ANXIETYGUARD: EMPOWERING EARLY INTERVENTION WITH A RESILIENT, STREAMLINED, AND EVER-ACCESSIBLE CLOUD-BASED MACHINE LEARNING PREDICTION MODEL FOR ANXIETY DETECTION

Submitted By:

MEHER DURDANA KHAN RAISA

192-35-2818

Department Of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

Supervised By:

AFSANA BEGUM

Assistant Professor & Coordinator M.Sc

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering Spring-2023

ACKNOWLEDGEMENT

I want to start by expressing my gratitude to Almighty God for bestowing His divine favor and allowing me to complete my undergraduate thesis.

I would like to express my gratitude and deep amount of respect to my supervisor, Afsana Begum, **Assistant Professor** in the "Department of Software Engineering" at "Daffodil International University" in Dhaka. His profound knowledge and guidance in the segment on "**Machine Learning**" help me a lot to complete this entire thesis work. It has been made possible by his never-ending empathy, academic leadership, continuous motivation, regular and vigorous monitoring, constructive criticism, helpful counsel, reviewing numerous subpar manuscripts, and fixing them at every level.

I wish to extend my sincere appreciation to **Dr. Imran Mahmud, Head of the "Software Engineering" Department**, Faculty of Science and Information Technology, as well as to the other professors, faculties, and personnel of the SWE Department of "Daffodil International University" for their considerate assistance in accomplishing my work.

Last but not least, I must respectfully thank my parents for their unwavering love and patience.

ABSTRACT

Anxiety disorders significantly impact the health and academic performance of students, necessitating objective and data-driven detection strategies. This research study employed an online survey to collect data for the university students of Bangladesh, including the GAD-7 anxiety severity scale and demographic information. The dataset underwent preprocessing and cleaning, and several classification methods (LDA, KNN, Naive Bayes, Decision Tree, RF, Logistic Regression, and SVM) were analyzed and compared using classification metrics. The findings revealed that Support Vector Machine (SVM) and Random Forest (RF) achieved the highest accuracy rates. To improve upon these individual methods, a hybrid model, AnxietyGuard, was developed, achieving an accuracy of 84%, surpassing the performance of separate methods. These results underscore the potential of machine learning-based systems for accurate anxiety identification among university students. The hybrid approach shows promise in enhancing identification and intervention measures, offering hope for improved support and intervention for individuals with anxiety disorders.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Nomenclatures

KNN - “K- Nearest Neighbor”.

SVM - “Support Vector Machine”.

RF - “Random Forest”.

LDA - “Linear discriminant analysis”.

LR - “Linear Regression”.

NB - “Naive Bayes”.

DT - “Decision Tree”.

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND:

Anxiety is a prevalent mental health condition that affects a significant number of people worldwide, including university students. In the context of Bangladesh, the increasing prevalence of anxiety among undergraduates has become a cause for concern. This research aims to address the urgent need for effective identification and treatment of anxiety among Bangladeshi undergraduate students, taking into account the unique sociocultural factors that influence its manifestation.

1.2 MOTIVATION OF THE RESEARCH:

The detrimental impact of anxiety disorders on individuals' well-being, academic performance, and overall quality of life underscores the importance of developing effective methods for early detection and intervention. By leveraging machine learning techniques and analyzing comprehensive datasets, we can improve our understanding of anxiety in the context of Bangladeshi undergraduates, identify key risk factors, and develop accurate prediction models to guide personalized interventions.

Problem Statement: Despite the high prevalence of anxiety among Bangladeshi undergraduate students, there is a lack of comprehensive research focusing on the predictive abilities of machine learning algorithms specifically in this population. This research aims to address this gap by investigating the use of machine learning algorithms in predicting anxiety among Bangladeshi students, considering the sociocultural context and individual-level characteristics that contribute to anxiety manifestations.

1.3 RESEARCH QUESTIONS:

- What are the key sociodemographic, psychosocial, and scholastic factors associated with anxiety among Bangladeshi undergraduate students?
- How effective are machine learning algorithms in predicting anxiety among Bangladeshi undergraduate students?
- Can accurate prediction models be developed using machine learning techniques to identify individuals at risk for anxiety?

1.4 RESEARCH OBJECTIVES:

- The research objectives are as follows:
- To identify the factors contributing to anxiety among Bangladeshi undergraduate students through comprehensive data analysis.
- To evaluate the performance of machine learning algorithms in predicting anxiety levels among Bangladeshi undergraduate students.
- To develop accurate prediction models utilizing machine learning techniques for identifying individuals at risk for anxiety.
- To provide evidence-based insights and recommendations for tailored interventions and support systems aimed at reducing anxiety symptoms and promoting well-being among Bangladeshi undergraduate students.

1.5 RESEARCH SCOPE:

This research focuses specifically on anxiety among undergraduate students in Bangladesh. The study utilizes machine learning algorithms to analyze comprehensive datasets encompassing sociodemographic factors, scholastic

achievement, psychosocial variables, and other relevant indicators. The research does not specifically consider the impact of the COVID-19 pandemic on anxiety levels among students.

1.6 THESIS ORGANIZATION:

The thesis is organized into several chapters. Chapter 1 provides an introduction to the research, including the background, motivation, problem statement, research questions, objectives, scope, and organization of the thesis. Chapter 2 presents a comprehensive literature review on anxiety disorders, machine learning algorithms, and previous studies related to anxiety prediction. Chapter 3 outlines the methodology employed, including data collection, preprocessing, feature selection, and the implementation of machine learning algorithms. Chapter 4 presents the results and analysis of the study. Finally, Chapter 5 concludes the thesis by summarizing the findings, discussing their implications, and providing recommendations for future research and interventions aimed at addressing anxiety among Bangladeshi university students.

Chapter 2: Literature Review

2.1 INTRODUCTION:

In the division of the literature review, we as researchers assessed prior work, research, conference papers, books, journals, etc. Through it, we discovered what prior study had been done on the topic, provided a broad summary of it, and noted any gaps in their work. After doing an investigation, we concentrated on limitations and discovered workarounds to enhance outcomes.

2.2 PREVIOUS LITERATURE:

The topic of anxiety disorders has been the subject of numerous research projects. During the first wave of the COVID-19 pandemic, a research study discovered that university students in Bangladesh were depressed and anxious. The purpose of this study was to examine how well different machine learning algorithms performed at predicting sadness and anxiety in Bangladeshi university students during the initial COVID-19 pandemic wave (Nayan, et al., 2022). The techniques tested included support vector machine (SVM), logistic regression (75.63%), random forest (89.76%), linear discriminate analysis (75.31%), K-nearest neighbors (89.38%), and naive bayes (70.79%). The findings show that Naive Bayes had the lowest accuracy, followed by Random Forest and K Nearest Neighbors. These results show the potential utility of random forest and K-nearest neighbors in mental health screening during difficult periods like a pandemic, suggesting that they are promising algorithms for predicting sadness and anxiety in this group. To validate these findings and investigate the generalizability of these algorithms to various contexts and populations, additional study and validation are required.

We examined another paper on anxiety and depression. A machine learning strategy was used in a research to forecast the outcome of therapy in a digital mental health intervention for depression and anxiety (Hornstein, Forman-Hoffman, Nazander, Ranta, & Hilbert, 2021). The algorithms put to the test were random forest (AUC = 0.64, SD = 0.04), naive Bayes (AUC = 0.60, SD = 0.04), logistic regression (AUC = 0.61, SD = 0.03), and support vector machines

(AUC = 0.63, SD = 0.06). The findings showed that random forest, support vector machine, logistic regression, and naive Bayes had the highest area under the curve (AUC) for predicting therapeutic outcome. According to these results, random forest may be a suitable algorithm for estimating the success of therapy in the context of online mental health interventions for depression and anxiety. To validate and generalize these findings in broader and more diverse groups, additional study is required.

In a different study, the outcome of therapy in a digital mental health intervention for depression and anxiety was predicted using machine learning (Tennenhouse, Marrie, Bernstein, Lix, & others, 2020). The algorithms put to the test were random forest (AUC = 0.64, SD = 0.04), naive Bayes (AUC = 0.60, SD = 0.04), logistic regression (AUC = 0.61, SD = 0.03), and support vector machines (AUC = 0.63, SD = 0.06). The findings showed that random forest, support vector machine, logistic regression, and naive Bayes had the highest area under the curve (AUC) for predicting therapeutic outcome. According to these results, random forest may be a suitable algorithm for estimating the success of therapy in the context of online mental health interventions for depression and anxiety. To validate and generalize these findings in broader and more diverse groups, additional study is required.

The performance of several screening cutoffs was examined in an observational machine learning study with a focus on optimizing mental health screening techniques within a dementia screening and risk factor app (Kuleindiren, et al., 2022). They discovered that the PHQ-9/GAD-7 screening cutoffs had considerably superior predictive performance than the PHQ-2 screening cutoffs (with a difference in the area under the curve of 0.04, 95% CI 0.00-0.08, P=.02). The PHQ-9/GAD-7 and GAD-2 cutoffs did not significantly differ in terms of performance, but (with a difference in the area under the curve of 0.00, 95% CI -0.02 to 0.03, P=.42). For the PHQ-9 (R2 = 0.655, mean absolute error = 2.267) and GAD-7 (R2 = 0.837, mean absolute error = 1.780), regression models were created to accurately predict total scores. These results imply that regression models can accurately predict total scores for both PHQ-9 and GAD-7, while the optimized PHQ-9/GAD-7 screening cutoffs can offer better

prediction results compared to PHQ-2.

Using a modelling strategy based on machine learning, a study was conducted. The primary aim of this experiment was to forecast the general anxiety levels during the COVID-19 epidemic (Albagmi, Alansari, Al Shawan, AlNujaidi, & Olatunji, 2022). We examined the classification results of J48 Decision Tree and Support Vector Machine (SVM). The findings indicated that, with 10 features, the SVM classifier outperformed the J48 Decision Tree, achieving 100% classification accuracy, 1.0 precision, 1.0 recall, and 1.0 f-measure. These findings imply that the SVM classifier did a better job of predicting generalized anxiety levels during the COVID-19 pandemic when the chosen set of attributes was used. It is critical to underline that additional validation and testing in different datasets are necessary to determine whether our findings are generalizable.

The machine learning method was applied in a study, which examined anxiety and sleep problems during the COVID-19 lockdown (Anbarasi, et al., 2022). A Random Forest (RF) model was used by the researchers to look into aspects that can lead to stress in college students, notably their anxiety levels. Anxiety scores and the identified stress-induced components showed a clear correlation in the Random Forest model. These results indicate that the RF model may be useful in Analysing and determining the elements causing anxiety in college students during the COVID-19 lockdown. To investigate more factors and confirm that the model's performance may be applied to a variety of populations and situations, more study is nonetheless required.

Using machine learning approaches, focuses on the detection and categorization of anxiety among university students (Bhatnagar, Agarwal, & Sharma, 2023). The algorithms put to the test were Decision Tree (71.05%), Support Vector Machine (75.55%), Random Forest (78.90%), and Naive Bayes (71.05%). According to the findings, Random Forest, Support Vector Machine, Decision Tree, and Nave Bayes all performed best at classifying anxiety. These results imply that Random Forest and Support Vector Machine algorithms may be able to identify and categories worry in college students. However, additional investigation is

required to verify these findings and examine the generalizability of the models across various student demographics.

Researchers looked into the lifestyle factors that predicted depression and anxiety during the COVID-19 pandemic using a machine learning approach (Simjanoski, et al., 2022). On several test datasets from Brazil and Spain, the prediction models were assessed. Elastic Net technique had accuracy of 0.78 for the Spain test dataset, while Random Forest (RF) and XGBoost (XGB) had accuracy of 0.78 and 0.77, respectively. The Elastic Net algorithm had an accuracy of 0.765 in the Brazil test dataset, while Random Forest (RF) and XGBoost (XGB) had accuracy values of 0.74 and 0.75, respectively. According to these results, machine learning algorithms, in particular Elastic Net, Random Forest, and XGBoost, have a chance to accurately forecast depression and anxiety during the COVID-19 pandemic in various nations. To validate these findings and investigate the generalizability of these models across other groups and circumstances, additional study and validation are required.

The researchers did a study to evaluate the viability of a machine learning-based smartphone application in identifying sadness and anxiety in a predominantly geriatric population (Lin, et al., 2022). They used a Natural Language Processing (NLP) algorithm. Both the positive and negative groups had 61% of the procedure completed. The positive group reported a usage rate of 27% and the negative group a usage rate of 9% over the norm. Compared to the PHQ-8 and GAD-7 with a threshold score of 10, the Ellipsis Health App showed encouraging results, with an area under the curve (AUC) of 0.82 for the combined groups. The application was shown to function well not just for older participants but also for people of all ages. These results show the potential of the machine learning-based smartphone app in accurately identifying anxiety and depression in the older population as well as in other age groups.

Multiple machine learning algorithms were used in the prospective cohort study that attempted to develop a novel digital biomarker for identifying Generalized Anxiety Disorder (GAD) using non-identifiable smartphone data (Choudhary, et al., 2022). GAD-7a multiclass Random Forest (48%), GAD-7a binary Random Forest (74%), GAD-7 multiclass K-nearest

neighbors (29%), and GAD-7 binary K-nearest neighbors (60%), were the test algorithms, and their corresponding accuracy levels. GAD-7 multiclass logistic regression (38%), GAD-7 binary logistic regression (55%), GAD-7 multiclass XGBoost (50%), and GAD-7 binary XGBoost (76%), among others, were taught algorithms and their accuracy levels. These findings show how different machine learning methods perform differently when diagnosing GAD using data from smartphones, with binary models typically performing more accurately than multiclass models.

The ecological momentary assessment study used a machine learning model based on the Markov Model to investigate the shift in social media app usage that occurred during the COVID-19 lockdown and its association with clinical anxiety symptoms (Ryu, et al., 2021). In 10-fold cross-validation, the model had an accuracy of 62.30% on average (with a standard deviation of 16%) and an AUC of 0.70 (with a standard deviation of 0.19) for its receiver operating curve. Based on these findings, the model's ability to identify the clinical anxiety group based on patterns of social media app activity during the lockdown period is only of a moderately accurate kind.

Another study examined anxiety and depressive symptoms during the COVID-19 epidemic using a machine learning model based on XGBoost (Hueniken, et al., 2021) . With a r2 value of 0.39, the model was successful in explaining a sizable percentage of the variance in emotional discomfort. Increased financial problems, concerns of getting COVID-19, and younger age were the three variables that had the greatest impact on elevated emotional distress (SHAP=0.17, 0.17, and 0.13, respectively). These results emphasize the important role of age, COVID-19-related anxieties, and financial concerns on emotional suffering during the pandemic, as indicated by the XGBoost model.

Researchers wanted to determine the likelihood of psychiatric illnesses among players from Asian countries who participated in Player Unknown's Battlegrounds (PUBG) by employing supervised machine learning strategies in a separate study (Aggarwal, Saluja, Gambhir, Gupta, & Satia, 2020). The purpose of this research was to determine whether or not these

gamers were at risk for the game. When it came to accurately predicting various psychological problems, the Decision Tree algorithm obtained an accuracy of 84.90%. This shows that Decision Tree modelling can be useful in identifying and forecasting the likelihood of psychiatric illnesses among PUBG players from Asian countries.

The application of several machine learning algorithms for behavioral modelling in the context of mental health was explicitly investigated in a different study (Srividya, Mohanavalli, & Bhalaji, 2018). The algorithms tested had accuracy values of 0.84 for Logistic Regression, 0.73 for Naive Bayes, 0.89 for Support Vector Machine, 0.81 for Decision Tree, 0.89 for K Nearest Neighbors, 0.89 for Ensemble (Bagging), and 0.90 for Tree Ensemble (Random Forest). According to these results, Support Vector Machine and K Nearest Neighbors came in second and Ensemble and Tree Ensemble algorithms (Bagging and Random Forest, respectively) came in third with the greatest accuracy in behavioral modelling for mental health.

To distinguish panic disorder from other forms of anxiety disorders, researchers concentrated on creating machine learning algorithms (Na, Cho, & Cho, 2021). The accuracy of the Random Forest (RF) model was 0.649, the accuracy of the Gradient Boosting Machine (GBM) model was 0.676, the accuracy of the Support Vector Machine (SVM) model was 0.784, the accuracy of the Artificial Neural Network (ANN) model was 0.73, and the accuracy of the Logistic Regression model was similarly 0.784. These algorithms underwent testing. According to these results, it seems that SVM, ANN, and LR were more accurate than RF and GBM in differentiating panic disorder from other anxiety disorders.

In an alternate research study, researchers used a multimodal machine learning technique to distinguish between major depression (MD) and generalized anxiety disorder (GAD) (Hilbert, Lueken, Muehlhan, & Beesdo-Baum, 2017). The evaluation of clinical, hormonal, and structural MRI data was part of the analysis. The findings revealed that while questionnaire data performed poorly in disease classification accuracy (56.58%), it performed well in case classification accuracy (96.40%). Cortisol data and imaging data, on the other hand, showed

the opposite trend, with better accuracy in disorder-classification (68.05% for imaging, 74.60% for cortisol) compared to case-classification (58.71% for imaging, 38.02% for cortisol). The overall accuracy for case classification was 90.10%, and for disorder classification it was 67.46% when all data modalities were merged. According to these findings, multimodal machine learning algorithms may be able to discriminate between GAD and MD, and the use of many data sources will increase classification precision.

The screening of anxiousness and depression among mariners was the primary emphasis of the research, which made use of machine learning technologies (Sau & Bhakta, 2019). Several different algorithms, such as CatBoost, Random Forest, Logistic Regression, Naive Bayes, and SVM, were utilized in this process. The accuracy rates that were reached were 76.10% for SVM, 82.60% for CatBoost, 81.20% for Random Forest, 77.80% for Logistic Regression, and 75.80% for Naive Bayes. The findings suggest that machine learning algorithms can be useful in diagnosing anxiety and depression among sailors, with CatBoost displaying the greatest level of accuracy among the algorithms that were investigated for this purpose.

A study looked at pregnant and postpartum women in Cyprus in order to assess the psychometric qualities and factor structure of the Generalized Anxiety Disorder-7 (GAD-7) in Greek (Vogazianos, Motrico, Domínguez-Salas, Christoforou, & Hadjigeorgiou, 2022). The internal consistency and factor structure of the GAD-7 were assessed using reliability coefficients, Cronbach's Alpha and McDonald's Omega, and factor analysis, both exploratory and confirmatory, on a total of 457 Cypriot women in the perinatal period. The findings demonstrated that the GAD-7 had a single component structure and strong internal consistency, demonstrating its validity and reliability as a screening tool for anxiety symptoms in Greek Cypriot women who are pregnant or recently gave birth. However, more research is required to establish the cut-off value and maximize the sensitivity and specificity of the scale. The use of healthcare specialists is advised to use GAD-7 as a standard instrument for screening anxiety symptoms in this population.

During the COVID-19 shutdown, different research on anxiety intended to evaluate the mental health of university students in Bangladesh (Alam, Ali, Banik, Yasmin, & Salma, 2021) . The data was collected online from 509 students, and K-means clustering and confirmatory factor analysis were used to examine it. The findings demonstrated that a sizable proportion of students experienced mental health problems, and that factors like having close family members who were infected with the COVID-19 virus, experiencing insecurity, using social media, and smoking increased mental health imbalances while stressing out about schoolwork, spending more time with family, and performing household chores decreased them. It was established that getting enough sleep and carrying out home duties were preventive strategies. The authors advise putting mental health programming into place and fortifying relationships with family members to improve mental health status.

Examining the frequency and risk factors for depression and anxiety among Bangladeshi school-aged children was the study's goal (Islam, Rahman, Moonajilin, & van Os, 2021). A cross-sectional study included 563 Dhaka City students between the ages of 13 and 18 who attended the chosen schools. There were 26.5% and 18.1%, respectively, of those who reported having moderate to severe depression or anxiety. In contrast to depression being connected to being between the ages of 15 and 16, not having positive connections with peers, and anxiety, depression was linked to poor sleep, smoking, and anxiety. The findings highlight the necessity of assessing, preventing, and treating mental health problems among Bangladeshi school-aged adolescents.

In order to better understand how the COVID-19 pandemic affected Australians' mental health during the acute phase of the outbreak, a study was conducted (Dawel, et al., 2020). Depression, anxiety, and psychological well-being were evaluated using standardized questionnaires. The study discovered that despair and anxiety symptoms were significantly worse, even in those without a history of mental illness. Exposure to COVID-19 had little link with mental health outcomes, in contrast to the pandemic's impairments in employment and social functioning, which were strongly associated with greater dejection and anxiety

symptoms as well as decreased psychological wellbeing. Another important predictor of worse mental health, aside than job loss per se, was financial difficulty brought on by the epidemic. The study emphasizes the importance of lessening disruptions to social and occupational functioning and increasing community access to mental health services to lessen the epidemic's consequences on mental health and wellbeing.

A machine learning algorithm was used to identify risk indicators in another anxiety-based investigation, and the proportion of likely anxiety and probable insomnia among undergraduate students during the COVID-19 pandemic was evaluated (Ge, Zhang, Wu, & Mu, 2020). Freshmen participated in a web survey two months after submitting baseline data for the study to gauge their psychological health. Anxiety and sleeplessness were shown to have probability prevalence rates of 12.49% and 16.87%, respectively. The XGBoost algorithm correctly predicted likely anxiety and insomnia with 97.3% accuracy and 96.2% precision, respectively. In contrast, the most important factors in determining the likelihood of insomnia were suicidal ideation, sleep symptoms, and a history of anxiety symptoms. The most important factors in determining the likelihood of anxiety were aggression, psychotic experiences, suicidal ideation, and romantic relationships. The study demonstrates that undergraduate students who display symptoms of anxiety and insomnia need urgent psychological assistance while taking these risk factors into account.

A study concluded by evaluating the validity and dependability of the GAD-7 questionnaire among Bangladeshi university students (Dhira, Rahman, Sarker, & Mehareen, 2021). The GAD-7 questionnaire exhibited outstanding reliability, and using both EFA and CFA analyses, a modified one-factor model was determined to be appropriate for the sample. In addition, the study found that the GAD-7 questionnaire had good convergent validity because it correlated positively with both the PHQ-9 and PHQ-ADS. The study demonstrates that the modified unidimensional form of the GAD-7 questionnaire is suitable for use in this population and is a valid and reliable method for measuring anxiety in university students in Bangladesh.

An internet-based cognitive behavioral treatment (iCBT) Programme was used in a study to examine anonymized data from 54,604 people who engaged in treating their anxiety and depression symptoms (Chien, et al., 2020). Five distinct patient categories were developed based on the longitudinal variability of iCBT involvement patterns. The study found that more effective therapy outcomes might come from focused medicines for specific engagement subtypes. Whether or not clinician support and supporters use machine learning insights effectively may depend on their well-informed choices. The findings suggest that people with depression and anxiety may gain from individualized therapy.

Although Bangladeshi university students are especially susceptible to stress and mental health problems, the effects of the COVID-19 pandemic on their wellbeing have not been thoroughly researched (Faisal, Jobe, Ahmed, & Sharker, 2022). In April 2020, 874 university students from Bangladesh were given access to three online mental health assessment tools by researchers to close this disparity. According to their findings, 53% of participants had a moderate to poor mental health state, 40% of participants had moderate to severe anxiety, and 72% of participants had depressive symptoms. A path analysis showed that anxiety and a moderate to poor mental health state were predicted by worrying about COVID-19 and knowing about the virus, while depressive symptoms were predicted by understanding and believing that COVID-19 was a serious problem in Bangladesh. The implications of these results are crucial for comprehending how the pandemic has impacted students' mental health.

The research used machine learning to forecast the recovery from anxiety disorders within two years (Bokma, et al., 2022). In order to predict recovery from anxiety disorders and all other common mental disorders at the 2-year follow-up, the researchers chose 887 patients with anxiety disorders and used a wide range of baseline predictors from five domains. 54.6% of patients had recovered from their anxiety conditions at follow-up, according to the findings. When using all of the predictor domains to forecast recovery from anxiety disorders, the random forest classifiers obtained a cross-validated area-under-the-receiving-operator-

characteristic-curve of 0.67. The research discovered that while depression features were primarily responsible for recovery from all other common mental disorders, anxiety features were primarily responsible for predicting recovery from anxiety disorders.

An investigation of depression and anxiety among Bangladeshi university students was conducted during the COVID-19 pandemic to determine their prevalence and related causes (Islam, Barna, Raihan, Khan, & Hossain, 2020). According to a cross-sectional web-based study with 476 participants, 15% of students reported having moderately severe depression, while 18.1% reported having significant anxiety. In older students and those who had provided private lessons before the pandemic, depression was more prevalent, according to a binary logistic regression analysis. The study suggests that the government and institutions should collaborate to resolve academic delays and financial concerns in order to minimize depression and anxiety among university students.

Generalized anxiety disorder (GAD) is a common mental health condition that affects older people who live in communities (Gonçalves, Pachana, & Byrne, 2011). A study was conducted to identify GAD prevalence and look at its associated conditions. The sample was made up of 3035 individuals, with ages ranging from 55 to 85. A weighted 12-month prevalence of GAD was found in 2.8% of the population, according to the results. Elderly age, functional limitations, prior depression comorbidity, worries of acquiring a major illness despite doctor's assurances, and family history of anxiety or depression were the most significant predictors of 12-month GAD in older people. These findings emphasize the need for increased clinical awareness of GAD among older adults because to its high prevalence and close ties to functional limitations, mental illness, and higher medication usage.

Although understudied and lacking operationalized criteria, generalized anxiety disorder (GAD) is a relatively dependable anxiety syndrome (Rapee, 1991). High trait anxiety is equivalent to GAD, the "basic" anxiety disorder. More study of GAD may be beneficial for all anxiety disorders. Few studies have focused on the etiological factors that contribute to

GAD, however recent research has identified possible supporting factors. An illustration of GAD maintenance is given from the viewpoint of information processing. Treatment research has shown promise, despite the restrictions they now face. The importance of future GAD research is emphasized as this review outlines some of the issues that require further investigation.

During the early phases of the COVID-19 epidemic, research was conducted to gauge the level of fear and worry among the larger Bangladeshi population (Islam, Ferdous, & Potenza, Panic and generalized anxiety during the COVID-19 pandemic among Bangladeshi people: An online pilot survey early in the outbreak, 2020). A cross-sectional online survey was conducted in Bangladesh with 1311 community residents between the ages of 13 and 63. Estimates of panic episodes and general anxiety disorders were found to be 79.6% and 37.3%, respectively. All of the factors that statistically predicted panic included age, education level, marital status, and living with a large family. Generalized worry was statistically predicted by being a woman, being older, having a higher degree, being married, and working for a non-governmental organization. The study's flaws were the use of self-reported measurements and possible bias in the sampling. The findings highlight the need for increased surveillance, evidence-based intervention programmes, and supporting services to address panic and generalized anxiety in Bangladeshi individuals during the COVID-19 epidemic.

In order to identify any potential underlying causes as well as the prevalence of depression and anxiety among first-year undergraduate students at Jahangirnagar University in Bangladesh, a study was conducted on these issues (Islam, Akter, Sikder, & Griffiths, 2020). 400 students, aged 18 to 23, took part in a cross-sectional study. The prevalence rates for mild to very severe depression and anxiety were 69% and 69.5%, respectively, with no significant gender variations. The two main risk factors for depression were insufficient physical activity and inadequate sleep. The main source of anxiety risk was too much internet use. According to the research, first-year students are a population that is particularly susceptible to mental diseases, hence Bangladeshi university students need intervention courses and appropriate

supportive services.

In order to assess the reliability of the Hospital Anxiety and Depression Scale (HADS), often known as the HADS, it was decided to use it as a screening tool for both anxious and depressive illnesses in research on anxiety (Brennan, Worrall-Davies, McMillan, Gilbody, & House, 2010). According to the findings of the meta-analytic pooling analysis of 41 studies, major depressive disorders had a sensitivity of 0.82 and a specificity of 0.74 at a cut point of 8, and a sensitivity of 0.56 and a specificity of 0.92 at a cut point of 11. Based on the fact that the cut point was 8, these conclusions were drawn. The HADS was found to be a useful screening tool for recognizing emotional distress in nonpsychiatric patients, but it did not outperform other screening tools in the process of diagnosing certain mental illnesses in scenarios involving physical health.

Mental health problems including stress, anxiety, and depression are more prevalent in today's fast-paced world (Priya, Garg, & Tigga, 2020). In research, employed and unemployed participants from various cultures and groups utilized machine learning algorithms to forecast their levels of stress, anxiety, and sadness. The information was gathered using the Depression, Anxiety and Stress Scale (DASS 21) questionnaire. Using five different algorithms, the degrees of stress, sadness, and anxiety were predicted. The study found that the accuracy of the models was affected by the imbalanced classes of the confusion matrix. Consequently, using the f1 score measurement, it was found that the Random Forest classifier was the most accurate model among the five techniques. It was shown that the algorithms were sensitive to unfavorable outcomes based on the specificity parameter.

The study intended to compare the subjective quality of life of older adults with GAD to that of other worried and nonpsychiatric groups (Bourland, et al., 2000). According to the findings, older persons with GAD had a worse quality of life than samples from nonpsychiatric categories. Their levels of life satisfaction, on the other hand, were comparable to those of younger adults who experienced social anxiety. The study also looked at the quality-of-life determinants in the GAD group. It was discovered that the degree of sadness

and anxiety both predicted a lower level of life satisfaction and a worse level of quality of life. On the other side, optimism indicated more life pleasure. Overall, the data imply that older persons with GAD have a worse quality of life, and that other factors such as their degree of melancholy and anxiety, as well as their optimism, have a role. The goal of treating GAD should be to enhance the quality of life of elderly persons who suffer from the illness.

An article investigates the application of machine learning (ML) in the healthcare industry's huge data processing, specifically for the diagnosis of mental illness (Cho, Yim, Choi, Ko, & Lee, 2019). It highlights common misunderstandings regarding machine learning algorithms, such as the belief that deep learning is the only ML approach or that it can solve the problem of small sample sizes. This paper examines five standard ML techniques: Support Vector Machines, Gradient Boosting Machines, Random Forest, Naive Bayes, and K-Nearest Neighbor. They are often used in mental health research and are organized and summarized carefully. The results reveal that, despite the fact that each ML approach has distinct advantages, many researchers fail to explain why they are using it, and some even employ it without fully knowing the data's features. When dealing with limited data, researchers should be aware of the features and restrictions of ML algorithms, as well as the consequences of these algorithms, according to the paper.

From 1990 to 2017, a study provides a thorough analysis of the incidence and illness burden of mental diseases in each of the states of India (Sagar, et al., 2020). The research calculated the prevalence of mental illnesses, years lived with disability (YLDs), and disability-adjusted life years (DALYs) caused by these diseases using data from various sources. 197.3 million individuals in India had mental illnesses in 2017, including 45.7 million people with depression and 44.9 million people with anxiety disorders, the survey said. The percentage of DALYs in India caused by mental illnesses increased from 25% in 1990 to 47% in 2017. The paper also identifies important differences in the patterns of these differences over time and in the prevalence of various mental diseases between states, which may serve as a guide for appropriate policies and actions from the health system to more effectively address the burden

of mental disorders in India.

Data collection may now be done secretly and continually thanks to smartphones and other pervasive and personal sensing devices (Garcia-Ceja, et al., 2018). In order to predict contextual information about people, like their location, mood, and physical activity, researchers have applied machine learning. Recently, there has been interest in applying these technologies to address mental health issues, especially for automated continuous monitoring of mental health issues including depression, anxiety, and stress. In a study, machine learning-based mental health monitoring systems based on sensor data are examined. The authors look at the basic phases of mental health monitoring systems and provide a categorization taxonomy to guide the evaluation of relevant research. Additionally, they highlight the challenges of doing research in the field and the promise for future advancements in mental health care due to ubiquitous sensing technologies.

Strategies based on machine learning are being used in the area of physical healthcare at an increasing pace (Tiffin & Paton, 2018). In the context of mental health, we discuss some of the possible advantages of this method in an article, as well as some of the difficulties and limitations that are connected with it. We present a number of different circumstances in which machine learning may show to be more advantageous than more traditional statistical models.

As the field of physical healthcare makes quick progress towards the implementation of these methods, approaches that are predicated on the concept of machine learning are rapidly becoming more widespread in the sector. In the context of mental health, it was analyzed in an article some of the prospective benefits of employing this technique, in addition to some of the challenges and limits that are associated with it. This is done in comparison to the setting of mental illness. It offered a number of different scenarios in which machine learning may prove to be more beneficial than more traditional statistical models. Its needed complete control over the level of complexity that each of these scenarios has.

The objective of the systemic review was to compile and evaluate research that used machine

learning and natural language processing (NLP) methods to address mental health issues (Le Glaz, et al., 2021). 327 papers in total, 58 of which were included in the evaluation, were found in medical databases. The investigations concentrated on identifying symptoms, categorizing sickness severity, evaluating the efficacy of various therapies, supplying psychopathological hints, and questioning conventional nosology. Medical records and social media were the sources of the data, which underwent typical NLP preprocessing techniques as well as the extraction of unique identifiers from medical texts. Python was the most often utilized platform, and efficient classifiers were chosen. However, the review points out that rather of producing new knowledge, machine learning and natural language processing (NLP) models frequently corroborate clinical theories, and that before applying them to mental health care, ethical concerns should be taken into account.

The quality of life for patients is improved by early detection of mental health issues in youngsters. Techniques for machine learning have shown promise in the diagnosis of medical data, particularly issues with mental health (Sumathi, Balasubramanian, & Ramaraj, 2017). Eight machine learning techniques were tested in a study to see how well they diagnosed five common mental health issues. A data collection of sixty cases was gathered for the study for testing and training purposes. Following the identification of 25 attributes as critical to the problem's diagnosis, feature selection algorithms were used to condense the attribute data set. The outcomes demonstrated that the Multilayer Perceptron, Multiclass Classifier, and LAD Tree classifiers generated the most accurate results, with just a tiny difference in their performances over the complete attribute set and the selected attribute set. This study emphasizes the potential of machine learning methods for identifying children's mental health issues.

2.3 SUMMARY:

In our research, we experimented with to address as many of the deficiencies of the previously cited studies as possible. From the aforementioned studies, it can

be deduced that a variety of techniques or algorithms were used to identify anxiety disorders among university students. They processed the data and then applied the required algorithm that they concentrate on.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 INTRODUCTION:

We conducted a survey with university students, collecting data on anxiety levels using a Google Form. After preprocessing and cleansing the data, we utilized various machine learning techniques to analyze and predict anxiety severity. Our approach involved exploring the dataset, selecting and engineering features, and developing a hybrid model for improved accuracy. The workflow encompassed data collection, model development, and rigorous assessment, providing valuable insights into anxiety levels among university students.

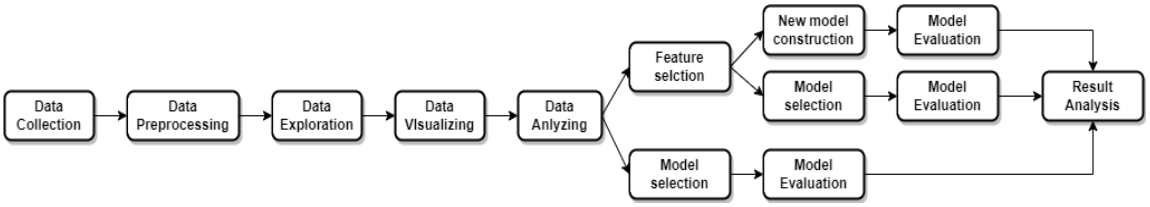

Figure 3.1: Research Flow

3.2 DATASET COLLECTION:

For the purposes of producing data for the thesis and gathering information for the survey, a google form questionnaire containing all of the required information was created. Following that, the survey was administered to all of the Bangladeshi university students. In total, there were 2552 responses to the question. Eighty-three percent of the students who took part were between the ages of 18 and 23, fifteen-point eight percent were between the ages of 24 and 28, and only one point nine four percent were older than 28. On the other hand, there were 999 women who participated in the survey, which is equivalent to 391% of the total, and there were 1552 men, which is equivalent to 60% of the total. To

put it another way, the survey received the most responses from female respondents. The survey also has one participant who identifies as a third gender, which accounts for 0.1% of all replies.

3.3 DATA EXPLORATION & PREPROCESSING:

I conducted a thorough analysis of the dataset by using data exploration and processing tools such as Pandas, Matplotlib, and Seaborn. the treatment of outliers and the identification of data that are missing. I always employed hot encoding to make sure that all of the data were represented in numerical form. I worked on the data to execute feature engineering, which consisted of encoding numerical features, encoding categorical variables, and converting the data. The methods of data visualization were utilized in order to get insights into the structure and content of the dataset as well as to grasp the interactions that exist between the variables. The dataset was then split into a training dataset and a testing dataset. Following completion of this task. The dataset is presently ready to be used in the construction of an accurate machine learning model.

3.4 ANALYZING OVERFITTING AND UNDERFITTING:

During the course of the investigation, I carried out a thorough investigation and assessment of the possible complications that may arise from our machine learning model being overfit or underfit. In order to identify instances of overfitting, I first segmented the dataset into training and testing sets before using a number of different machine learning techniques. In addition, we examined the compatibility of the model, monitored the relevance of the features, and evaluated the performance of the model on the validation set in order to strike a balance between the model's level of complexity and its ability to generalize its findings.

In addition, we paid great attention to underfitting and successfully tuned our model to capture the underlying patterns in the data. This allowed us to avoid overfitting. The results of the model's performance on a variety of assessment measures were greatly improved as a direct result of this exhaustive examination.

3.5 FEATURE SELECTION & ENGINEERING:

The decision to use any of feature selection and engineering techniques for my thesis research depended on a number of variables, including the characteristics of the data, the nature of the classification problem, and the analysis algorithms chosen. To determine the most informative characteristics for predicting anxiety severity levels among university students, I employed a variety of methods. These included statistical techniques such as correlation analysis for identifying variables with significant relationships to the objective variable. In addition, I investigated techniques such as forward/backward feature selection and recursive feature elimination to iteratively select the most pertinent features based on their effect on the model's performance. In addition, I took domain knowledge and expert insights into account when designing new features that could potentially capture key aspects of student apprehension. This entailed constructing interaction terms, polynomial features, or deriving ratios between variables to provide a more complete data representation. By meticulously selecting and engineering features, I intended to improve the model's accuracy, reduce dimensionality, and ensure that the most pertinent information was used to predict anxiety severity levels with high precision.

3.6 MODEL SELECTION:

The choice to use these machine learning algorithms in my study was made in

light of the fact that they have shown effectiveness in classification tasks in a number of previous studies, particularly in the area of medical diagnostics. Each approach has unique characteristics that make them suitable for detecting anxiety in the classification-based dataset. The use of linear discriminant analysis (LDA) to handle multicollinearity and reduce dimensionality is well known. The proximity of data points affects the non-parametric K-nearest Neighbors (KNN) method. The foundation of naive Bayes is Bayesian probability, which assumes independence between attributes. The decision tree algorithms provide a set of comprehensible categorization criteria. Random Forest (RF) combines many decision trees in order to improve accuracy and handle complex interactions. Logistic regression is often used for binary classification issues. The Support Vector Machine (SVM), which has a strong theoretical underpinning and can handle high-dimensional data by maximizing the margin between classes, is the last example. I used this diverse variety of machine learning approaches to assess their success and choose the one that would be most suitable for precisely identifying anxiety in the dataset.

3.7 DEVELOPMENT OF A NEW HYBRID MODEL:

There are various ways that distinct models can function. Anxiety levels were predicted in this study using a variety of categorization techniques. These techniques include Linear Discriminant Analysis (LDA), Naive Bayes, Decision Tree, K Nearest-Neighbors (KNN), Random Forest, Logistic Regression, and Support Vector Machine (SVM). SVM and Random Forest each had accuracy results that were 83% and 78%, respectively, when the performance of these models was evaluated. Even though each model performed rather well on its own, the intention was to combine the advantages of the two algorithms to

increase the prediction capability. SVM and Random Forest were used together to produce a novel hybrid model. The hybrid model seeks to capitalize on each base algorithm's distinct traits and methods for making decisions in order to potentially increase predictive performance. In order to merge the models, ensemble learning methods were used to combine the outputs of both models. In particular, a Voting Classifier, which combines multiple models by taking into account their combined predictions, was used. The soft voting technique, which considers the anticipated probabilities from each model, was incorporated into the hybrid model's architecture. Utilizing SVM and Random Forest as its basic estimators, the hybrid model was developed utilizing training data from the individual models. The hybrid model attempted to increase the overall predicted accuracy by taking into account the combined predictions of both models. Predictions were made on the test dataset once the hybrid model had been trained. The projected anxiety levels and the actual test set labels were compared in order to assess the hybrid model's accuracy. The evaluation showed that the hybrid model was more accurate than both the SVM and the Random Forest models taken separately. This shows that combining the advantages of Random Forest and SVM can result in better performance when predicting anxiety levels. Based on each model's excellent accuracy in prior anxiety prediction experiments or research, SVM and Random Forest were chosen as the components of the hybrid model. The hybrid model seeks to capture a more thorough picture of the underlying data patterns by taking advantage of the complementary capabilities of two methods, which could potentially result in increased accuracy. In order to improve the predictive performance for anxiety level prediction, the hybrid model was developed using an ensemble learning approach that combined SVM and

Random Forest.

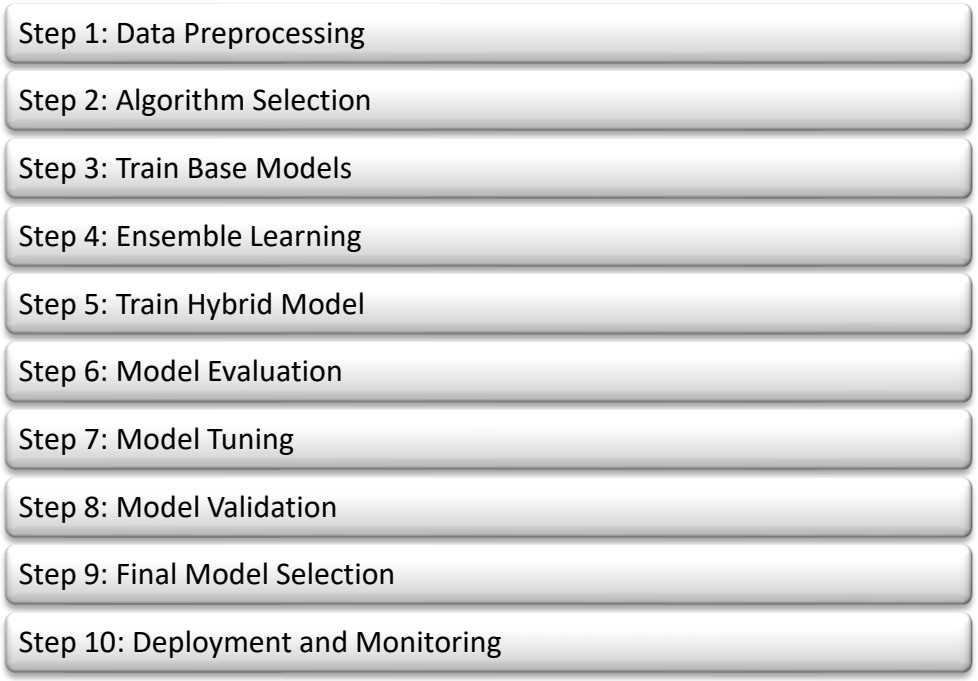

Figure 3.2: Steps of preparing the hybrid model

CHAPTER 4: RESULTS AND DISCUSSION

4.1. INTRODUCTION:

Following the stage of data gathering and preparation, we proceeded to describe the method for the model's actual implementation. In this section, we will discuss the final outcome of the model once it has been trained.

4.2. RESULT:

4.2.1 Data Analysis

The distribution of each variable is shown in tables 1,2 & 3. We performed extensive activities for data analysis, including analyzing the data distribution and computing key metrics like count, mean, standard deviation, minimum, 25th percentile, 50th percentile, and 75th percentile, as well as max-min. By analyzing these measurements, we were able to learn important things about the dataset's properties, gauge how full the data was, and comprehend central tendency and spread. Through careful investigation, we also succeeded in finding outliers and missing value. We were able to fully comprehend the shape and skewness of the data distribution via our calculations of percentiles. We now have a strong foundation for our next data analysis, modeling, and decision-making processes thanks to this work.

	GAD-1	GAD-2	GAD-3	GAD-4	GAD-5	GAD-6	GAD-7	Anxiety_Severity_level
count	2552	2552	2552	2552	2552	2552	2552	2552
mean	1.410658	1.565439	1.692006	1.338558	1.237853	1.545846	1.487069	1.30838558
std	1.009098	1.053837	1.043153	1.003226	1.03839	1.053012	1.092287	0.942548101
min	0	0	0	0	0	0	0	0
25%	1	1	1	1	0	1	1	1
50%	1	1	2	1	1	1	1	1
75%	2	3	3	2	2	3	3	2
max	3	3	3	3	3	3	3	3

Table 1.1

	PHQ-1	PHQ-2	PHQ-3	PHQ-4	PHQ-5	PHQ-6	PHQ-7	PHQ-8	PHQ-9	Depression_Severity_level
count	2552	2552	2552	2552	2552	2552	2552	2552	2552	2552
mean	1.316223	1.571317	1.49373	1.561912	1.235502	1.498433	1.36442	1.064263	1.068574	2.056818182
std	0.93367	0.997256	1.034471	0.982843	1.028688	1.048789	1.047124	1.005173	1.079912	1.263850954
min	0	0	0	0	0	0	0	0	0	0
25%	1	1	1	1	0	1	1	0	0	1
50%	1	1	1	1	1	1	1	1	1	2
75%	2	2	2	2	2	2	2	2	2	3
max	3	3	3	3	3	3	3	3	3	4

Table 1.2

	Where_Are_You_Currently_Living?	Marital_status	Gender	Age	University_Type?	When_do_you_usually_go_to_sleep?	How_many_hours_do_you_sleep_da	How_much_time_do_you_usually_spend_daily_in_front_of_the_s
count	2552	2552	2552	2552	2552	2552	2552	2552
mean	0.306818182	1.928683386	1.386363636	0.170455	3.34678683	0.636363636	1.838166144	1.284090909
std	0.461263652	0.27220687	0.487815254	0.398376	2.02668339	0.831689456	0.930379853	0.936860921
min	0	0	0	0	0	0	0	0
25%	0	2	1	0	1	0	2	0
50%	0	2	1	0	4	0	2	1
75%	1	2	2	0	5	1	2	2
max	1	2	2	2	6	2	3	3

Table 1.3

We have determined the association between any variable and the goal. For instance, Figure 4.4 shows the relationship between the target variables and the degree of anxiety in a scatterplot. Using a histogram as a graphical representation, Figure 4.5 depicts the distribution of GAD variables. The correlation matrices for all the variables have been created. Figure 4.6. We looked at the relationship between all of them as we had both continuous and categorical variables.

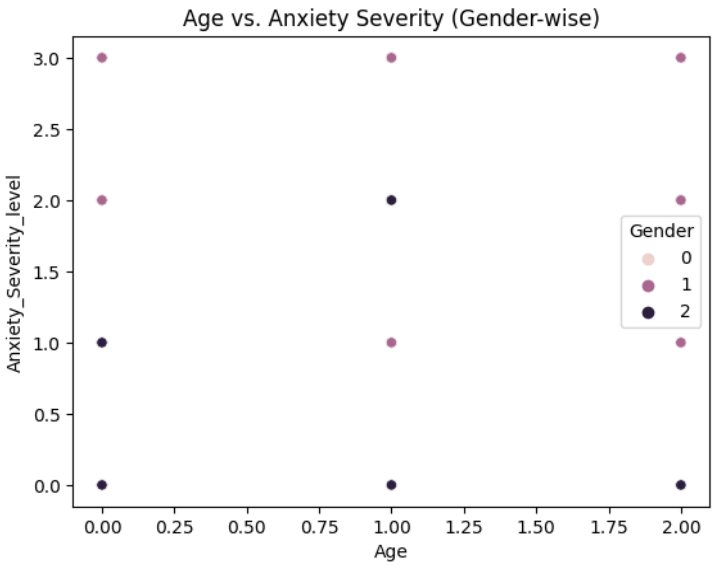

Figure 4.1: Scatter-Plot (Age vs. Anxiety Severity (Gender wise))

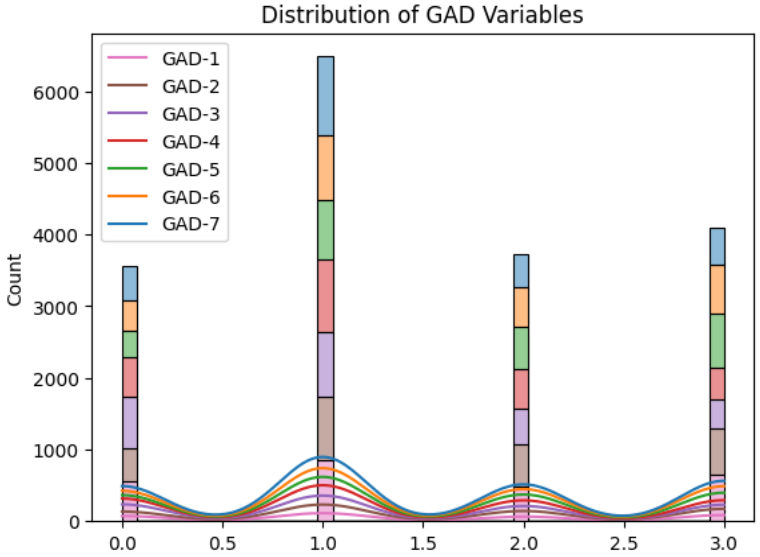

Figure 4.2: Distribution of GAD Variable

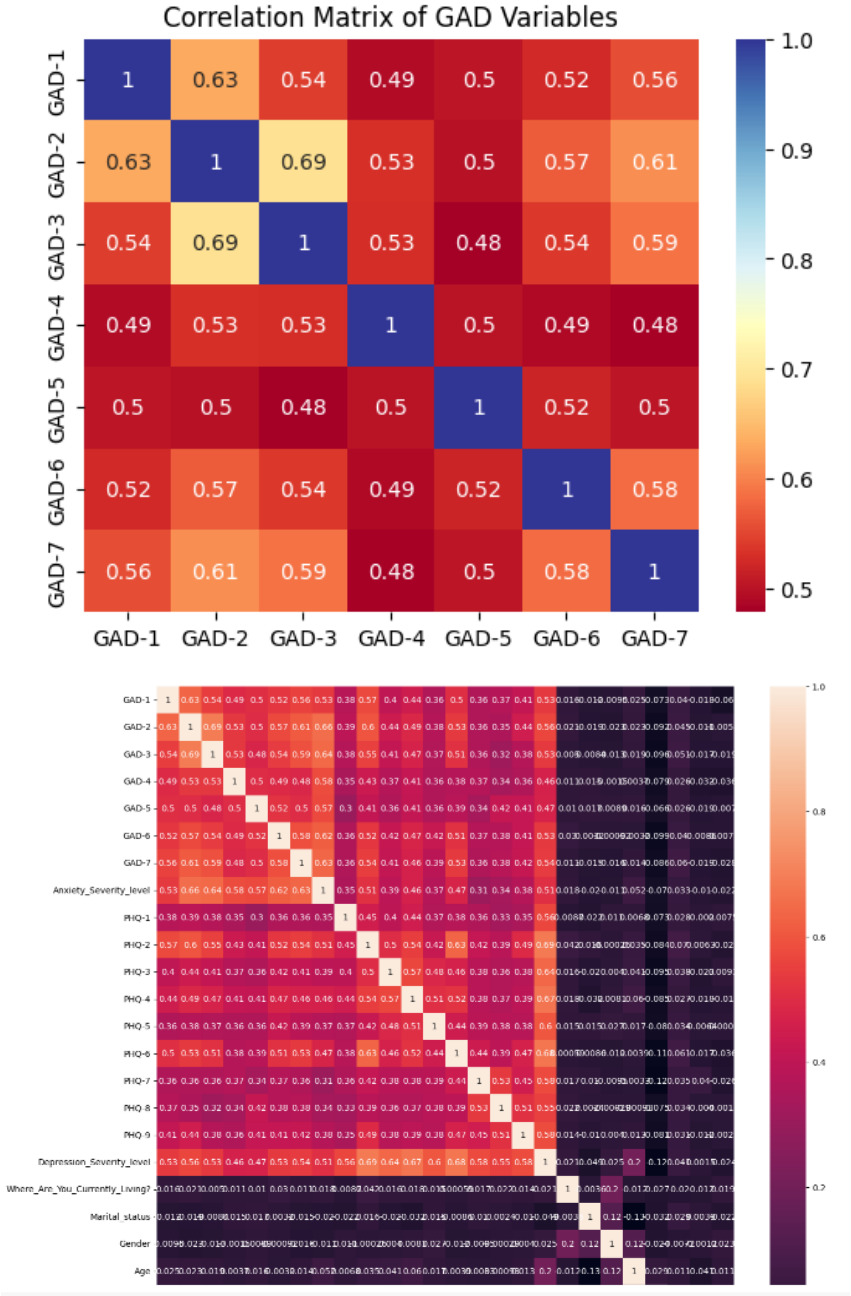

Figure 4.3: Correlation Matrix

4.2.2 Existing model implementation:

In order to diagnose anxiety disorder, we used LR, RF, LDA, SVM, NB, DT, and KNN models respectively. The effectiveness of each model to identify an anxiety condition is seen in Figure 4.7. KNN we to 071, LDA we got 0.74, LR we got 0.74, NB we got 0.72, RF we got 0.78, SVM 0.83 we got, DT we got 0.64 of

accuracy. The categorization of the algorithms that were actually implemented may be seen in Figure 4.7.

KNN Regression Classification Report:

	precision	recall	f1-score	support
0	0.92	0.66	0.77	116
1	0.68	0.80	0.74	198
2	0.65	0.69	0.67	138
3	0.70	0.59	0.64	59
accuracy			0.71	511
macro avg	0.74	0.69	0.70	511
weighted avg	0.73	0.71	0.72	511

LDA Classification Report:

	precision	recall	f1-score	support
0	0.90	0.72	0.80	116
1	0.70	0.84	0.76	198
2	0.69	0.62	0.65	138
3	0.71	0.71	0.71	59
accuracy			0.74	511
macro avg	0.75	0.72	0.73	511
weighted avg	0.75	0.74	0.74	511

LR Classification Report:

	precision	recall	f1-score	support
0	0.90	0.77	0.83	116
1	0.71	0.85	0.77	198
2	0.70	0.58	0.63	138
3	0.72	0.73	0.72	59
accuracy			0.74	511
macro avg	0.76	0.73	0.74	511
weighted avg	0.75	0.74	0.74	511

NB Regression Classification Report:

	precision	recall	f1-score	support
0	0.76	0.80	0.78	116
1	0.72	0.69	0.70	198
2	0.73	0.59	0.66	138
3	0.55	0.81	0.65	59
accuracy			0.70	511
macro avg	0.69	0.72	0.70	511
weighted avg	0.71	0.70	0.70	511

RF Classification Report:

	precision	recall	f1-score	support
0	0.94	0.78	0.85	116
1	0.78	0.83	0.81	198
2	0.70	0.75	0.72	138
3	0.75	0.71	0.73	59
accuracy			0.78	511
macro avg	0.79	0.77	0.78	511
weighted avg	0.79	0.78	0.79	511

SVM Classification Report:

	precision	recall	f1-score	support
0	0.98	0.82	0.89	116
1	0.85	0.84	0.84	198
2	0.78	0.83	0.80	138
3	0.66	0.80	0.72	59
accuracy			0.83	511
macro avg	0.82	0.82	0.82	511
weighted avg	0.84	0.83	0.83	511

DT Classification Report:

	precision	recall	f1-score	support
0	0.73	0.73	0.73	116
1	0.64	0.66	0.65	198
2	0.59	0.54	0.56	138
3	0.61	0.64	0.63	59
accuracy			0.64	511
macro avg	0.64	0.65	0.64	511
weighted avg	0.64	0.64	0.64	511

Figure 4.4: Classification of implemented algorithm

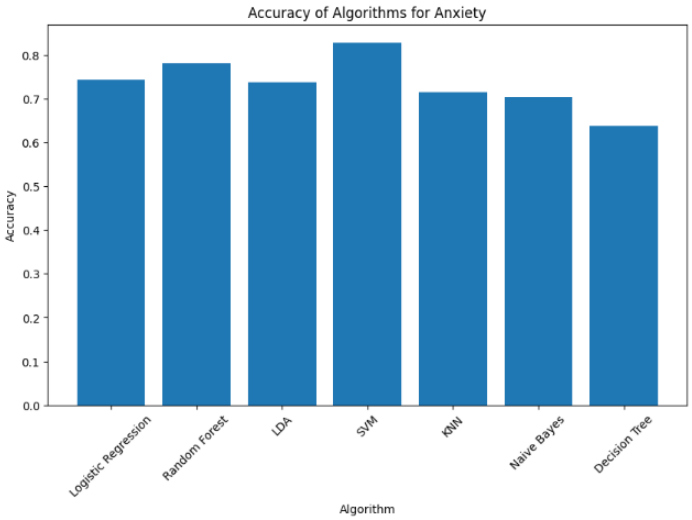

Figure 4.5: Bar chart of algorithms accuracy

4.2.3 Hybrid model implementation:

We decided to design a new model in order to improve the level of accuracy of our findings when detecting anxiety disorders. Which results in a model that is a mix of SVM and RF. Both of these models have the greatest degree of accuracy in comparison to the others. These two models were selected for the process of constructing a hybrid model primarily for this one reason, which was the rationale stated above. In addition, this model carried out well for the study that we conducted. A value of 0.84 was obtained for the accuracy of the hybrid model. Figure 4.9 provides us with information on the correctness of the hybrid model as well as the results of its different metrics.

Hybrid Classification Report:

	precision	recall	f1-score	support
0	0.96	0.84	0.90	116
1	0.86	0.84	0.85	198
2	0.76	0.83	0.79	138
3	0.75	0.80	0.77	59
accuracy			0.84	511
macro avg	0.83	0.83	0.83	511
weighted avg	0.84	0.84	0.84	511

Figure 4.6: Classification report of hybrid algorithm

4.3. DISCUSSION:

As we predicted, the newly developed hybrid model is performing much better. The accuracy was more than the previously implemented algorithm models. We choose the models which gave highest accuracy among the implemented models. That's why it's showing that much accuracy. Comparatively, our hybrid model is showing much more accuracy than our models which have already been implemented.

4.4. SUMMARY:

In this segment, we discussed the final results of our proposed-model. The hybrid is used for the prediction for anxiety disorder in our dataset.

CHAPTER 5: CONCLUSION & RECOMMENDATION

Comparing machine learning algorithms for identifying anxiety disorders in college students was the goal of this thesis. We examined a dataset that was compiled via surveys in order to test the effectiveness of several algorithms in accurately predicting the degrees of anxiety intensity. All of the tested algorithms performed well, with Random Forest displaying the highest degree of accuracy. This research contributes to our understanding of how to recognize anxiety in college students by employing data-driven techniques. The findings emphasize the possibility for early detection and treatment of anxiety-related issues using machine learning techniques. Specific support systems for enhancing the mental health and wellbeing of university students may be developed on the basis of additional research.

5.1 FINDINGS & CONTRIBUTIONS:

The research paper "AnxietyGuard: Empowering Early Intervention with a Resilient, Streamlined, and Ever-Accessible Cloud-based Machine Learning Prediction Model for Anxiety Detection" presents findings and contributions in the field of anxiety detection among university students. Through the analysis of survey data from 2552 participants, the study identified patterns and indicators of anxiety, contributing to the understanding of its prevalence and early detection in this population. A hybrid model combining multiple machine learning algorithms was designed, resulting in improved accuracy compared to individual algorithms. Additionally, the development of a cloud-based prediction model offers increased

accessibility, scalability, and real-time monitoring. The research emphasizes the importance of early intervention by detecting anxiety levels at an early stage, aiming to mitigate its negative impact on academic performance and well-being. By integrating resilience measures, the study expands the understanding of anxiety in a holistic manner, paving the way for interventions that promote resilience-building among university students. Overall, the findings and contributions of this research advance the field of anxiety detection, providing effective tools, empowering early intervention, and contributing to the well-being of university students.

5.2 RECOMMENDATIONS FOR FUTURE WORK:

Future research should concentrate on performing longitudinal studies to monitor changes in university students' anxiety levels over time. Predictive accuracy and generalizability will be improved by investigating advanced feature engineering approaches, combining different data sources, and verifying models in various cultural settings. For ethical implementation and prompt action, real-time monitoring systems development and ethical issues must also be addressed. These initiatives will advance the diagnosis of anxiety disorders and improve university students' wellbeing.

References

Aggarwal, S., Saluja, S., Gambhir, V., Gupta, S., & Satia, S. P. (2020). Predicting likelihood of psychological disorders in PlayerUnknown's Battlegrounds (PUBG) players from Asian countries using supervised machine learning. *Addictive behaviors, 101*, 106132.

Alam, M. K., Ali, F. B., Banik, R., Yasmin, S., & Salma, N. (2021). Assessing the mental health condition of home-confined university level students of Bangladesh due to the COVID-19 pandemic. *Journal of Public Health*, 1–8.

Albagmi, F. M., Alansari, A., Al Shawan, D. S., AlNujaidi, H. Y., & Olatunji, S. O. (2022). Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach. *Informatics in Medicine Unlocked, 28*, 100854.

Anbarasi, L. J., Jawahar, M., Ravi, V., Cherian, S. M., Shreenidhi, S., & Sharen, H. (2022). Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. *Health and Technology, 12*, 825–838.

Bhatnagar, S., Agarwal, J., & Sharma, O. R. (2023). Detection and classification of anxiety in university students through the application of machine learning. *Procedia Computer Science, 218*, 1542–1550.

Bokma, W. A., Zhutovsky, P., Giltay, E. J., Schoevers, R. A., Penninx, B. W., Van Balkom, A. L., . . . Van Wingen, G. A. (2022). Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach. *Psychological Medicine, 52*, 57–67.

Bourland, S. L., Stanley, M. A., Snyder, A. G., Novy, D. M., Beck, J. G., Averill, P. M., & Swann, A. C. (2000). Quality of life in older adults with generalized anxiety disorder. *Aging & Mental Health, 4*, 315–323.

Brennan, C., Worrall-Davies, A., McMillan, D., Gilbody, S., & House, A. (2010). The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *Journal of psychosomatic research, 69*, 371–378.

Chien, I., Enrique, A., Palacios, J., Regan, T., Keegan, D., Carter, D., . . . others. (2020). A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA network open, 3*, e2010791–e2010791.

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation, 16*, 262.

Choudhary, S., Thomas, N., Alshamrani, S., Srinivasan, G., Ellenberger, J., Nawaz, U., & Cohen, R. (2022). A machine learning approach for continuous mining of nonidentifiable smartphone data to create a novel digital biomarker detecting generalized anxiety disorder: prospective cohort study. *JMIR Medical Informatics, 10*, e38943.

Dawel, A., Shou, Y., Smithson, M., Cherbuin, N., Banfield, M., Calear, A. L., . . . others. (2020). The effect of COVID-19 on mental health and wellbeing in a representative sample of Australian adults. *Frontiers in psychiatry, 11*, 579985.

Dhira, T. A., Rahman, M. A., Sarker, A. R., & Mehareen, J. (2021). Validity and reliability of the Generalized Anxiety Disorder-7 (GAD-7) among university students of Bangladesh. *PloS one, 16*, e0261590.

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

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing, 51*, 1-26.

Ge, F., Zhang, D., Wu, L., & Mu, H. (2020). Predicting psychological state among Chinese undergraduate students in the COVID-19 epidemic: a longitudinal study using a machine learning. *Neuropsychiatric disease and treatment*, 2111–2118.

Gonçalves, D. C., Pachana, N. A., & Byrne, G. J. (2011). Prevalence and correlates of generalized anxiety disorder among older adults in the Australian National Survey of Mental Health and Well-Being. *Journal of Affective Disorders, 132*, 223–230.

Hilbert, K., Lueken, U., Muehlhan, M., & Beesdo-Baum, K. (2017). Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain and behavior, 7*, e00633.

Hornstein, S., Forman-Hoffman, V., Nazander, A., Ranta, K., & Hilbert, K. (2021). Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach. *Digital Health, 7*, 20552076211060659.

Hueniken, K., Somé, N. H., Abdelhack, M., Taylor, G., Marshall, T. E., Wickens, C. M., . . . others. (2021). Machine Learning–Based Predictive Modeling of Anxiety and Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study. *JMIR Mental Health, 8*, e32876.

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

Islam, M. S., Ferdous, M. Z., & Potenza, M. N. (2020). Panic and generalized anxiety during the COVID-19 pandemic among Bangladeshi people: An online pilot survey early in the outbreak. *Journal of affective disorders, 276*, 30–37.

Islam, M. S., Rahman, M. E., Moonajilin, M. S., & van Os, J. (2021). Prevalence of depression, anxiety and associated factors among school going adolescents in Bangladesh: Findings from a cross-sectional study. *Plos one, 16*, e0247898.

Islam, S., Akter, R., Sikder, T., & Griffiths, M. D. (2020). Prevalence and factors associated with depression and anxiety among first-year university students in Bangladesh: a cross-sectional study. *International Journal of Mental Health and Addiction*, 1–14.

Kuleindiren, N., Rifkin-Zybutz, R. P., Johal, M., Selim, H., Palmon, I., Lin, A., . . . Mahmud, M. (2022). Optimizing existing mental health screening methods in a dementia screening and risk factor app: observational machine learning study. *JMIR Formative Research, 6*, e31209.

Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., . . . Lemey, C. (2021, May 4). Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J Med Internet Res, 23*, e15708.

Lin, D., Nazreen, T., Rutowski, T., Lu, Y., Harati, A., Shriberg, E., . . . Aratow, M. (2022). Feasibility of a machine learning-based smartphone application in detecting depression and anxiety in a generally senior population. *Frontiers in Psychology, 13*.

Na, K.-S., Cho, S.-E., & Cho, S.-J. (2021). Machine learning-based discrimination of panic disorder from other anxiety disorders. *Journal of Affective Disorders, 278*, 1–4.

Nayan, M. I., Uddin, M. S., Hossain, M. I., Alam, M. M., Zinnia, M. A., Haq, I., . . . others. (2022). Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among university students in Bangladesh: A result of the first wave of the COVID-19 pandemic. *Asian Journal of Social Health and Behavior*, *5*, 75.

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

Rapee, R. M. (1991). Generalized anxiety disorder: A review of clinical features and theoretical concepts. *Clinical Psychology Review*, *11*, 419–440.

Ryu, J., Sükei, E., Norbury, A., H Liu, S., Campaña-Montes, J. J., Baca-Garcia, E., . . . Perez-Rodriguez, M. M. (2021). Shift in Social Media App Usage During COVID-19 Lockdown and Clinical Anxiety Symptoms: Machine Learning–Based Ecological Momentary Assessment Study. *JMIR mental health*, *8*, e30833.

Sagar, R., Dandona, R., Gururaj, G., Dhaliwal, R. S., Singh, A., Ferrari, A., . . . others. (2020). The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017. *The Lancet Psychiatry*, *7*, 148–161.

Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*, *16*, 100228.

Simjanoski, M., Ballester, P. L., da Mota, J. C., De Boni, R. B., Balanzá-Martínez, V., Atienza-Carbonell, B., . . . others. (2022). Lifestyle predictors of depression and anxiety during COVID-19: a machine learning approach. *Trends in Psychiatry and Psychotherapy*, *44*.

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral modeling for mental health using machine learning algorithms. *Journal of medical systems*, *42*, 1–12.

Sumathi, P., Balasubramanian, V., & Ramaraj, R. (2017). Prediction of mental health problems among children using machine learning techniques. *International Journal of Engineering Science and Technology*, *9*, 1286–1293.

Tennenhouse, L. G., Marrie, R. A., Bernstein, C. N., Lix, L. M., & others. (2020). Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *Journal of psychosomatic research*, *134*, 110126.

Tiffin, P. A., & Paton, L. W. (2018). Rise of the machines? Machine learning approaches and mental health: opportunities and challenges. *The British Journal of Psychiatry*, *213*, 509–510.

Vogazianos, P., Motrico, E., Domínguez-Salas, S., Christoforou, A., & Hadjigeorgiou, E. (2022). Validation of the generalized anxiety disorder screener (GAD-7) in Cypriot pregnant and postpartum women. *BMC Pregnancy and Childbirth*, *22*, 841.

