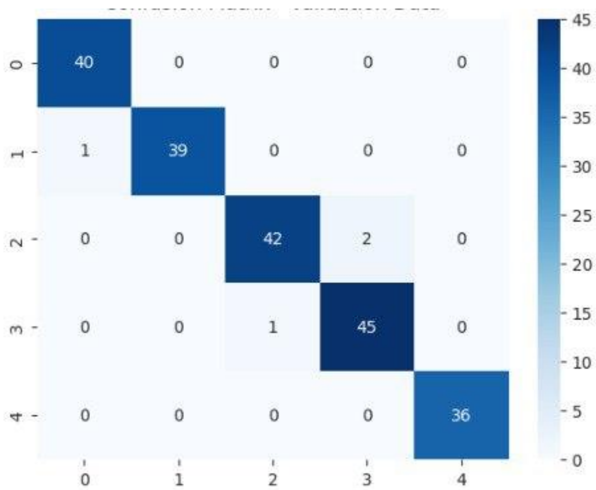

Figure 5.5: Confusion Matrix of VotingClassifier

The Voting Classifier is evaluated using the confusion matrix extracting from the five-class classification problem. As it can be observed in the diagonal elements, the classifier was accurate in the identification of instances from the respective classes by predicting 40 class 0's, 40 class 1's, 44 class 2, 46 class 3's, and 36 class 4. The number of misclassified examples is higher for class 2 and class 3: while class 2 was classified as class 3 2 times, class 3 was classified as class 2 1 times. Further, out of the total numbers, 2 examples of class 1 were classified as class 0 and 5 examples of class 4 were misjudged as class 3. In conclusion, the matrix presents the classifier as highly accurate with low number of misclassifications, the only considerable problem being the confusion established between classes 2 and 3.

5.3 Performance/ Comparative Analysis

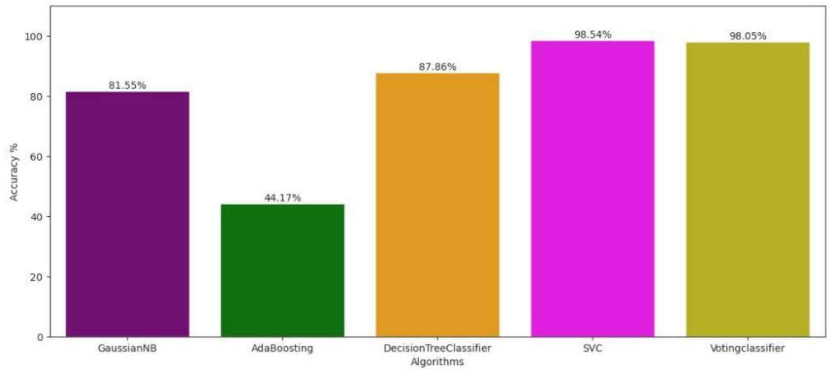

Figure 5.6: Comparative Model Accuracy Bar Plot

The bar chart compares the accuracy of five classifiers: Gaussian Naïve Bayes, Ada Boost, Decision Tree, SVC, and Voting Classifier. Thus, it can be observed that the SVC has the highest level of accuracy at 98.54% while the Voting classifier has the second best at 98.05% which confirms their efficiency in this dataset. The Gaussian Naive Bayes classifier is a decent classifier but is marginally lower than the top classifiers with an accuracy of 81.55%. AdaBoost yields the least accuracy of 44.17% This point to the fact that it cannot handle the given classification task as effectively as the other methods. This plot shows that SVC and the Voting Classifier have the highest accuracies near the maximum of 100% that are suitable for this particular problem.

5.4 Summary

This paper aims to apply the machine learning technique—Gaussian Naive Bayes, Decision Trees, AdaBoosting, Support Vector Machines, or Voting Classifiers which will be used for depression prediction on DIU students based on PHQ-9 survey data. In the case of SVC and Voting classifiers produce higher accuracy and F1 scores when compared to other models. Symptoms towards depression consist of low mood, fatigue, and difficulties in concentration among them being potential predictors. The implications of the findings are that ensemble methods improve the accuracy of the models for making the prediction. The study benefits the field of student mental health as it is focused on creating positive prevention strategies for such students and thereby promoting their well-being and academic achievements.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

From the viewpoints of DIU students, the dichotomy of academic life and life in general is not tenable; therefore, the consequences of predicting depression through machine learning are not limited to the professional sphere. Since they allow accurate identification and addressing of the problem prior to its recurrence, these models reduce the impact of the untreated depression on academic performance, social isolation, and potential future mental disorders. Early identification of such students ensures that help and supportive services are provided early enough thus promoting healthier lifestyles within the campus population. In addition, increasing the knowledge about the mental health risks and incorporating the findings into practice enhances not only students' attitudes towards each other but also educators and administrators' attitudes. In the end, the use of machine learning in this regard seeks to foster general well-being, strength, and prodigal success in learners and, therefore, the learning process and their subsequent lives.

6.2 Impact on Society & Environment

Therefore, the areas of social and environment also benefit from the process of using machine learning in predicting depression among DIU students. Thus, these models help in decreasing the social costs of untreated mental disorders due to enhanced early diagnosis and management methods. Improving the state of students' well-being equates to healthier populace and better productivity of workforce in future. Also, creating such a culture within education institutions assists in raising mental health literacy and coping mechanisms among future generations. In that context, getting intake of mental health services offsets the secondary environmental effects of unaddressed depression including reduced efficiency and additional utilization of health care products. In sum, incorporating machine learning in mental health programs not only improves people's lives but also creates a solid and humane future.

6.3 Ethical Aspects

There is a critical ethical factor when it comes to predicting depression among the DIU students using machine learning. Privacy and confidentiality of students' data is a sensitive matter and hence, proper consent must be sought obeying all the regulations laid by the data protection agencies. Gaining the participants' and stakeholders' trust requires proper disclosure of the data collection, storage, and utilization processes. Managing bias at the data gathering and using it for decision-making by the algorithm is relevant in avoiding discriminating actions to different students' groups. Furthermore, there is a duty to eliminate possible risk factors to prevent depression among students by helping the at-risk learners get the necessary support they need. Policies and standards of ethical practice should inform the steps in collection of data and development of models, as well as the strategies for deploying and making sense of the results; to the end that the study maintains the highest standards of ethical, fairness and respect to the learners 'interests in the course of the entire study.

6.4 Sustainability Plan

The sustainability plan of predicting depression among DIU students by applying the machine learning approach has been carefully mapped down from the present aspect to a future one keeping the goal in mind. Some of these are the following; The models should be updated frequently so that faults that appear with time can be rectified. Trials that involve mental health workers and teachers guarantee the incorporation of the essential practices and support techniques constantly. Standard procedures regarding ethical practices and data confidentiality are observed as a way of safeguarding the students' welfare and to build credibility. Conducting seminars or sessions for students and faculty or awareness programmes help in creating a positive environment in the institution related to mental health. By integrating such practices, the project has the goal to establish long-term prevention-focused frameworks for students in regards to mental health problems influencing positive and lasting impact of the project on learners' wellbeing and academic achievement.

6.5 Summary

Predicting depression among the students of DIU, the papers and studies that used machine learning to predict employ tries various algorithms such as Gaussian Naive Bayes, Decision

Trees, AdaBoosting, Support Vector Machines, and Voting Classifiers. SVC and Voting Classifiers are recognized as the best performers; their advantage is pointed out regarding accuracy and efficiency. Essential facets as low feelings and flying tiredness are fundamental markers. This research provides a boost to early identification procedures, which in turn promotes timely actions and assistance. Therefore, by enhancing awareness of mental health issues and encouraging the prevention strategies among students, it intends to advance the general student well-being together with academic performance. Thus, algorithm choice and ensemble strategy remain critical factors that underpin the creation of efficient tools in addressing mental health issues in learning environments.

CHAPTER 7
CONCLUSION AND FUTURE WORLD

7.1 Conclusions

The analysis of forecasting depression among university students given Machine Learning is done on the dataset obtained from DIU in the context of Depression Inventory and the use of SVC and Voting Classifiers proved the potential of the model to make accurate prediction from symptoms collected from PHQ-9 Survey successfully. These models can help in the early detection of the vulnerable students so that appropriate measures can be taken to address the problem before it deepens, hence eradicating the effect of the untreated depression on student's health. That is why various elements of the symptom assessment, including low mood, fatigue, and difficulties concentrating, are included in the computations of predictive scores. Thus, the further use of these models in institutional support frameworks will contribute to students' better well-being and outcomes. Future studies could look at other factors and empirical studies to determine other variables that can be used to predict mental disorders and the different ways machine learning can help to provide timely solutions for clients within educational facilities to have holistic, preventive mental health strategies.

7.2 Further Suggested Works

The matter could be expanded with more recommendations regarding the uplift of the predictive models through adding more socio-demographic factors and/or behavioral markers, which would increase the models' precision and diversification. Research could also try and come up with better intervention mechanisms based on risk assessment of the individuals through machine learning. Future longitudinal research could explore the effectiveness of such interventions focusing on the changes of retention, persistence, performance and given mental health indicators among the students. In addition, expansion on whether or not the use of real-time monitoring systems or even applications for students to obtain and submit support or feedback could prove to improve the use of proactive self-care for mental health concerns in educational settings is doable. These avenues are to be pursued with an interest in enhancing instruments and approaches that contribute equally to the students' health and the overall human resilience.

7.3 Limitations/ Conflict of Interests

Limitations of this study include the following; the data was based on the PHQ-9 survey data whereby the respondents were asked to self-report and could have responded in a biased way or interpreted the questions differently from one another. The participants of this study DIU students and their characteristics maybe another source of weakness in terms of generalization of the study. Also, to reduce prejudice and discrimination in algorithmic decision-making, some measures were applied, while prejudice and discrimination in data collection as well as in model training might affect the outcomes. The duty may be in conflict with the interests depending on the relationships with organizations advocating for particular interventions or technologies in mental health. Thus, the conflicts can be minimized and the study's results and objectivity safeguarded through transparent reporting and compliance with the ethical standards.

REFERENCES

[1] Rois, R., Ray, M., Rahman, A., & Roy, S. K. (2021). Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. Journal of Health, Population and Nutrition, 40, 1-12.

[2] Serin, N. B., Serin, O., & Özbaş, L. F. (2010). Predicting university students' life satisfaction by their anxiety and depression level. Procedia-Social and Behavioral Sciences, 9, 579-582.

[3] Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. Procedia Computer Science, 152, 349-353.

[4] Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., ... & Uher, R. (2021). Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychological Medicine, 51(16), 2742-2751.

[5] Nayan, M. I. H., Uddin, M. S. G., Hossain, M. I., Alam, M. M., Zinnia, M. A., Haq, I. & Methun, M. I. H. (2022). Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic. Asian Journal of Social Health and Behavior, 5(2), 75-84.

[6] Choudhury, A. A., Khan, M. R. H., Nahim, N. Z., Tulon, S. R., Islam, S., & Chakrabarty, A. (2019, June). Predicting depression in Bangladeshi undergraduates using machine learning. In 2019 IEEE Region 10 Symposium (TENSYMP) (pp. 789-794). IEEE.

[7] Aleem, S., Huda, N. U., Amin, R., Khalid, S., Alshamrani, S. S., & Alshehri, A. (2022). Machine learning algorithms for depression: diagnosis, insights, and research directions. Electronics, 11(7), 1111.

[8] Yadav, S., Kaim, T., Gupta, S., Bharti, U., & Priyadarshi, P. (2020, December). Predicting depression from routine survey data using machine learning. In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 163-168). IEEE.

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

[10] Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science, 167, 1258-1267.

[11] Usman, M., Haris, S., & Fong, A. C. M. (2020, December). Prediction of depression using

machine learning techniques: A review of existing literature. In 2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS) (pp. 1-3) IEEE.23

[12] Alsubaie, M. M., Stain, H. J., Webster, L. A. D., & Wadman, R. (2019). The role of sources of social support on depression and quality of life for university students. International Journal of Adolescence and Youth, 24(4), 484-496.

[13] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences, 18, 43-49.

[14] Faisal, R. A., Jobe, M. C., Ahmed, O., & Sharker, T. (2022). Mental health status, anxiety, and depression levels of Bangladeshi university students during the COVID-19 pandemic. International journal of mental health and addiction, 20(3), 1500-1515.

[15] Wahed, W. Y. A., & Hassan, S. K. (2017). Prevalence and associated factors of stress, anxiety and depression among medical Fayoum University students. Alexandria Journal of edicine, 53(1), 77-84.

[16] Cassady, J. C., Pierson, E. E., & Starling, J. M. (2019, February). Predicting student depression with measures of general and academic anxieties. In Frontiers in Education (Vol. 4, p. 11). Frontiers Media SA.

[17] Narayanrao, P. V., & 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.

[18] Islam, M. A., Barna, S. D., Raihan, H., Khan, M. N. A., & Hossain, M. T. (2020). Depression and anxiety among university students during the COVID-19 pandemic in Bangladesh: A web-based cross-sectional survey. PloS one, 15(8), e0238162.

[19] Ahmmed, M. M., Rahman, M. M., Bhowmik, A., & Siddiqua, A. (2021). Impact of COVID-19 on academic and psychological aspects of undergraduate students in Bangladesh: a case study. AIUB Journal of Science and Engineering (AJSE), 20(3), 63-69.

[20] Yue, C., Ware, S., Morillo, R., Lu, J., Shang, C., Bi, J., ... & Wang, B. (2020). Automatic depression prediction using internet traffic characteristics on smartphones. Smart Health, 18, 100137.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: Predicting Depression in DIU Students using Machine Learning

Student ID: 203-15-14496 & 203-15-14495

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge Predicting Depression detection problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (With Knowledge Profile)	References
1.	EP1: Depth of Knowledge required	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	Our paper requires data collection from Survey form based on PHQ-9 and we have to know the design of a machine learning-based model which covers **K3** and **K4.**	**Page no:** [12-14] **Section:** [3.2] **Page no:** [1] **Section:** [1.1]
				Our paper requires the integration of different components and Utilizing machine learning algorithms to analyze academic performance and behavior patterns for early detection of depression in university students and implementing proactive mental health interventions and support systems within engineering design and practice curricula to address and mitigate depression risks among university students that cover **K5** and **K6.**	**Page no:** [12-14] **Section:** [3.2] **Page no:**

					[15-16] **Section:** [3.4]
				We have studied existing models with similar goals **(K8).**	**Page no:** [5-10] **Section:** [2.2, 2.3]
2.	EP2: Range of Conflicting Requirements	Yes	CO2, and CO7	In our paper, we encountered challenges in locating specific regulations governing the collection of data from university students on Survey Form (PHQ-9). When we collect data manually, many people don't want to give us data because they don't want to share their personal information. They feel insecure because we will have their email and some information.	**Page no:** [11] **Section:** [2.4, 2.5]
3.	EP3: Depth of analysis required	Yes	CO2, and CO6	Finding a clear-cut solution for our paper is challenging, primarily because of the abundance and diversity of manually or google form media data. We conducted an in-depth analysis to carefully choose a specific algorithm from numerous alternatives.	**Page no:** [17] **Section:** [4.2]
4.	EP4: Familiarity of Issues	Yes	CO8	To understand and study mental health and depression for our paper, we need to look into various aspects we were not very familiar with. Specifically, we have to figure out how to measure the level of depression	**Page no:** [5-11] **Section:** [2.2, 2.4]

5.	EP5: Extends of application codes	No	CO5	As part of our paper, we have visited and met with a psychologist. The purpose is to understand how we can measure depression based on a person's activities. This meeting will guide us in determining the specific type of data we need to use for our analysis.	N/A
6.	EP6: Extends of stakeholders involved and conflicting requirements	No	CO8	N/A	N/A
7.	EP7: Interdependence	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in data collection from (PHQ-9) **EP-7.**	**Page no:** [12-16] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	EA1: Range of resources	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, Machine Learning frameworks, annotated datasets, and ethical considerations to ensure systematic research for depression prediction.	**Page no:** [14-15] **Section:** [3.3]
2.	EA2: Level of interaction	No		N/A	N/A
3.	EA3: Innovation	No		N/A	N/A
4.	EA4: Consequences for society and the environment	Yes		This project contributes to society by Depression Detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for data privacy.	**Page no:** [25-26] **Section:** [6.1, 6.2, 6.4]

5.	EA-5: Familiarity	Yes		This project expands upon existing research by examining a novel approach in Depression Detection through machine learning model demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [10] **Section:** [2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [15-16] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced classification technologies which comply with **CO9.**	**Page no:** [24] **Section:** [5.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [12-17] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

TT2024

ORIGINALITY REPORT

12%	7%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.fedoa.unina.it Internet Source	1%
2	Submitted to Daffodil International University Student Paper	1%
3	Submitted to Universidad De Manila Student Paper	1%
4	www.researchgate.net Internet Source	1%
5	Ana Claudia Oliveira Melo. "Gaussian Mixture Models for Supervised Classification of Remote Sensing Multispectral Images", Lecture Notes in Computer Science, 2003 Publication	<1%
6	Shumaila Aleem, Noor ul Huda, Rashid Amin, Samina Khalid, Sultan S. Alshamrani, Abdullah Alshehri. "Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions", Electronics, 2022 Publication	<1%
7	bc.pollub.pl	