DAFFODIL INTERNATIONAL UNIVERSITY

PROJECT ON

Mental health detection and treatment using machine learning

BY

HABIBA RAHMAN ID: 201-16-520

Supervised by

Ms. Sonia Nasrin

Lecturer

Department of Computing and Information System (CIS)
Daffodil International University

Declaration

I hereby declare that; this project has been done by me under supervision of **Ms Sonia Nasrin** department of Computing and Information System (CIS) of Daffodil International University. I am also declaring that this project or any part of there has never been submitted anywhere else for the award of any educational degree like, B.Sc., M.Sc., Diploma or other qualifications.

Supervised By

Sonia
13/01/25

Ms. Sonia Nasrin
Lecturer
Department of CIS
Daffodil International University

Submitted By

Habiba

Habiba Rahman
ID: 201-16-520
Department of CIS
Daffodil International University

APPROVAL

This Project titled "Mental Health Detection and Treatment using Maching learning .", Submitted by Habiba Rahman, ID No: 201-16-520 to the Department of Computing and Information Systems, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computing & Information Systems and approved as to its style and contents. The presentation has been held on 13-01-2025.

BOARD OF EXAMINERS

Md Sarwar Hossain Mollah — **Chairman**
Associate Professor and Head
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Nasimul Kader — **Internal Examiner**
Assistant Professor
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Mehedi Hassan — **Internal Examiner**
Lecturer (Senior Scale)
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Ahmed Saif Reza — **External Examiner**
Managing Director & Chief Technology Officer
Medico Bio Limited

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty God for His divine blessing that made it possible to complete the final year project successfully.

I am really grateful and wish our profound indebtedness to Supervisor Ms. Sonia Nasrin Department of CIS, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine Learning " to carry out this project. Her 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.

I would like to express my heartiest gratitude Faculty of Science and Information Technology, DIU, for his valuable support and advice in finishing my project and also my heartiest thanks to other faculty members and the staff of the Department of CIS, Daffodil International University.

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

Finally, I must acknowledge with due respect the constant support and patience of my parents

Abstract

Student mental health is an important aspect of overall student well-being, but it is a less explored area with respect to the Bengali participants. We are using machine learning methods to predict the mental health of Bengali students and this study is also reflects on the factors that affect students' mental health. Using data from student responses, we trained and compared four machine learning models: Logistic Regression, Random Forest, XGBoost, and Support Vector Classifier (SVC). Each model was evaluated based on performance metrics like accuracy, precision, recall, and F1 score.

Logistic Regression proved to be the best model out of all of them, with the highest accuracy (94%) and well-balanced metrics; 86% precision, 100% recall, 92% F1 score. Both Random Forest and SVC models showed competitive performance, providing more descriptive analysis about the mental health classification, while XGBoost slightly lagged behind but maintained a balanced score. Thus, these results could shine a light on how data can be used to identify and mitigate mental health issues in educational settings.

While the study is promising regarding the potential of machine learning in predicting mental health, the researchers acknowledged limitations such as dataset surround and survey potential bias. Further studies could increase sample size, improve survey measures for sociological parameters, and create monitoring systems that aim to increase prediction accuracy in real-time.

This work establishes a groundwork for early mental health interventions and provides educators, guardians, and mental health professionals tools to promote the well-being of Bengali students. Through the integration of data-driven insights into educational structures, the objective of this research is to promote early detection and intervention strategies for mental health challenges in academic settings

Table of Contents

List of Tables: viii

Chapter 1 1

1.1 Introduction 1

1.2 Motivation 3

1.3 Research Gap 4

1.4 Research Question 4

1.5 Objectives 5

1.6 Expected Outcome 5

1.7 Scope of the study 6

1.8 Methodology 7

1.9 Organization of the Thesis Paper 8

Chapter 2 10

Literature Review 10

2.1 Overview 10

2.2 Related work 10

2.3 Limitations of Existing Work 14

2.4 Contribution to our Studies 14

Chapter 3 15

Methodology 15

3.1 Overview 15

3.2 Proposed Model 17

3.3 Data Collection 17

3.3.1 Survey Questionaries: 18

3.3.2 Survey Response 19

3.4 Data Preprocessing 22

3.4.1 Tabular Data Oversampling using SMOTE 23

3.5 Data Preprocessing 24

3.6 Model Train 25

3.7 Model Evaluation 26

3.8 Deploy 27
Chapter 4 29
4.1 Overview 29
4.2 Tools and Technology 29
4.2.1 Python 30
4.2.2 Sci-Kit Learn 30
4.2.3 Google Form 31
4.2.4 Google Colab 31
4.3 Machine-Learning Techniques 32
4.3.1 Logistic Regression 32
4.3.2 Support Vector Classifier 34
4.3.3 Random Forest Classifier 35
4.3.4 XGBoost 37
Chapter 5 40
5.1 Overview 40
5.2 Experimental Result 40
5.2.1 Result of Logistic Regression 41
5.2.2 Result of Support Vector Classifier 42
5.2.3 Result of Random Forest 44
5.2.4 Result of XGBoost 47
5.3 Comparative Analysis of Model 48
5.4 User Interface 49
5.4.1 Students User Interface 50
5.4.2 Teacher User Interface 51
5.5 Limitation to our work 52
5.6 Future Work 53
5.7 Conclusion 53
Reference 54
Plagiarism report 57

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1: Flow Chart of the Project	08
Figure 3.1: Methodology of the Project	15
Figure 3.2: System workflow chart of current version	16
Figure3.3: Correlation Heatmap of Dataset	17
Figure 3.3.2: Student response	20
Figure 3.3.3: Percentage of student stable and unstable in different education level	21
Figure 3.3.4: Percentage of Stable and Unstable students by their quality	22
Figure 4.1: Logistic Regression Model	34
Figure 4.2: Support Vector Classifier	35
Figure 4.3: Random Forest Classifier	37
Figure 4.4: XGBoost model	39
Figure 5.1: Confusion Matrix of logistic regression	41
Figure 5.2: Confusion Matrix of SVC	44
Figure 5.3: Confusion Matrix of Random Forest	46
Figure 5.4: Confusion Matrix of XGBoost	48
Figure 5.5: Student Interface Page	51
Figure 5.6: Teacher Interface	52

LIST OF TABLES

TABLE NAME	PAGE NO
Table 5.1: Comparison between Models	49

CHAPTER 1
INTRODUCTION

1.1 Introduction

Physical health and mental health constitute the well-being of an individual. I speak plenty about the essentials of physical health, but mental health remains under the radar and very much stigmatized. This stigma, in addition to a lack of education, causes many mental health issues to go unnoticed. This means people have to battle on without ever understanding exactly how mental health issues can decrease their overall level of life enjoyment. Your mental health determines how you think, feel, and act in your life. it also affects your ability to cope with stress, maintain relationships, and work productively as well as allows you to make certain decisions. In addition to these dangerous situations, the neglect of this fundamental dimension of health is also a cause for great concern, especially in contexts where people are exposed to constant pressure and expectations: academic facilities.

Why Mental Health Is So Important

Introduction: Mental health is a crucial aspect of the well–being of an individual. It includes emotional, psychological, and social dimensions of life that shape how a person sees the world and how we interact with it. Good mental health is achieved not only by avoiding mental illness, but also by supporting resilience, coping with emotions, and keeping a balanced state of mind. Mental health problems are often viewed in many cultures as a personal failure and with that judgment or discrimination can follow. When we are unable to talk about and discuss our problems then it results in an environment of absolute secrecy and no one is able or willing to even seek help. But mental health issues are normal and can affect anyone—even the young with seemingly perfect pasts and demon-free bank accounts. Meeting these challenges head-on will require us to take a preventative approach, especially in environments where mental health issues are likely to be more common (i.e., students).

Mental Health Challenges Among Students

For students, especially, the pressures of academia make them especially susceptible to mental health problems. Students may feel high levels of stress, anxiety, and even depression as a result of their coursework, exams, and the competitive nature of the academic environment. The expectations for academic success are increasing, and with this comes the feeling that we are not meeting these expectations. Personal and social pressures like dealing with relationships, juggling extracurriculars, and thinking of life after university further add to this pressure. Students are a high-risk mental health group anyway; a mix of academic and personal stressors is not meant to be conducive to emotional stability.

Students' mental health can come in many forms, from low academic achievement to disappearing socially or feeling fatigued and struggling to sleep. However, a lot of students do not seek help for these things because of mental health stigma as well as lack of accessibility to such services. In turn, they only spiral further into the darkness because it takes a devastating toll not just on their academic career but also on their mental health overall. Taken in this light, there is a necessity for platforms that are truly dedicated to early warning, alert, and action — affording students the support they require before their mental health dips lower.

Leveraging Predictive Modeling

In this paper, I've proposed a system using the new advanced predictive modeling methods to understand and alert the mental health problems of Bengali students. In our system, the Logistic Regression algorithm analyses responses to questionnaires that document studentprovided data, and it allows us to make predictions as to whether or not students have mental health difficulties. Logistic regression is a popular statistical method that focuses on binary classification tasks and here this can be used for predicting if, the student may need mental health support based on their responses.

Our goal is to produce a preliminary screening mechanism for educators and mental health professionals by employing this system. The platform is user-friendly and anonymous to similarly inspire students to openly express opinions, thoughts, and ideas without fear of

judgment. Detection through a system of this kind allows timely action that can prevent issues from worsening. Further, the data-driven insights enabled by the system offer instructors with intel to help struggling students in a more personalized way.

Allowing Educators to be Empowered and Supported

Our system not only puts mental health challenges on the map, it enables educators to do something about them as well. Frequent contact with teachers, who see changes in behavior and performance which might be signals of mental health problems Unfortunately, it can be tough for educators to tell the difference between typical academic stress and signs that a student may need more serious mental health help, especially if they lack the proper tools or training. Our system gives educators the data required to identify when and with whom intervention is warranted and provides information on how that intervention can be developed.

I'm working towards a culture of openness and understanding by adopting a more holistic approach to education, in which mental health is recognized as being just as important (if not more) than academic success. A student who is supported mentally and emotionally will perform better academically. Mental health is not an addition or accessory to the student life portion of your institution, but a part of the academic framework itself which will enable students to manifest resilience that embodies their reality both in academia and real adulting.

Against this backdrop, we delve into the prominent nature of mental health among students and introduce an efficient model for predictive analytics for the timely identification of symptoms. Our system enables educators to intervene early so that students receive timely support and can succeed not only academically but also personally. This promotes good well-being for students and is more likely to lead toward a system of education that is comprehensive, responsive as well as capable.

1.2 Motivation

This project is motivated by a strong commitment to addressing the pressing issue of mental health among Bengali students and is inspired by the growing global awareness of the challenges related to mental well-being. Because educational settings present special

stressors, they have a particularly negative effect on students' mental health, which is why we are committed to creating a welcoming and encouraging environment. The need to lessen the stigma associated with mental health and give students a safe area to express their emotions without worrying about criticism is at the heart of our drive. I am motivated by the conviction that teachers, who have a significant impact on students' lives, can be equipped with the knowledge and skills necessary to recognize and address possible mental health issues. Given its critical function, early intervention is a major priority. Given its critical role in reducing the long-term effects of mental health issues, early intervention is a major area of focus. My thesis attempts to close current gaps in mental health care within educational institutions by utilizing technology to address real-world problems. In the end, this research is driven by a commitment to making a significant contribution to the academic community by offering perspectives and workable solutions that might guide future efforts in the field of mental health in educational settings.

1.3 Research Gap

Little is known in the area of student-focused psychological support and there is scant research exploring this within educational contexts influenced by Bengali experiences. Although there have been some studies on mental health concerns, the lack of specialist systems for students remains somewhat troubling. Additionally, there is a crucial vacuum in the agreed-upon questions that can diagnose the mental condition of the students, particularly when one speaks of Bengali culture as well as linguistic space. One not have been able to tell if these are good questions for figuring out the mental health of students in this age group because that hasn't been established yet (at least, according to previous research). Further, the a lack of participative frameworks between students, teachers, parents, and psychological professionals: Fey et al. (2014) However, we argue that insufficient attention has been placed on the interconnected dynamics between these three central stakeholders in mental health support systems and this has hampered the evolvement of cohesive, collaborative and comprehensive system responses. The present research aims to bridge these essential gaps by suggesting the development of a system specific for Bengali students, enhancing the evaluation of mental states by culturally sensitive questions, and

promoting collaborative activities of all related stakeholders towards a more efficient and inclusive student mental health approach.

1.4 Research Question

The existing research reveals a startling lack of an inclusive and systematic mechanism in place for the evaluation and prescription of student mental health. The central research question guiding this study is: Is there any system that can be used to find out the mental state of students and recommend teachers or guardians, to help them This kind of student care and support for mental health in Bangladesh? This is the main research question that represents the core purpose of this study which intends to explore whether a customized system capable of identifying and intervening in students' mental well-being proactively has been implemented in Bangladeshi specific social-cultural context. Investigating such a system is essential because it will lead to an improved understanding of how to better support student mental health within the education sector.

1.5 Objectives

This study will aim to improve mental health scanning for students with data-driven methodologies, and new algorithms. The purposes of this agreement are as follows:

- ✓ Build a Real-time system for students to respond to questions and questionnaires and have results stored in a database securely. Teachers can then view at a web interface taskbar if students are stable or not during the answer to each question.
- ✓ Utilize several machine learning models (Logistic Regression, SVC, Random Forest, and XGBoost) to effectively classify student mental health statuses, and allow for recognition of students who may require extra support.
- ✓ Visual representation of the dataset and also a Visualization on how well the model is able to predict the mental health trends among students are wing, in addition, to assessing your model’s performance as well.
- ✓ An easy-to-use platform provides teachers with data-driven insights, enabling them to reach out early and provide interventions through their identified and at-risk students for mental health support in a timely manner.

- ✓ Integrate mental health checks within academic institutions to improve the educational experience. By creating responsive learning environments that support and screen stud ✓ ent well-being.

1.6 Expected Outcome

The intended results of this analysis are likely to lead to transformative breakthroughs in the specter attached to mental health neighboring for students among the background Bangladeshis. Developing and implementing a data-driven system in real-time has the ability to change everything we know about how to identify and manage mental health concerns in students, providing timely recommendations for intervention. This is expected to further strengthen the bond between students, teachers, and guardians, thus creating a wide base of student care that should simmer down into a system designed to ensure the holistic well-being of students. The development of this comprehensive support network is intended to further institutionalize the intervention and proactive management of student mental health issues by increasing its perspective from one that is narrowcasted and limited to small numbers of at-risk students to a broader concern about creating a healthier, more robust population. Vice-versa, this will guide the implementation of checking mental health within online learning with an expectation that it will help in bettering the online learning experience by emphasizing the interlinked nature of being mentally and academically fit. Together, the intended impacts recovered making an effort to trigger a psychological revolution when it comes to the handling and management of student mental illness paving the ways for a more understanding, knowledgeable, and interconnected educational terrain in Bangladesh.

1.7 Scope of the study

These hypotheses therefore define the ambit of this study which is to make use of a machine learning framework that aims at creating/predicting an outcome, in our case the mental health status of students in an educational landscape. The main objective is to develop a system that can predict the feasibility of the education system for different students by predicting future mental states. The study is expected to predict the mental well-being of

students in an effective manner, which can throw light on their emotional and psychological conditions by using sophisticated machine learning algorithms. This extends to how these predictions can be incorporated into the educational process — allowing teachers and educators to adapt their guidance and support for all students given that they each have individual states of mind. The study is also aimed at the reconstruction of collective protocols based on teachers, students and possible caregivers that help to establish a systemic approach linking mental health with teaching. The goal is to assist in the development of a a more personalized and compassionate learning experience that benefits the students directly through personal wellness and academic achievement.

1.8 Methodology

The methodology used for this resource project is based on a careful and diligent process of systematically gathering information with the intention for building a strong framework for understanding the model for assessing student mental health. It begins from the building of one very sophisticated Google form, with a plethora of questions focusing on different educational levels (high school to undergraduate) which helps understand how students are feeling in their minds and then re-use this data that becomes the cornerstone for our paper. This guarantees a common and comprehensive technique to measure the intricate nature of students' mental health conditions. After collecting all the necessary data, a process phase starts to ensure the accuracy and application of collected information. This raw data, which has been carefully processed and normalized becomes the basis on which machine learning models are trained. This is done using an 80–20 split, with 80% of data used to train the model and the remaining 20% designated for testing but tested on severely so that it reflects the proficiency of the model in prediction and gauges its learning overall new data points to make the loaded data generalizable. This structured methodology ends in the final phase where-on the evaluation and evolution of different machine learning models unfold. A study of their performance metrics is performed to establish which model is the optimal and most accurate one. This process is an iterative one, with the last model determined to best meet the goals of the project. This is followed by Deploying for any model that was selected — and therefore signing off on the completion of the project mission: using machine learning

to effectively identify indicators for student mental health. This methodology helps make the resulting tool strong and valid, building a vast amount of science to what we know about student mental health within the complex domain of education.

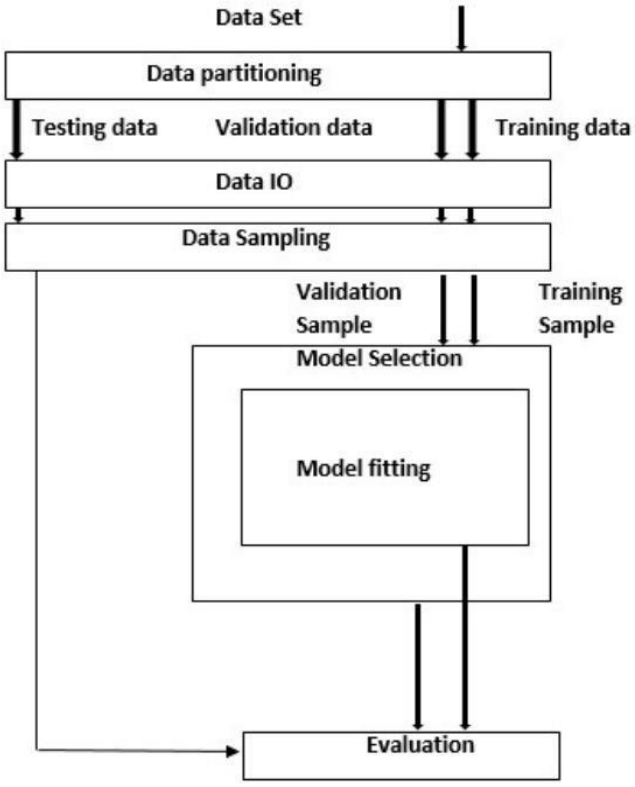

Figure 1.1: Flow Chart of the Project

1.9 Organization of the Thesis Paper

- ✓ **Chapter 1- Introduction:** This chapter establishes the framework for the thesis by offering an outline of the study's objectives, motivation, and scope. It covers the research topics and techniques that will be used, as well as a preview of the overall project structure.
- ✓ **Chapter 2 - Related Work**: Existing research on the study's issue is reviewed. The chapter acknowledges limitations in previous research and emphasizes the unique contributions this study makes to the area**.**

- ✓ **Chapter 3 - Proposed Model:** This chapter finds into the particulars of the research model. It discusses data gathering methods, preprocessing approaches, and feature transformations. The model process training and selection presented resulted in findings concerning the model's design.

- ✓ **Chapter 4 - Model Development:** This chapter concentrates on the technologies and tools used in the study, such as Google Form, Sci-Kit Learn, Python, and Google Colab. It also discusses data processing, model validation, and the machine learning methods used.

- ✓ **Chapter 5 - Result Analysis and Discussion:** The final chapter summarizes and evaluates the experimental outcomes. It outlines the study's shortcomings and proposes areas for future research, before finishing with a summary of the significant results and contributions

CHAPTER 2
Literature Review

2.1 Overview

Building on the core viewpoints provided in the previous chapter, the next section seeks to explore deeper into the vast terrain of related work on the subject. This chapter provides a comprehensive summary of current studies on the prediction of mental health among Bengali students. It conducts a thorough assessment and analysis of previous studies, offering light on their techniques, conclusions, and limitations. This chapter establishes the groundwork for describing our study's unique contributions by methodically identifying gaps in current research efforts. As we navigate the amount of knowledge acquired in this subject, we identify specific areas where breakthroughs are required . The findings from this comprehensive review establish the framework for the novel approach and significant contributions described in the next sections of this thesis. This chapter acts as a key bridge, linking the current body of knowledge to the distinctive contributions that our study brings to the forefront.

2.2 Related Work

The linked research includes a wide range of studies that investigate many aspects of mental health in different groups, providing useful insights into the varied nature of psychological well-being. Priscilla Rose Prasath et al. (2022) study the interrelationship of psychological capital, well-being, and distress among overseas students in US colleges during the COVID19 era, focusing on the mediating role of well-being and moderation in the link with depression [1]. Suen Mein-Woei investigates the spiritual well-being and mental health of Indonesian students, finding a significant correlation between the two with no gender disparities [2].

Priscilla Rose Selvaraj and Christine Suniti Bhat (2018) highlight the positive correlation between the development of psychological capital and favorable mental health outcomes by

concentrating on the predictive ability of psychological capital in assessing college students' mental health [3]. According to Feng, Zhouyang Li2's (2019) evaluation of music therapy's impacts on college students' mental health, it has a favorable impact on symptom relief, selfacceptance, and self-worth [4]. A Bayesian prediction system for substance use and mental health symptoms is presented by Anthony T. Fojo et al. (2020), demonstrating its high predictive value in clinical situations [5].

The Positive Mental Health Questionnaire (PMHQ) is a reliable and stable tool for measuring positive mental health among Spanish university nursing students, according to Roldan-Merino et al. (2017) [6]. The 12-item General Health Questionnaire (GHQ-12) is used to measure positive mental health, according to Yong Jian Hu et al. (2007), who demonstrate that the GHQ-12's two-factor structure distinguishes between "symptoms of mental disorder" and "positive mental health" [7]. Women and the unemployed had higher rates of mental health issues, according to Michael H. Banks et al. (1980), who evaluated the psychometric properties of the GHQ-12 and recommended its use in employment studies [8].In order to evaluate self-management in people with bipolar illness, depression, or anxiety, Simon Coulombe et al. (2015) provide the Mental Health Self-Management Questionnaire (MHSQ), emphasizing its positive psychometric qualities [9]. In a crosssectional study of college and university students in Bangladesh, Abid Hasan Khan et al. found that the COVID-19 lockdown caused 28.5% of participants to experience stress, 33.3% to report anxiety, and 46.92% to experience depression, ranging from mild to extremely severe [10]. M. Srividya's study investigates how machine learning algorithms might be used to forecast mental health conditions in various age groups. Using a variety of classifiers, including SVM and KNN, and clustering approaches, the study produces encouraging findings with 90% accuracy.

Mendis E.S. et al. used machine learning and image processing to create a smartphone application for mental health in Sri Lanka. Using the General Self-Efficacy Scale Decision Tree Classifier, it detected stress, anxiety, and depression levels with 99% accuracy. Happiness was the easiest emotion to identify, however accuracy varied. Although the app's functionality and pricing are limits, its goal is to increase awareness and offer help [11]. Nor Safika Mohd Shafiee et al. investigate mental health concerns among Malaysian university students, pointing to elements such a lack of social support, monetary difficulties, and

demanding classroom settings. It examines current machine learning methods for forecasting mental health issues. The most widely used algorithm, Support Vector Machine (SVM), has a high accuracy rate of 70% to 96%. To address students' mental health issues, the study intends to close research gaps and offer guidance for applying computational models [12]. Decision Trees, Random Forest, and Naïve Bayes are among the categorization algorithms used by Mrs. M. Kavitha et al. to predict mental health. The working class is the target audience for the encoding and accuracy analysis of data from internet sources. With an accuracy rating of 82%, the Decision Tree model offered users recommendations and probability information related to mental health. If a lot of data is gathered, future plans call for improving the system to forecast particular mental diseases. By incorporating these prediction technologies into websites, people can get easily accessible mental health help and self-care advice.[13]. We employed a Chi-Square test after discovering that over half of the variables in the data set were consistently missing, and that the accuracy of the five most significant predictors for predicting any kind of mental health condition varied. This study fills in a knowledge vacuum on the impact of COVID-19 on students' mentalities and the application of machine learning and deep learning techniques in this field. [14]

Using a variety of machine learning methods, including KNN, Random Forest, and CatBoost, this study investigates the prediction of suicide ideation among undergraduate students in south Bangladesh. Using the KNN model, we were able to attain the maximum accuracy score, which was approximately 91.45%. Since 19.9% of students reported having suicidal thoughts, we must screen for and monitor early warning signs and interventions. This study's cross-sectional design is problematic because it made it impossible to establish a link between risk factors and suicidal thoughts. The study clarifies how machine learning models can forecast suicide attempts for preventative purposes, but also delves further into the other risk factors that may have played a role. Information gathering: A straightforward random sample method was used to get the data [15].

This study examines the foundations of mobile connectivity and specifically examines the psychological effects on Bangladeshi students' mental health, which will be further investigated by depression detection methods—in our case, machine learning algorithms. According to the results, 31% of students had severe to very severe depression symptoms, 25% had mild to moderate depression symptoms, and 44% had no depressive symptoms

other than minor deregulations in positive affective and increased somatic complaints. Machine learning models have demonstrated high detection accuracy of depression, and the kind of feature engineering method was practically required to significantly increase its accuracy. In this work, eight machine-learning algorithms and cross-sectional survey data were used. As a result, more research into feature engineering techniques is anticipated to improve the model's generalization for features in other disciplines, necessitating a far wider range of applications [16].

This study significantly advances the assessment and prevention of mental health issues by using machine learning classification algorithms to predict the suffering of college freshmen in Bangladesh. After screening the keyword feature variables using Pearson's chi-squared test, the stacking classifier with 24 characteristics outperformed the other six classifiers. However, studies indicate that depression is rather common among students, and in this regard, having prediction models tailored to their needs may be beneficial. We lack particular knowledge regarding depression-specific predictors, nevertheless, and further study is necessary to advance our understanding of this topic and create thorough models of vulnerability [17]. A Bayesian prediction system for substance use and mental health symptoms is presented by Anthony T. Fojo et al. (2020), demonstrating its high predictive value in clinical situations [18]. This study examines how machine learning (ML) techniques can be used to predict students' stress levels. The more complex explanation is that it interprets a number of signals and then provides varying recommendations based on whether it believes you are under stress. Even though the study wasn't intended just for Bangladeshis, Researchers emphasized how crucial it is to genuinely keep an eye on students' mental health in learning environments. Students, this is a blatant sign that you need to work on your mental health. With implications for more effective mental health interventions, this study highlights the potential of machine learning (ML) for identifying hidden relational structures in psychological data. It suggests that more sophisticated student mental health systems that can anticipate stress and offer tailored feedback are required [19].

Using information from yearly health surveys, the study investigates the application of machine learning (ML) to forecast mental health issues in students. The study shows that behavioral and demographic characteristics are important indicators of mental health

problems. A comparison of machine learning algorithms, including logistic regression, elastic net, random forest, XGBoost, and LightGBM, revealed that questions about campus life significantly affected prediction accuracy. However, the model's performance was not significantly impacted by variables pertaining to answering time. The findings highlight the need for additional research to improve mental health prediction through better model optimization and revised health survey items [20].

2.3 Limitations of Existing Work

Limitations of many existing student mental health work make for a Scottish view. One of the major problems is that there is a lack of integrated systems that encourage student and teacher collaboration in some parts. Fewer platforms for the same manifest as a lack of true assistance in dealing with our core issues surrounding mental health. Furthermore, there is a lack of research on student mental health in the region of the Indian subcontinent which limits the generalisability of the findings to socio-cultural context there. Furthermore, there was little consistency in the instruments used which makes cross-study comparisons and generalizability of results challenging. Most importantly, the cross-sectional design of most studies precludes any conclusions on temporality or changes over time. The limitations together highlight the need for a greater level of coordination and regional specifics assessment tool, as one size model does not fit all countries and more longitudinal data to significantly improve our understanding of mental health students and their support systems.

2.4 Contribution in our Studies

The new system will represent a substantial added value in the student mental health support landscape. First and foremost, it has brought in a novel idea that brings students and teachers on a common platform. The partnership allows us to develop a full-scope understanding of student well-being, and to rally around providing support in a coordinated way. In addition, the system deals with an important limitation in research, especially in the Indian subcontinent, where few initiatives have focused on student mental health. With the use of sophisticated machine learning models, such as logistic regression, in combination with current data, several predictions on mental health could be made more accurate. This predictive functionality supports proactive intervention — teachers and parents can plan with pinpointed individual support based on each student. A carefully designed survey

questionnaire and the use of established psychological instruments build reliability and validity into the system's assessments. In conclusion, the proposed system is shown to be a holistic and pioneering solution for fostering the spirit of collaboration, becoming a datacentric, culturally-sensitive initiative to tackle with intricate issue of student mental health.

CHAPTER 3
Methodology

3.1 Overview

In this chapter, we present our approach to forecasting the mental health of students in Bangla. In this chapter, I will talk about collecting data with tailored survey questionnaires, putting survey responses together and the fundamental steps to be taken before building the model such as data preprocessing (including Tabular Data Oversampling with SMOTE). In feature transformation, model selection, and deployment strategy, the chapter provides a summary of how our approach differs as introductory information to the detailed exploration of similar steps for the model implementation and its outcomes.

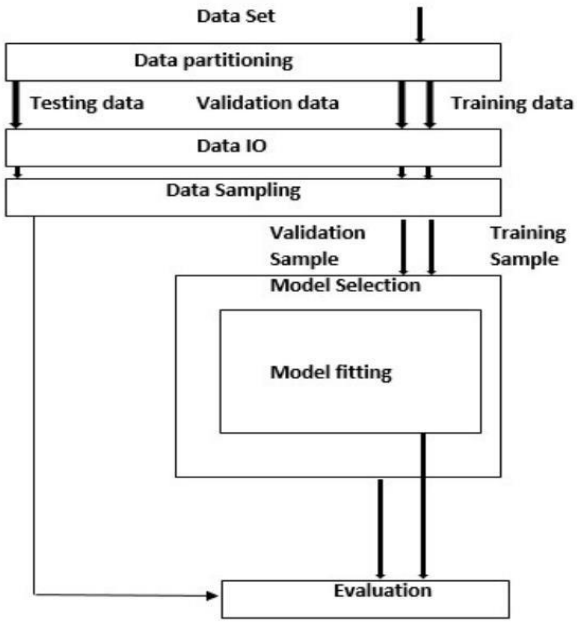

Figure 3.1: Methodology of the Project

3.2 Proposed Model

This made us decide, as a progressive approach employed when brainstorming our proposed model, to use logistic regression instead of other possible contenders like SVC, RFC, and XGBoost as in the context of our study its performance was better. At the core of our first module is data provided by students themselves, through their responses to basic base-level questions. Crucial features come in the form of these responses, which decide whether a pupil is stable or unstable; as such they represent the labeled training dataset.

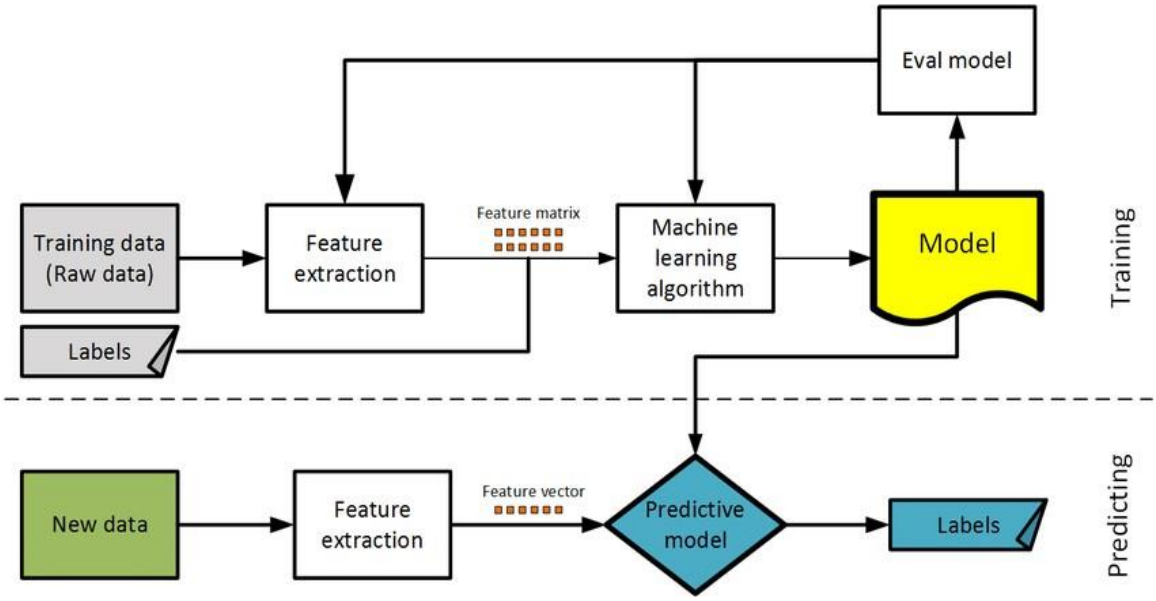

Figure 3.2: System workflow chart of the current version

Processed data is the most important part of model training which undergoes a rigorous preprocessing phase. This preprocessing includes sophisticated methods like feature extraction and mandatory encoding to organize the data in a way that helps for efficient learning by the logistic regression model. This approach makes the model interpretable as well as increasing its predictive accuracy. Our model aims to explain the mental health of Bengali students using logistic regression as our primary predictive framework with a strong preprocessing. A targeted and informed intervention response to identified mental health challenges is made more possible through this approach, not only in terms of the technical aspects. This entire model design is congruent with our broad aim of promoting holistic

conceptualization and preventive management of student mental health in the educational domain.

3.3 Data Collection

Data for this study was collected through an online survey which was distributed to high school, college, and university students. This survey was thoughtfully planned and delivered through Google Forms for an easy & convenient experience for participants. For this reason, with this online survey methodology, we hoped to reach a broad range of responses that represent the experiences and perspectives of students at different academic stages. Google Forms featured in helping to collect raw data which made it possible for a thorough investigation into the constituents related to the mental health of Bengali students. This wider net is consistent with the goals of the study to capture a comprehensive picture of mental health within an inclusive education system.

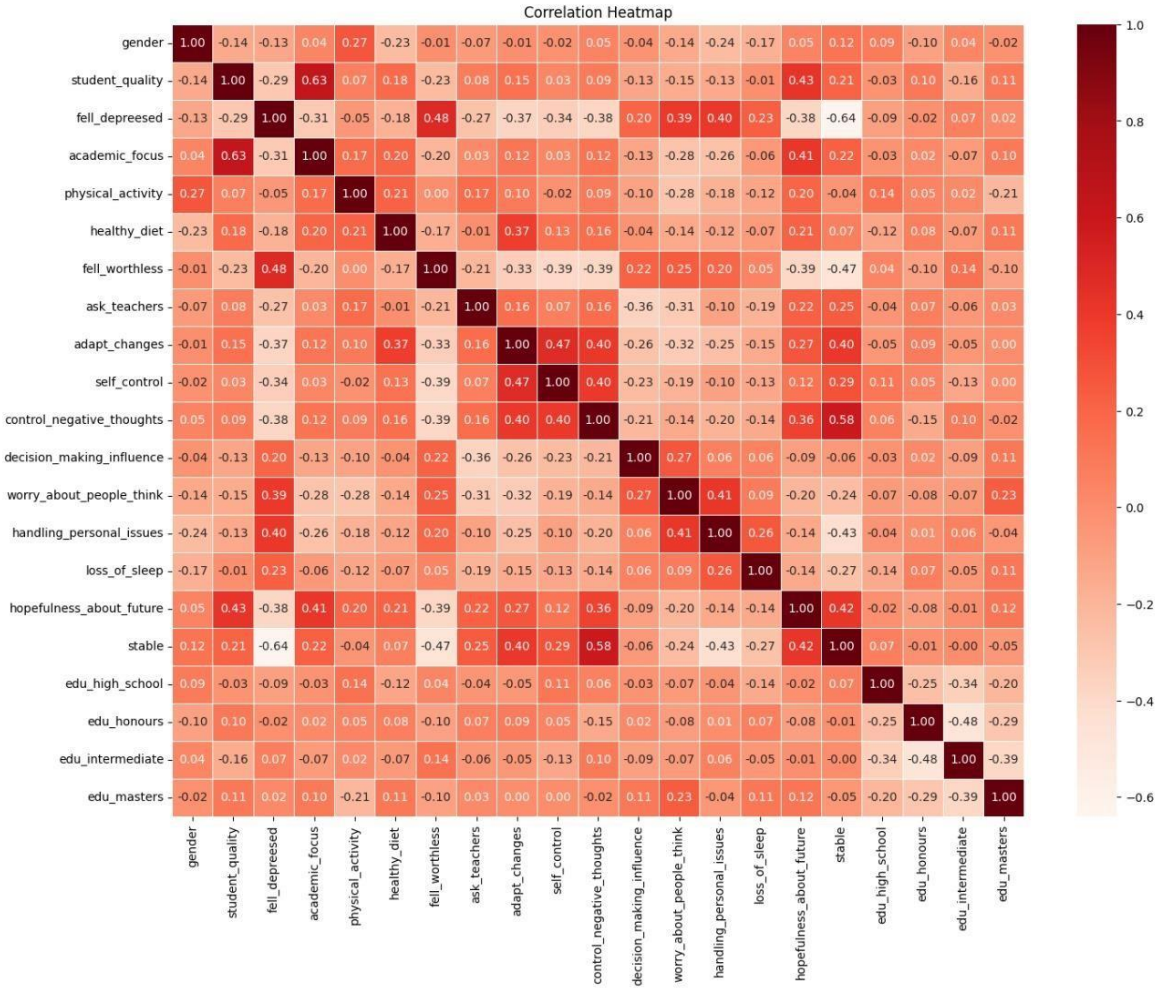

Figure 3.3: Correlation Heatmap of Dataset

3.3.1 Survey Questionaries

The survey employed a succinct yet comprehensive set of questions, carefully curated from prior research, to delve into various facets of a student's mental well-being. The questionnaire covered a spectrum of inquiries, encompassing aspects such as emotional states, feelings of depression, the ability to study attentively, gender identity, engagement in physical exercise, and adherence to a healthy diet. These questions were thoughtfully crafted to capture nuanced insights into the diverse dimensions that contribute to a student's mental health. The survey aimed to create a holistic understanding by exploring both emotional and lifestyle factors, acknowledging the interconnected nature of mental well-being. Here are the questions:

- ✓ আপনি কি পুরুষ নাকি মহিলা?

  Are you male or female?

- ✓ আপনি কোন ক্লাসে পড়েন?

  What class are you in?

- ✓ আপনি শিক্ষার্থী হিসেবে নিজেকে কেমন মনে করেন?

  How do you feel as a student? [5]

- ✓ আপনি কি নিজেকে বিষন্ন মনে করেন?

  Do you find yourself depressed? [7]

- ✓ আপনি কি পড়াশোনায় মনোযোগ ধরে রাখতে পারেন?

  Can you concentrate on your studies? [7]

- ✓ আপনি কি খেলাধুলা বা শারীরিক পরিশ্রমে নিযুক্ত?

  Do you engage in sports or physical activity? [9]

- ✓ আপনার কি স্বাস্থ্যকর খাদ্যভাস আছে?

  Do you have a healthy diet? [9]

- ✓ আপনি কি নিজেকে অযোগ্য এবং অকেজো বোধ করেন?

  Do you feel unworthy and useless? [7]

- ✓ আপনি পড়লেখার সমস্যা স্যারকে শেয়ার করতে পারেন কি?

  Can you share the academic problem sir? [5]

- ✓ আপনার চারপাশে পরিবর্তন হলে আপনি কি নিজেকে তার সাথে খাপ খাইয়ে নিতে পারেন?

  Can you adapt yourself to changes around you? [5]

- ✓ কোন দ্বন্দ্বে আপনি কি নিজেকে নিয়ন্ত্রণ রাখতে পারেন?

  Can you control yourself in any conflict? [5]

- ✓ আপনি কি নেতিবাচক চিন্তা থেকে নিজেকে নিয়ন্ত্রণ করতে পারেন?

  Can you control yourself from negative thoughts? [5]

- ✓ সিদ্ধান্ত নেওয়ার সময় আপনি কি অন্যের মতামত দ্বারা প্রভাবিত হন?

  Are you influenced by the opinions of others when making decisions? [7]

- ✓ লোকে কি ভাববে এটা নিয়ে আপনি কি খুব চিন্তিত থাকেন?

Do you worry too much about what people will think? [5]

- ✓ আপনার সমস্যাগুলো প্রায়ই আপনাকে আবদ্ধ করে রাখে?

  Do your problems often hold you back? [5]

- ✓ আপনি কি সম্প্রতি উদ্বেগের জন্য গভীর ঘুম হারিয়েছেন?

  Have you recently lost deep sleep due to anxiety? [7]

- ✓ আপনি কি আপনার ভবিষ্যত সম্পর্কে আশাবাদী?

  Are you optimistic about your future? [5]

- ✓ আপনি কি মানসিকভাবে স্থিতিশীল?

  Are you mentally stable? [5]

3.3.2 Survey Response

Results from 80 students in the survey completion phase highlighted a broad spectrum of opinions about their self-perceived mental health. Importantly, the findings reported an equivocal distribution with feelings of insecurity and security being experienced by 41.8 % and 58.2 % of students amongst others respectively. This distribution highlights the diversity of experiences in the mental health of those surveyed overall for this survey and emphasizes how important it is that we think carefully about the many challenges and well-being issues facing our students. The survey data is a starting point for understanding the trend of mental health problems among Bengali students which will be validated in the next stages towards building an efficient model using this research initiative.

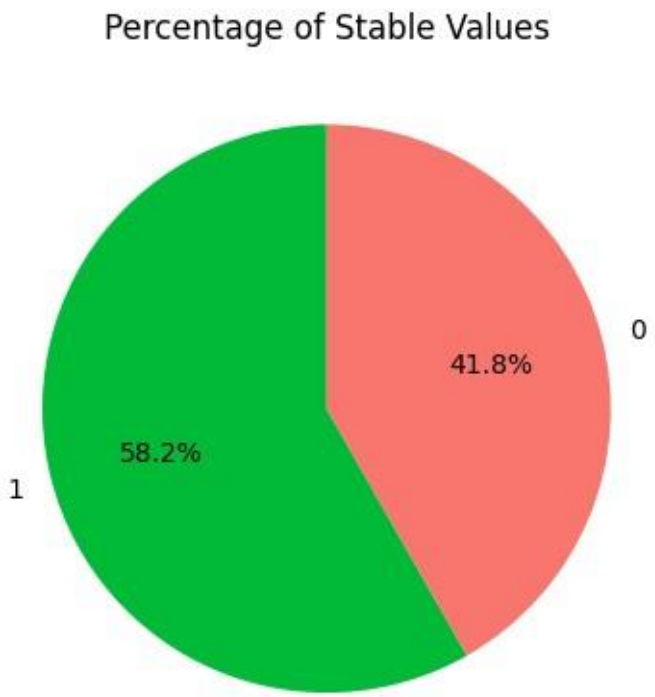

Figure 3.3.2: Student response

In the course of the EDA, an interesting pattern came up showing that as education increases among students, there is a proportional increase in mental instability along with it. Results: The number of mentally unstable students increased progressively through the student's educational level. Among all other categories of students, high school and honors level students showed fewer diagnostic instabilities as compared to college level, and up till master's the highest number was seen for instability. Another key finding came up about how students whose self-worth is connected to specific aspects of themselves face some mental well-being challenges. In other words, students wanted more professors who thought they were bright and assigned easier-to-correct grades even though this means those very same students with the rose-colored glasses placed NALs on more unstable peckerdannies. The tandem discovery not only isolates the complicated structure of mental health and education levels but it also highlights how self-image plays an essential part in impacting these states of mind among students at different educational decades. These results from exploratory data analysis provide the foundation for understanding more detailed

information about what contributes to mental health for students in the Bengali educational context.

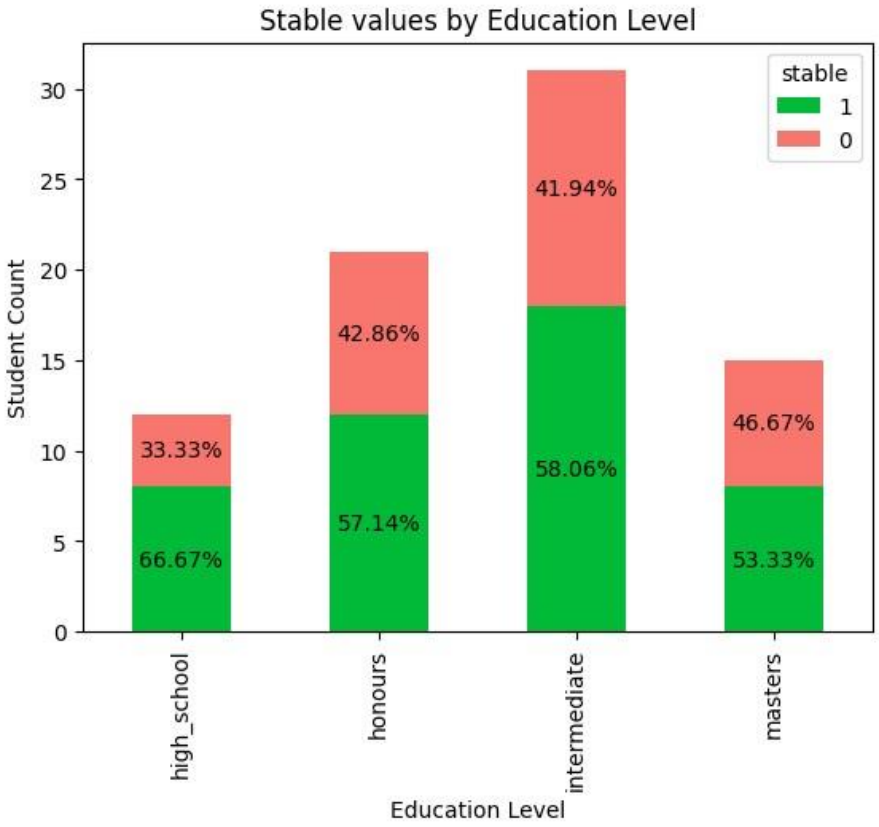

Figure 3.3.3: Percentage of student stable and unstable in different education level

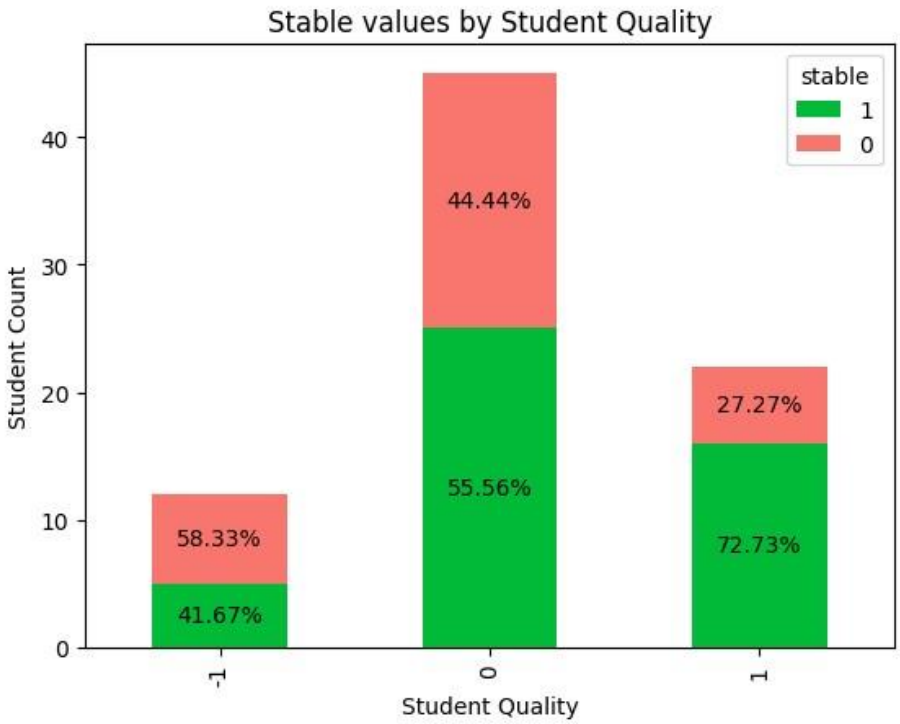

Figure 3.3.4: Percentage of Stable and Unstable students by their quality

3.4 Data Preprocessing

The data had to be preprocessed in a detailed manner and cleaned so that it is efficient enough for model training. Firstly, some data columns were removed, making the dataset to have fewer features as we want only essential features to predict mental health but not unimportant columns. Then, a binal encoding was proposed in two stages. Furthermore, tackle the problem of changes in categorical variables — to convert them by means of onehot encoding and binary embedding where the model learns different types of information from data! Given the class imbalance regarding mental health that was just described, the list of student records was passed through an SMOTE variation known as the Synthetic Minority Over-sampling Technique (SMOTEN) to enhance the representation of both stable and unstable mental health classes. This oversampling leads to a stronger model that can handle more imbalance in class distributions. In addition, training-testing partitions of 80% versus 20% respectively resulted in evaluating the model performance. This split truncates the producer down to only half of the entire trait space so that we can, in essence, evaluate online its generalizability to new (unseen) data and get a holistic sense of its

predictive performance. Together, these data preprocessing procedures work to improve the quality of the dataset while also preparing it for compatibility with later stages of model development and assessment.

3.4.1 Tabular Data Oversampling using SMOTE

Handling imbalanced datasets is an important aspect of the data science and machine learning workflow especially when it comes to tabular data sediments. Here, in this project at the Tabular Data Oversampling stage SMOTE (Synthetic Minority Oversampling Technique) type approach is applied for Tabular Data Augmentation to efficiently handle class imbalance problems.

This stems from a concept called class imbalance which just means that one class in the dataset is less represented than others. This project is concerned about specifically mental health states not likely to be as accurately represented among the student body. SMOTE helps fix such imbalance to make the model really efficient.

SMOTE works by generating synthetic examples in the data set, primarily concentrating focus on the minority class. SMOTE picks the D-dimensional points belonging to the M minority class items in the feature space Instead of just resampling the minority class existing data points (which are not enough), SMOTE synthetically augments the new examples closer to these points. This strategic addition enhances the dataset to get an even distribution of instances for all the classes.

SMOTE, on the other hand, is preferred mainly due to its generalization properties and a method to solve class imbalances without making assumptions of overfitted complexity. When a model learns the training data too well —preparing itself very accurately to detect the noise and idiosyncrasies of the sample—, that phenomenon occurs: overfitting. SMOTE generates synthetic instances that mimic existing minority class data points, so the oversampled class balances the original one in a meaningful way, and not by introducing artificial patterns which might result in overfitting.

The base dataset, augmented with synthetic samples, is used at the beginning of subsequent analysis and model. Adding these synthetic examples gives the machine learning algorithms

a better grasp of the minority class, which helps them to make reliable predictions. It helps especially with cases in which a label is highly skewed like predicting for some specific mental health state of the student.

Additionally, the use of SMOTE helps in making the machine learning model less prone to failure. A more balanced dataset makes our model generalize better for new data i.e. unseen test data. Indeed, it is the crucial part of any predictive model because if it does not generate value on new unseen instances then that model has no utility in practice.

Similarly, SMOTE results are key to the fairness of students' mental health status classification. It helps in tackling class imbalances hence ensuring that the model does not always make predictions based on majority classes but gives reasoned predictions across all classes. In the case of mental health predictions, the importance of fair decisions cannot be overstated as sustainable and effective solutions cannot leave anyone behind.

3.5 Feature Transformation

The feature transformation phase of our study aimed to improve the performance, especially with categorical data encoding optimization, in this case-friendly approach for model training. This required a structured method designed specifically for the properties of the categorical variables in our dataset.

At first, we tackled the "education level" variable by one-hot encoding. This technique separated the categorical variable into separate binary columns making it compatible for use in machine learning algorithms. This step was done in order to allow the algorithm to properly read and understand the different levels of education present in the dataset.

Based on this, we then worked on transforming the column "student quality". To do this, we converted their categorical data (Bad, Medium, Good) into numerical equivalents by assigning a value of -1 to bad students and a positive 1 to good students, ie — we standardized it. This assignment to a number reads as an easy way to understand the range of student quality in predicting mental health.

The binary encoding was used for the binary columns that are inherent in nature from our dataset. Overall, this made encoding binary features a much easier process making the model

both more interpretable and computationally efficient. These feature transformation steps were extremely careful to create a more unified and standard way of representing a variety of categorical features in the data.

So basically, all these feature transformation steps cumulatively were very important to make the dataset model-ready. When writing the function it was made to work with, the procedure involved setting any good practice for handling categorical variables (e.g. label encoding) as well as preparing them appropriately into a form readable by the model (this would depend on your machine learning implementation). This had served to better train on the data and so led to an increase in the predictive skills of said models.

Finally, in the feature selection step, we removed the column for student mental health state (our target variable). This final split encapsulates the complete feature engineering steps and establishes a base for further model training and testing. When we narrowed down to the target feature we knew that we had a well-defined primary goal of predicting mental health states which is important for our research goal.

3.6 Model Train

In our study, the model training phase was performed with careful data preparation. The dataset was converted into a suitable format carefully, utilizing one hot encoding for the categorical data and binary encoding for the binary columns. The preprocessing goal was to make the data compatible and effective for machine learning algorithms.

The target column which is the stability of the mental health of students was removed from the supervised learning framework. The training dataset for all those machine learning models that we utilized to predict mental health statuses is on the basis of this column.

The models used here for machine learning were unique: Logistic Regression, Support Vector Classifier (SVC), Random Forest Classifier (RFC), and XG Boost as all these models have their own characteristics and are used in different scenarios. For each, models were selected based on their expected contribution to prediction performance. This diverse set of models allowed us to cover most algorithmic approaches along the spectrum of mental health prediction.

Hyperparameter tuning was performed on 80% of the dataset, with the final 20% used for model evaluation. This split provided a thorough evaluation of the generalization performance of the models on new, unseen data. One of which, the Logistic Regression, was trained using an iterative model to fine-tune its parameters for better predictions. Our use of the iterative refinement process has been very useful for consistently perfecting and adapting with our model parameters which has helped a lot in that regard.

On the other hand, The SCC, RRC, and XGBoost models had intensive training and parametric fine-tuning. A deep dive into their predictability was performed considering the particular setting in relation to forecasting students' mental health among Bengalis. The iteration nature of the training process and an acceptance that tweaks were ongoing due to the complexity of the dataset and mental health prediction nuances.

During training, the following comprehensive evaluation metrics were used such as accuracy, precession, recall and f1-score. These metrics allowed for a very comprehensive, quantitative evaluation of the performance of each model. This multi-metric approach allowed a detailed and nuanced consideration of how well the models performed in different elements of prediction accuracy and reliability.

ResultsThe logistic regression model was the best-performing predictive model for our study given a systematic search of a variety of algorithms, iterative optimization approaches and evaluation metrics. This model was chosen because it consistently performed the best upon the evaluation metrics used.

3.7 Model Evaluation

In our work, the model evaluation phase plays a key role in rigorously testing the efficiency of predictive models for identifying mental health status among Bengali students. In this study, we adopted a multifaceted approach to the construction methodology employing an extensive set of evaluation metrics like precision, recall, f1-score and accuracy for a deep comprehension of how well the models performed.

One of the most important metrics is Precision, which determines how many of our positive predictions were actually correct. This tells us how well the model can avoid false positives

— incorrect positive predictions as such it is key to have accurate predictions on positive instances. Conversely, recall demonstrates how many true positive instances are identified by the model and is a proxy for sensitivity with respect to mental health conditions detection.

Together, these two metrics provide a somewhat balanced insight into the model's ability to predict positive developments.

The F1-score is a measure of both the precision and recall of the classification model, and it is defined as the harmonic mean of the precision and recall. Useful especially in the case of imbalanced datasets, the F1-score is situated between precision and recall—thus allowing us to gain an insight into a model's ability to catch both false positives and negatives. This is critical for assessing models in real-world contexts where the costs of misclassifications are often severe.

One of the key metrics is accuracy which shows how many of the instances were predicted right over all predictions. Accuracy gives an overall idea of how this is the model but including precision, recall and F1-score will give a more detailed idea about performance. The focus on multiple metrics provides a clearer picture of both the advantages and limitations of each model.

By examining each of these variables in detail, we learned interesting things about how accurate each predictive model was. Throws Made By About 60%: Let's dissect this using the key dimensions of our analysis. (dimension 1- the method used to predict) · Random Forests were by far the best · No major differences for any other methods. This methodology helped us make a more informed decision when selecting the ideal model for our use case. The evaluation method was a diagnostic approach because we considered false positives and false negatives, which provides higher generalizability for our predictive system to assess Bengali students' mental health.

3.8 Deploy

With our study trained, evaluated, and tested, we now finish deploying the best-performing model in a user-friendly web framework. The best-suited model to predict the mental health of Bengali students is saved and can be integrated into the platform. How does this work:

This deployment has a tech stack of Python Flask for backend processing with a frontend that uses HTML and CSS to make the web page look cool, JavaScript for animations.

One — the web platform, provides a literal solution you can give to students to enter their information and see almost-at-once estimates of when they should seek help. And this smooth amalgamation of ML flavour will provide the user-empowering mode to aid educators and student stakeholders in their endeavor for insight discovery about psychological wellness in the Bengali educational system.

That is where our deployment process comes into play — turning the results of scientific research into a usable and accessible instrument. Real-time predictions through the platform allow for timely proactive mental health support, thereby nurturing a more comprehensive approach to student well-being. Our study contributes the development of systems for supporting student's mental health with applicability and impact, due to an intuitive web interface coupled with sophisticated machine learning algorithms.

CHAPTER 4
MODEl DEVELOPMENT

4.1 Overview

It should be noted that what will be done in this chapter is just a lightning overview of the basic tools and technologies, which are underpinning our project. This chapter provides an in-depth overview of the fundamental building blocks that are applied at various phases during a project lifecycle. A foundational component mentioned is the versatile Python programming language, which this tutorial focuses on with regard to data collection, processing, analysis and model development. Moreover, We will delve into the mighty Scikit-learn library and discuss how important it is in applying machine learning algorithms and methods.

The chapter extends from the role of Google Forms in data collection — its application to collect useful information needed for the study. It also covers the nuances of machine learning algorithms with the Logistic Regression Algorithm as a notable demonstration. This investigation details the rigorous methods employed as part of the modeling.

An extensive examination of the main part: data processing and analysis, emphasizing changes made to enhance it for modeling. It illustrates the importance of such preprocessing steps in this chapter as a quality assurance that suits the data to be utilized in any later procedures. In addition, this chapter explains the model evaluation process and discusses performance metrics selection with careful attention to detail. This discussion will be on precision, recall, F1-score, and accuracy — shedding more light on how we measured the effectiveness of our models.

At a nascent stage, It provides much-needed details on the undisputable tools and technologies used at the project been strategically chosen for its success from this chapter to readers This full dive into Python, Scikit-learn, Google Forms, and machine learning approaches will help to provide a broader picture of how the technology and logic behind the project work.

4.2 Tools and Technology

Various tools and technologies were used to build this project. Data manipulation and machine learning: As the primary programming language, python provided a wide-to-real foundation for data manipulation and machine learning. Scikit-learn is a very popular library for training, evaluation and development of machine learning models. For data collection, we used Google Forms because it allowed us to simplify the process of filling out important surveys. Moreover, we used Google Colab as a cloud-based Python programming environment that allowed us to work jointly and train models efficiently, therefore speeding up getting insight during project phases.

4.2.1 Python

Python is central to our all-encompassing study case because it represents a powerful and flexible programming environment that well suits the subtleties of machine learning as well as web development. At these stages and for these tasks, Python is the most prominent language mainly due to an extraordinary number of libraries and frameworks (Pandas, NumPy, Scikit-learn which offers Logistic Regression… etc) that are required to prepare, transform and analyze our dataset. Its clear readability and straightforward usage can allow anyone to better understand the intricacies of even some of the most complicated algorithms available, all while exploring their data on a deeper level. Additionally, we use the Flask framework from Python to deploy our most performant model onto a web platform and offer machine learning information through a user-friendly interface. Python has flexibility and scalability that extends into all parts of our analysis pipeline: from data preprocessing, and model training to web deployment. Python, with its thriving ecosystem, huge community support, and versatility, serves as a principal aspect of our objective of evaluating and improving the mental health of Bengali students using groundbreaking machine-learning applications.

4.2.2 Sci-Kit Learn

Among the other pieces of this intricate puzzle forming our study, sci-kit-learn stands out as a key player, providing an extensive and accessible machine-learning library in the Python environment (SciPy ecosystem). This impressive tool integrates effortlessly into our data wrangling and model-building pipelines, offering much of the functionality we need from

feature transformation through heavy-duty machine learning. Because of sci-kit-learn's well-thought-out and simple API, it is easy to experiment with different models quickly and compare how well they perform. Due to the benefit of functionalities, for example, one-hot encoding, binary encoding, and model choice, sci-kit-learn gives an instead elegant constant approach to our study objectives. Support for all types of machine learning algorithms, more notably logistic regression, has further augmented our model training process and allowed us to orient towards the nuances of the given dataset. Using Scikit-learn as the backbone of our research, we hope to exploit its flexibility to understand and leverage this unwieldy mental health problem among Bengali students, thereby steering educational institutions towards creating a better-informed and more proactive support ecosystem.

4.2.3 Google Form

Google Forms acts as an instrumental tool for collecting data in our study, providing an easy-to-use platform for aggregating critical data regarding the psychological well-being of Bengali students. They allow us to create a customized survey that includes many different types of questions about the emotional health, study, and lifestyle behaviors, our students are engaging in on Google Forms, which is a popular survey-making tool. Its intuitive interface ensures that data collection is smooth; it also makes it incredibly simple for students across all educational levels, from high school to college or university, to provide their responses. The flexibility of Google Forms allowed us to have both categorical and binary responses, which fit perfectly with the kind of data required for our study. The primary feature of this tool is that it generates regular and structured data, which makes the later set of preprocessing steps that would be executed on the dataset with high rates of readiness for training machine learning models. Using Google Forms also epitomizes our dedication towards inclusivity and expediency, enabling a standard yet easy approach to understanding the complex landscape of student mental health in the Bengali educational milieu.

4.2.4 Google Colab

Google Colab is an amazing tool in our study and it really becomes essential, especially when we work together to analyze the same data and make our projects a reality. Having this integration with Google Drive, we hook up our dataset in no time and now we have a

warehouse that everyone can access where data manipulation is going to seamlessly be performed by multiple people at the same time. Google Colab is a cloud-based platform that enables researchers to collaborate in real-time, meaning you can efficiently share code and insights. Google Colab itself is an amazing platform that allows us to perform so many tasks starting from data pre-processing, here we will be dealing with the like Feature Transformation and Over Sampling using SMOTE to Model selection and Evaluation. Having integration with everyday Python libs like pandas, sci-kit-learn, and xgboost means we can train (fit) and evaluate machine learning models collaboratively. Google Colab is cloud-based and scalable, meaning we have a free pass to resources and hardware limitations in order for us to analyze the state-of-the-art and improve the mental health prediction system of Bengali students.

4.3 Machine-Learning techniques

For this project, logistic regression was chosen as the model to predict the mental health status of Bengali students. Due to the simplicity and interpretability of logistic regression, this algorithm is a great solution when dealing with categorical features as it provides a clear indication of what affects mental well-being. This model is based on Python and Scikit learn, which adheres to our open-sourced and efficient approach — offering a robust instrument for mental health estimation pertinent to Bengali students in the educational context.

4.3.1 Logistic Regression

Bookends in this project aimed at making sense of how features relate to a binary heart attack risk involved logistics being a basic and explainable classification algorithm, which is used as the last resort. We will apply this model to the preprocessed dataset and mark our first step in establishing an elementary understanding of insights that those features have over data well in advance.

Simple and Interpretable: I'm choosing logistic regression as it is simple and easy to interpret for this dataset. In cases where the decision boundary between classes is assumed to be linear (like early heart attack detection), the model works at its best. The algorithm uses the logistic function to model the probability that an instance belongs to a particular class, i.e., ensures that the probabilities predicted are between 0 and 1. In the context of

predicting heart attack risk, a trait like this is especially desirable because class separability is crucial for discerning clear groupings.

Work Principle: Logistic Regression works on the basis of analyzing the relationships between those features (one at a time) like age, blood pressure, and lifestyle that underwent some changes leading to heart attack. These relationships are then mathematically modeled and the algorithm will give you some common examples of what a certain type of person (having a set of features) is likely to get a heart attack by converting these into probability statistics. The logistic regression function squashes the linear combination of risk factors to a probability score, allowing us to better appreciate how each feature positively/negatively contributes to a heart attack.

Interpretability: Since Logistic Regression tends to have a simple parameterization, its interpretability is great. The coefficients given to each feature in this model provide us with measurable answers as to how an individual factor contributes to the probability of having a heart attack. This is important to establish transparency so that the data can create linear separability and basic intuition in making predictions based on different features.

Base Benchmark Model: Logistic Regression is a base benchmark model and one can use Logistic regression to get the baseline performance that can be expected by using more complex algorithms in further analysis steps This simplicity provides a straightforward basic understanding of the structure and separability of the pool enabling structured benchmarking with more complex models. Such a benchmarking process is essential to assess the efficacy of more advanced algorithms and improve the predictive ability of the overall system.

In summary, Logistic Regression, with its simplicity and interpretability as well as its linear relationship modeling ability is used for laying the foundation structure in detecting complex patterns within the pre-processed dataset. More importantly, it not only works as a prediction module but also serves as an important basal component of the follow-on analyses and benchmarking efforts which concrete to the independence for any part of heart attack risk prediction system.

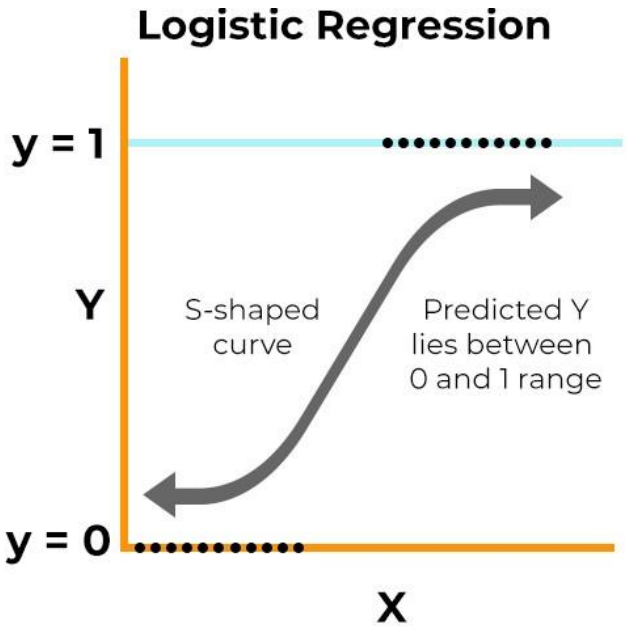

Figure 4.1: Logistic Regression Model

4.3.2 Support Vector Classifier

The Support Vector Classifier (SVC) was used on our preprocessed dataset to take advantage of its capabilities in managing complex decision boundaries. SVC works by identifying a hyperplane that best separates instances of distinct classes within a highdimensional feature space.

.

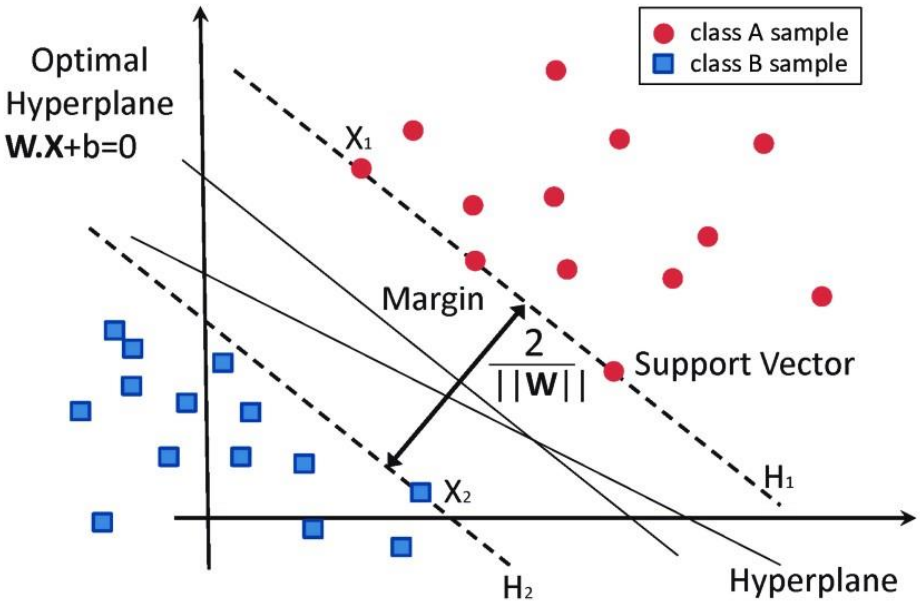

Figure 4.2: Support Vector Classifier

This is especially helpful in our dataset, where the relationship between different features and mental health risk may not be linear. SVC uses a kernel trick to project the data into a higher-dimensional space, making it easier to find complex patterns that may not be visible in lower dimensions. By providing flexibility in class boundary delineation, SVC is wellsuited to capture subtle relationships between a variety of student attributes and the probability of mental health issues. SVC's ability to handle non-linearities and adjust to complex data structures makes it a powerful tool for our mental health prediction framework, increasing the predictive power of the model.

4.3.3 Random Forest Classifier

The addition of the Random Forest Classifier into our study marks a purposeful move aimed at capitalizing on its versatility and resilience in managing intricate interactions within our preprocessed dataset for mental health predictions according to students' data.

Ensemble Learning Dynamics: The Random Forest Classifier stands out as a versatile and effective ensemble learning technique. It works by building a large number of decision trees during the training phase and then combining their predictions using a voting mechanism

during testing. This ensemble learning strategy improves the classifier's ability to handle complicated interactions and reduce overfitting, making it ideal for the complexities inherent in our mental health prediction framework.

Decision Tree Diversity: In the context of my dataset, which includes a wide range of student traits, the Random Forest excels at capturing complex feature relationships. Each decision tree in the ensemble learns distinct parts of the data, providing a wide range of perspectives. The ensemble's collective decision-making process helps to eliminate individual biases, resulting in a more thorough understanding of the complicated links between numerous student qualities and their mental health conditions.

Robust Predictions for Mental Health: In the context of my dataset, which includes a diverse set of student characteristics, the Random Forest excels at capturing complex feature interactions. Each decision tree in the ensemble learns different bits of the data, resulting in a variety of perspectives. The ensemble's collective decision-making process helps to minimize individual biases, leading in a more detailed knowledge of the complex relationships between a variety of student characteristics and their mental health issues.

Bootstrapping and Feature Selection: Random Forest uses bootstrapping, a resampling approach, for creating individual decision trees. This adds variety to the training process, improving the model's capacity to generalize to previously unseen data. Furthermore, the approach uses random feature selection for each decision tree, which diversifies the learning process and adds to the model's robustness.

Improving Predictive Accuracy: Random Forest tries to improve forecast accuracy by harnessing the combined strength of several decision trees. This is critical in our effort of

creating an effective and dependable framework for recognizing pupils' mental states. The ensemble approach assures that the model can detect patterns and relationships in the data holistically, yielding more accurate predictions.

In essence, incorporating the Random Forest Classifier into our mental health prediction system indicates a deliberate decision to capitalize on the benefits of ensemble learning. Its ability to handle complexity, mitigate overfitting, and generate robust predictions complements the multifaceted nature of our dataset, significantly increasing the richness and effectiveness of our model for mental health predictions among students.

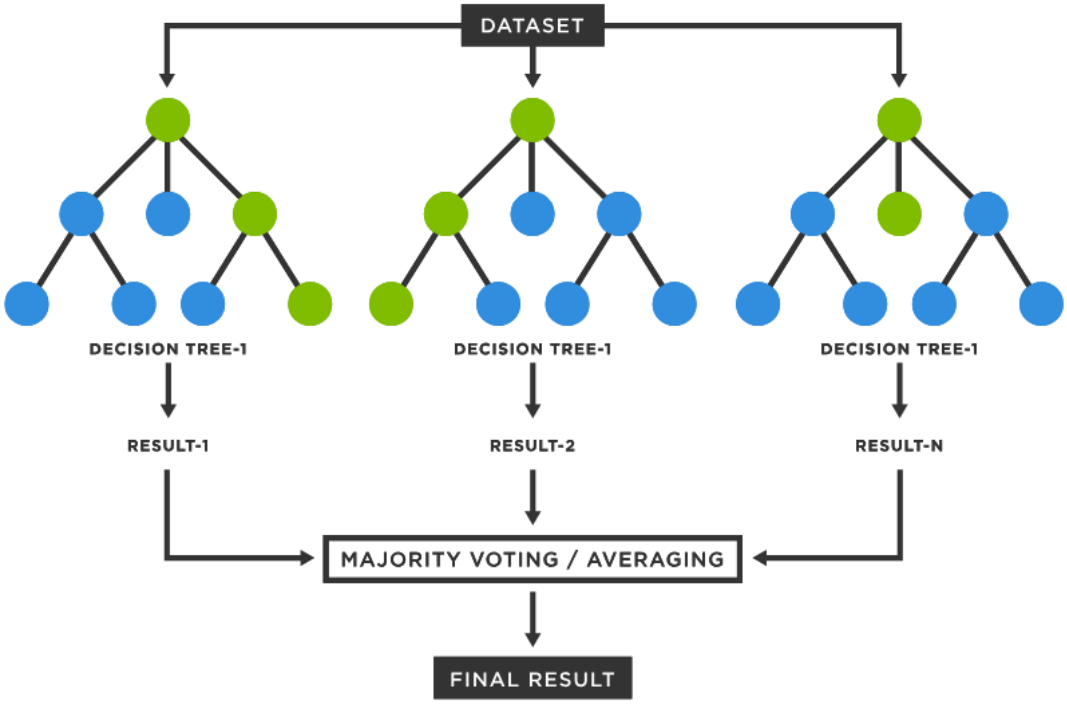

Figure 4.3: Random Forest Classifier

4.3.4 XGBoost

The application of an XGBoost Classifier to our research provides a more complex and accurate method of improving predictions regarding mental health issues from student data.

xgboost: a truly sophisticated gradient boosting / xgboostines Gradient boosting dynamics is one thing and implementing it is another. While traditional ensemble methods grow decision trees independently (e.g., Random Forest), XGBoost (as in Extreme GradientBoosting) grows an ensemble of weak learners (decision trees) systematically. This involves building m dels and correcting the errors that these preceding models make because this contributes to the inefficient representation of complex relationships in our preprocessed dataset.

Efficiency and predictive accuracy: XGBoost is an optimized gradient boosting model which allows it to use much more complex base learners, thus providing the ability to capture other complicated patterns even with smaller datasets. This part is more important in our case study as we are predicting the mental states of students from a wide variety of features. In the case of gradient boosting, and XGboost as implemented above, the model learns from its mistakes stained by strain as it is sequential and iterative in nature in order to improve upon its predictive accuracy.

Regularization: Both simple and effective regularization techniques are used in XGBoost, shrinkage (a.k.a learning rate), column/feature weights scaling, or unique feature importance based on model size at each step of boosting. This is important to prevent the model from being too complicated and fitting too well the training data at risk of losing its ability to generalize on another unseen empirical set. This is a very stable part of our mental health prediction framework.

Dealing with Big Data: This is one of the advantages of XGBoost as it can effectively manage inventing data having diverse features. Given that XGBoost is very capable, and our data set is comprehensive and represents a wide array of student attributes; it looks encouraging. XGBoost has shown that the model can adapt to complex patterns and feature interactions in our dataset, which is why XGBoost is a good option for constructing a predictive model with questionnaire responses (as we can observe from how the software works in practice).

Ability to Learn Complex Patterns: One of the strengths of XGBoost in our study is its ability to learn complex patterns and feature interactions. The salient features are most evident when predicting mental health of course, as they involve kind of nuanced relationships between the variables, but that ability to discern and thus capture these subtleties is what helps show the richness of our predictive model. Such an aspect adds to the model's ability to detect varied features, which are key for detecting the underlying complex nature of student mental states being influenced by different factors.

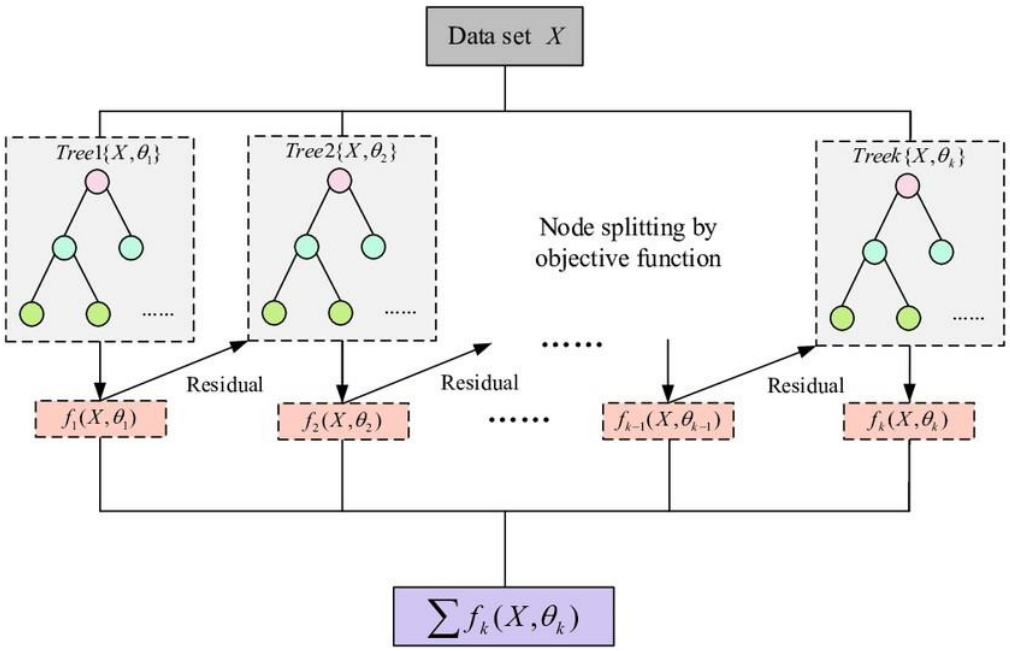

Figure 4.4: XGBoost model

CHAPTER 5

RESULT ANALYSIS AND DISCUSSION

5.1 Overview

Result Analysis and Discussion This chapter is the most important one as we are going to talk about the results of our project. What I emphasize is that the Logistic Regression model plays a key role in categorizing the status of mental health for students. In parallel, we evaluate the limitations of the system in a similar manner to better understand its behavior in in-the-wild conditions. In addition, we discuss new directions for future work and potential improvements and advances that could be made in the system. This way of summarizing the entire story is that this section, tied to the execution of the project but much broader in fact — acts as a sort of bridge and uncovers accompanying achievements AND areas for future work.

5.2 Experimental Result

Results: We evaluated the performance of the models on the testing dataset and calculated important performance measures like accuracy, precision, recall, and F1 score to rank the models based on their predictive outcomes. The metrics obtained are shown here to get an idea of how well the models have been able to predict the mental health status of Bengali students.

5.2.1 Result of Logistic Regression

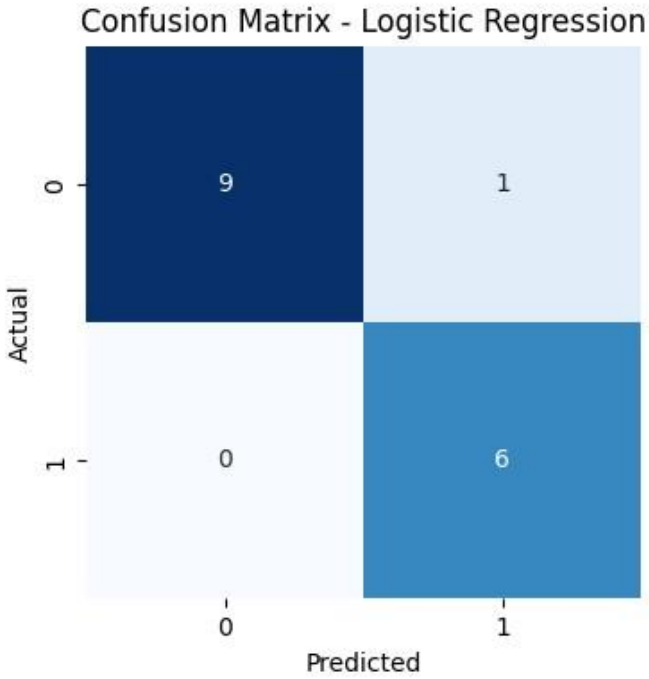

Figure 5.1: Confusion Matrix of logistic regression

The classification report for the Logistic Regression model provides a thorough assessment of how well it predicts Bengali students' mental health state, including detailed information on a number of important criteria that together demonstrate the model's efficacy.

Overall Accuracy (94%):

The model has an average accuracy of 94%, which means it can successfully predict instances with the corresponding mental problem. This high level of accuracy suggests how good the model is at making accurate predictions, which adds to its overall reliability when used for mental health assessment.

Precision (86% for Unstable Mental Health):

A crucial criterion, Precision, for the unstable mental health class reaches an impressive rate of 86%. This precision rate illustrates the system's competence to correctly identify those students as truly mentally ill thus leading to a significant reduction in false positives. Finally,

our model has a positive predictive value of 86%, meaning that if the model predicts student x to have unstable mental health, it is correct 85% of the time.

Recall (100% for Unstable Mental Health):

Since the recall metric is getting 100% for one of its classes, the unstable mental health class. This indicates how sensitive the model is to detecting all real cases of unstable mental health. Ultimately, when a student is actually experiencing a true mental illness, the model will correctly identify — and have a low chance of falsely identifying them as such (yielding no false negatives).

F1-Score (92%):

It has an F1-score of 92% which is a really good blend mean between Precision and Recall. This is a mean of false positives and false negatives and thus provides an equal-weight average score. A high F1-score demonstrates that the model can maintain a good balance between Precision and Recall (80%), compared to models with lower F1-scores which make more trade-offs of either one, or both along with the ability to Aire low false positive rates at the expense in accepting actual unstable individuals.

Overall Robust Performance:

So, all in all, Logistic Regression is performing very well with the higher value of differentdifferent Evaluation Metrics. Of particular interest about its performance is that it provides a trade-off with balanced evaluation in terms of true positives, meaning both the precision and recall values are representative according to the sensitivity of identifying mentally unstable cases. The ability of the model to balance precision and recall provides relatable reassurance about its dependability in understanding the mental health status among Bengali students. The success of this model should provide educators, guardians, and mental health professionals another tool in our effort to better understand and intervene with the intricate nature of student mental health.

5.2.2 Result of Support Vector Classifier

The classification report for the Support Vector Classifier (SVC) offers insights into how well it is able to predict Bengali students' mental health status, based on key metrics together which gives overall effectiveness of the model.

Overall Accuracy (81%):

The best performance by the SVC model is with an overall accuracy of 81% and very good precision across both the stability classes (stable mental health state and unstable mental health state) suggesting that, in case, it misclassifies its confidence will be low as well. This serves as a basic measure of the correctness of the model in general, determining its efficacy at making accurate predictions amidst the complicated landscape that is student mental health.

Specificity (100% for Stable mental health, 67% for Unstable Mental Health) — how many people who were expected to be really healthy and sick are classified correctly?

In terms of identifying class, the model shows a precision of 100% for stable mental health and 67% for unstable mental health which results in fewer false positives. That is, for example, the predicted probability of belonging to each class that the model assigns as stable or unstable mental health groups conditional on a student belonging to one group (from 100% accuracy at stability prediction rate to 67% at instability — the task with very high level of precision in classification).

Precision (100% for Stable, 67% for Unstable Mental Health):

For stable mental health, the recall is 70% and it is 100% for unstable mental health indicating that the model identifies many true cases among real ones in both states of mental health. Their model is particularly good at identifying and sorting out students who are mentally stable or unstable, with an emphasis on not being 100% sure that a student is okay—with extremely high recall statistics for the unstable mental health class in particular.

F1-Scores (Stable: 82%, Unstable Mental Health: 80%):

The F1-scores, which are the harmonic mean of precision and recall, are 82% in stable mental health and 80% in unstable mental health. In these scores, the precision and recall are treated equally meaning that the model is not penalized for false negatives any more or

less than it is penalized for false positives. The high F1 scores further indicate that the model can balance precision and recall effectively, suggesting its general reliability in predicting mental health.

Contributions and Insights:

To sum up, the mental health status of Bengali students of this approximation shows the best-established results tried to quite acceptable values for the Support Vector Classifier (SVC) model in accuracy, precision-recall, and F1 score. The DEVISED model is a sensitive model that which operational definitions maintain different levels of information which provides enhanced insights into the psychological well-being of students representing a reliable multidimensional measurement tool serving educators, guardians, and neuropsychiatric services to work through the complex stages related to student mental petitionary process. This is a testament to the all-round performance of the SVC model and its importance in delivering holistic verdicts for intervention strategies and preemptive support in an educational domain.

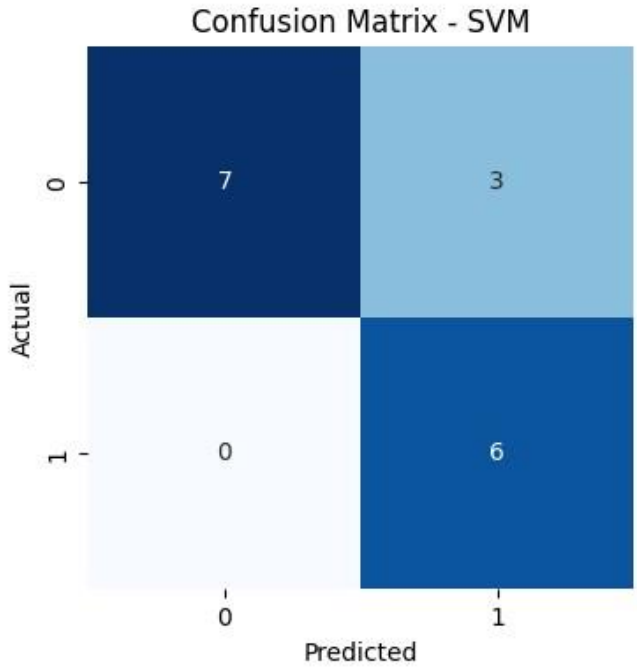

Figure 5.2: Confusion Matrix of SVC

5.2.3 Result of Random Forest

The Random Forest model's classification report offers a comprehensive assessment of how well it predicts Bengali students' mental health conditions, together with in-depth information on the main performance indicators that together define the model's success.

Overall Accuracy (81%):

With an overall accuracy of 81%, the Random Forest model demonstrates its ability to accurately categorize occurrences across both stable and unstable mental health conditions. This statistic is a fundamental indicator of the model's accuracy, demonstrating its ability to make precise predictions in the complex field of student mental health.

Precision (89% for Stable, 71% for Unstable Mental Health):

The model's 71% precision value for unstable mental health and 89% precision value for stable mental health highlight how well it detects each class while reducing false positives. Practically speaking, there is a high likelihood that the model's forecast of a student's mental health status will be correct (89% for stable, 71% for unstable), demonstrating the model's accuracy in classification.

Recall (80% for Stable, 83% for Unstable Mental Health):

The model's sensitivity in capturing real-world occurrences of both mental health states is demonstrated by its recall values of 80% for stable mental health and 83% for unstable mental health. The model's ability to prevent false negatives and capture a significant percentage of actual cases of mental health issues highlights how well it can detect and categorize kids with stable or unstable mental health.

F1-Scores (84% for Stable, 77% for Unstable Mental Health):

For stable mental health, the F1 scores—which are the harmonic mean of precision and recall are 84%, whereas for unstable mental health, they are 77%. With both false positives and false negatives taken into account, these scores highlight the model's balanced performance. The model's reliability in predicting mental health issues among Bengali students is highlighted by the strong F1 scores, which further support its capacity to effectively balance precision and recall.

Contributions and Insights:

In summary, the Random Forest model predicts Bengali students' mental health condition with respectable accuracy, precision, recall, and F1 score. Its sophisticated performance across a variety of indicators offers educators, parents, and mental health experts a trustworthy tool for comprehending and managing the complex dynamics of student mental health, as well as insightful information on the mental health of students. The Random Forest model's balanced performance highlights how important it is for providing thorough insights for proactive assistance and intervention tactics in the educational setting.

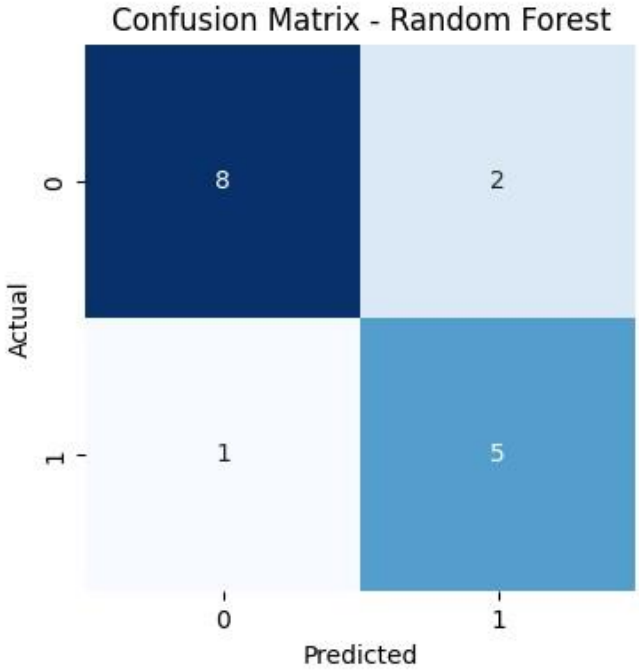

Figure 5.3: Confusion Matrix of Random Forest

5.2.4 Result of XGBoost

The XGBoost model's categorization report offers a thorough assessment of how well it predicts Bengali students' mental health conditions, providing a deep grasp of the important criteria that together characterize its efficacy.

Overall Accuracy (75%):

With an overall accuracy of 75%, the XGBoost model demonstrates its ability to accurately categorize occurrences across both stable and unstable mental health conditions. This metric functions as a basic indicator of the model's accuracy, demonstrating its ability to make precise predictions in the intricate field of student mental health.

Precision (80% for Stable, 67% for Unstable Mental Health):

With precision values of 67% for unstable mental health and 80% for stable mental health, the model effectively minimizes false positives while accurately recognizing each class. Practically speaking, there is a high likelihood that the model's forecast of a student's mental health status—80% for stable and 67% for unstable—will be correct, demonstrating the model's accuracy in classification.

Recall (80% for Stable, 67% for Unstable Mental Health):

The recall values of 80% for stable mental health and 67% for unstable mental health show the model's sensitivity in capturing actual instances of both mental health states. The model excels at correctly identifying and classifying students with stable or unstable mental health, emphasizing its ability to avoid false negatives and capture a significant proportion of true instances of mental health conditions.

F1-Scores (80% for Stable, 67% for Unstable Mental Health):

The model's balanced performance, taking into account both false positives and false negatives, is demonstrated by the F1 scores, which are the harmonic mean of precision and recall and are 80% for stable mental health and 67% for unstable mental health. The model's ability to maintain a trade-off between precision and recall indicates its reliability in predicting mental health conditions among Bengali students.

Contributions and Insights:

In summary, the XGBoost model is a valuable tool for educators, guardians, and mental health professionals seeking to understand the diverse mental well-being of Bengali students. It has commendable accuracy, precision, recall, and F1 score, and while its accuracy may be slightly lower than other models, its balanced performance across various metrics contributes to a comprehensive understanding of student mental health.

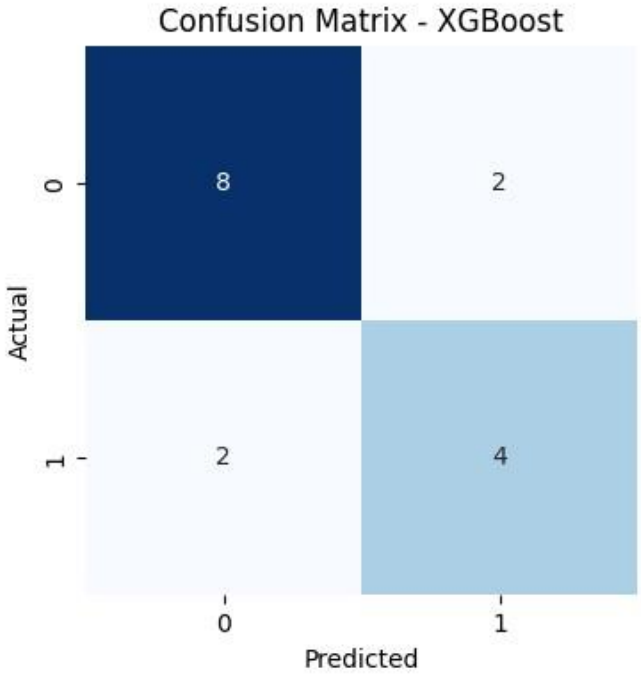

Figure 5.4: Confusion Matrix of XGBoost

5.3 Comparative Analysis of Model

The comparative examination of the deployed machine learning models suggests that Logistic Regression is the most effective model for predicting Bengali students' mental health state. The performance measures clearly show that it outperforms other models,

stressing the balanced trade-off between precision and recall, which is crucial for valid mental health evaluation.

Logistic Regression:

- *Accuracy:* 94%
- *Precision:* 86%
- *Recall:* 100%
- *F1-Score:* 92%

Among all the models, logistic regression performs the best in terms of accuracy, demonstrating its admirable capacity to accurately categorize cases across both stable and unstable mental health states. The model's accuracy in detecting each class while reducing false positives is demonstrated by its 86% precision. The model's sensitivity to capture every real case of unstable mental health is indicated by its perfect recall of 100%. The 92% F1 score indicates a well-balanced performance that balances recall and precision.

Table 5.1: Comparison between Models

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression	**86%**	**100%**	**92%**	**94%**
SVC	100%	70%	82%	81%
Random Forest	89%	83%	84%	81%
XGBoost	80%	80%	80%	75%

The tabulated format takes the essential performance metrics of each model and presents them in a readable, easy-to-understand form. Above graph tells us Logistic Regression is the top-performing model and provides the best Values on both F1-score. A clear structure of the presentation allows easy and rapid comparison, helping decision-makers retrieve their best-selected model to predict mental health status among Bengali students.

5.4 User Interface

Different Interfaces for Students and Teachers (The system has two interfaces – one for students and the other for teachers) Student Interface: a space for students to anonymously answer questions pertaining to their mental health, creating an open and personable platform. These responses are then sent to the server for further processing with the help of our forecasting model. Meanwhile, the teacher interface is meant to provide teachers with a window into what their students are feeling. The transformed data can be accessed by teachers so that they have an overall picture of the mental well-being of students. This double interface system helps to promote collaboration between students and their teachers where they will receive the necessary help from proactive-based mental health prediction made by the server.

5.4.1 Students User Interface

The student interface is developed to work interactively with students in an easy manner for online input of data. This interface with elements created by a framework developed with HTML, CSS and JavaScript, built to provide an intuitive and visually pleasing experience for students. On the UI, there was a list of a few selected questions that would be displayed to the student and he or she needed to answer those questions on how stable or unstable they are feeling from inside. What makes this tool interesting is that it allows students to write down not only their answers to specific questions in an easy-to-find way but they can also write down their thoughts and emotions. Once the user finishes questioner, data is sent directly to the Flask server for additional information processing. The student interface is carefully designed to promote candid responses; ultimately creating a friendly environment perfect for accurate mental health predictions. It does this in an accessible way which allows users, students in particular, to easily interact with the system and generate useful data for further analysis.

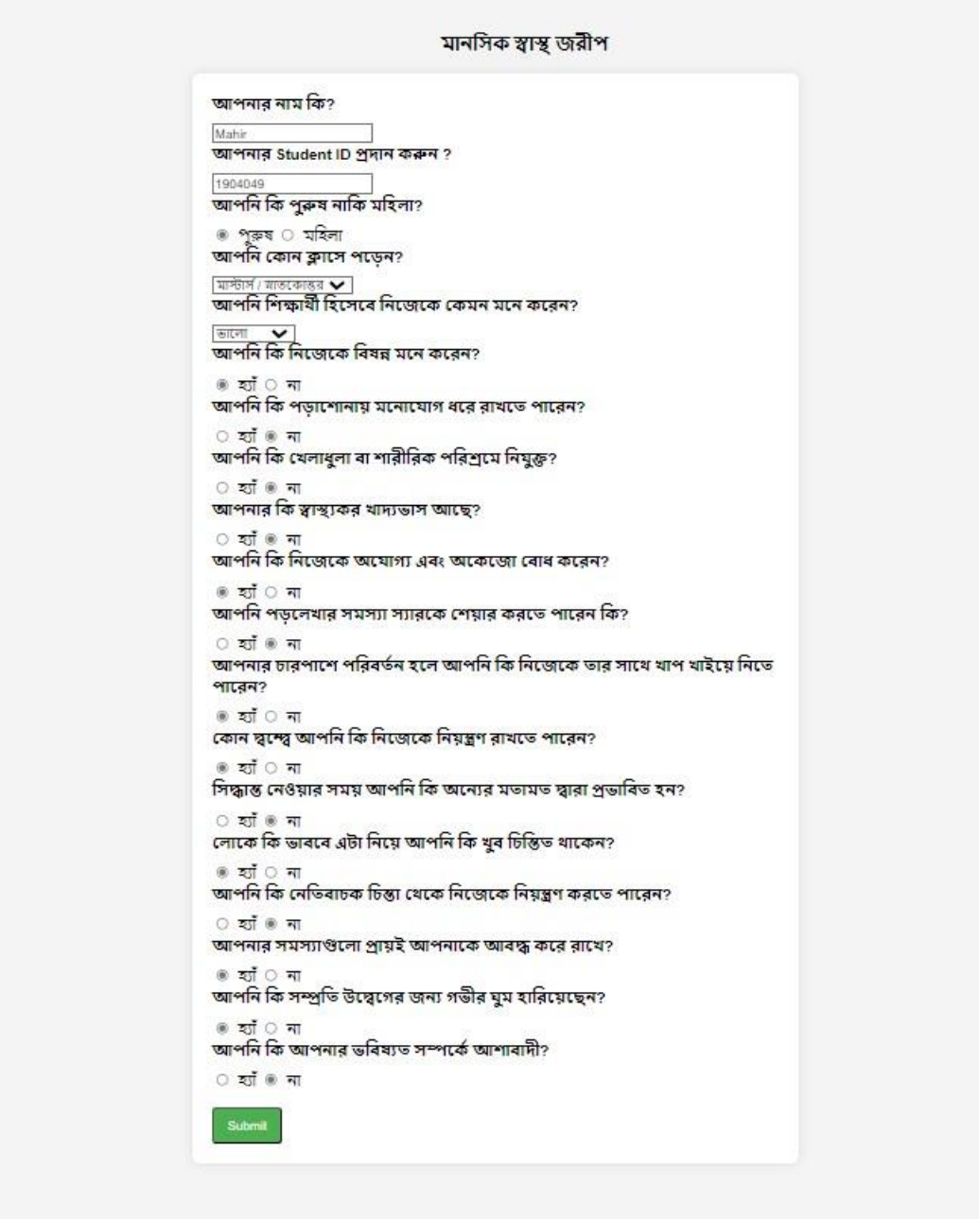

Figure 5.5: Student Interface Page

5.4.2 Teacher User Interface

It is a bespoke, teacher-only platform providing educators with valuable insight into the mental health and wellbeing of their students Using this elegantly designed tool, teachers are able to see a dashboard with the names, IDs, and mental statuses of all students. This information is safely held on the server and helps teachers to get a complete picture of the

mood and emotions of each student. The Teacher Dashboard provides an easy access point for teachers to identify students in need of additional support or intervention by analyzing mental health data. Incorporating information on student identifiers and mental states offers a lot broader view for teachers to adjust their approaches or care better for their students. Through this, the collaborative feature is boosted up in the system which helps build a more proactive support eco-system within the education context.

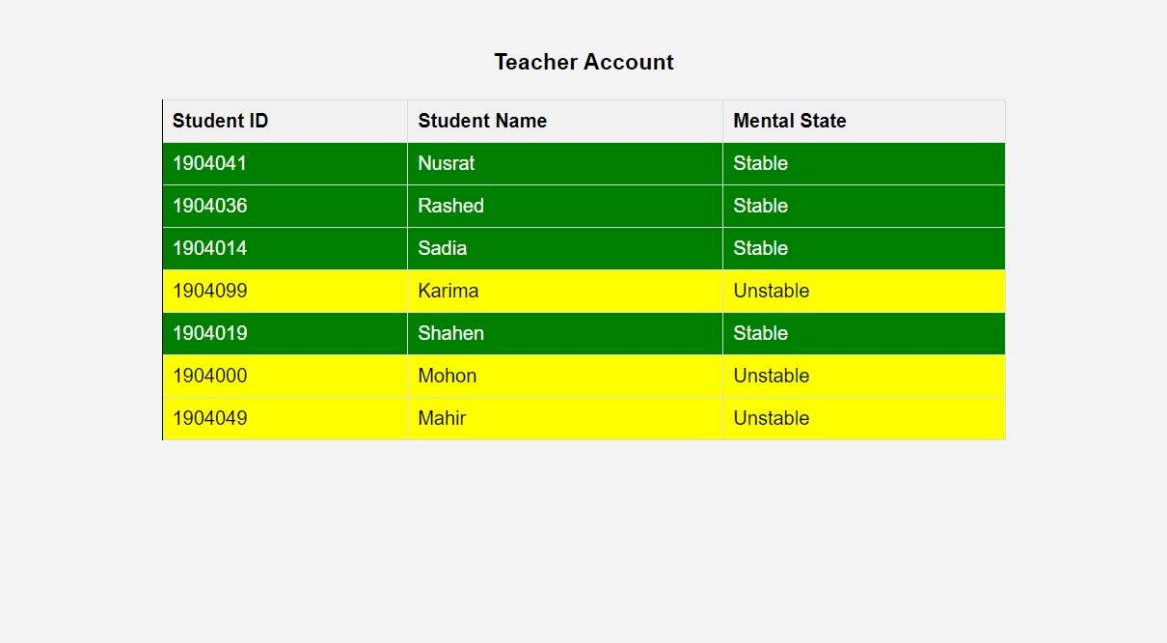

Figure 5.6: Teacher Interface

5.5 Limitations to Our Work

Acknowledgment The present study, though providing important information on predicting the mental health of Bengali students is not free from limitations. To begin with, the study is based on a small dataset which might affect the representability of its results to wider communities. Additionally, even though the dataset is well-curated the target demographic for mental health might be more varied than covered in the dataset. This survey is also open in nature creates scope for bias as people self-select to participate, and the sample may not be completely representative of all students. Also, because this research focuses on the students who belong to the Bengali context it is doubtful that the model derived from these

data applies in all other cultures or regions. The inbuilt limitations of the study, despite all efforts to address biases and methodological robustness, highlight that one needs to interpret with caution and put it into perspective within the specified context.

5.6 Future Work

Future work for this study can be made by extending the present results to newer domains as indicates several avenues of exploration and grounding. Initially, scaling up the dataset by including a wider and longer-range set of Bengali students across different schools and socio-economic classes would make the predictive model more generalizable. Moreover, by adding longitudinal data instead to the list of potential hypotheses this could help explain variation in students' well-being as events change in time. Especially further refinement of survey questions might enable all aspects of mental health even better than working with mental health experts on the questions that should be used. Another approach to further pursue in this study was the use of real-time monitoring tools with wearable devices through which we might gain constant knowledge regarding students' cognitive status. Additionally, it would be beneficial if such a system was co-created with educational institutions and mental health services and empirically evaluated in practice to determine its effectiveness as well as to allow for necessary improvements. Together, these pathways provide a roadmap for future work that promises further elucidation of student mental health and additional support on the journey toward nurturing an inclusive educational environment.

5.7 Conclusion

Finally, this study supports the complex psychosocial pathways of the ontogeny of mental health among Bengali students that might be helpful in building a predictive model. In this paper, I used several machine learning techniques like Logistic Regression, Random Forest, XGBoost and Support Vector Machine to understand the various features that are responsible for shaping the mental well-being of students. Although some constraints are inherently based on the dataset and survey biases, these results reveal the opportunities in using predictive models to go further beyond screening for student mental health. So, the Logistic Regression model with the highest values of accuracy, precision, recall, and F1score comes out to be the best in this case. Acknowledging that these are positive initial

steps, future efforts must either expand the dataset to include more platforms and granular data, improve survey questions in order to isolate sociological variables from platformspecific factors or develop real-time monitoring tools to achieve a more comprehensive understanding of information-based disinformation. Collectively, the work paves the way for early intervention mental health support in Bengali educational contexts and in particular highlights how data-driven insights can augment student well-being.

References

[1]. Prasath, P. R., Xiong, Y., Zhang, Q., & Jeon, L. (2022). Psychological capital, well-being, and distress of international students. International Journal for the Advancement of Counselling, 44(3), 529-549.

[2]. Mein-Woei, S. (2020, January). Spiritual well-being and mental health of students in s(ACPCH 2019) (pp. 149-152). Atlantis Press.

[3]. Selvaraj, P. R., & Bhat, C. S. (2018). Predicting the mental health of college students with psychological capital. Journal of Mental Health, 27(3), 279-287.

[4]. Feng, Y., & Li, Z. (2020). Empirical analysis on the intervention of music therapy on mental health of college students. Revista Argentina de ClínicaPsicológica, 29(2), 1272.

[5]. Roldán-Merino, J., Lluch-Canut, M. T., Casas, I., Sanromà-Ortíz, M., Ferré-Grau, C., Sequeira, C., ... & Puig-Llobet, M. (2017). Reliability and validity of the Positive Mental Health Questionnaire in a sample of Spanish university students. Journal of psychiatric and mental health nursing, 24(2-3), 123-133.

[6]. Fojo, A. T., Lesko, C. R., Benke, K. S., Chander, G., Lau, B., Moore, R. D., ... & Zeger, S. L. (2020). A learning algorithm for predicting mental health symptoms and substance use. Journal of psychiatric research, 134, 22-29.

[7]. Hu, Y., Stewart-Brown, S., Twigg, L. I. Z., & Weich, S. (2007). Can the 12-item General Health Questionnaire be used to measure positive mental health?. Psychological medicine, 37(7), 1005-1013.

[8]. Banks, M. H., Clegg, C. W., Jackson, P. R., Kemp, N. J., Stafford, E. M., & Wall, T. D. (1980). The use of the General Health Questionnaire as an indicator of mental health in occupational studies. Journal of Occupational Psychology, 53(3), 187-194.

[9]. Coulombe, S., Radziszewski, S., Trépanier, S. G., Provencher, H., Roberge, P., Hudon, C., ... & Houle, J. (2015). Mental health self-management questionnaire: Development and psychometric properties. Journal of Affective Disorders, 181, 41-49.

[10]. Abid Hasan Khan, Mst. Sadia Sultana, Sahadat Hossain, M. Tasdik Hasan, Helal Uddin Ahmed, Md. Tajuddin Sikder, "The impact of COVID-19 pandemic on mental health & wellbeing among home-quarantined Bangladeshi students: A cross-sectional pilot study", ELSEVIER, Journal of Affective Disorders, Vol. 277, pp. 121-128, Dec 1, 2020.

[11]. Rajib Ahmed Faisal, Mary C. Jobe, Oli Ahmed & Tanima Sharker, "Mental Health Status, Anxiety, and Depression Levels of Bangladeshi University Students During the COVID19 Pandemic", Springer Link, Vol. 10, Jan 4, 2021.

[12]. Md. Saiful Islam, Md. Safaet Hossain Sujan, Rafia Tasnim, Md. Tajuddin Sikder, Marc N. Potenza, Jim van Os, "Psychological responses during the COVID-19 outbreak among university students in Bangladesh", PLOS ONE, Vol. 10, Dec 31, 2020

[13]. Sabina Yesmin, Rajon Banik, Shorif Hossain, Md. Najmul Hossain, "Impact of COVID-19 pandemic on the mental health of children in Bangladesh: A cross-sectional study", ELSEVIER,Vol. 117, Oct 2020.

[14]. Srividya, M., Subramaniam Mohanavalli, and Natarajan Bhalaji. "Behavioral modeling for mental health using machine learning algorithms." *Journal of Medical Systems* 42 (2018): 1-12.

[15]. Chung, Jetli, and Jason Teo. "Mental health prediction using machine learning: taxonomy, applications, and challenges." *Applied Computational Intelligence and Soft Computing* 2022 (2022): 1-19.

[16]. Mendis, E. S., et al. "Mobile Application for Mental Health Using Machine Learning." *2022 4th International Conference on Advancements in Computing (ICAC)*. IEEE, 2022.

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

[18]. Kavitha, Mrs M., et al. "CLASSIFICATION ALGORITHM BASED MENTAL HEALTH PREDICTION USING DATA MINING." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 13.2 (2022): 1168-1175.

[19]. Sailaja, N. Venkata, et al. "Happiness Index Prediction of Students Using Machine

Learning." *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023)*. Vol. 18. Springer Nature, 2023.

[20]. Srividya, M., Subramaniam Mohanavalli, and Natarajan Bhalaji. "Behavioral modeling for mental health using machine learning algorithms." *Journal of medical systems* 42 (2018): 1-12.

[21]. Md, Monir., Ahammod, Bin., Md, Ilias, Bappi., Kyungbeak, Kim., Kwanghoon, Choi., Md., Martuza, Ahamad., Khondaker, Masfiq. (2024). 1. Impact of Covid-19 on Bangladeshi university students mental health: ML and DL analysis. medRxiv, doi: 10.1101/2024.05.17.24307476

[22]. Sabiha, Shirin, Sara., Md., Asikur, Rahman., Riaz, Rahman., Ashis, Talukder. (2024). 3. Prediction of suicidal ideation with associated risk factors among university students in the southern part of Bangladesh: Machine learning approach.. Journal of Affective Disorders, doi: 10.1016/j.jad.2024.01.092

[23]. Muntequa, Imtiaz, Siraji., Ahnaf, Rahman., Mirza, Muntasir, Nishat., Md., Abdullah, Al, Mamun., Fahim, Faisal., Lamim, Ibtisam, Khalid., Ashik, Ahmed. (2023). 5. Impact of mobile connectivity on students' wellbeing: Detecting learners' depression using machine learning algorithms. PLOS ONE, doi: 10.1371/journal.pone.0294803

[24]. Stephen, C., Newman. (2022). 8. Machine Learning Classification Algorithms for Predicting Depression Among University Students in Bangladesh. doi: 10.1007/978-981-16-7597-3_6

[25]. Fojo, A. T., Lesko, C. R., Benke, K. S., Chander, G., Lau, B., Moore, R. D., ... & Zeger, S. L. (2020). A learning algorithm for predicting mental health symptoms and substance use. Journal of psychiatric research, 134, 22-29.

[26]. (2022). 18. Predictive Analysis of Student Stress Level Using ML. International Journal of Advanced Research in Science, Communication and Technology, doi: 10.48175/jars-5354.

mental health

ORIGINALITY REPORT

17%	3%	2%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Islamic University of Technology Student Paper	15%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to George Bush High School Student Paper	1%
4	Mokheleli, Tsholofelo Diphoko. "A Comparison of Machine Learning Techniques for Predicting Mental Health Disorders.", University of Johannesburg (South Africa), 2024 Publication	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		