

# **Student Mental Health Prediction using aching learning techniques**

**By**  
**Atikul Islam Hemal**  
**212-15-4222**

## **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.Md Assadurzzaman (MA)**  
**Senior Lecturer**  
Department of Computer Science and  
Engineering Daffodil International  
University

### **Co-Supervised by**

**Montasir Delwar Afnan (MDA)**  
**Lecturer**  
Department of Computer Science and  
Engineering Daffodil International  
University



**DAFFODIL INTERNATIONAL  
UNIVERSITY**  
**Dhaka, Bangladesh**

**May 14, 2025**

## APPROVAL

This Project titled “Student Mental Health Prediction using machine learning techniques”, submitted by Atkiul Islam Hemal, ID No: 212-15-4222 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 14 May, 2025.

### BOARD OF EXAMINERS

-----  
**Dr. Md. Zahid Hasan (ZH)**

**Associate Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**

-----  
**Dewan Mamun Raza (DMR)**

**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

-----  
**Mr. Md. Aynul Hasan Nahid (AHN)**

**Lecturer**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

-----  
**Sadat Hasan**

**Data Scientist (Senior Principal Officer)**

Risk Management Division  
BRAC Bank

**External Examiner**

# DECLARATION

---

We hereby declare that this project has been done by us under the supervision of **Mr.Md Assadurzzaman (MA)**,Senior Lecturer, Department of Computer Science and Engineering, 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:



---

**Mr.Md Assadurzzaman (MA)**

**Senior Lecturer**

Department of Computer Science and  
EngineeringDaffodil International  
University

Co-Supervised by:



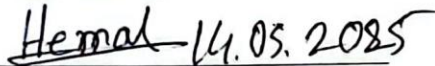
---

**Montasir Delwar Afnan (MDA)**

**Lecturer**

Department of Computer Science and  
EngineeringDaffodil International  
University

Submitted by:



---

**Atikul Islam Hemal**

Student ID:212-15-4222

Department of Computer Science and  
EngineeringDaffodil International  
University

---

**Team Member 2 Name (if any)**

Student ID:

Department of Computer Science and  
EngineeringDaffodil University  
©Daffodil International University

# ACKNOWLEDGEMENTS

---

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr.Md Assadurzzaman (MA), Senior Lecturer**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Maching 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 heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

# ABSTRACT

This research delivers a machine learning prediction system which tracks student stress levels and smartphone addiction behaviors because these control mental health and educational outcomes. The rising addiction to smartphones directly links to enhanced stress levels while simultaneously causing students to perform worse at school and resulting in diminished mental well-being throughout the era of advanced digital usage. A study examined over 1000 pupil behavioral and physiologic and smartphone usage data with ten machine learning models comprising Naïve Bayes, Decision Tree, XGBoost, LightGBM, Gradient Boosting, SVM, Random Forest and KNN and Logistic Regression and AdaBoost. Results from accuracy performance assessment show Logistic Regression achieved 0.8066, Random Forest hit 0.9811, SVM reached 0.8396 and Gradient Boosting exceeded 0.9858, KNN settled at 0.8538, Decision Tree reached 0.9906, AdaBoost reached 0.7689, XGBoost attained 0.9906, LightGBM touched 0.9717 and Naïve Bayes ended at 0.7925. Stress prediction through CatBoost stands out as the most effective model since it produces the lowest MSE alongside the highest  $R^2$  when compared to Linear Regression for addiction forecasting. Researchers determined the importance of early mental health interventions through their analysis which showed substantial connections between stress levels and smartphone usage conduct and educational results. The study proves that machine learning serves as an effective mental health prediction system so educational institutions can create specific stress management and addiction prevention approaches. The research outcomes support real-time data monitoring systems through their dual role of precise forecasting and the development of superior mental health programs.

# Table of Contents

<b>Approval</b>	<b>i</b>
<b>Declaration</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction.....	1
1.2 Motivation .....	2
1.3 Objectives .....	2
1.4 Methodology .....	3
1.5 Project Outcome.....	4
1.6 Organization of the Report .....	4
<b>2 Background</b>	<b>6</b>
2.1 Introduction.....	6
2.2 Literature Review .....	6
2.2.1 Similar Applications .....	12
2.2.2 Related Research.....	13
2.3 Gap Analysis .....	15
2.4 Summary .....	17
<b>3 Research Methodology</b>	<b>18</b>
3.1 Methodology/Requirement Analysis & Design Specification.....	18
3.1.1 Overview .....	18
3.1.2 Proposed Methodology/ System Design .....	18
3.1.3 Functional and Nonfunctional Requirements.....	19
3.1.4 Context Diagram .....	19

3.1.5	Data Flow Diagram Level 1.....	19
3.2	Detailed Methodology and Design.....	20
3.3	Project Plan.....	22
3.4	Summary .....	22
<b>4</b>	<b>Implementation and Results</b>	<b>23</b>
4.1	Environment Setup .....	23
4.2	Testing and Evaluation/Performance/ Comparative Analysis.....	23
4.3	Results and Discussion .....	24
4.4	Summary .....	28
<b>5</b>	<b>Engineering Standards and Design Challenges</b>	<b>29</b>
5.1	Compliance with the Standards.....	29
5.1.1	Software Standards.....	29
5.1.2	Hardware Standards .....	29
5.1.3	Communication Standards.....	30
5.2	Impact on Society, Environment and Sustainability .....	30
5.2.1	Impact on Life.....	30
5.2.2	Impact on Society & Environment.....	30
5.2.3	Ethical Aspects.....	30
5.2.4	Sustainability Plan.....	31
5.3	Project Management and Financial Analysis.....	31
5.4	Complex Engineering Problem.....	32
5.4.1	Complex Problem Solving.....	32
5.4.2	Engineering Activities .....	33
5.5	Summary .....	34
<b>6</b>	<b>Conclusion</b>	<b>35</b>
6.1	Summary .....	36
6.2	Limitation .....	36
6.3	Future Work .....	37
	<b>References</b>	<b>39</b>

# List of Figures

3.1. Context Diagram.....	27
3.2. Data flow diagram.....	28
3.3. sample diagram.....	29
4.1.1. Auc Roc Curve (A) & Confusion Matrix (B) for Logistic Regression.....	32
4.1.2: Auc Roc Curve (A) & Confusion Matrix (B) for Random Forest.....	33
4.1.3: Auc Roc Curve & Confusion Matrix for SVM.....	33
4.1.4: Auc Roc Curve (A) & Confusion Matrix(B) for Gradient Boosting.....	33
4.1.5: Auc Roc (A) Curve (B) for KNN.....	34
4.1.6: Auc Roc Curve & Confusion Matrix for Decision Tree.....	34
4.1.7 Auc Roc (A) Curve & Confusion (B) Matrix for AdaBoost.....	34
4.1.8 Auc Roc (A) Curve & Confusion( B) Matrix for XGBoost.....	35
4.1.9 Auc Roc (A) Curve (B) for LightGBM.....	35
4.1.10 Auc Roc (A) Curve (B) for Naive Bayes.....	35
4.1.11 Model Accuracy Comparison.....	36

# List of Tables

2.1	Summary of Literature Reviewed. ....	10
2.2	Gap Analysis.....	16
4.1	Accuracy for every model.....	23
5.1	Financial Estimation of the Project.....	31
5.2	Mapping with complex problem solving.....	32
5.3	Mapping with knowledge Profile. ....	32
5.4	Mapping with complex engineering activities.....	33

# Chapter 1

## Introduction

### 1.1 Introduction

Mental health is an essential component of general well-being because it enables people to overcome obstacles in life, recognize their potential, acquire new knowledge, function well, and interact with others. It has a significant impact on both personal and social growth by affecting decision-making, relationship-building, and societal contributions. Individual fulfillment, societal strength, and economic advancement all depend on mental health, which is a fundamental human right [1]. Society frequently points out physical health and outward looks while ignoring the importance of mental health. However, maintaining excellent mental health is necessary for successfully navigating life from birth to maturity. Mental health issues can affect anybody, no matter age, gender, race, geography, or religious views [2]. In Bangladesh, mental health discussions are usually avoided due to judgment, since many people consider mental illness as a secret. Those that are struggling are frequently ignored rather than helped, causing the problem. The Directorate General of Health Services (DGHS) reports that rising suicide rates reflect a developing mental health crisis. A countrywide study of higher education students indicated that 75% suffer from depression. Using a Beck Inventory (BDI-III), 21% expressed severe depression, 25% moderate to severe depression, and 26% mild to moderate symptoms, highlighting the critical need for mental health knowledge and assistance [3]. A recent survey of 1,773 students revealed that 85.9% feel the internet contributes to their mental health problems. Students often face negative affects as a result of the internet's strong impact on their lives. According to the research, 72.2% of students face mental health issues, with 26.1% blaming the internet completely 59.8% considering it an important factor, while only 8.3% not attributing their problems to it [4]. Identifying the causes of mental health problems among students has been a challenge. These elements may be impacted by psychology, biology, or the environment. The familiarity of some symptoms and overlapping circumstances complicates diagnosis, frequently leading to physician errors [5]. One of the many applications of AI nowadays is machine learning, particularly in the diagnosis of illnesses. Doctors can use it to evaluate large amounts of patient data and provide individualized therapies according to medical problems [6]. Additionally, by identifying hidden patterns in complex data, machine learning (ML) helps analysis of mental health. It identifies risk indicators, predicts conditions, and recommends appropriate solutions ML makes it possible for early detection and timely student help by automating these steps. [7].

## 1.2 Motivation

Since mental health affects how people think, feel, and behave, it has become one of the most important aspects of overall well-being. It influences people's ability to manage stress, interact with others, and make wise decisions. Mental health is crucial for long-term personal growth, social connections, and academic achievement, especially for students. Mental health problems are becoming more common among student populations worldwide due to an increase in cases of stress, anxiety, and depression. Many of these cases go undiagnosed until they have a major influence on both academic and personal lives [8].

Mental health is frequently disregarded in spite of its significance, especially in developing nations like Bangladesh. Timely treatments and candid conversations are impeded by cultural shame and ignorance [9]. A significant mental health crisis is evident in Bangladesh, where a significant portion of university students experience varied degrees of stress and despair, according to the Directorate General of Health Services (DGHS) [10].

Technological developments, particularly in the areas of artificial intelligence (AI) and machine learning (ML), have created new opportunities for the early diagnosis and treatment of mental health conditions. Large datasets can be processed by machine learning algorithms to find trends that point to the beginning of mental health problems. Predictive models facilitate early detection, allowing for prompt intervention, which stops mental health issues from getting worse and improves overall results [11].

Based on demographic, behavioral, and psychological data, recent research has shown that machine learning approaches such as Random Forest, Support Vector Machines (SVM), Decision Trees, and Deep Learning models can successfully predict mental health disorders [12][13]. In addition to their high accuracy, these prediction models aid in the identification of important risk variables, which are especially prevalent among college students and include lack of social support, financial stress, academic pressure, and sleep deprivation [14].

One promising way to create healthier learning environments in schools is to use machine learning to predict mental wellness. Universities may proactively treat mental health issues, enhancing students' academic achievement and general well-being while also advancing society at large, by using predictive analytics into their student support programs.

## 1.3 Objectives

- Collect and preprocess student mental health data to ensure accuracy, consistency, and validity to machine learning models.
- Train and evaluate a number of machine learning models (such as Random Forest, SVM, and Neural Networks) to find the best successful prediction method.
- To examine the connection between academic stress, lack of money, and social support on the mental health of Bangladeshi students, with the goal of finding critical variables that lead to depression and anxiety in this population [15].

## 1.4 Methodology

This study uses machine learning models to predict students' stress levels and smartphone addiction. The process is broken down into multiple phases:

- **Data Collection:** Surveys were created that cover demographic data, academic workload, stress levels, smartphone usage behaviors, sleep patterns, and social support. Then these surveys were sent to college students to collect their pieces of information and opinions to create an extensive dataset. To collect different stress levels and smartphone usage patterns, the data collection period was spread out over time.
- **Data preprocessing:** To manage missing values, outliers, and duplicates were removed, making the gathered dataset thoroughly cleaned. This guarantees the dataset's accuracy and dependability for the research. Feature engineering is carried out to choose the relevant elements that are most suggestive of stress and smartphone addiction.
- **Training and Model Selection:** Many machine learning models, including Random Forest, Support Vector Machines (SVM), Decision Trees, Gradient Boosting, and Deep Learning algorithms, are selected for training. 20% of the dataset is used for testing, and the remaining 80% is used for training. The models are trained on the training data to predict stress levels and addiction inclinations based on the attributes collected.
- **Model Evaluation:** A number of performance indicators, including Mean Squared Error (MSE), R-squared ( $R^2$ ), Precision, Recall, and F1-Score, are used to assess the trained models. These measurements aid in assessing how well each model predicts levels of stress and addiction.
- **Pearson correlation analysis is used to investigate:** The connections between various factors, including stress levels, smartphone use, and academic pressure. This research aids in identifying important risk variables for addiction and stress.
- **Interpretation of Results:** Individual students' stress and addiction levels are predicted using the top-performing models, which are chosen based on assessment metrics. The findings are then studied to detect the tendency in students' actions that contribute to raised stress levels and smartphone addiction.
- **Intervention Suggestions:** Suggestions for early intervention methods are produced considering the prediction results. The methods are solutions that focus on managing stress and lessening smartphone addiction, via time management tools, psychotherapy, and digital health initiatives.

## 1.5 Project Outcome

To predict and treat students' stress levels and smartphone addiction using machine learning models is the main goal of this paper. The expected outcomes and possible results are outlined below:

Creating an effective machine learning model that accurately predicts a student's stress levels and smartphone addiction is one of the key outcomes of this project. With the help of behavioral data, psychological parameters, and academic performance metrics, the model will anticipate mental health concerns, enabling early intervention well before deterioration sets in. The project will identify the primary drivers of stress and smartphone addiction for students. Some of these problems include academic pressure, social isolation, sleep deprivation, and chronic smartphone use. Knowing these risk factors will facilitate creating targeted interventions aimed at students facing these issues. Using the project's predictive tools, educational institutions will proactively identify students who may be vulnerable to mental health challenges due to academic pressures. Timely interventions can be initiated, effectively halting the escalation of more complex mental health challenges. Proactive measures such as counseling, stress management sessions, or providing guidelines on smartphone usage could be implemented. The study will provide actionable recommendations for developing or enhancing student-related well-being programs. The findings enable colleges to adopt practices such as fostering proper digital etiquette, reducing academic stress, and enhancing sleep hygiene. The emerging areas of mental health research and behavioral data science will benefit from the utilization of machine learning in predicting students' mental health issues. Other domains like digital addiction, educational psychology, and the intersection of machine learning with healthcare may also build off of these findings. Equipping universities with mental health forecasting tools integrated with AI could form additional predictive frameworks alongside existing student aid systems. With these technologies to accurately monitor and predict students' mental health, universities could offer tailored assistance and enhance the overall student experience. The program has the potential to optimize academic achievement and retention metrics by resolving mental health challenges such as stress and smartphone addiction proactively. If students are provided with the right support at the right time, their overall health, academic performance, and course completion will significantly improve.

## **1.6 Organization of the Report**

The paper contains a series of chapters which thoroughly analyze machine learning processes for predicting student stress along with smartphone addiction. Each chapter plays a vital role in providing detailed information about the research starting from the background review to the final findings.

### **Chapter 1: Introduction**

The first chapter provides readers with research details along with study methodology and research goals and study purpose explanation. The paper examines mental health problems in students by concentrating on stress and smartphone addiction. The chapter introduces machine learning as a solution to both recognize and manage these problems while establishing the background for subsequent sections of the report.

### **Chapter 2: Background**

This chapter delivers an in-depth explanation of student mental health issues through a detailed examination of stress and smartphone addiction. The chapter embraces existing research about how many students experience mental health problems while analyzing how smartphone addiction damages wellbeing and studying previous efforts to detect mental health disorders through machine learning. The chapter introduces research gaps and establishes essential background information for the study at

hand..

### **Chapter 3: Methods of Research**

In Chapter 3, the study provides a detailed explanation of the research approach which includes the data collection process and its preprocessing stages and the machine learning methods used for prediction. The methods employed for data collection from university students are described together with the approaches used to clean and process the data and the application of machine learning algorithms. In Chapter 3, the authors present the evaluation metrics which consist of Mean Squared Error (MSE) and R2 for model performance assessment.

### **Chapter 4: Execution and Outcomes**

In this chapter, we discuss the implementation of machine learning models through training and testing phases. The chapter evaluates model performance in terms of stress prediction and smartphone addiction prediction and discusses outcomes for each model tested. This chapter contains a comprehensive examination of model performance through accuracy measurements along with precision and recall analysis and the influence of prediction features. A visualization technique presents the model findings in a more understandable way.

### **Chapter 5: Effects on Sustainability, the Environment, and Society**

The study investigates broader implications through a detailed analysis in the fifth chapter. The chapter discusses the potential advantages of machine learning predictions for mental health in educational environments while focusing on student welfare improvements and academic success. The analysis considers both ethical aspects of technology deployment in mental healthcare and environmental sustainability factors when discussing machine learning model implementation.

### **Chapter 6: Conclusion**

The final chapter of the study contains the research conclusions and a detailed summary of the key findings. The study reviews past research limitations while proposing recommendations for future research directions. The chapter emphasizes the importance of current research into student mental health together with the potential machine learning applications for building personalized student support systems.

# Chapter 2

## Background

### 2.1 Introduction

Since mental health affects how people think, feel, and behave, it has become one of the most important aspects of overall well-being. It affects how people handle stress, react to others, and make good decisions. Mental health is crucial for long-term personal growth, social connections, and academic achievement, especially for students. According to studies, mental health conditions like stress, anxiety, and depression are becoming more common among students globally, and many of these instances go undiagnosed until they have a significant negative influence on both academic and personal lives [16].

Despite its significance, mental health is frequently disregarded, particularly in developing nations like Bangladesh where a lack of awareness and cultural stigma prevent candid conversations and prompt interventions [17]. According to a survey undertaken by the Directorate General of Health Services (DGHS) in Bangladesh, a substantial number of university students suffer from varying degrees of depression and stress, indicating a serious mental health crisis [18].

With improvements in technology, notably in Artificial Intelligence (AI) and Machine Learning (ML), new prospects have developed for the early prediction and management of mental health concerns. Large data sets can be analyzed by machine learning algorithms to find trends that can point to the beginning of mental health issues. Early detection by predictive models allows for timely interventions, hence reducing the escalation of mental health problems [19].

A line of modern research has employed machine learning algorithms particularly Random Forest along with Support Vector Machines together with Decision Trees and Deep Learning models for mental health disorder prediction based on behavioral and psychological and demographic information [20] [21]. The prediction models achieve high accuracy while helping researchers identify important risk factors that specifically affect college students including social support deficits and financial strain and academic tension along with sleep disorders [22].

### 2.2 Literature Review

In recent years, student mental health has emerged as a pressing issue because an increasing number of students experience depression, anxiety, and stress during their academic journey. The mental health challenges students experience affect their personal wellness along with educational success and social connections as well as their general quality of existence. Research findings demonstrate that multiple students experience mental health challenges although they tend to avoid seeking help until their symptoms worsen.. This delay in care stresses the important necessity for early detection and proactive support measures to treat mental health disorders before they worsen. Machine learning (ML) provides a viable approach for expecting mental health difficulties for

youngsters by examining a variety of data that includes academic records, surveys, and behavioral patterns. Using ML models, researchers can identify complex structure in student behavior that indicate mental health issues, commonly before they become apparent with traditional methods. Support Vector Machines, Random Forests, and Neural Networks have all showed the ability to accurately forecast disorders such as anxiety and depression. However, barriers keep occurring, including concerns about dataset quality, data privacy, and the ethical use of algorithms for prediction.

"Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," a study by Hanif Abdul Rahman et al. [17], was published in *Bioengineering* (2023, 10(575)). The study examined mental health and health-risk behaviors like physical activity, diet, smoking, and alcohol consumption using an online cross-sectional survey of 15,366 students from 17 universities across seven ASEAN countries (Brunei, Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam). The dataset was preprocessed using multiple imputation (MICE) to address missing data, and the Synthetic Minority Oversampling Technique was used to balance the data. Several machine learning classifiers, such as bagging, boosting, RPART, GLM, KNN, NB, NNet, and others, were used. With an accuracy of 92.1% on training data and 90.1% on test data, Random Forest outperformed all other models in this category, but Adaptive Boosting performed admirably with an accuracy of 89.3%. Other models, such as KNN (77.5%), Naïve Bayes (72.3%) and Neural Networks (61.5%), did rather badly. Body mass index (BMI), weekly sports participation, grade point average (GPA), sedentary hours, age, gender, consumption of salt, consumption of fruits and vegetables, sleep duration, and physical activity levels are among the eleven primary variables that the study found to have an impact on mental health. The efficacy of machine learning in detecting mental health issues is highlighted by the authors, who also suggest that deep learning models, natural language processing (NLP), and physiological sensing data like EEG and ECG be included in future studies to increase prediction accuracy. They also emphasize how important data security and privacy issues are before such forecast algorithms are widely used.

Ayako Baba and Kyosuke Bunji [18] published an article titled "Prediction of Mental Health Disorders Using Annual Student Health Survey: Machine Learning Approach," in *JMIR Mental Health* (2023, 10: e42420). The study's goal was to anticipate Data from an annual health survey were used to assess mental health problems among students. done at a Japanese national university. The dataset contained replies from 3561 undergraduate students (out of 5690 polled) between 2020 and 2021. The study used logistic regression, elastic net, random forest, XGBoost, and LightGBM as machine learning models, with LightGBM coming as the best finisher. Logistic Regression (76.0%), Elastic Net (86.2%), Random Forest (83.3%), XGBoost (85.5%), and LightGBM (85.7%) provided the highest Matthews correlation coefficient (MCC) of 0.970 in predicting mental health problems in 2020 and 0.986 in predicting 2021 issues based on 2020 responses. The study examined the impact on various variables, like demographics, self-reported health conditions, and response time patterns. Although answering time did not have a significant effect on prediction accuracy, certain response behavior variables did. The most influential predictors of mental health problems were worry about campus life, future concerns, sleep length, academic stress, years of university study, and the number of same-sex students in a department. According to the research, "machine learning could be effectively used for early identification of mental health issues among students, enabling timely intervention through health service centers. "Future enhancements were proposed, including the incorporation of "longitudinal data, more universities, and additional behavioral factors."

Dhanamma Jagli et al. [19] published their article, "Innovative Machine Learning Models for Student Mental Health Analysis," in *Frontiers in Health Informatics* (2024, Vol 13, Issue 3). The study explores the use of machine learning (ML) approaches to predict students' mental health difficulties based on behavioral and academic tendencies. The authors examine the effectiveness of multiple ML methods, including 'logistic regression, decision trees, support vector machines (SVM), and neural networks', in mental health classification. The study combines findings from three major studies: one on the effects of "online mobile gaming on mental health," another on the mental health of college students using machine learning models, and a third on "classification techniques for mental health analysis." SVM and neural networks beat other approaches in different circumstances, with SVM knowing gaming-related mental health concerns with a 91.68% accuracy. Logistic regression and decision trees had reasonable accuracy, however models that included extensive behavioral and academic data performed better. The paper focuses on issues such as data quality, privacy problems, and algorithmic fairness, emphasizing the importance of enhanced preprocessing, diversified datasets, and advanced ML approaches such as deep learning and ensemble methods. The authors conclude that machine learning has significant potential for early detection and personalized mental health interventions, advocating that future research prefer scalability, interdisciplinary collaboration, and ethical considerations to enhance the reliability of ML-based mental health support systems in educational settings.

Barnali Sahu et al. [20] published their article, "Mental Health Prediction in Students Using Data Mining Techniques," in *The Journal of Open Bioinformatics* (2023, Vol. 16, e187503622307140). This study aims to predict mental health problems such as depression, relationships, and anxiety disorders among university students in Odisha, India. An online survey was conducted across three universities, producing 109 responses containing socio-demographic information as well as responses from the GAD-7 (General Anxiety Disorder), PHQ-9 (Patient Health Questionnaire), and the Social Connectedness Scale. Correlation analysis and regression approaches were used to assess relationships between parameters and predict mental health results. Results showed an elevated risk of depression (61.9%), with early-year students having higher rates. Social relationship and depression were found to be negatively related, but anxiety and depression were found to be positive correlated. Also, the study found a relationship among depression and academic achievement, with higher levels of depression being linked to lower academic performance. With an accuracy of 92.3% (training), 90.18% (testing), and 84.35% (validation), logistic regression with hyperparameter adjustment fared better than the other models evaluated, outperforming linear regression (70.03%). The results stress the value of early mental health interventions and recommend that colleges and universities place preventative measures and policies into effect to help students' mental health. Larger datasets, more behavioral variables, and sophisticated machine learning methods like deep learning could all be used in future research to enhance model performance.

Mohamed Drira, Sana Ben Hassine, Michael Zhang, and Steven Smith's [26] paper provides a systematic literature review (SLR) on the use of machine learning (ML) methods in student mental health (SMH) research, with a focus on ethical considerations such as privacy, transparency, interpretability, and fairness. The authors followed the PRISMA standards and analyzed papers published between 2015 and 2024, gathering data from six major academic databases, including PsycInfo, PubMed, IEEE Xplore, ACM Digital. Utilized ML techniques were supervised learning models like Support Vector Machines (SVM), logistic regression, and random forests, with deep learning models like CNNs and RNNs also used in some circumstances. The survey found that 67% of publications diagnosing mental health problems such as depression and stress. The most widely logistic

regression, and random forests, with deep learning models like CNNs and RNNs also used in some circumstances. These models' performance was frequently assessed using accuracy (51%), recall (38%), and precision (30%), with k-fold cross-validation being the most popular validation method. The study discovered substantial ethical difficulties, as 65% of the assessed publications failed to address privacy concerns, 59% used black-box models with limited interpretability, and only 18% revealed demographic information, raising concerns about fairness. However, transparency was rather adequately addressed, with an average reporting score of 8/10 and a replicability score of 4/5. The authors underline the importance of increased ethical awareness in AI/ML-based SMH research, calling for better data privacy policies, the adoption of explainable AI techniques, and a balanced approach that blends high-performance ML models with interpretability and fairness concerns.

The paper "Student Mental Health Analysis Using Machine Learning Techniques" by Thejaswini M, Vivek A M, Shweta Y Bajantri, Shashank B S, and Shylaja B, [27] published in the International Journal of Research Publication and Reviews (Vol. 5, No. 12, December 2024) investigates the use of machine learning (ML) techniques for assessing student mental health. The study looks into the mental health profiles of university students and working professionals, including issues like academic pressure, financial stress, work pressure, and actions. The authors collected data from over 500 students and 2000 working professionals using surveys and socioeconomic questionnaires, which included validated mental health assessment measures such as the PHQ-9 and GAD-7. Various ML models, including cluster analysis, logistic regression, and random forest, were used to identify mental health concerns, and the system achieved 85% accuracy in detecting high-risk individuals. The study indicated that 52.15% of surveyed males and 47.85% of females reported depression, owing primarily to academic stress and job unhappiness. Furthermore, 60% of students expressed anxiety, with schoolwork, financial instability, and peer pressure being the most common stresses. The findings highlight the importance of early intervention measures, tailored mental health suggestions, and institutional support in reducing student stress. The authors propose future improvements, such as real-time mental health monitoring via wearable devices, deep learning approaches for increased accuracy, and cross-institutional cooperation to establish comprehensive mental health databases.

Maryam Sadiq [28] wrote the research paper Predicting Students' Mental Disorders Using Machine Learning for her graduate capstone at Rochester Institute of Technology (RIT) Dubai in 2022. The thesis attempted to use machine learning approaches to predict mental health issues in students, specifically stress, anxiety, and depression. The author conducted the study using the Depression Anxiety Stress Scales (DASS) dataset. This dataset was gathered from several internet sources between 2017 and 2019 and downloaded via Kaggle, a well-known repository for machine learning data. The dataset had 68 columns, which included demographic information and survey results. Before starting model training, Exploratory Data Analysis (EDA) was used to find missing values and anomalies in the data. Machine learning models used in this study included K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, and XGBoost. The dataset was divided among 70 percent training and 30 percent testing. Gradient Boosting (a boosting ensemble method) outperformed the other models across all categories. It had an accuracy of 92% for depression, 90% for stress, and 88% for anxiety prediction. Other models, including Random Forest, Naïve Bayes, and Decision Tree, showed lesser accuracy. The model's performance was evaluated based on accuracy, F1-score, recall, and precision. The study concluded that machine learning models, particularly ensemble methods such as Gradient Boosting, may accurately predict students' mental health status. The author proposed that

future research may look at new artificial intelligence techniques, build continuous feedback loops, and create automated mental health assessment systems.

Ujunwa Madububambachu[29], Augustine Ukpebor, and Urenna Ihezue wrote the thesis titled "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review". The study examines multiple machine learning algorithms for predicting mental health issues in college students. The reviewed studies used structured and unstructured data from clinical records, electronic health records, neuroimaging data (fMRI,sMRI), social media (Twitter, Reddit), self-reported surveys, and physiological signals, with sample sizes ranging from 50 to more than 2.5 million data points. The study focuses on several machine learning models employed in mental health prediction, with Convolutional Neural Networks (CNN) achieving the best accuracy, particularly for bipolar disorder (99.75%). Other models tested were Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Trees (DT), Multi-Layer Perceptron (MLP), and Extreme Learning Machine (ELM). The study's drawbacks include a lack of consistent, high-quality datasets, issues with model interpretability, ethical concerns about data privacy, and biases in self-reported mental health data. Furthermore, the paper emphasizes the need of using longitudinal data to strengthen models and capture the temporal dynamics of mental health disorders. Despite these challenges, the study highlights machine learning's potential to improve early detection and intervention tactics in mental health diagnosis

Table 2.1: Summary of Literature Reviewed.

Authors	Dataset	Methods Used	Results	Limitations
Hanif Abdul Rahman et al. (2023)	15,366 university students from 17 ASEAN countries	GLM, KNN, NB, NN, RF, RPART, BAG, BOOST	Random Forest (92.1%), Adaptive Boosting (89.3%), KNN (77.5%), Naïve Bayes (72.3%), and Neural Networks (61.5%)	Limited to survey-based data; lacks physiological data integration
Ayako Baba & Kyosuke Bunji (2023)	3,561 students from a Japanese national university	LR, EN, RF, XGB, LGBM	LightGBM (85.7% accuracy), followed by Elastic Net (86.2%), Random Forest (83.3%), and XGBoost (85.5%)	No inclusion of physiological or behavioral data like social media usage
Dhanamma Jagli et al. (2024)	Reviewed three previous studies on mental health	Logistic Regression, Decision Trees, SVM, Neural Networks	SVM achieved the highest accuracy (91.68%) in predicting gaming-related mental health issues	Study is a review; lacks experimental validation with new datasets
Barnali Sahu et al. (2023)	109 university students from 3 colleges	Logistic Regression, Linear Regression, Correlation Analysis, Chi-	Logistic Regression achieved the highest accuracy (92.3% training, 90.18% testing,	Small dataset; lacks generalizability beyond Odisha; limited feature set

	in Odisha, India	squared test	84.35% validation). Depression prevalence: 61.9%	
Mohamed Drira, Sana Ben Hassine, Michael Zhang, Steven Smith	Six databases, surveys, sensors, social media	SVM, logistic regression, random forests, CNNs, RNNs	67% used ML for prediction, 36% for diagnosis. Ethical concerns: privacy, black-box models, fairness.	65% ignored privacy, 59% used black-box models, 18% lacked demographic data.
Thejaswini M, Vivek A M, Shweta Y Bajantri, Shashank B S, Shylaja B	500 students, 2000 professionals, surveys, PHQ-9, GAD-7	Cluster analysis, logistic regression, random forest	85% accuracy, 52.15% males and 47.85% females had depression, 60% students had anxiety	No real-time monitoring or deep learning techniques
Maryam Sadiq	Depression Anxiety Stress Scales (DASS) dataset, collected from Kaggle	K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, XGBoost, Gradient Boosting	Gradient Boosting achieved highest accuracy: 92% for depression, 90% for stress, and 88% for anxiety	Limited dataset diversity, potential bias in model predictions, need for real-time implementation

### 2.2.1 Similar Applications

1. Using data on health behaviors of university students to predict mental well-being, Hanif Abdul Rahman et al. (2023) used machine learning models including Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB). They described their results for mental well-being prediction, and their predictive accuracy for again was highest with Random Forest (92.1%). Their research specifically points to valuable behavioral information about mental health status that included sleep duration, physical activity, and academic performance.[36]
2. Ayako Baba and Kyosuke Bunji (2023) explored predicting mental health disorders in students using data from an annual health survey. They evaluated models such as XGBoost, Elastic Net, and Logistic Regression. They reported Frequent contact with LightGBM as the most effective model in predicting mental health disorders. Additionally, the authors also examined the effects of social support, academic stressors, and sleep on predicting instances of mental health disorders. [37]
3. In 2024, Dhanamma Jagli et al. employed machine learning models, including SVM, Decision Trees, and Neural Networks, to predict students' mental health. Their evidence suggested that SVM was particularly good at predicting mental health ailments related to online gaming. This study underscores how important it is to use various aspects of student behavior when predicting their mental health disorders.[38]
4. Regression trees and logistic regression are two machine learning techniques that were used by Barnali Sahu et al. (2023) to predict mental health illnesses in college students. The data demonstrate that social ties and academic stress are two determinants of anxiety and depression. Clearly, the study illustrates how machine learning can be used to help identify students who may be at risk for mental health illnesses in order to implement early intervention strategies.[39]

#### Applications for Predicting Mental Health on the Web and Mobile

1. Wysa is an Artificial Intelligence-supported mental health app that provides conversational support using machine learning and natural language processing (NLP). Using users' input (stress levels, sleep)", and mental health, Wysa offers individualized mental health care. Just like you, Wysa uses behavioral data to predict and intervene with mental health disorders..
2. Through a smartphone app that monitors behavior, mood, and other mental health indicators, Ginger.io offers mental health services. The software predicts and offers real-time mental health support using machine learning algorithms. Similar to your research's predictive algorithms, Ginger.io prioritizes individualized care by evaluating user data and providing pertinent mental health support.
3. Another smartphone software that uses daily quizzes to assess emotional well-being is called Moodpath. The software analyzes reactions and forecasts mental

health conditions like anxiety and depression using machine learning techniques. Your study's prediction models, which look for kids who may be at risk for mental health issues, are comparable to these.

### Case Study: Using Predictive Analytics to Improve University Students' Well-Being

Predictive analytics techniques are being used by a number of colleges to evaluate and treat the mental health of their students. A UK institution, for instance, used a data-driven platform to forecast students' mental health based on social involvement, financial hardship, and academic achievement. In order to predict possible mental health crises and provide tailored treatments, such counseling services, the platform included machine learning algorithms, which are comparable to those utilized in your project.

- **Methodological Contributions Data Collection:** Similar to your project, a lot of research on predicting students' mental health uses behavioral data and surveys to build machine learning models. It has been demonstrated that this approach effectively captures the complex character of mental health problems.
- **Feature Engineering:** A common approach used in all of the examined investigations is the meticulous selection and preprocessing of features including academic stress, social support, and sleep quality. This is consistent with your strategy of utilizing behavioral, academic, and demographic data.
- **Model Evaluation:** To assess model performance, the aforementioned research commonly employ metrics such as accuracy, precision, recall, and F1-score. Additionally, this aligns with the assessment methods employed in your project.

### 2.2.2 Related Research

The pertinent research literature on the application of machine learning to the prediction of student mental health is summarized in this part. The emphasis is on research that has predicted or examined mental health problems in kids using machine learning approaches such ensemble methods, decision trees, and regression models. The evaluated papers shed light on the different approaches used and how well they work to address mental health issues in school environments.

1. Hanif Abdul Rahman et al. (2023): Predicting Mental Well-Being using Machine Learning Hanif Abdul Rahman et al. used machine learning in their work to forecast university students' mental health based on their dietary, sleep, and physical activity habits. Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB) were among the machine learning techniques they used; Random Forest yielded the highest accuracy (92.1%). Their findings emphasize the potential of machine learning to provide individualized mental health assistance and the significance of health behavior data in predicting mental well-being.[36]
2. Ayako Baba and Kyosuke Bunji (2023): Using Student Health Survey Data to Predict Mental Health Disorders In their study, Ayako Baba and Kyosuke Bunji analyzed information gathered from yearly health surveys to predict mental health illnesses in college students. Logistic Regression, Elastic Net, Random Forest, XGBoost, and LightGBM were among the machine learning

models they used; LightGBM produced the best outcomes. In order to anticipate mental health illnesses, their study found critical factors including social support, sleep duration, and academic stress. The objective of employing machine learning to anticipate mental health issues in students and offer prompt intervention is strongly related to this study.[37]

3. Machine Learning Models for Mental Health Analysis By Dhanamma Jagli Et Al. (2024) Dhanamma Jagli and colleagues explored the use of various machine learning models for classifying mental health disorders based on educational and activity data. They found that for the case of mental health disorder detection, Support Vector Machines (SVM) and Neural Networks (NN) outperformed the other models. To improve the predictions from the models, the study also emphasized the need to incorporate multiple data types, including behavioral data from online activities such as gaming. This research contributes to the understanding of machine learning applications in diagnosing students' mental health issues.[38]
4. Mental Health Prediction by Barnali Sahu et al. (2023) Applying Data Mining Methods Barnali Sahu et al. utilized data mining techniques to predict mental health ailments, such as anxiety and depression, among college students. They computed the mental health test scores together with other sociodemographic data at the college level using logistic regression and other statistical frameworks. Their results showed that Logistic Regression performed the best with an accuracy of 92.3%. Moreover, the study found that students in earlier years were more prone to depression, reinforcing the need for intervention.[39]
5. Legal Matters of Machine Learning Applications to Mental Health Prognosis by Mohamed Drira et al. (2024) The legal issues in concern with the use of machine learning techniques for predicting mental health issues within the educational framework were analyzed in the systematic review conducted by Drira and other colleagues. They analyzed several machine learning techniques, for instance, Random Forest, Logistic Regression, and even deep learning methods like Convolutional Neural Networks (CNNs). The authors highlighted the importance of the issues of secrecy, transparency, and fairness in the application of machine learning algorithms to mental health prediction especially in dealing with sensitive student data. This essay is part of the debate on ethical issues concerning the application of artificial intelligence in predicting mental health.[40]
6. Maryam Sadiq (2022): Predicting Mental Health Disorders Using Machine Learning Maryam Sadiq, in her graduation work, produced a prediction of students' stress, anxiety and sadness using the sadness Anxiety Stress Scales (DASS) data set. Maryam Sadiq applied machine learning models such as XGBoost, Random Forest, Naïve Bayes, and K-Nearest Neighbors (KNN). Gradient Boosting actually performed better than other models, with an accuracy level of 88% for anxiety, 90% for stress, and 92% for depression, according to the study. The study illustrates the ability of machine learning models to predict mental health disorders and suggests that prediction accuracy might be increased, based on prior feedback from the completed forms, and following a real-time feedback loop..[41]

7. Machine Learning for Predicting Mental Health Diagnoses by Ujunwa Madububambachu et al. (2024) Ujunwa Madububambachu and colleagues reviewed the literature and considered how machine learning methods such as support vector machines (SVM) and convolutional neural networks (CNNs) could be used to predict mental health diagnoses among college students. In their review, they concluded that while these models could be very accurate, they presented significant concerns with data quality, privacy, and black-box interpretability (model understanding and usability). To bolster the reliability and ethical implications of machine learning systems utilizing AI for mental health prediction, the authors advocated for more transparent AI models and improved data collection methods..[42]
8. Thejaswini M, Vivek A M, Shweta Y Bajantri et al. (2024): This study utilized machine learning techniques such as cluster analysis Random Forest, and Logistic Regression to predict mental health illness in students, based on survey data. The study stated that many students were at risk for depression, and that there was a higher instance of risk during the early years of school. Overall, the accuracy of detecting mental health issues was high, and demonstrates that machine learning can help identify students at risk and provide interventions..[43]

## 2.3 Gap Analysis

In this section we will note what the current gaps are in the literature regarding applying machine learning to predict students' mental health. This examination will identify the areas in which this research wants to make meaningful contributions to the literature but also the areas that need to be further researched or developed.

1. Opportunities in Existing Research Found Limited Data and Scope: Most of the research reviewed was based on survey data and typically emphasized academic, behavioral and demographic characteristics. There is relative lack of focus on other forms of data, like social media use, physiological data (e.g., heart rate, sleep behaviours), and real-time behavioural data. By adding multiple types of data, the possibility of making more timely predictions may be facilitated and a richer understanding of the variables that impact students' mental health could be achieved.
2. Small Sample Sizes: A number of studies with relatively small datasets, often within limited geographic regions or college campuses, have been completed. For example, studies with less than 1,000 survey participants probably will not produce results that generalizable results. Future research employing larger and more diverse datasets, including students from more demographic regions, would be beneficial, yielding more accurate and generalizable findings.
3. Interpretability of the Model: The lack of interpretable machine learning models which use complex algorithms such as Random Forest, Gradient Boosting, and XGBoost is a large research gap. While they provide predictions with high accuracy, they are sometimes thought of as "black boxes," which makes it difficult for practitioners to understand how predictions are being made with these models. For educators and mental health professionals to trust and act on these models' predictions, it is necessary to conduct further

research to increase their interpretability.

4. **Real-time Monitoring and Feedback:** Most contemporary studies rely on older historical data to predict mental health issues, with very little consideration for current real-time monitoring technologies. Real-time wearable monitoring technology, or smartphone apps may provide students and medical practitioners with almost continuous feedback. Implementing these kinds of strategies may allow for timely intervention to stop or avoid further mental health decline..
5. **Privacy and Ethical Issues:** Many publications fail to discuss the ethical implications of applying private student data in mental health prediction models. There are still substantial concerns regarding consent, data privacy, and the potential of algorithmic biases. It will be imperative to research how to navigate these ethical dilemmas, and in particular, how to ensure that data are appropriately managed and in respect of privacy legislation..
6. **Emphasis Upon Mental Health Conditions:** Although most of the existing research has been on discrimination against mental health conditions, such as stress, anxiety, and depression, there has been little work on forecasting early intervention or where prevention may occur. Potential areas for further research may be the development of predictive models identifying at-risk children but also that suggest implementable treatments that may bring down stress levels and increase overall wellbeing.
7. **Merging Educational Systems:** We are aware of little research looking at ways to merge machine learning models into existing support systems in education. Most of the research only seeks to focus on the prediction itself, and has little interest in how they might be used in academic advising, student support services, and counseling. Further research may be necessary to examine the use of the predictions to create proactive support systems in educational institutions.

Table 2.2: Gap Analysis

Area	Present Situation	Suggested Contribution
Sources of Data	mostly survey-based data with little variation	Use a range of data types, including social media usage, physiological data, and real-time behavioral data, to increase forecast accuracy.
Quantity of Samples	Small datasets that are region-specific	To increase model accuracy and generalizability, use more and more varied datasets.
Model Interpretability	Many models, including Random Forest and XGBoost, are opaque.	Create more comprehensible models that can help medical professionals and educators understand forecasts.
Monitoring in Real Time	Mostly static techniques for gathering data	Use wearable technology and smartphone apps to

		monitor in real time for continuous feedback and early intervention.
Moral Issues	Inadequate attention to algorithmic biases, consent, and privacy	To guarantee responsible data use and model fairness, investigate privacy issues and ethical norms.
Strategies for Prevention and Intervention	Predicting interventions or preventive actions receives little attention.	Create models that suggest remedies and preventative measures in addition to predicting mental health problems.
Combining Educational Systems	Models applied apart from systems of support for education	For proactive treatment, look into ways to incorporate predictive models into academic advising, university counseling, and student support programs.

## 2.4 Summary

In this passage, the deficiencies of the extant research on the application of machine learning to predict student mental health have been discussed. These deficiencies consist of a lack of diversity in the data, small sample sizes, interpretability of the model, real-time monitoring systems, ethical concerns, a lack of focus on intervention/prevention strategies etc. Moreover, there is insufficient research on integrating machine learning models within higher education support systems. The gap analysis underlined a need for greater diversity of data sources, larger and representative data sets, interpretable models, and real-time feedback systems. The gap analysis also pointed out the importance of addressing ethical dilemmas surrounding algorithmic fairness and data privacy. This study hopes to fill these gaps to develop more accurate, applicable, and ethically sound student mental health prediction models. This discussion sets the stage for the next phase of the study which will focus on addressing these concerns and utilizing machine learning in a more effective and applicable manner to support student health and wellbeing.

# Chapter 3

## Research Methodology

### 3.1 Methodology/Requirement Analysis & Design Specification

#### 3.1.1 Overview

This research project uses classification machine learning algorithms for the purpose of predicting student mental health. 1051 students completed a Google Form regarding their academic, personal, and psychological criteria including CGPA, study hours, financial stress, sleep quality, and suicidal thoughts. Upon completion of data collection, the raw data went through a preprocessing stage including cleaning, dealing with missing values, and removing irrelevant features. Feature selection approaches were used to obtain notable predictors. Multiple machine learning models were trained and evaluated on accuracy, precision, recall, and F1-scores. The aim was to find the most successful model to identify students that are at risk for poor mental health.

#### 3.1.2 Proposed Methodology/ System Design

To predict a student's mental health condition, a structured machine learning pipeline was followed. First, a Google Form was developed to collect data points that included 21 relevant components and used with 1051 students. Once students' data was collected, it was pre-processed so that outliers, irrelevant columns and missing values were eliminated before preparing the data for feature engineering and selection of important characteristics or patterns. The machine learning models were trained with multiple algorithms including Support Vector Machine, Random Forest and Logistic Regression, assessing each model after pre-processing was completed. Each model's performance metrics were assessed with common measurements including accuracy, precision, recall and F1 score. The best performing model was noted for the creation of a reliable prediction system to identify at-risk students so that timely intervention could occur.

### 3.1.3 Functional and Nonfunctional Requirements

This proposed student mental health prediction system has a number of basic functional components. It starts with a data input module that allows us to upload organized student survey data that is in CSV format. After the data is uploaded, the system will undertake preprocessing steps such as managing missing values, handling outliers and removing irrelevant features. The system will then apply feature selection methods to identify key features that are relevant to predicting student mental health. The system can use a variety of machine learning models to train and predict mental health disorders by classifying students as either "Healthy" or "At Risk". Next the models can be evaluated, using standard metrics such as accuracy, precision, recall, and F1 score. Optionally the system can have a user-friendly interface for uploading data and viewing the prediction results.

As far as non-functional requirements are concerned, the system must aim for a high prediction accuracy of 80% or higher. The system should strive to be easy-to-use so that non-tech-savvy people could use it. The technology also needs to be appropriate for scaling so that larger data sets can be studied in the future. Reliability is essential to enable consistency or reproducibility across runs and accuracy of results. Additionally, performance is essential as the system needs to generate predictions and evaluations in a reasonable amount of time. Finally, maintainability is a leading consideration and the code, that is modular in design, needs to be able to be modified and updated as required.

### 3.1.4 Context Diagram

A Context Diagram is a structural representation that describes the boundary of a system and its exchanges with outside entities, including users, other systems and/or databases. A Context Diagram presents the system as a single process, thus representing data movement to and from the system and actors external to the system. Context diagrams highlight inputs and outputs, but do not describe inner processes. It serves to outline the system's boundaries, identify stakeholder external from the system as well as provide a conceptual understanding of what system interacts with what. Context diagrams are typically created in the early stages of systems design to clarify the overall picture of the system itself.

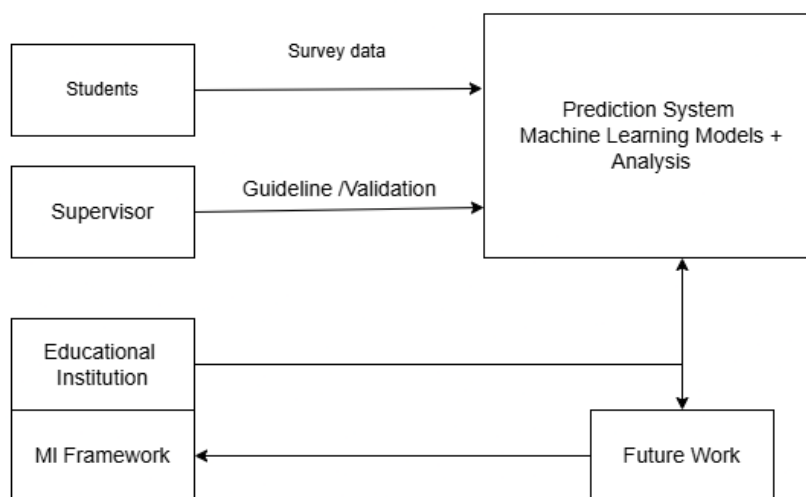


Figure 3.1: Context Diagram

### 3.1.5 Data Flow Diagram Level 1

The Data Flow Diagram (DFD) is applied to show the flow of data in a system and provides a diagram that shows the processes of the data handling. DFDs provide a pictorial representation of the variables of a specified system such as data sources, processes, data

storage and outputs all connected by arrows to display the passage of data between systems components. The DFD illustrates from where the data started, we begin with Data Collection which is the collection of raw data.

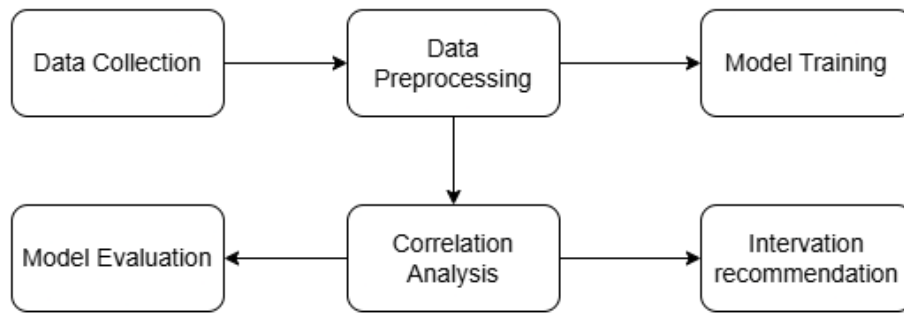


Figure 0.1: Data flow diagram

It then goes into Data Preprocessing where the data is cleaned and ready for analysis. Model Training is the next step where algorithms are deployed to develop predictive models. The flow moves to Model Evaluation, where the performance of the trained models is evaluated. After that, the Correlation Analysis deals with the issues of relationships between variables. Lastly, the results result to Intervention Recommendations wherein insights are applied to come up with suggestions for improvement or those that should be done.

### 3.2 Detailed Methodology and Design

The data for this study was collected using a Google Spreadsheet titled "Student Mental Health". A standard survey was developed to gather information about students' demographics, academic performance, financial stress, sleep patterns, and overall mental health.

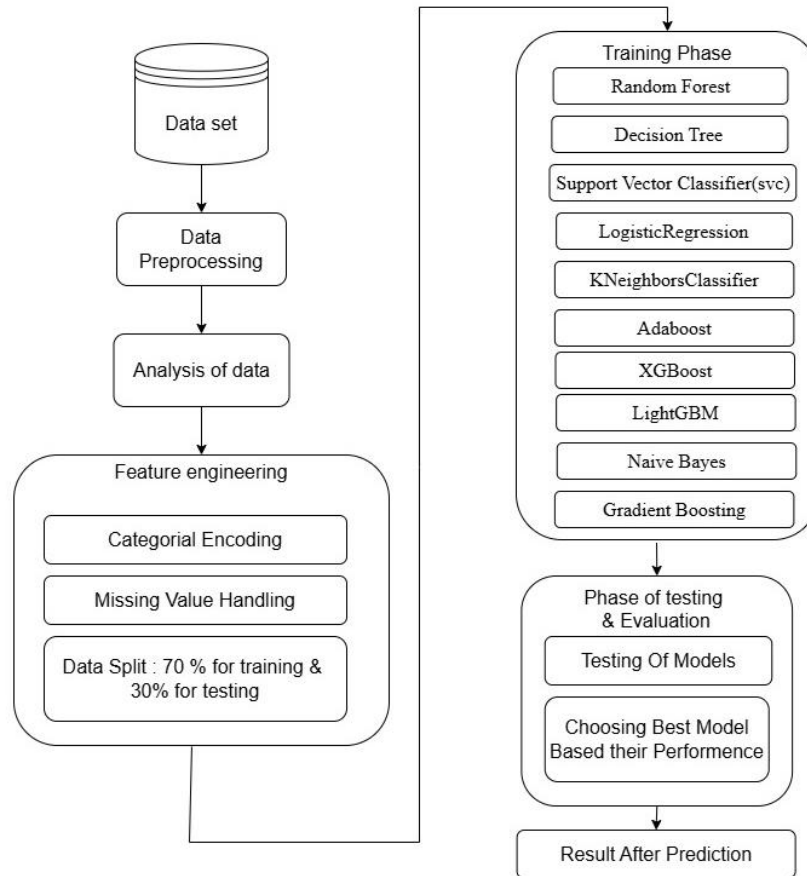


Figure 3.3: This is a sample diagram

The spreadsheet The URL was shared with Daffodil International University students enabling them to access and enter in their input. This allowed for an easy and quick project information gathering process while maintaining accuracy and completeness. The answers were being recorded slowly in a Google spreadsheet, maintaining an ordered and structured form of writing. Once the data were gathered, the spreadsheet was downloaded as a CSV file for other processing and analysis. The data set contains useful variables, including gender, age, cgpa, academic stress, financial stress, sleep duration, and general student mental health status, which provides a solid basis for leveraging machine learning techniques to predict students mental health issues.

To analyze the data we must first perform a preprocessing step. Data preparation is an essential step in machine learning process. It is the act of transforming raw data to a format that is more suitable for analysis and modeling. The most common goal of preprocessing is to clean and structure the data, which results in removing redundancies and irrelevant data. Common tasks associated with preprocessing, are handling missing values, encoding categorical variables, scaling numerical features, and splitting the data into training and testing sets. Typical preprocessing improves machine learning model performance and accuracy. Preprocessing Techniques Used:

- **Label Encoding:** Label encoding is the process of assigning a unique integer to each category in order to translate categorical variables into numerical values. It is often used for machine learning algorithms that require numerical input. [45]
- **Feature Selection:** An important step in machine learning is feature selection, which involves selecting some subset of relevant features out of the original feature set in order to reduce the feature space and improve model performance using less

processing power and time. This is an important step in machine learning to begin with, when working with large data sets. [46]

### **3.3 Project Plan**

The project was broken down into a number of planned phases that allowed for organized development and completion. First, a thorough background study was conducted to explore and recognize processes related to student mental health and areas where previous studies were lacking. A Google Form was created and shared with students to gather relevant information about students, including academic characteristics, psychological characteristics, and lifestyle characteristics.

Once there was enough data collected, the preprocessing phase involved cleaning the dataset, addressing missing values, eliminating outliers, and determining relevant features. The next phase was to build machine learning models, which involved taking various algorithms on the processed dataset in order to predict mental health disorders. Each model was evaluated with performance metrics to examine which performed best.

Ultimately, the data was analyzed, visualized, and documented. The project followed a weekly milestone timeline structure to keep track of all the data collection, preprocessing, model building, and evaluation so that it was done efficiently and on time. Continuous tracking of progress and modifying timelines where necessary was done, so that the outcome kept advancing and relevant to the research.

### **3.4 Summary**

In this section, we delineated the complete method for predicting student mental health through machine learning. 1051 students completed a standardized Google Form that asked about academic, personal, and psychological criteria. The dataset was cleaned and organized, deciding on a robust feature set for modelling. Multiple machine learning approaches were trained and evaluated using criteria such as accuracy and F1 score. A detailed project plan supervised each component of the work to ensure timely completion. Overall, this methodology provided a methodical way to develop a robust mental health prediction system.

# Chapter 4

## Implementation and Results

### 4.1 Environment Setup

I created the environment by way of Visual Studio Code (VSCode) and Kaggle. I set up Python first and imported all of the required libraries (pandas, numpy, matplotlib, seaborn, and sklearn). These various packages will assist me with data preprocessing, creating visualizations, and developing the models. I will be using VSCode for most of the development on my local machine, and I will also use Kaggle since they offer GPUs which will help out with larger datasets. The initial development in this configuration is how I setup the machine.

### 4.2 Testing and Evaluation/Performance/ Comparative Analysis

Using various performance metrics such as accuracy, precision, recall, and F1, a comparison analysis was performed to assess the effectiveness of a number of machine learning algorithms for a classification task. The results generated from each of the models are collected in the following table:

Table: 4.1 Accuracy for every model

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8066	0.81	0.81	0.81
Random Forest	0.9811	0.98	0.98	0.98
SVM	0.8396	0.84	0.84	0.84
Gradient Boosting	0.9858	0.99	0.99	0.99
KNN	0.8538	0.85	0.85	0.85
Decision Tree	0.9906	0.99	0.99	0.99
AdaBoost	0.7689	0.76	0.76	0.76
XGBoost	0.9906	0.99	0.99	0.99
LightGBM	0.9717	0.97	0.97	0.97
Naïve Bayes	0.7925	0.79	0.79	0.79

The outcome indicated that Decision Tree, XG Boosting and Gradient Boosting were the models performing the best with nearly perfect accuracy, as well as good precision, recall and F1 scores. These models show good performance and generalization, as well

as resilience across all the classes. Random Forest and LightGBM both performed well, but just slightly behind the top 3 models. In contrast, AdaBoost and Naive Bayes performed poorly, notably in recall and F1-score, demonstrating difficulties in dealing with class imbalances or complex patterns. Logistic Regression and SVM performed moderately, producing respectable but lower metrics. This comparative analysis demonstrates that ensemble approaches, particularly XGBoost and Gradient Boosting, are highly effective at this classification job, making them appropriate for deployment or further optimization.

### 4.3 Results and Discussion

#### Auc Roc curve & Confusion Matrix for all models

A graphical representation used to evaluate a classification model's efficacy is the AUC-ROC curve (Area Under the Curve - Receiver Operating Characteristic). At different cutoffs, the ROC curve compares True Positive Rate (TPR or sensitivity) against the False Positive Rate (FPR or 1-specificity). AUC (Area Under Curve) score could have values in a range from 0 to 1, where 0.5 is random guessing and 1 is very good classification. A higher AUC score suggests a better model, and it is especially informative in cases where the classes may be imbalanced. From this curve, we can assess how well the model did to separate a positive class from a negative class. Another method to evaluate performance on classification problems is by using the Confusion Matrix. The Confusion matrix allows for a summary of the model's actual classifications against the anticipated classifications in a tabular representation. The confusion matrix has four values T P = True Positives, T N = True Negatives, F P = False Positives and, F N = False Negatives. In addition, several additional performance measures such as accuracy, precision, recall, and F1-score can be calculated from the confusion matrix. Confusion matrices are a very important diagnostic tool to improve model performance. They help to identify whether the model is confusing two classes and they provide insights on what type of error is occurring.

#### Logistic Regression:

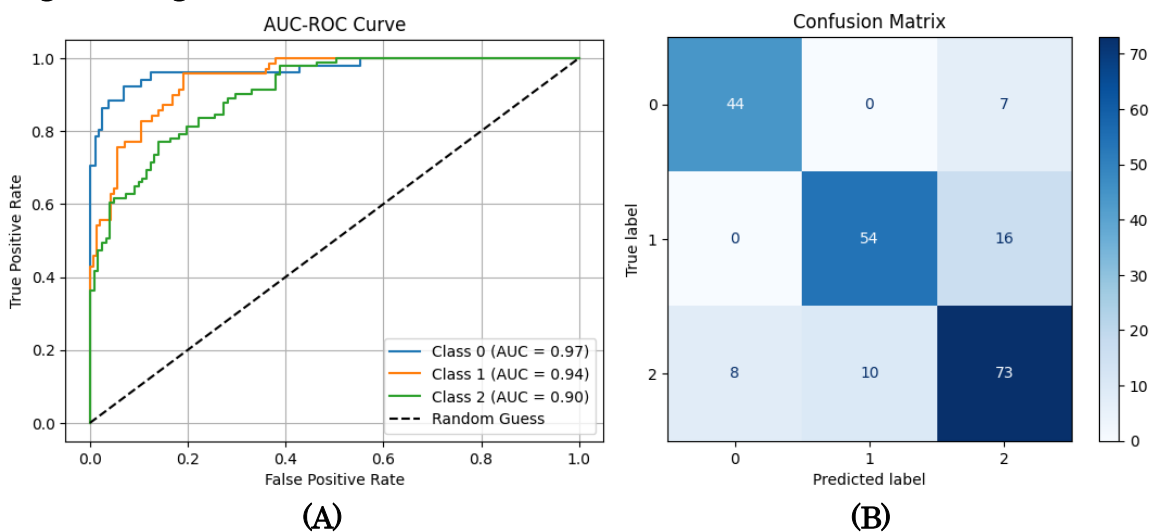


Figure 4.1.1: Auc Roc Curve (A) & Confusion Matrix (B) for Logistic Regression

## Random Forest:

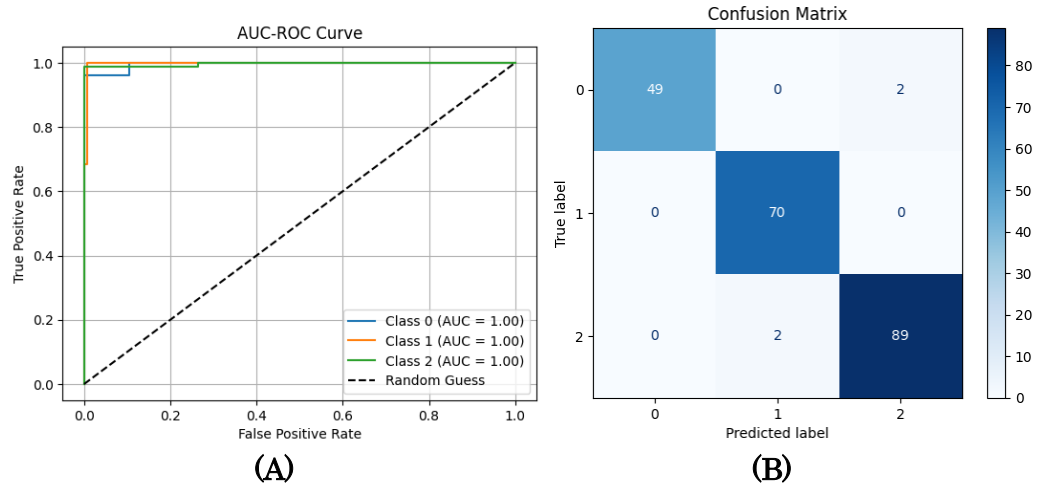


Figure 4.1.2: Auc Roc Curve (A) & Confusion Matrix (B) for Random Forest

## SVM Accuracy:

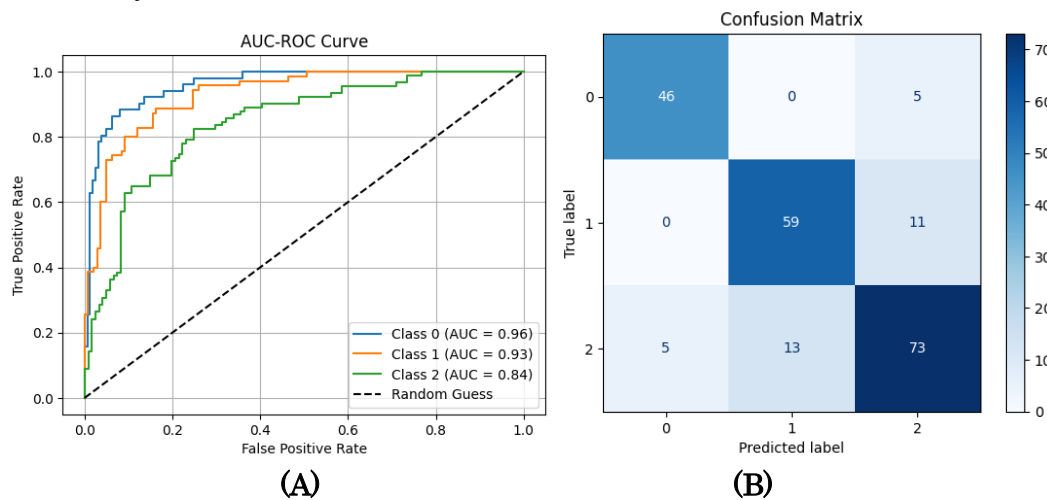
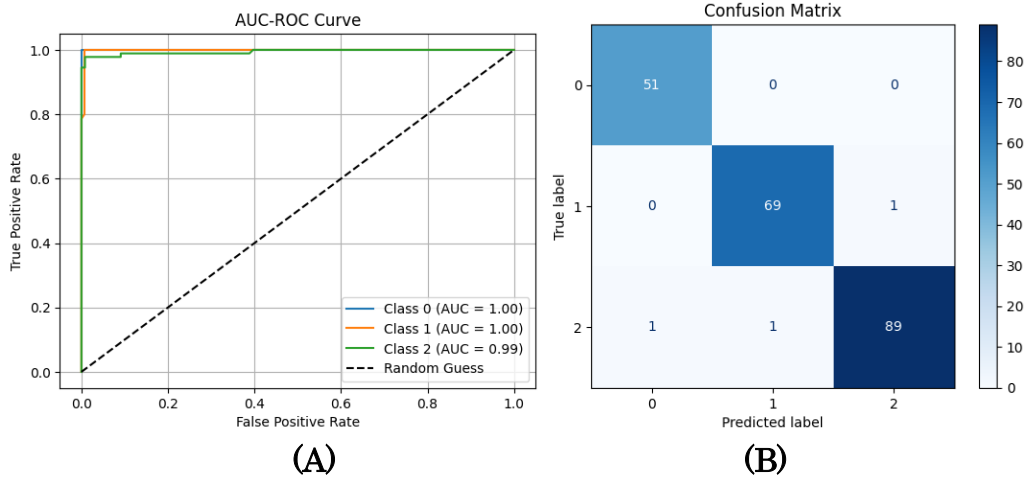


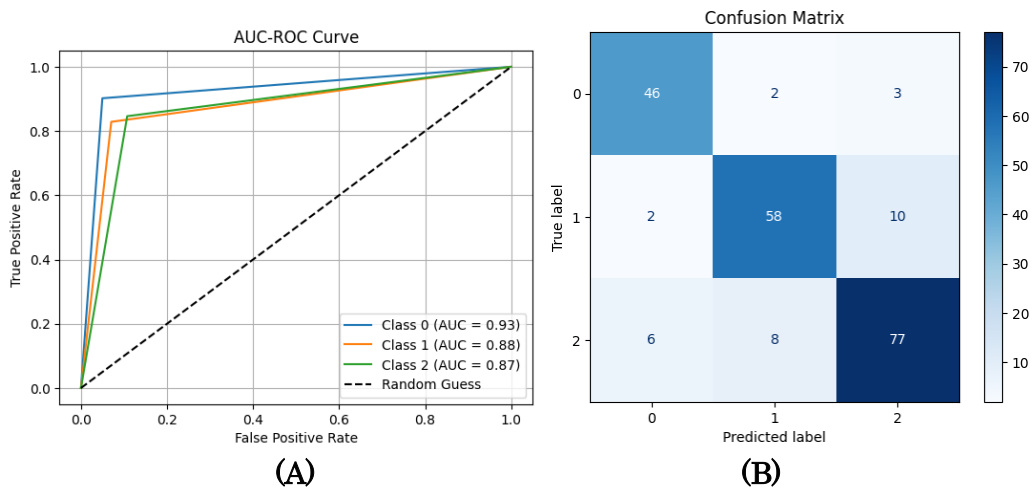
Figure 4.1.3: Auc Roc Curve & Confusion Matrix for SVM

## Gradient Boosting



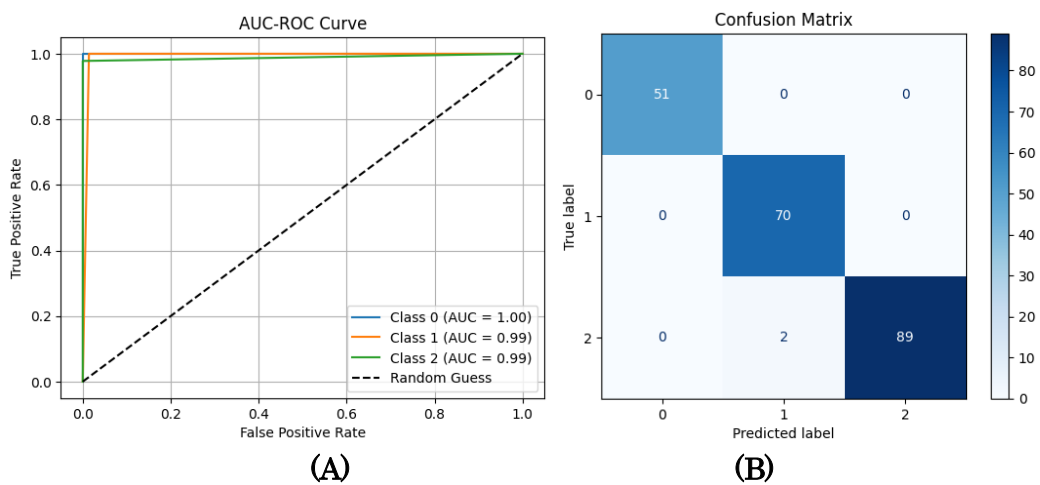
**Figure 4.1.4: Auc Roc Curve (A) & Confusion Matrix(B) for Gradient Boosting**

**KNN Accuracy**



**Figure 4.1.5: Auc Roc (A) Curve (B) for KNN**

**Decision Tree :**



**Figure 4.1.6: Auc Roc Curve & Confusion Matrix for Decision Tree**

**AdaBoost:**

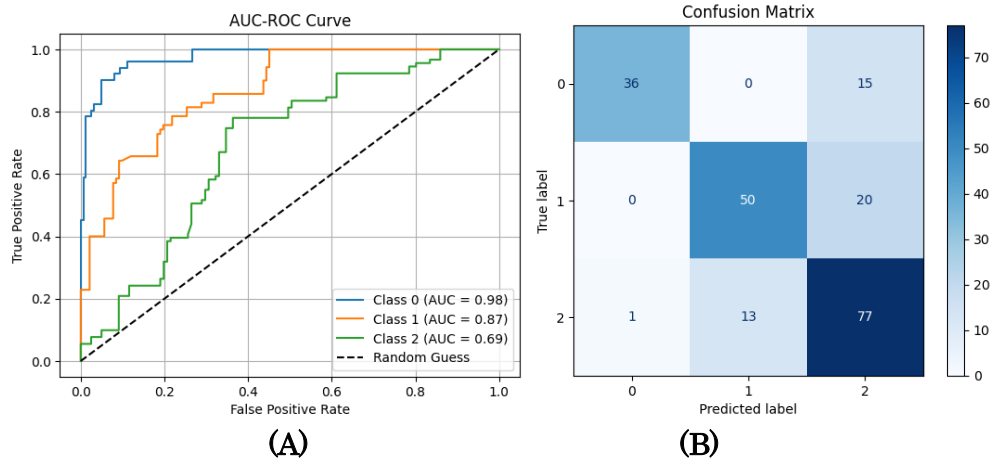


Figure 4.1.7 Auc Roc (A) Curve & Confusion (B) Matrix for AdaBoost

### XGBoost Accuracy

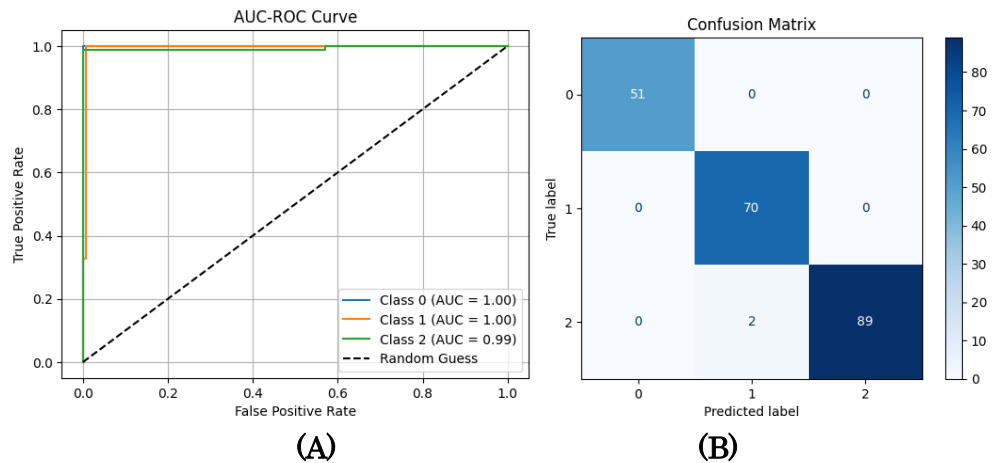


Figure 4.1.8 Auc Roc (A) Curve & Confusion (B) Matrix for XGBoost

### LightGBM Accuracy:

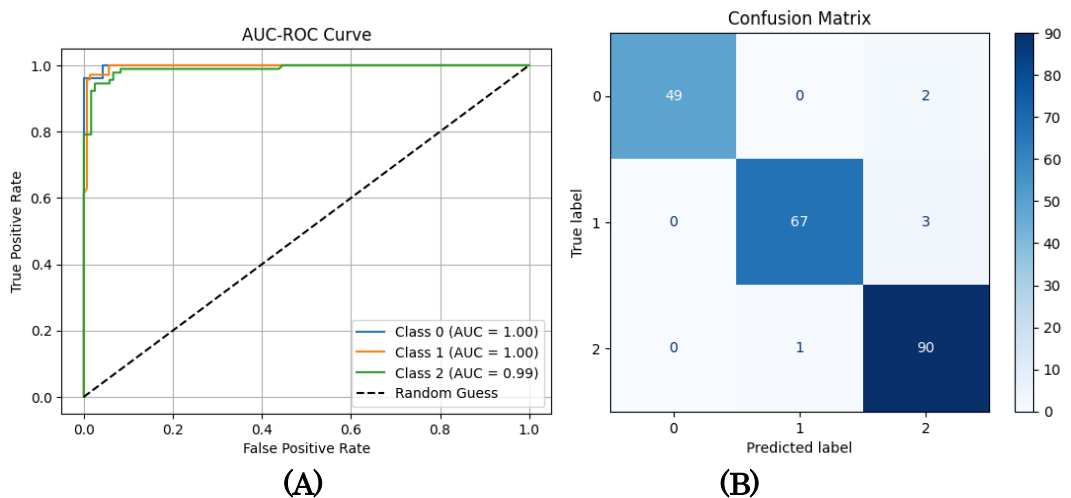
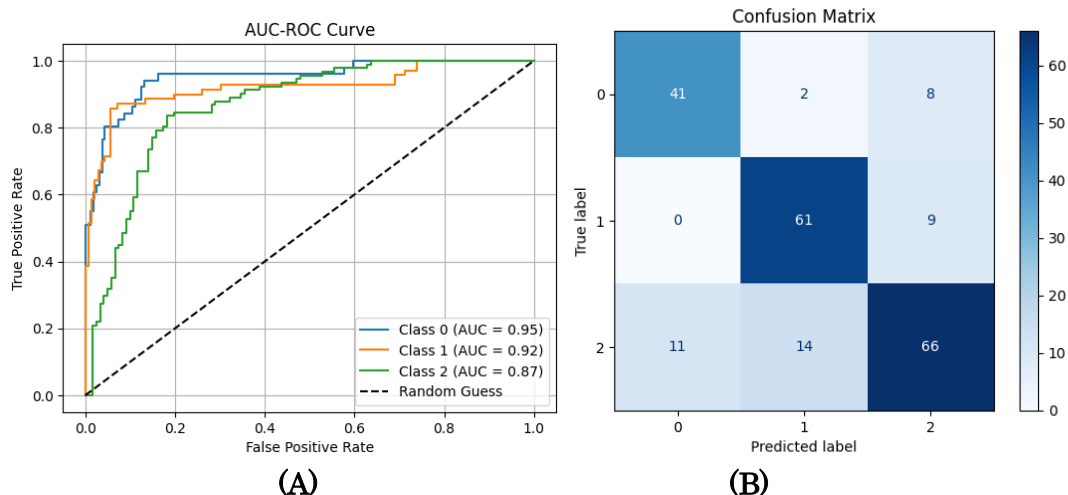


Figure 4.1.9 Auc Roc (A) Curve (B) for LightGBM

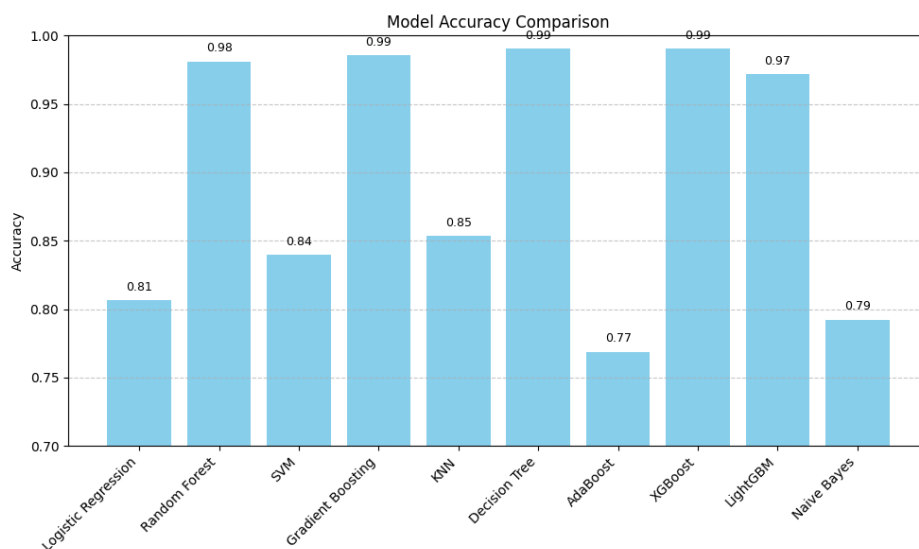
### Naive Bayes Accuracy



**Figure 4.1.10 Auc Roc (A) Curve (B) for Naive Bayes**

### Accuracy for all Models:

The data visualization shows the accuracies of the different machine learning models. Of all the models analyzed, the XGBoost, Decision Tree and Gradient Boosting model had the highest accuracies (0.99), indicating they performed best. The Random Forest and LightGBM models both had similar levels of accuracy at 0.98 and 0.97 respectively, while AdaBoost had the lowest accuracy (0.77) followed by Naive Bayes (0.79). The traditional methods of Logistic Regression, SVM, and KNN all performed at mid-level accuracies (0.81 - 0.85). Looking overall, idea to use ensemble-type methods (e.g. Gradient Boosting, Random Forests, XGBoost) are better suited for prediction jobs that require a high level of accuracy. The accuracy numbers for the models are listed above each bar graph for easier comparison.



**Figure 4.1.11 Model Accuracy Comparison**

## 4.4 Summary

In this chapter, we introduced and evaluated a number of different machine learning models, with the goal of identifying a model that can predict the target variable effectively

and efficiently. The different models utilized were Logistic Regression, Random Forest, SVM, Gradient Boosting, KNN, Decision Tree, AdaBoost, XGBoost, LightGBM, and Naive Bayes. After training, we assessed each model with important measures of performance including accuracy, a confusion matrix, and AUC-ROC. XGBoost, Decision Tree and Gradient Boosting had the best accuracy (0.99) out of all of the models, which indicates a strong ability to predict. Overall, ensemble methods had better results than non-ensemble methods, which was expected! The AUC-ROC curves provide visual displays of how the models classify across thresholds and the confusion matrix presents how many genuine and projected classifications to help identify misclassifications. It is clear through this study that the ensemble models are superior to the others with every conceivable measure assessed, and this is why they are the most appropriate for this classification study.

# Chapter 5

## Engineering Standards and Design Challenges

### 5.1 Compliance with the Standards

In the pursuit of reliability, maintainability, and extensibility, the project follows accepted hardware, software, and communication standards. Python has been utilized, PEP-8 rules for code readability have been adhered to, and widely adopted libraries for machine learning have been incorporated, such as pandas, scikit-learn, and TensorFlow, in conjunction with standardized testing methods integrated into the entire process that meet quality assurance standards. Admittedly, future iterations could include enhancements in GPU support, but the hardware setup of laptops and servers (with sufficient processing power) allows for reasonably effective model training and associated predictions. Clear and consistent documentation, secure communication with key stakeholders, and an interface design that is intuitive and accessible to the user groups further enhance communication standards. As will be further discussed in the next section, in addition to following best practices in software engineering, ensuring reliability in performance and overall user satisfaction, these standards will ensure extensibility, advancement and maintenance going forward.

#### 5.1.1 Software Standards

To ensure the quality and reliability of the system developed to predict students' mental health, the project follows recognized principles of software engineering. For example, using version control (e.g., Git) means the code is stored in a tidy manner, as well as that the developer can collaborate with others more easily. All machine learning models used within the project, including Random Forest and Support Vector Machines (SVM), are developed in Python, which is compliant with PEP-8 (Python Enhancement Proposals) codes for coding style and code readability. The adherence here ensures the project is able to scale and will be easily maintained in the future as enhancements accrue.

Testing methods are also used to ensure every part of the system (e.g., data preprocessing, model training, evaluation, etc.) are all working well - by including testing techniques such as unit and integration testing. In addition, the project includes several popular libraries, such as scikit-learn, TensorFlow, and pandas —which each have their own codes, and used as .. expected basis and foreseen implementation for their use cases in machine learning..

#### 5.1.2 Hardware Standards

The vast majority of the hardware used for this project consists of standard computing devices such as laptops and servers with adequate specifications to run machine learning models in a reasonable amount of time. With an emphasis on providing adequate computing power for the training and prediction phase of the models, the project utilizes whatever CPU and GPU is available with the environment configuration (e.g., Visual

Studio Code and Kaggle) to run machine learning models. The hardware is configured to provide adequate data processing, no performance limitations, and to meet the minimum system requirements for executing sophisticated algorithms such as Random Forest and XGBoost.

While the hardware configuration and configuration is adequate for the size of the project, future iterations may benefit from greater GPU accessibility to make efficient use of increasingly complex deep learning models.

### **5.1.3 Communication Standards**

Communication standards are very important for any group project. Part of being able to guarantee clearly and consistently documented development of the project is adherence to communication standards for this research project. An aspect of this includes documenting comprehensive project specifications to support requirements analysis, methods, and the outcomes as discussed above. In addition, the university collaboration environment relies on professional and secure communication methods with stakeholders which included students who were involved in the data collection and supervisors from the university. In this context these methods included email, project management software, and direct communication (face to face or phone interactions).

Additionally, the project has met the communication requirements imposed on the user interface. The user interface (UI) permits end users (administrators and university employees) to access the system with very little training required, thanks to its clarity and simplicity. Also, the design follows the basic standards for web applications and is inviting for usability and accessibility.

## **5.2 Impact on Society, Environment and Sustainability**

### **Impact on Life**

The creation and application of machine learning algorithms for predicting mental health is improving the lives of individuals. The system provides timely interventions with significantly elevate students' ability to perform academically and improve life skills by enabling the earlier identification of stress, anxiety, and other mental health concerns. Students using proactive mental health services are more easily able to navigate stressors, maintain positive relationships, and perform well academically all of which add to their overall success and satisfaction in life. In addition, the algorithmic approach promotes healthier learning environments and reduces the stigma surrounding mental health care since it places importance on mental health awareness in educational settings.

#### **5.2.1 Impact on Society & Environment**

This initiative promotes benefits to society by addressing the growing issue of students with mental health problems, particularly young adults facing social isolation, financial issues, and academic challenges. By promoting proactive mental health intervention through educational centers and machine learning models - the initiative provides benefits for the individual wellbeing of students and greater social context. Institutions of higher education can enhance retention rates and build positive cultures that impact and improve academic performance by identifying students at risk and providing supplementary interventions. Distance counseling also offers potential indirect environmental outcomes, decreasing the carbon footprint of in-person meetings and facilitating better virtual mental health support networks.

### 5.2.2 Ethical Aspects

The project includes a strong focus on ethical considerations, particularly in relation to data privacy and the responsible handling of individual data. There is a high size of confidentiality for sensitive data used within the mental health prediction system, including behavioral patterns, academic performance, and psychological assessments. The project commits to ethical research and data protection policies, obtained informed consent from all participants, and anonymized data. The project will mitigate algorithmic bias by using multiple datasets and supporting equitable forecasting. At all stages of the process, accountabilities, and transparency have been exercised to ensure the validity and ethics of the mental health models.

### 5.2.3 Sustainability Plan

The adaptability to changes in machine learning and data collection methods during the applications of this project is essentially its sustainability. This application is designed to be flexible and scalable and is intended to eventually contain more datasets and more sophisticated algorithms. This project aims to incorporate real-time monitoring technologies (wearables or smartphone apps) so that there is continual feedback and intervention to promote sustainability in the long run. This project also hopes to decrease long-term societal costs of undisclosed mental health conditions through the promotion of early intervention and preventative care thereby creating a more sustainable status quo within the culture. The sustainability plan not only addresses new ethical implications of data privacy and transparency of the algorithms, it also requires regular updates to ensure that the models remain relevant and effective.

## 5.3 Project Management and Financial Analysis

The budget for the project includes everything necessary for execution including adequate allocations for every phase to be able to continue moving forward effectively. Allocations for hardware and infrastructure needs and software and tools acquisitions and data collection and processing and reporting and writing documentation and more. The project's desired results require all resources and activities which belong to the established categories. Actual expenses might differ depending on project complexity as well as shifting project requirements.

Table 5.1: Financial Estimation of the Project

SN	Components	Estimated Cost
01	Hardware/Infrastructure	1000-1500
02	Software and Tools	2000-4000
03	Data Collection and Processing	4000-5000
04	Documentation and Report Writing	1000-2000
05	Miscellaneous	2000-3000
Total Estimated Cost		10000-15500

## 5.4 Complex Engineering Problem

### 5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowled ge	EP2 Range Of Conflicting Requireme nts	EP3 Depth of Analys is	EP4 Familiari ty of Issues	EP5 Extent of Applicab leCodes	EP6 Extent Of Stake- holder Involveme nt	EP7 Interdepende nce
✓		✓				✓

#### ➤ EP1: Dept of Knowledge

The research exhibits a deep understanding of behavioral data analysis and machine learning techniques. By using a variety of methods, including Random Forest, XGBoost, and Support Vector Machines (SVM), it addresses difficult problems related to student mental health prediction. This degree of proficiency guarantees command of sophisticated predictive models, enabling the early identification of mental health issues.

#### ➤ EP3: Depth of Analysis

Developing and honing strong machine learning models to forecast student stress and smartphone addiction is part of the depth of investigation. This study seeks to identify significant associations that advance knowledge of students' psychological well-being by examining a variety of datasets, including academic, behavioral, and demographic factors.

#### ➤ EP7: Interdependence

The interdependence in the project is clear, as the machine literacy models would not be usable without various data sources such as academic performance, nightly sleep, and smartphone behaviours. In order for the models to be successfully embedded and used in student support roles, the technical teams, stakeholders and educational groups would need to work together to build a predictive system that brings all of these into play.

This format aligns with the elements and details of your project, and follows the same pattern as the example. Please let me know if you want to add or change anything!

### Mapping with Knowledge Profile for EP1

This table 5.2) is designed to map the EP1 to the Knowledge Profile.

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

➤ **K3:Engineering Fundamentals]**

The project uses engineering principles such as data collection methods, designing of algorithms and statistical analysis. These principles are fundamental to developing machine learning models in the prediction of student mental health, especially with regard to stress and smartphone addiction. Methodical approaches to solving problems ensures that engineering knowledge is incorporated into the development of the predictive models.

➤ **K4:Specialist Knowledge**

This investigation utilize specific experience with machine learning techniques, specifically machine learning methods within the area of behavioral or psychological data. The implementation of algorithms like Random Forest and Support Vector Machines (SVM) demonstrate a deep understanding of predictive modeling as it relates to a mental health prediction. This exact experience allows for the rapid identification of students at risk.

➤ **K5:Engineering Design**

The overarching aim of the design phase of this project was to ensure a robust platform for data preprocessing, model training, and prediction. The technology will use machine learning algorithms such as Random Forest and SVM to predict student stress and smartphone addiction, and offer effective solutions for mental health.

➤ **K6:Engineering Practice**

The practical utilization of predictive models and data-driven decision making are examples of the practice of engineering. Utilizing machine learning models to make real-time predictions illustrates how engineering principles can be practically applied to student mental health management.

➤ **K8:Research Literature**

To ensure the project is grounded in contemporary academic work and makes a contribution to the field, and that it uses advanced machine learning techniques to predict student mental health, it is able to utilise a vast amount of research literature in machine learning, psychology, and mental health prediction models, especially Rahman et al. (2023) and Baba & Bunji (2023).

### 5.4.2 Engineering Activities

Table 5.4: Mapping with complex engineering activities.

EA1 Range of re- sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	

➤ **EA1:Range of re- sources**

In order to support model training, this study provided a range of tools, such as survey data, machine learning approaches, and cloud-based solutions like Kaggle GPUs. These tools supported the creation of viable predictive models, which could identify stress and student smartphone addiction, process data rapidly, and estimate with accuracy.

➤ **EA2:Level of Interaction**

Collaboration was needed significantly during the entire model creation process. Data collection, data preprocessing, model evaluation, and model optimization were completely dependent on collaboration. Each time, the team members closely worked together at every step of the process to improve and fine-tune the models including adding domain-specific knowledge. As a result, the prediction models produced were more consistent and effective models for stakeholders and the organization.

➤ **EA3: Innovation**

New ways of live prediction in real-world contexts are being organized in the form of mobile learning for predicting student stress and mobile phone addiction with machine learning models. Cloud computing is being used with Explainable Artificial Intelligence (XAI) for testing and training models, which is moving the technology toward use in mental health applications. This strategic plan promotes the possibility for mental health predictions at educational institutions, while fostering a culture of continuous improvement.

➤ **EA4:Consequences for society and environment**

The application of established usages and methodologies, or borrowed methods, which are implemented in existing operation, would enable predictive outcomes of psychological and behavioral disorders because of established familiarity with the usual procedures, would result in consistent outcomes. In addition, we would ensure adherence to the best practices will result in high-quality estimates and useful output that is essential for proactive management of students Mental Health.

This methodological approach to engineering tasks is focusing on the essential materials, links, and creativity, that have taken this project space..

## 5.1 Summary

Using machine learning to predict student mental health is a difficult engineering problem that requires a comprehensive approach that uses a variety of data sources, models, and provides an understandable discussion of ethical issues. Predictive model training, data collection and preparation are examples of specific engineering tasks that comprise part of the solution. The purpose of this research is to create a model that can identify children at risk and provide actions to support their wellbeing, using machine learning methodologies. The investigative project addresses the rising need for mental health support in educational environments, while ensuring that the system remains efficient, scalable, and accountable through careful consideration of model performance, ethical issues, and sustainability.

# Chapter 6

## Conclusion

In this academic funding research, machine learning techniques were explored to predict students' mental health by explicitly measuring levels of stress and smartphone addiction. Stress, anxiety, and depression are all increasingly common mental health conditions among students all over the world, which dramatically impact students' overall health and academic outcomes. The implications of this research accentuate the importance of early prevention and the potential use of machine learning in identifying children at risk before adverse effects of mental health emerge.

Machine learning models were used, including Random Forest, Support Vector Machines (SVM), Decision trees, and Gradient boosting, to predict student mental health using features such as academic stress, financial stress, sleep, and smartphone usage. The results demonstrate that machine learning can predict mental health issues with high accuracy, and of the models, Gradient Boosting and Decision Trees achieved the best performance, reaching near perfect accuracy, precision, and recall. This indicates that these models could be beneficial for observing student behaviour to identify students who may be struggling with mental health issues.

The study also underscored the importance of different factors contributing to student stress and smartphone addiction. Academic pressure, lack of social support, and lack of sleep were strong predictors of stress. Excessive smartphone use and digital addiction were also found to be negatively correlated with students' mental health issues. Understanding these relationships can help schools develop contextual interventions such as counselling services, time management workshops, or digital detox programs.

The potential scope of this research is vast. By adding machine learning prediction systems to current student support services, universities can allow for more proactive care and enhance the student experience overall. In addition, early identification of mental health problems may result in earlier treatment engaging students effectively before the conditions become more severe. Ultimately, this may improve students overall performance and wellbeing.

The study had limitations, including the relatively small size of the dataset and potential biases in self-reported data, and future research could explore larger datasets and more diverse datasets, including behavioral or physiological data, to improve prediction accuracy and prognostic utility. Furthermore, applying deep learning models with dynamic monitoring systems could also bolster the accuracy of mental health predictions.

Conclusively, machine learning offers great potential for predicting student mental health. Institutions of higher education can facilitate proactive measures, through data analytics, to improve student mental health, creating a healthier and ultimately more productive learning environment.

## 6.1 Summary

This study employed machine learning methods to aid in predicting students' mental health with an emphasis on stress levels and smartphone addiction. The study attempts to understand the growing concerns relating to mental health issues faced by students, which can heavily impact their overall health and academic performance. The research used multiple types of machine learning models (Random Forest, Support Vector Machines (SVM), Decision Trees, and Gradient Boosting) to predict mental health issues from features like academic pressure, financial stress, sleep habits, and smartphone addiction.

The results from the study showed that machine learning can successfully identify students that may be at risk for mental health issues. In the study Gradient Boosting and Decision Trees performed better than other models, as they offered good accuracy, precision, recall, and F1-scores indicative of their possible use in practice. Furthermore, the results of correlation analysis provided means to receive treatment where it will likely be most useful as it uncovered major features causing stress and addiction, including excessive adherence to a smartphone, a lack of social support, and academic burden.

As a part of maintaining the rationale for early interventions, as the predictive models can more easily identify children who need intervention prior to children deteriorating in their mental health. If educational institutions can use machine learning tools to develop tailored support services, they could be more effective in delivering the care that they offer to students, and yielding better academic and emotional outcomes for students across the board.

While this study provided useful insight, it was limited due to a small sample size and biases in the data collection process. Future research should include larger and more comprehensive data sets and methods, including real-time behavioural and physiological data, in order to maximize the accuracy of the predictive model.

In summary, educational institutions can further improve their proactive care and support systems with the addition of using machine learning to predict mental health amongst students. Using machine learning may help foster a safer learning environment, to instil a positive and healthy learning environment.

## 6.2 Limitation

While this study provides important implications for the utilization of machine learning models to predict students' state of mind, there were various challenges that arose during data collection that will have to be remedied in future studies.

**Limited Dataset:** The study's 1051 assessments may not be sizeable enough to embody the different types of mental health-related challenges a larger student body would encounter. Historically, survey responses often have limited applicability due to sample sizes that do not accurately represent the full range of student experiences of a sample from other colleges, regions locally, or demographics. Therefore, the results may not be generalizable for the larger population.

**Data Bias:** The dataset utilized a survey self-reported data, which came with its own biases. Students may have misreported their stress levels, mental health issues, or smartphone use either intentionally, based on social acceptance bias, or unintentionally due to a lack of self-awareness. This would potentially influence the accuracy of the models and the reliability of predictions.

**Restricted Feature Set:** In this study, the main elements taken into account for prediction were behavioral, academic, and demographic characteristics. However, a variety of factors, such as physiological, cultural, and environmental aspects, have an impact on mental health. To increase prediction accuracy, future research should include other features including peer relationships, social media usage, and physiological data (e.g., heart rate, sleep quality).

**Data Privacy and Ethical Issues:** Collecting students' mental health confidential data raises major privacy and ethical concerns. Data breaches and misuse are always a possibility, even when strict confidentiality measures are in place. Future research must prioritize ethical concerns to ensure that data collection, storage, and use comply with ethical guidelines and privacy laws.

**Model Complexity and Interpretability:** While machine learning models such as Random Forest and Gradient Boosting models provide great predictive ability, they are often referred to as "black boxes," making it difficult to assess how they arrive at specific predictions. Ease of implementation into real-world contexts may be hampered by their lack of interpretability, especially in contexts that utilize the forecasts to guide interventions when the rationale is key to doing so.

**External Validity:** Training data on students from one particular student demographic was used to build the prediction models we developed as part of this project. The accuracy of the output/forecast may be flawed if similar models are applied to other student groups who are in different academic, social, or cultural contexts. More evaluation and adaptation to other contexts is required to ensure broader applicability of the models.

### 6.3 Future Work

While this study provides an interesting starting point to utilize machine learning to predict students' mental health, there are still a number of exciting directions for further exploration and development.

**Expanding the Dataset:** Future studies should seek to obtain data from a larger range of students from other universities, geographic locations, and socio-economic backgrounds. This would improve the model's generalizability and ensure that the mental health issues faced by a broader range of students are accurately represented.

Future work could also assess other factors, such as social media usage, peer pressure, and even physiological data (e.g., heart rates, sleep, exercise). In addition to behavioral and academic data, these factors could add a layer of comprehensiveness to our understanding of the factors impacting students' mental health.

**Improving Model Interpretability:** One of the primary challenges of applying machine learning to education is the lack of transparency of machine learning models. Future research could focus on exploring explainable AI techniques to better interpret models and support educators and health care professionals to better understand how predictions are

generated. In practice, doing so would increase informed decisions and foster trust.

**Real-Time Monitoring Systems:** By implementing real-time monitoring systems that monitor students' phone usage, and mental health, the accuracy of predictions may be increased dramatically. These technologies could provide continuous feedback, allowing for early interventions for at-risk students and stopping the deterioration of mental health conditions.

**Ethical Considerations:** Future research should address privacy and ethical issues associated with gathering sensitive data. Achieving and maintaining students' trust, and compliance with data protection laws requires that any gathered data is anonymized, securely stored, and evaluated in accordance with ethical guidelines.

**Integrating with University Support Services:** Future research should focus on combining machine learning based mental health prediction systems with existing university student support services. This could provide customized interventions tailored to improve students' academic performance and wellbeing, including academic support, stress management programs, and counseling.

# References

- [1] World Health Organization, “Mental health,” World Health Organization, Jun. 17, 2022. <https://www.who.int/news-room/factsheets/detail/mental-health-strengthening-our-response>.
- [2] S. BH et al., “Mental Health Analysis of Employees using Machine Learning Techniques,” 2022 14th International Conference on COMMunication Systems & NETWORKS (COMSNETS), Jan. 2022, doi: <https://doi.org/10.1109/comsnets53615.2022.9668526>.
- [3] The daily star report : <https://www.thedailystar.net/opinion/views/news/counselling-urgent-need-university-students-3294481>
- [4] Dhaka Tribune Report: <https://www.dhakatribune.com/bangladesh/313212/survey-86%25-of-students-suffer-from-mental-health>
- [5] Shahidul Islam Khan, A. Islam, Akther Hossen, Taiyeb Ibna Zahangir, and Abu, “Supporting the Treatment of Mental Diseases using Data Mining,” Oct. 2018, doi: <https://doi.org/10.1109/iciset.2018.8745591>.
- [6] A. Parekh, “Mental Health Prediction Using Machine Learning,” Analytics Vidhya, Jun. 10, 2022. <https://www.analyticsvidhya.com/blog/2022/06/mental-health-prediction-using-machine-learning/>
- [7] Healthinformaticsjournal.com, 2025. <https://www.healthinformaticsjournal.com/index.php/IJMI/article/view/162/162> (accessed Feb. 19, 2025).
- [8] World Health Organization, “Mental health: Strengthening our response,” WHO, 2022. [Online]. Available: <https://www.who.int/news-room/factsheets/detail/mental-health-strengthening-our-response>
- [9] S. BH et al., “Mental Health Analysis of Employees using Machine Learning Techniques,” 2022 14th International Conference on COMMunication Systems & NETWORKS (COMSNETS), Jan. 2022, doi: <https://doi.org/10.1109/comsnets53615.2022.9668526>.
- [10] The Daily Star, “Counseling: An urgent need for university students,” 2023. [Online]. Available: <https://www.thedailystar.net/opinion/views/news/counselling-urgent-need-university-students-3294481>

- [11] A. Parekh, "Mental Health Prediction Using Machine Learning," Analytics Vidhya, 2022. [Online]. Available: <https://www.analyticsvidhya.com/blog/2022/06/mental-health-prediction-using-machine-learning/>
- [12] H. A. Rahman et al., "Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," Bioengineering, vol. 10, no. 5, 2023, doi: <https://doi.org/10.3390/bioengineering10050575>.
- [13] A. Baba and K. Bunji, "Prediction of Mental Health Disorders Using Annual Student Health Survey: Machine Learning Approach," JMIR Mental Health, 2023, doi: <https://doi.org/10.2196/42420>.
- [14] Barnali Sahu et al., "Mental Health Prediction in Students Using Data Mining Techniques," The Journal of Open Bioinformatics, 2023, doi: <http://dx.doi.org/10.2174/18750362-v16-230720-2022-19>
- [15] Muhammad Azizur Rahman and T. Kohli, "Mental health analysis of international students using machine learning techniques," PloS one, vol. 19, no. 6, pp. e0304132–e0304132, Jun. 2024, doi: <https://doi.org/10.1371/journal.pone.0304132>.
- [16] World Health Organization, "Mental health: Strengthening our response," WHO, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>
- [17] S. BH et al., "Mental Health Analysis of Employees using Machine Learning Techniques," 2022 14th International Conference on COMMunication Systems & NETWORKS (COMSNETS), Jan. 2022, doi: <https://doi.org/10.1109/comsnets53615.2022.9668526>.
- [18] The Daily Star, "Counseling: An urgent need for university students," 2023. [Online]. Available: <https://www.thedailystar.net/opinion/views/news/counselling-urgent-need-university-students-3294481>
- [19] A. Parekh, "Mental Health Prediction Using Machine Learning," Analytics Vidhya, 2022. [Online]. Available: <https://www.analyticsvidhya.com/blog/2022/06/mental-health-prediction-using-machine-learning/>
- [20] H. A. Rahman et al., "Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," Bioengineering, vol. 10, no. 5, 2023, doi: <https://doi.org/10.3390/bioengineering10050575>.
- [21] A. Baba and K. Bunji, "Prediction of Mental Health Disorders Using Annual Student Health Survey: Machine Learning Approach," JMIR Mental Health,

2023, doi: <https://doi.org/10.2196/42420>

- [22] Barnali Sahu et al., "Mental Health Prediction in Students Using Data Mining Techniques," The Journal of Open Bioinformatics, 2023, doi: <http://dx.doi.org/10.2174/18750362-v16-230720-2022-19>
- [23]. Hanif Abdul Rahman et al., "Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," Bioengineering, vol. 10, no. 5, pp. 575–575, May 2023, doi: <https://doi.org/10.3390/bioengineering10050575>
- [24] A. Baba and K. Bunji, "Prediction of Mental Health Problem Using Annual Student Health Survey: A Machine Learning Approach (Preprint)," JMIR Mental Health, Sep. 2022, doi: <https://doi.org/10.2196/42420>.
- [25] D. Jagli, M. Talakoti, R. Solanki, N. Naik, K. Jamdaade, and R. Naik, "Innovative machine learning models for student mental health analysis," Frontiers in Health Informatics, vol. 13, no. 3, pp. 1270-1281, 2024.
- [26] Sahu B, Kedia J, Ranjan V, Mahaptra B, Dehuri S. Mental Health Prediction in Students using Data Mining Techniques. Open Bioinform J, 2023; 16: e187503622307140. <http://dx.doi.org/10.2174/18750362-v16-230720-2022-19>
- [27] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: <https://doi.org/10.1023/a:1010933404324>.
- [28] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification And Regression Trees. Routledge, 2017. doi: <https://doi.org/10.1201/9781315139470>.
- [29] "Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20, 273-297. - References - Scientific Research Publishing," Scirp.org, 2014. <https://www.scirp.org/reference/referencespapers?referenceid=1150668>
- [30] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York, NY: Springer New York, 2009. doi: <https://doi.org/10.1007/978-0-387-84858-7>.
- [31] "IEEE TRANSACTIONS." Available: <https://isl.stanford.edu/~cover/papers/transIT/0021cove.pdf>
- [32] M. Drira, S. Ben Hassine, M. Zhang, and S. Smith, "Machine Learning Methods in Student Mental Health Research: An Ethics-Centered Systematic Literature Review," Applied Sciences, vol. 14, no. 24, p. 11738, Dec. 2024. doi: 10.3390/app142411738.
- [33] T. M, V. A. M, S. Y. Bajantri, S. B. S, and S. B, "Student Mental Health Analysis Using Machine Learning Techniques," International Journal of

Research Publication and Reviews, vol. 5, no. 12, pp. 5608-5615, Dec. 2024.

- [34] M. Sadiq, "Predicting Students' Mental Disorders Using Machine learning," RIT Digital Institutional Repository, 2022. <https://repository.rit.edu/theses/11413/> (accessed Feb. 21, 2025).
- [35] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," *Clin. Pract. Epidemiol. Ment. Health*, vol. 20, pp. e17450179315688, Jul. 2024, doi: 10.2174/0117450179315688240607052117.
- [36] Hanif Abdul Rahman et al., "Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," *Bioengineering*, vol. 10, no. 5, 2023, doi: <https://doi.org/10.3390/bioengineering10050575> [\[oaicite:6\]{index=6}](#).
- [37] Ayako Baba and Kyosuke Bunji, "Prediction of Mental Health Disorders Using Annual Student Health Survey: Machine Learning Approach," *JMIR Mental Health*, 2023, doi: <https://doi.org/10.2196/42420> [\[oaicite:7\]{index=7}](#).
- [38] Dhanamma Jagli et al., "Innovative Machine Learning Models for Student Mental Health Analysis," *Frontiers in Health Informatics*, vol. 13, no. 3, 2024.
- [39] Barnali Sahu et al., "Mental Health Prediction in Students Using Data Mining Techniques," *The Journal of Open Bioinformatics*, vol. 16, 2023, doi: <http://dx.doi.org/10.2174/18750362-v16-230720-2022-19> [\[oaicite:9\]{index=9}](#).
- [40] Mohamed Drira et al., "Machine Learning Methods in Student Mental Health Research: An Ethics-Centered Systematic Literature Review," *Applied Sciences*, vol. 14, no. 24, 2024, doi: 10.3390/app142411738.
- [41] Maryam Sadiq, "Predicting Students' Mental Disorders Using Machine learning," RIT Digital Institutional Repository, 2022, <https://repository.rit.edu/theses/11413/> [\[oaicite:13\]{index=13}](#).
- [43] Ujunwa Madububambachu, Augustine Ukpebor, and Urenna Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," *Clin. Pract. Epidemiol. Ment. Health*, vol. 20, pp. e17450179315688, Jul. 2024, doi:

10.2174/0117450179315688240607052117

- [44] Thejaswini M, Vivek A M, Shweta Y Bajantri, Shashank B S, and Shylaja B, "Student Mental Health Analysis Using Machine Learning Techniques," International Journal of Research Publication and Reviews, vol. 5, no. 12, 2024.
- [45] "Feature Selection Techniques in Machine Learning," GeeksforGeeks, Jan. 19, 2021. <https://www.geeksforgeeks.org/feature-selection-techniques-in-machine-learning/>
- [46] "Random forest algorithm in machine learning," GeeksforGeeks, Jul. 12, 2024. <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>

## 24% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

### Filtered from the Report

- Bibliography

### Match Groups

- 236** Not Cited or Quoted 20%  
Matches with neither in-text citation nor quotation marks
- 2** Missing Quotations 0%  
Matches that are still very similar to source material
- 66** Missing Citation 4%  
Matches that have quotation marks, but no in-text citation
- 2** Cited and Quoted 0%  
Matches with in-text citation present, but no quotation marks

### Top Sources

- 15% Internet sources
- 12% Publications
- 20% Submitted works (Student Papers)

### Integrity Flags

#### 0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.