ANXIETY DISORDER DETECTION USING MACHINE LEARNING APPROACH

BY

Sharmin Sultana Mim

ID: 201-15-13639

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

Supervised By

Md Umaid Hasan
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project/internship titled **"Anxiety Disorder Detection Using Machine Learning Approach"**, submitted by Sharmin Sultana Mim, ID No: 201-15-13639 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *25 January, 2024*.

BOARD OF EXAMINERS

Chairman

Dr. Md. Fokhray Hossain (MFH)
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Md. Sadekur Rahman (SR)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Most. Hasna Hena (HH)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. S. M. Hasan Mahmud (SMH)
Assistant Professor
Department of Computer Science
American International University-Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md Umaid Hasan, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

22.01.24

Md Umaid Hasan
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Sharmin Sultana Mim
ID: -201-15-13639
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project/internship successfully.

I sincerely thank **Md Umaid Hasan, Lecturer**, Department of CSE Daffodil International University, and express my deep gratitude. To complete this project, my supervisor has deep knowledge & keen interest in the field of "*Artificial Intelligence*". His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

I would like to thank our entire coursemates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

The objective of this research is to tackle urgent problems in the early identification of Anxiety Disorder, specifically in Bangladesh. The country is renowned for its high prevalence of this mental disorder, which is often overlooked. The study aims to predict disorders using machine learning techniques, specifically targeting young adults. It's a mental health condition characterized by nervousness, panic, fear, sweating, and rapid heartbeat. The study uses machine learning models to predict anxiety disorders in individuals of all ages. The study employs various algorithms like CatBoost, Random Forest, SVM, Bagging, XGBoost, Logistic Regression, Decision Tree, and Naive Bayes to analyze data from surveys with a focus on university students. The data was processed using Python and Pandas library, with preprocessing performed, including handling null values and manual review. The survey results show a strong correlation between symptoms and anxiety disorder, indicating that machine learning can effectively identify potential anxiety cases. The anxiety dataset was analyzed using various ML algorithms, with CatBoost achieving the highest accuracy of 92.81%, followed by RF, SVM, Bagging, LR, XGBoost, DT, and GNB. The approach is revolutionary in Bangladesh, where data-driven healthcare methods are rare. The study's findings are relevant in both academic and practical circles since they provide a more cost-effective and efficient technique of mental diagnosis than previous approaches. Completing the dataset, refining the models to accommodate a wider range of demographics, and incorporating these results into public health initiatives are possible future directions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Modern technologies have encouraged people to push themselves, which has resulted in tension, worry, despair, frustration, and discontent. While anxiety is a typical stress response, anxiety disorders are a prevalent mental health illness marked by tense, uneasy, and anxious thoughts. They can induce bodily changes such as high blood pressure and disrupt daily activities such as job performance, schoolwork, and relationships. While some anxiety is normal, anxiety disorders involve continuous concern and fear, which affects people's general well-being and capacity to operate efficiently in daily life. Individuals may be affected differently by anxiety disorders, such as panic, social, generalized, and phobia-related forms.

Anxiety disorders are prevalent mental health issues affecting nearly 30% of adults, with the exact cause of these disorders remaining unknown. It can be acquired via ancestry and is caused by defective neurological systems controlling fear and emotions. Environmental stressors including violence, mortality, abuse or neglect of children, and additional mental health conditions might also raise risks. As per the World Health Organization (WHO), over 69 lakh individuals in Bangladesh suffer from anxiety disorders, with Bangladesh ranking first in the Southeast Asia Region. Anxiety disorder impacts people of all ages and professions, producing ongoing issues due to a lack of understanding and messing with their everyday lives.

The research presents a machine-learning strategy for identifying anxiety problems in patients using questionnaires gathered from Bengali patients of various ages since conventional diagnostic scales for common mental illnesses are frequently employed in Bangladesh. Machine learning has been used to create a novel approach for early detection and intervention in medical domains dealing with anxiety disorders. The

objective was to determine the most effective way of implementing for improving the support given to medical professionals in managing anxiety disorders.

1.2 Objective

Machine-learning algorithms are increasingly used in symptom analysis, as technology is crucial for our existence and for dealing with the constantly changing structure of our lives. AI can help predict diseases at an early stage, making this technological era intertwined with daily life. Our major goal was to predict anxiety disorder patients using questionnaires and machine learning. Our research aims to use machine learning to detect diseases using various types of questions, with a focus on important topics for our paperwork, despite exploring other topics for further examination.

Our goal is to use machine learning to detect anxiety disorders, enhancing diagnostic accuracy and efficiency. Individual anxiety disorder detection takes time, and further diagnosis methods struggle in the absence of prior data. For this reason, a platform has been developed to enable the detection of anxiety disorders utilizing this machine-learning technology, overcoming the problems encountered by other diagnostic tools. Through improved precision and effectiveness in the identification of anxiety disorders, timely actions, treatment plans, and improved mental health results for clinicians, researchers, and medical professionals, the study seeks to close the distance between technological innovation and mental health care.

1.3 Motivation

In the modern world, having good psychological conditions is essential since it influences our attitudes, behaviors, and relationships. Overcome obstacles, develop connections, and make wise judgments, it affects our emotional, psychological, and social well-being. From childhood to maturity, mental health is crucial, and taking care of one's mental health can increase happiness and success. The increasing self-adaptation brought about by modern technological breakthroughs has resulted in a variety of mental diseases,

including anxiety. People are not resolving anxiety disorder issues, which is impacting their everyday lives, despite a lack of knowledge. Earlier identification of anxieties is critical in healthcare facilities, but traditional methods can be difficult due to the identical appearance of many mental illnesses and the period, effort, and expertise necessary for manual diagnosis.

This is why a method to detect illnesses early and help physicians treat them on time has been developed. By utilizing machine learning techniques, the system will save time, reduce human effort, and minimize effort by identifying issues. This will make it possible for people to identify the condition early and take the necessary steps before it worsens.

1.4 Rationale of the study

AI is transforming several industries by giving robots the ability to acquire knowledge, evolve, and carry out tasks while also increasing productivity and creativity. AI is a broad field of technological advancement that allows scientists and engineers to address a range of issues, including the identification of diseases, in a variety of sectors. Because of its abilities, it plays a big role in technological development and research, which makes it important in the modern age. Machine learning is an important technique for diagnosing anxiety disorders, which have grown more difficult to diagnose because of their level of severity and complexity.

1.5 Expected Output

There are many advantages of using AI to diagnose anxiety disorders, as well as anticipated results. Here are a few of them:

- Develop a robust model that can determine if a patient has an anxiety disorder.
- Make it easier to detect the disorder at early stages.
- Help professionals to recognize the disease in a faster way.
- On the test to see how accurate the model can predict.

1.6 Research questions

I’m presently working on my first professional project, which is proving to be a considerable challenge for us. Before starting my research, my head was frequently filled with many questions. The major concerns addressed by this research are stated here.

- What dataset should I use?
- What is Bangladesh's current anxiety condition?
- Which question should I use for identifying the disease?
- What are the limitations of this field?
- Using collected raw data can detect the disease?
- How effective are machine-learning algorithms in accurately detecting the disease?
- What are the key challenges and limitations in applying ML methods for disorder and how can these challenges be addressed to improve the accuracy?
- Who will benefit from our research, specifically?

1.7 Layout of the Report

This work helps clinical professionals identify the anxiety disorders of the patients.

- In Chapter 1, I have already talked about the research papers that have been initiated, with an objective, reasons for working on the subject, problem statements, expected results, research question, and research methods discussed, along with the chosen anxiety disorder.

- In Chapter 2, I'll discuss the study's history in this section, tasks, summarization, difficulties, and evaluation against selected concepts & also discuss the challenges faced in completing the task effectively and evaluating against the chosen concepts.

- In Chapter 3, The research methodology which includes system architecture, data collection, preprocessing procedure, proposed methodology, and outputs will be discussed in this section.
- In Chapter 4, the experimental findings and discussion will be covered in this chapter. We will discuss the model accuracy, algorithm techniques, and results.
- In Chapter 5, summary, comparison, and analysis, the advantages of our application will be covered.
- In Chapter 6, the conclusion and future research will be discussed.
- Chapter 7, will contain all the references I used for this research.

CHAPTER 2
BACKGROUND

2.1 Introduction

The diagnosis of anxiety diseases is critical for addressing global psychological challenges, as anxiety disorders are common and unpleasant. Conventional diagnostic techniques, like subjective evaluations and clinical observations, frequently result in underdiagnosis and delaying discovery, which has an impact on day-to-day functioning. Through the analysis of a variety of datasets, including physiological indicators, behavioral trends, and textual data, machine learning technologies are increasing the efficiency and accuracy of the diagnosis of anxiety disorders. Through a knowledge of the drawbacks of present procedures, this study is aimed at contributing to the development of novel and trustworthy ways to the identification of anxiety disorders. This part offers a thorough examination of the actual circumstance and the associated paperwork idea, outlining our experience, benefits, and uses.

2.2 Related works

The implementation of different machine learning has transformed several industries, including healthcare and agriculture. There are several machine-learning ongoing thesis works. From the angle of Bangladesh, the author evaluates the effectiveness of ML algorithms in predicting anxiety and depression in Bangladeshi university students during COVID-19. Six models were used, with RF showing the best accuracy (89%), while SVM achieved the highest accuracy (91.49%) [1].

The study used machine learning algorithms and psychological evaluations to identify mental diseases like anxiety and depression. Data from the University of Dhaka's Psychology Department and Rangpur Medical College and Hospital were used. It also utilized five algorithms, with CNN showing the highest accuracy rates for anxiety(96%)

and depression(96.8%), revealing that 15.6% of women aged 18-35 had depressive disorders and 7.4% experienced anxiety [2].

Machine Learning method was used to predict depression in a younger population struggling with studies, employment, relationships, drugs, and family issues, using KNN, RF, and SVM algorithms. The study demonstrated Random Forest as the most accurate method, but it has limitations due to limited data and fewer method implementations [3].

Depression is a severe psychiatric condition that can lead to various health issues, including suicide, and can significantly deteriorate relationships with friends and family. A hybrid approach to depression detection using textual post-analysis on Reddit's test data employs six deep learning algorithms, with Word2VecEmbed+Meta showing the highest F1 Score and precision [4].

The impact of lifestyle choices and social and cultural factors on university students' mental health. By using a hierarchical model, it found that being a woman and dissatisfied with her education, financial situation, and career significantly predicted psychiatric illnesses, increasing anxiety and depression among university students. The study found a 27.1% increase in anxiety and 22.5% likely depression among Bangladeshi university students, but its focus on a single university may limit its applicability.[5]

Researchers developed an adaptive neuro-fuzzy inference approach using a real-world dataset, employing the ANFIS approach and a hybrid optimization learning method. The model achieved a high accuracy of (98.67%), sensitivity(97.14%), and specificity metrics of (100%) respectively [6].

By using a survey dataset, an ML model was developed to assess students' stress levels. The model was trained using CNN and stacking algorithms(KNN, Random Forest, and CNN Adaboost), and Adaboost achieved a maximum accuracy of 88.86%. Further stress examination is recommended for mental health issues.[7]

Panic disorder is a common psychiatric illness, differentiated from other anxiety disorders by neurobiological and developmental factors using machine learning-based approaches. The study used five methods: RF, gradient boosting, SVM, ANN, and regularized logistic regression, with LR being the most accurate. However, small sample size and design shortcomings were noted[8].

Machine learning methods were used to study stress patterns in working people, using a survey dataset of 750 replies and 68 personal and professional characteristics. Six machine learning models, including Logistic Regression, KNN, DT, RF, Boosting, Bagging, and Boosting, Boosting achieved the highest accuracy. Still, their limitations are due to a small response dataset[9].

This study explores the use of machine-learning algorithms to predict psychological health disorders like anxiety, depression, and stress in 348 individuals. The study used KNN, DT, RF, Naïve Bayes, and SVM, with data collected via Google Forms. The data was analyzed using Google Forms, with random forest being the most effective model and Naïve Bayes achieving the highest accuracy[10].

A deep learning model was used to assess a user's mental health status from 633,385 posts on Reddit. Using XGBoost and CNN models, CNN models outperformed XGBoost models in accuracy across all subreddits. However, future improvements and categorization models were not considered[11].

ML algorithms like LR, KNN, DT, RF, and Stacking are used to predict a disorder using a survey dataset with 27 columns and 1259 items. The study's accuracy was highest with the Stacking approach (81.75%), but its limitations lie in its use of a small data set[12].

The study investigates machine learning methods for detecting anxiety and depression among 470 Indian sailors using a questionnaire dataset. CatBoost, LR, Naive Bayes, RF,

and SVM were used, with CatBoost achieving the highest accuracy and precision of 82.6% and 84.1% respectively.[13]

The neurodevelopmental disease known as ADHD, which causes impulse control, hyperactivity, and inattention, is becoming more and more common in diagnoses. However, ineffective therapy for ADHD causes issues for the immune system. Machine learning techniques were used to identify ADHD, a neurological condition causing attention deficit, hyperactivity, and impulsivity. They used six methods, including SVM, LR, Naive Bayes, Random Forest, DT, and K-NN, on a dataset of 69 patients with 27 variables and with a Decision Tree obtaining an accuracy rate of 85.507%. [14]

Modern lives create stress and strain, which leads to psychological problems and irregularities. The efficacy of these aberrations in differentiating Bipolar disorder patients among healthy individuals and stratifying individuals according to total disease load, though, has not yet been examined. ML was used to screen for bipolar disorder using MDQ and decision tree classification on a 983-person interview dataset. The model accurately assigns each example to a category with a precision of 88.07%. [15]

Childhood sadness has detrimental effects on kids, families, and the community as a whole. Early diagnosis and identification are essential to averting serious implications down the road. ML algorithms were used to predict depression in children and teenagers using an Australian survey dataset. They used Decision Tree, XGBoost, RF, and Gaussian Naive Bayes, with RF outperforming all other methods by 99% with 95% accuracy and 99%. [16]

Anxiety, depressive disorders, and addiction are just a few examples of the many variables that may affect someone's mental health, which is essential for their emotional, psychological, and social well-being. It also affects how they think, feel, and respond to events. Machine Learning techniques were used to develop Behavioral Modeling for

Mental Health on a survey dataset of 656 samples, 20 characteristics, and 3 categories employing techniques such as SVM, DT, Naive Bayes, K-NN, LR, Ensemble, and Tree Ensemble. SVM and KNN performed better individually, while Random Forest, Ensemble, and SVM achieved higher accuracy levels.[17]

Alzheimer's syndrome is a progressive neural illness that mostly affects the elderly. It is generally brought on by dementia that looks which slowly destroys brain cells and impairs cognition and reading. The researchers developed a machine learning system for Alzheimer's disease, using four algorithms: SVM, Logistic regression, DT, and RF. SVM outperforms other algorithms in accuracy, recall, the area under the curve, and the F1-score model with a score of 92.1%. [18]

This study developed a diagnostic tool for subclinical anxious and depressed patients using 125 individuals divided into 4 groups. The bagged decision tree classifier outperformed random forest algorithms in predicting HD and HA groups with 68.07% and 74.18% accuracy respectively. [19]

A machine-learning technique was presented for analyzing pupil dynamics information as a biomarker for ADHD detection. It uses binary classification methods, a Voting Ensemble model, Naive Bayes, Logistic regression, and Support Vector Machine models. The Ensemble Voting Model outperforms sparse data. [20]

Table. 2.1 Literature Review Table

S.I.	Paper Name	Author's Name	Dataset	Model	Best Accuracy
1	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.	Nayan, M. I. H., Uddin, M. S. G., Hossain, M. I., Alam, M. M., Zinnia, M. A., Haq, I., ... & Methun, M. I. H.	Online survey dataset	logistic regression, random forest (RF), support vector machine (SVM), linear discriminate analysis, K-nearest neighbors, Naïve Bayes	SVM(91.49%)
2	A machine learning approach to detect depression and anxiety using supervised learning.	Ahmed, A., Sultana, R., Ullas, M. T. R., Begom, M., Rahi, M. M. I., & Alam, M. A.	Two Anxiety & Depression datasets	CNN, SVM, Linear discriminant analysis, KNN Classifier, and Linear Regression.	CNN(96% for anxiety and 96.8% for depression)
3	Predicting depression in Bangladeshi undergraduates using machine learning.	Choudhury, A. A., Khan, M. R. H., Nahim, N. Z., Tulon, S. R., Islam, S., & Chakrabarty, A.	A survey dataset	K Nearest Neighbor, Random Forest, and Support Vector Machine	Random Forest
4	Early depression detection from social network using deep learning techniques.	Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M. H.	Reddit dataset	TrainableEmbed, Word2VecEmbed, GloveEmbed, Meta Features, Word2VecEmbed+Meta, FastTextEmbed+Meta	Word2VecEmbed+Meta showing the highest F1 Score(0.81) and precision(0.78)
5	Impacts of socio-cultural environment and lifestyle factors on the psychological health of university students in Bangladesh: a longitudinal study.	Hossain, S., Anjum, A., Uddin, M. E., Rahman, M. A., & Hossain, M. F.	Online survey dataset	Hierarchical multivariate linear regression models	27.1% in anxiety and 22.5% in depression
6	Development and use of a clinical decision support system for the diagnosis of social anxiety disorder.	Fathi, S., Ahmadi, M., Birashk, B., & Dehnad, A.	real-world dataset	hybrid optimization learning method	98.67%
7	Evaluation of mental stress using predictive analysis.	Pingle, Y.	survey dataset	CNN algorithm and stacking algorithms mainly –KNN, RANDOM FOREST, AdaBoost	AdaBoost(88.86%)
8	Machine learning-based discrimination of panic disorder from other anxiety disorders.	Na, K. S., Cho, S. E., & Cho, S. J.	survey dataset	random forest, gradient boosting machine, support vector machine, artificial neural network, and regularized logistic regression	logistic regression (0.784)
9	Machine learning techniques for stress prediction in working employees.	Reddy, U. S., Thota, A. V., & Dharun, A.	Survey dataset	Logistic Regression, KNN, Decision Trees, Random Forest, Boosting, Bagging	Boosting
10	Predicting anxiety, depression and stress in modern life using machine learning algorithms.	Priya, A., Garg, S., & Tigga, N. P.	Online survey dataset	Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine, and KNN.	naïve Bayes
11	A deep learning model for detecting mental illness from user content on social media.	Kim, J., Lee, J., Park, E., & Han, J.	Reddit dataset	XGBoost and CNN	CNN
12	Predicting mental health illness using machine learning algorithms.	Vaishnavi, K., Kamath, U. N., Rao, B. A., & Reddy, N. S.	Survey dataset	Logistic Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and Stacking	Stacking approach(81.75%)
13	Screening of anxiety and depression among seafarers using machine learning technology.	Sau, A., & Bhakta, I.	Survey dataset	CatBoost, Logistic Regression, Naïve Bayes, Random Forest, and SVM	CatBoost(82.6%)

In Table 2.2, a small portion of our Literature review is shown.

2.3 Research Summary

Through a review of previous study projects and publications, it helps to identify areas for future investigation and learn how ML techniques can be used to assist people with anxiety disorders. Since early identification is usually normal, people with anxiety disorders frequently go undiagnosed. Following a comprehensive review of several papers, I decided to use machine learning approaches. In this situation, it might be quite helpful in overcoming the limitations and the greatest accuracy reached using CatBoost after integrating multiple methodologies. Identifying mental disorders such as anxiety disorder, CatBoost beats existing classification and identification techniques with a success rate of 92.81%, surpassing both RF at 91.62% & SVM at 89.82% given suitable training.

2.4 Scope of a Problem

Mental sickness is caused by anomalies in brain chemistry, or psychological health includes both an individual's mental state and surroundings. Major illnesses such as anxiety disorder, and bipolar disorder develop over time, necessitating early discovery for successful treatment and care. Our goal is to develop a system that makes the process of identifying anxiety disorders easier by utilizing a categorization technique, which is quite useful in this regard. The system I use has been taught to identify illnesses using data from Bangladeshi individuals, which might be useful in intricate medical procedures. The lack of technological participation in healthcare makes mental health more difficult to manage. Our study results will be made public as an open-source project, which may help other analysts with their research and maybe provide better results.

2.5 Challenges

The research faced significant challenges due to the learning machine and dataset collection, as it was challenging to learn the algorithm from the dataset.

Data Collection

The strike blockade situation impacted the initial task of collecting raw information for a research project. Physically collecting data was challenging due to the lack of awareness about mental health conditions. The data was collected through an online survey, but the Bangladeshi anxiety disorder dataset was not widely available on Kaggle, so manual collection was necessary. Also faced challenges in gathering information, initially planning to collect data from hospitals but having to physically collect data from some patients due to their lengthy processes. Then preprocessed and cleaned the data, and are now conducting training and testing to establish freshness identification. To achieve better outcomes, it is recommended to wait for the final resolution.

Model Selection

Model selection is the approach's most crucial and effective component. The training and validation sets are used to simulate new data. Many machine-learning approaches appear to be accessible. Various kinds of classification models may be found. Choosing the right model from the group of them is somewhat difficult. Selecting the appropriate data and model simplifies tasks.

Comparison

Our system, initially developed using an ML model, was compared to ten different algorithms to determine its accuracy and effectiveness. Thus, ten different kinds of algorithms— Logistic Regression, DT, RF, Bagging, Gaussian Naive Bayes', SVM, XGBoost, and CatBoost are employed here. The CatBoost algorithm outperforms other algorithms on the dataset, while the DT method performs poorly but is still respectable despite facing numerous challenges.

CHAPTER 3
RESEARCH METHODOLOGY

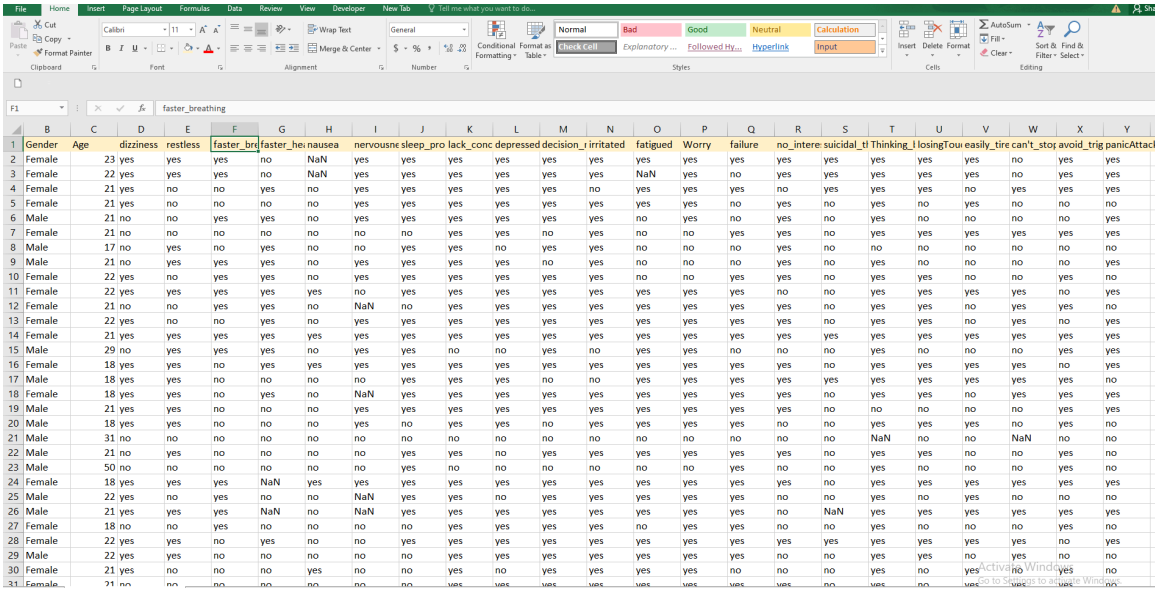
3.1 Research Subject and Instrumentation

The Method of research involves gathering data for decision-making, while research methodology is the process of obtaining information. Machine learning (ML) classification algorithms are used in prediction-based research to forecast results. Our research focuses on using ML methods to identify mental health issues like anxiety disorders, using machine learning methods to provide valuable insights.

In the 21st century, diseases are easier to treat but more complicated and costly. Common psychological diseases can worsen without early treatment, as people often overlook or ignore symptoms due to a lack of awareness. Anxiety disorder is a serious illness that requires treatment, but it is normal for most people to experience anxiety issues such as nervousness, lack of concentration, and fear in general life. Multiple anxieties can worsen the situation and potentially lead to an anxiety disorder. Mental illness commonly found in children, teenagers, adults, and students, diagnosed early can help maintain normalcy, though it cannot be permanently cured. The study proposes a model to predict anxiety disorders using health data and symptoms. Data is collected through Google Forms and manual methods, stored in MS Excel, trained using ML techniques, and applied to identify the disorder using appropriate algorithms.

3.2 Data collection

The sources of information were classified by location, age, and occupation, such as hospitals, schools, and universities. Variations in the signs and severity of anxiety disorders may develop owing to variances in people's ages, regions, and occupations.



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Gender	Age	dizziness	restless	faster_bre	faster_he	nausea	nervousne	sleep_pro	lack_conc	depressed	decision_i	irritated	fatigued	Worry	failure	no_intere	suicidal_tl	Thinking_I	losingTou	easily_tire	can't_stop	avoid_trig	panicAttacl
2	Female	23	yes	yes	yes	no	NaN	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes
3	Female	22	yes	yes	yes	no	NaN	yes	yes	yes	yes	yes	yes	NaN	yes	no	yes	yes	yes	yes	yes	no	yes	yes
4	Female	21	yes	no	no	yes	no	yes	yes	yes	yes	yes	no	yes	yes	yes	no	yes	yes	yes	no	yes	yes	yes
5	Female	21	yes	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	no	yes	no	yes	no	no	no
6	Male	21	no	no	yes	yes	no	yes	yes	yes	yes	yes	yes	no	yes	no	yes	no	yes	no	no	no	no	yes
7	Female	21	no	no	no	no	no	no	no	yes	yes	no	yes	no	no	yes	yes	no	yes	yes	yes	yes	yes	yes
8	Male	17	no	yes	no	yes	no	no	yes	yes	no	yes	yes	no	no	no	yes	no	no	no	no	no	no	no
9	Male	21	no	yes	yes	yes	no	yes	yes	yes	yes	no	yes	no	no	no	yes	no	yes	no	no	no	no	yes
10	Female	22	yes	no	yes	yes	no	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	yes	yes	no	no	yes	no
11	Female	22	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	yes	yes	no	yes
12	Female	21	no	no	yes	yes	no	NaN	no	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	no	yes	yes	no
13	Female	22	yes	no	no	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	no	yes	no	yes
14	Female	21	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
15	Male	29	no	yes	yes	yes	no	yes	yes	no	no	yes	no	yes	yes	no	no	no	yes	no	no	no	yes	yes
16	Female	18	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	no	yes
17	Male	18	yes	yes	no	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no
18	Female	18	yes	yes	no	yes	no	NaN	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	no	yes	yes	yes
19	Male	21	yes	yes	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes
20	Male	18	yes	yes	no	no	no	yes	no	yes	yes	no	yes	yes	yes	yes	no	no	yes	yes	yes	no	yes	no
21	Male	31	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	NaN	no	no	NaN	no	no
22	Male	21	no	yes	no	no	no	no	yes	yes	no	yes	yes	yes	yes	yes	yes	no	yes	yes	no	no	yes	no
23	Male	50	no	no	no	no	no	yes	no	no	no	no	no	no	no	yes	no	no	yes	no	no	no	yes	no
24	Female	18	yes	yes	yes	NaN	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
25	Male	22	yes	no	yes	no	no	NaN	yes	yes	no	yes	yes	yes	yes	yes	no	no	yes	no	yes	no	no	no
26	Male	21	yes	yes	yes	NaN	no	NaN	yes	yes	yes	yes	yes	yes	yes	yes	no	NaN	yes	yes	yes	yes	yes	yes
27	Female	18	no	no	yes	no	no	no	no	yes	yes	yes	yes	no	yes	yes	no	no	yes	no	no	no	yes	no
28	Female	22	yes	yes	no	yes	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes
29	Male	22	yes	yes	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	yes	no	no
30	Female	21	yes	no	no	no	yes	no	no	yes	no	yes	yes	yes	yes	no	no	no	yes	no	yes	no	yes	no
31	Female	21	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	no	yes	yes	yes	no

Figure. 3.1 Data Collection

Data collection entailed obtaining information from multiple sources to solve a research challenge. For training our system, we collected data from surrounding hospitals individually and also collected online data by using a basic Google Form. A generic Google form was used to collect data, and 553 individuals were interviewed individually. This data was gathered from 17 to 60-year-olds who worked in offices, hospitals, universities, and schools. For the survey basically used several questions like- whether someone is feeling restless, has faster breathing, has sleep problems, facing a lack of concentration, is feeling depressed, having a panic attack or not.

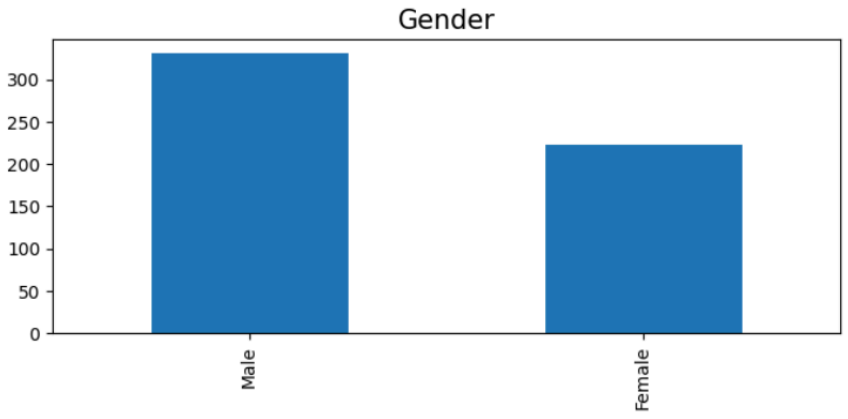


Figure.3.2 Bar Chart of Participants

Figure 3.2 shows the ultimate gathered responses from males and females. Though most of the participants were male, females also faced anxiety disorder frequently due to their surrounding situations. The obtained data was saved in Microsoft Excel, therefore Google Forms was an important part of the data-gathering process.

3.3 Data Preprocessing

Data preprocessing is the cleansing and organization of raw data in preparation for model creation and training. This process enhances data accuracy, quality, and efficiency by managing null values, class labeling, and label encoding. Data preparation is an important procedure that involves the use of ML algorithms, numerical approaches, and analytical expertise to handle data. It was separated into numerous research sections to secure the intended outcome. It assists in fixing basic errors such as incorrectly named data, which may be fixed by hand, and damaged data, which may be recovered and restored. Preprocessing is essential for successful data processing in our algorithms since it may alleviate procedural or technological challenges by utilizing a variety of elements.

Null value Handling

Null values are missing values that can impact the precision of machine learning models and their outcomes. Our dataset included some missing values so I applied basic imputation methods to complete the missing data. The dataset representing categorical data had only "Yes" and "No" responses, which is why it was filled with missing values utilizing fundamental imputation approaches.

Class Labeling

Machine learning models are trained to recognize and classify data samples through the human process of class labeling. The target field is another name for it. It involves carefully locating and labeling data, both labeled and unlabeled. To ensure the accuracy and proper categorization of each piece of information, data on the symptoms of anxiety disorders was gathered through strict identification and labeling. Clear and precise class

labels are required for the algorithm to acquire useful information and make correct predictions.

Label Encoding

Data encoding is an important step in data preprocessing, and utilized label encoding on the dataset. To make labels machine-usable, label encoding transforms them into numerical forms. I assigned a value of 1 to "yes" and a value of 0 to "no" for each question in the dataset that contains a response.

3.4 Proposed Methodology

System architecture

A machine learning system's architectural description is a detailed definition and depiction that includes the system's structure, component arrangements, and workflow. The machine learning architecture outlines data handling, model training, and prediction creation. It also includes system architecture characteristics like structure and phenomenon. The ultimate system implementation involves interconnected subsystems, which might be incorporated into a suggested system. The recommended system diagram is Figure. 3.3. This section provides an overview of the entire working process, detailing the visible system components and the research technique based on a diagram.

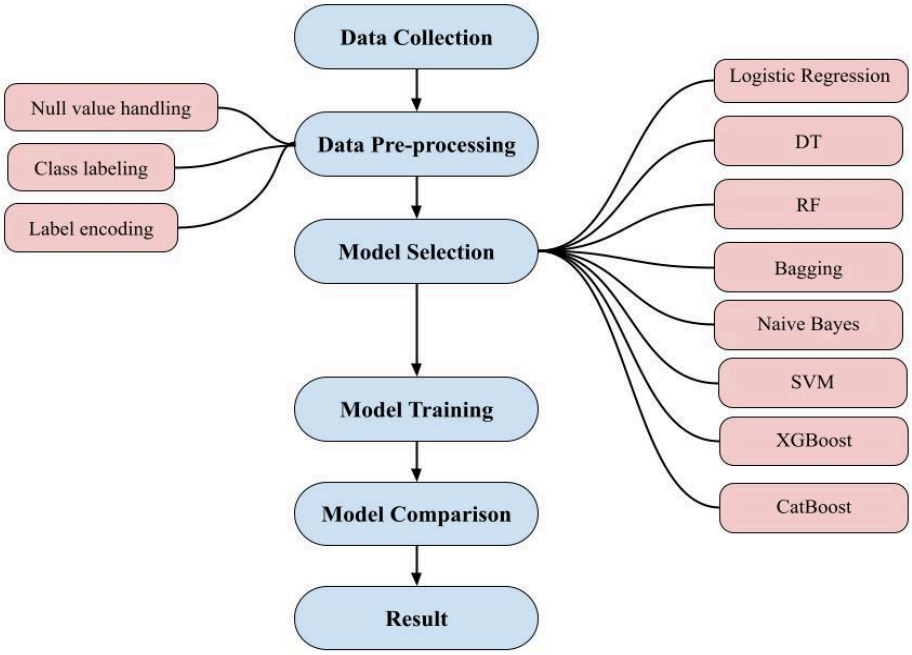


Figure.3.3 System Architecture

Classification techniques are used to categorize data. It can be possible to employ both ordered and disorderly information. I used eight various kinds of algorithms in our research: Logistic Regression, Bagging, Gaussian Naive Bayes, SVM, XGBoost, DT, RF, and CatBoost.

Algorithm classification

Logistic Regression

It's a predictive method of analysis that predicts a dependent variable by evaluating the connection between independent variables, simplifying the difficult calculation of probabilities. It is used to predict categorical dependent variables, with possible outcomes being either discrete or categorical, such as True, False, Yes, No, 0, or 1.

The accuracy of the Logistic Regression model is 89.82 %.

Decision tree

It's a method that predicts using a decision tree. It's a technique that breaks down complex data into manageable chunks, making it easier to understand and valuable for data analytics, particularly in predictive evaluation, data categorization, and regression. It is categorized according to feature values, with each vertex being a feature and each edge indicating a potential value in an instance for classifications.

The accuracy of the DT model is 83.83 %.

RandomForest

Random Forest is an ML algorithm that uses ensemble learning for classification and regression. It employs numerous decision trees on various datasets, improving predictive accuracy by using the average and avoiding overfitting, which results in better forecasts.

The accuracy of the RF model is 91.62 %.

Bagging

It's an ML ensemble meta-algorithm that improves the stability and accuracy of statistical classification and regression methods. Choosing a random sample of data for replacing and enabling the selection of particular data values several times, lowers variation and minimizes excessive fitting.

The accuracy of the Bagging model is 88.02 %.

Gaussian Naive Bayes

It's a machine learning method that classifies information using probabilistic methods using a Gaussian distribution based on Bayes' theorem. It computes the likelihood of all classes and chooses the one with the most likelihood, assuring independent predictions.

The accuracy of this model is 85.03 %.

SVM

It's a known supervised learning technique for classification and regression. It performs well in categorization by mapping data to a high-dimensional feature set, defining a separator across groups, and translating the information into a hyperplane-like when the

information is not linearly separate. This approach focuses on categorizing n-dimensional space, lowering classification risk, and allowing for rapid categorization of fresh data points in the future.

The accuracy of the SVM model is 89.82 %.

XGBoost

XGBoost is a technique that improves data interpretation and prediction, utilizing decision tree and optimization approaches such as random forest and gradient boost & especially useful for big, complicated datasets. It's recognized because of its effectiveness and widespread use, and it is also accessible as an open-source alternative.

The accuracy of the XGBoost model is 88.02 %.

CatBoost

CatBoost is an efficient gradient-boosting technique that works well with categorical data. It is the most recent and effective method for handling categorization issues. It also employs gradient boosting, an iterative process that adds different models to an ensemble, each training to remedy prior shortcomings, to enhance the findings of the preceding tree.

The accuracy of the CatBoost model is 92.81 %.

CatBoost performed the best result among all algorithm techniques.

Model Training

Machine learning model training supplies a method with training data to find and learn optimal values for features. By helping to discern patterns or make judgments from an unknown dataset, the software builds the most precise mathematical model of the connection between goal labeling and information attributes. It builds a working model that is validated, tested, and deployed. Based on how well the model performs while training, its applicability for end-user applications is determined.

The study highlights model training was an especially critical component of deployment. Our models are built using particular parameters, and the dataset is split into a 70:30

ratio, where 70% is used for training and 30% for testing data.

3.5 Implementation Requirement

To train the suggested approach, firstly need to collect and load data sources and ensure an uninterrupted procedure by cleaning the dataset. The dataset will be cleaned using various techniques, followed by feature engineering and translation of numerical data from categorical data. The training phase utilized 70% of the data, while the testing phase utilized 30%. The results were evaluated using various models. The CatBoost method achieves the highest accuracy, and the model with the highest accuracy, precision, recall, and F-1 score can be chosen for use.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Environment

Supervised classification trials were utilized to assess the performance of classifiers utilizing multiple markers of efficacy. The experiment is conducted on a personal computer where the microprocessor is - Intel(R) Core(TM) i3-3220 CPU @ 3.30GHz 3.30 GHz, 8.00 GB RAM, and Windows 10 operating system. The studies were carried out in a Python environment and executed on an open-source platform which is Google Collab.

Google Collab: Colab notebooks enable the integration of executable code, rich text, and images into a single document. It hosted the Jupyter Notebook service that offers free access to computing resources, making it an ideal tool for machine learning projects. The tool offers free GPU and TPU resources for speeding up model training and allows users to install popular machine-learning libraries for modeling purposes.

Python 3.5: Python is a popular programming language for a variety of applications such as developing software, machine learning, web development, statistical analysis, and more because of its community's support, paperwork, adaptability, and usefulness.

4.2 Experimental Results & Analysis

The study utilized ten machine-learning techniques and assessed the accuracy, recall, precision, confusion matrix, and F1 score of each approach.

Table 4.1.: Performance analysis of all Algorithms

Algorithm	Precision	F1_Score
CatBoost	92.81	0.93
RF	91.62	0.92
SVM	89.82	0.90
Logistic Regression	89.82	0.90
Bagging	88.02	0.88
XGBoost	88.02	0.88
Gaussian Naive Bayes'	85.03	0.85
Decision tree	83.83	0.84

The comparative performance analysis of all algorithms is shown in Table 4.1. In this table, we see the precision and F1 score for each method.

Here,

Precision: Illustrates the real-world scenarios for every precise affirmative prediction.

F1 Score: It determines test accuracy by analyzing Precision.

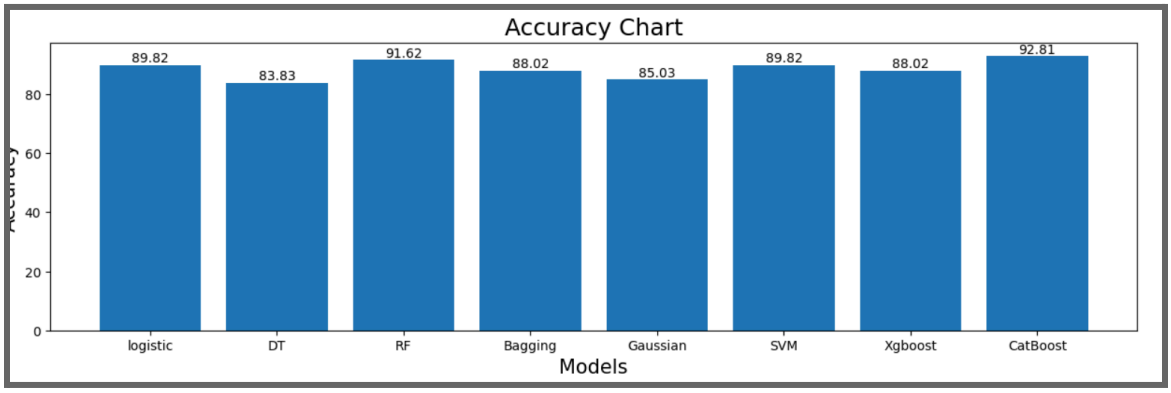


Figure 4.1.: Accuracy Bar

Analyzing test and training data helped determine how well the model performed. According to Figure 4.2., the accuracy scores of CatBoost, Random Forest, SVM, LR, Bagging, XGBoost, GNB, and DT are 92.81%, 91.62%, 89.82%, 89.82%, 88.02%,

88.02%, 85.03%, and 83.83%, respectively. According to Table 4.1.'s summary, CatBoost, RF, and SVM performed better than the other models. The top-performing algorithms were identified, while the GNB and DT models showed the lowest performance.

Confusion Matrix & Classification Report (Precision, Recall, F1-score)

The typical confusion matrix is a technique for analyzing a model's classification performance, and it provides a summary of predictions per class. It combines numerical ratings with a colored heatmap for easy understanding and finding issues. The classification report assesses the accuracy and dependability of the method meanwhile, the confusion matrix aids in identifying mistake kinds and causes. We have trained our model using the target class to get better performance. To assess our model's performance, I employed confusion matrices and classification reports. Here better performed three model reports are shown.

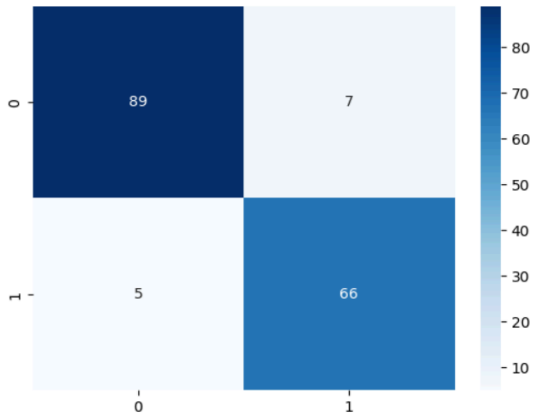


Figure 4.2.: Confusion Matrix of CatBoost

```
Model Accuracy:
92.81437125748504
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.93      0.94        96
           1       0.90      0.93      0.92        71

    accuracy                           0.93       167
   macro avg       0.93      0.93      0.93       167
weighted avg       0.93      0.93      0.93       167
```

Figure 4.3.: Accuracy, Precision, recall, and f1- score of CatBoost

Here, in Figures 4.2 and 4.3, we've displayed the CatBoost confusion matrix and classification report.

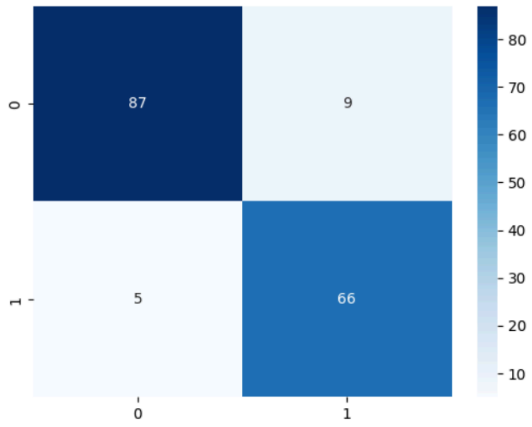


Figure 4.4.: Confusion Matrix of RF

```
Model Accuracy:
91.61676646706587
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.91      0.93        96
           1       0.88      0.93      0.90        71

    accuracy                           0.92       167
   macro avg       0.91      0.92      0.91       167
weighted avg       0.92      0.92      0.92       167
```

Figure 4.5: Accuracy, Precision, recall, and f1- score of RF

Here, in Figures 4.4 and 4.5, it displayed the confusion matrix and classification report of RF.

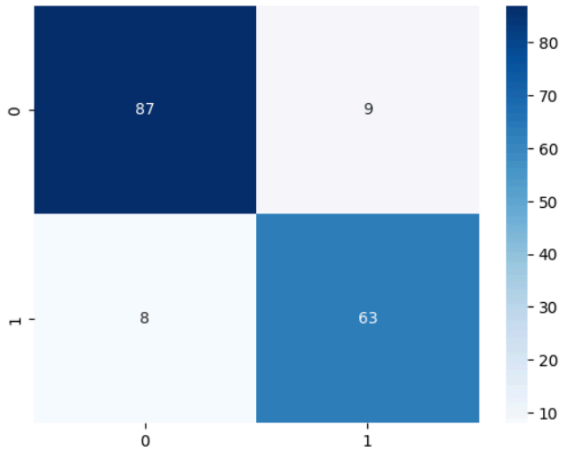


Figure 4.6: Confusion Matrix of SVM

```
Model Accuracy:
89.82035928143712
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.91      0.91        96
           1       0.88      0.89      0.88        71

    accuracy                           0.90       167
   macro avg       0.90      0.90      0.90       167
weighted avg       0.90      0.90      0.90       167
```

Figure 4.7: Accuracy, Precision, recall, and f1- score of SVM

Here, in Figures 4.6 and 4.7, it displayed the confusion matrix and classification report of SVM.

4.3 Result Discussion

Numerous researchers have utilized various algorithms, achieving varying precision and accuracy on different datasets. The accuracy of their work is primarily based on predicting multiple disorders like depression, anxiety, bipolar, and ADHD, primarily through implementation methods.

After reviewing various papers, gathering all the necessary data and building the model using the provided data, we implemented our system and found that CatBoost performs exceptionally well in detecting specific disorders following our study process. Model comparison tries to enhance machine learning performance and method selection, using confusion matrix outcomes to visualize algorithms' performance. It focuses on actual positives, misleading negatives, fake positives, and genuine negatives. CatBoost, Random Forest, and SVM models outperform all others, with the prediction rate indicating superior performance. The accuracy scores of CatBoost, Random Forest, and SVM are 92.81%, 91.62%, and 89.82%, respectively. According to Table 4.1's summary, CatBoost, RF, and SVM performed better than the other models. The top-performing algorithms were identified, while the Gaussian Naive Bayes' and DT models showed the lowest performance.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

This study explores the significant impact of machine learning on diagnosing anxiety in society. The integration of technology in healthcare can democratize access to mental health care, especially in resource-limited areas like Bangladesh. This can enable early detection of anxiety disorders, reduce long-term social and economic costs, and de-stigmatize mental health issues. Furthermore, this approach can catalyze innovation, leading to more advanced, patient-centric solutions.

5.2 Impact on Environment

Machine learning-based diagnostic tools have minimal environmental impact compared to conventional medical equipment. However, environmental stressors like trauma and childhood experiences can contribute to anxiety symptoms. Digital solutions are being used to identify environmental stressors, reducing the need for physical materials and resources commonly used in traditional diagnostic processes, which is crucial for recognizing individuals at risk. The reduction in physical visits to clinics or hospitals may lead to a decrease in transportation-related emissions. The energy consumption of running advanced computational models and maintaining data servers underscores the need for sustainable energy sources in computational healthcare solutions.

5.3 Ethical Aspects

Early detection of anxiety disorders requires ethical considerations, respecting individual autonomy and privacy, and informed consent, ensuring individuals are fully aware of the screening process's purpose and potential implications. Strict confidentiality measures are essential for protecting sensitive mental health information, preventing unauthorized access and disclosure, and building trust. Early detection of anxiety disorders raises

ethical concerns about mental health treatment accessibility, requiring timely and appropriate interventions, regardless of socio-economic status. Cultural sensitivity is crucial in early detection, requiring practitioners to comprehend diverse beliefs and values. Ethical practices necessitate minimizing stigmatization, ensuring equal access to mental health resources, and continuously refining screening methods to prevent overdiagnosis.

5.4 Sustainability Plan

The initiative's long-term sustainability necessitates a comprehensive plan, including funding for research, technology updates, and dataset expansion. Collaborations with healthcare providers, policymakers, and educational institutions are essential for widespread adoption. The project's impact, community outreach, and education are vital for maintaining trust in this innovative anxiety disorder diagnosis.

CHAPTER 6
SUMMARY, CONCLUSIONS, AND FUTURE WORK

6.1 Summary of the Study

The CatBoost algorithm demonstrated the highest accuracy, scoring 92.81%, outperforming other algorithms but with less accuracy.

Recently, there has been increased interest in applying artificial intelligence systems for detecting and forecasting illness probability. In this study, employed survey datasets and ML algorithms to identify people with anxiety disorders. The results of the study emphasize the value of early identification and diagnosis and have important ramifications for those who are struggling with anxiety. Increased effectiveness may result from early intervention. After preprocessing, separated training and evaluation, attaining the expected results. Our prediction algorithm can improve the identification rate of anxiety illness in settings with limited time and resources.

6.2 Conclusions

Anxiety disorder, a prevalent mental health issue, is growing rapidly, impacting daily activities and social life. It is particularly prevalent among Bangladeshi youth and children, causing significant challenges for their future. A system is being developed to improve the identification of anxiety disorders using machine-learning models. The CatBoost model has the highest classification accuracy of 92.81% accuracy rate, indicating significant results in predicting anxiety disorders. This approach can be further explored to better understand and treat anxiety disorders.

6.3 Future work

The study utilized a limited survey dataset, limiting its potential as a training model. The survey was conducted on individuals of various ages and occupations, but the limited responses were a major limitation. Customizing the survey format and increasing the

number of attributes based on relevance could improve the results. The work anticipated expansion to a larger dataset encompassing more classes in the future. Implement multiple ML and DL algorithms to assess their effectiveness when employing identical data, and study other features for classification.

REFERENCES

[1]. Nayan, Md Iqbal Hossain, M. Sheikh Giash Uddin, Md Ismail Hossain, Md Mohibul Alam, Maliha Afroj Zinnia, Iqramul Haq, Md Moshiur Rahman, Rejwana Ria, and Md Injamul Haq Methun. "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, no. 2 (2022): 75-84.

[2]. Ahmed, Anamika, Raihan Sultana, Md Tahmidur Rahman Ullas, Mariyam Begom, Md Muzahidul Islam Rahi, and Md Ashraful Alam. "A machine learning approach to detect depression and anxiety using supervised learning." In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1-6. IEEE, 2020.

[3]. Choudhury, Ahnaf Atef, Md Rezwan Hassan Khan, Nabuat Zaman Nahim, Sadid Rafsun Tulon, Samiul Islam, and Amitabha Chakrabarty. "Predicting depression in Bangladeshi undergraduates using machine learning." In *2019 IEEE Region 10 Symposium (TENSYMP)*, pp. 789-794. IEEE, 2019.

[4]. Shah, Faisal Muhammad, Farzad Ahmed, Sajib Kumar Saha Joy, Sifat Ahmed, Samir Sadek, Rimon Shil, and Md Hasanul Kabir. "Early depression detection from social network using deep learning techniques." In *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 823-826. IEEE, 2020.

[5]. Hossain, Sahadat, Afifa Anjum, Md Elias Uddin, Md Abdur Rahman, and Md Forhad Hossain. "Impacts of socio-cultural environment and lifestyle factors on the psychological health of university students in Bangladesh: a longitudinal study." *Journal of affective disorders* 256 (2019): 393-403.

[6]. Fathi, Sina, Maryam Ahmadi, Behrouz Birashk, and Afsaneh Dehnad. "Development and use of a clinical decision support system for the diagnosis of social anxiety disorder." *Computer Methods and Programs in Biomedicine* 190 (2020): 105354.

[7]. Pingle, Yogesh. "Evaluation of mental stress using predictive analysis." In *Int J Eng Res Technol (IJERT), NTASU—2020 conference proceedings, special*, no. 2021. 2021.

[8]. Na, Kyoung-Sae, Seo-Eun Cho, and Seong-Jin Cho. "Machine learning-based discrimination of panic disorder from other anxiety disorders." *Journal of Affective Disorders* 278 (2021): 1-4.

[9]. Reddy, U. Srinivasulu, Aditya Vivek Thota, and A. Dharun. "Machine learning techniques for stress prediction in working employees." In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-4. IEEE, 2018.

[10]. Priya, Anu, Shruti Garg, and Neha Prerna Tigga. "Predicting anxiety, depression and stress in modern life using machine learning algorithms." *Procedia Computer Science* 167 (2020): 1258-1267.

[11]. Kim, Jina, Jieon Lee, Eunil Park, and Jinyoung Han. "A deep learning model for detecting mental illness from user content on social media." *Scientific reports* 10, no. 1 (2020): 11846.

[12]. Vaishnavi, Konda, U. Nikhitha Kamath, B. Ashwath Rao, and NV Subba Reddy. "Predicting mental health illness using machine learning algorithms." In *Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012021. IOP Publishing, 2022.

[13]. Sau, Arkaprabha, and Ishita Bhakta. "Screening of anxiety and depression among seafarers using machine learning technology." *Informatics in Medicine Unlocked* 16 (2019): 100228.

[14]. Chen, Tianhua, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis, and Pan Su. "Automatic diagnosis of attention deficit hyperactivity disorder using machine learning." *Applied Artificial Intelligence* 35, no. 9 (2021): 657-669.

[15]. Jadhav, Ranjana, Vinay Chellwani, Sharyu Deshmukh, and Hitesh Sachdev. "Mental disorder detection: Bipolar disorder scrutinization using machine learning." In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 304-308. IEEE, 2019.

[16]. Haque, Umme Marzia, Enamul Kabir, and Rasheda Khanam. "Detection of child depression using machine learning methods." *PLoS one* 16, no. 12 (2021): e0261131.

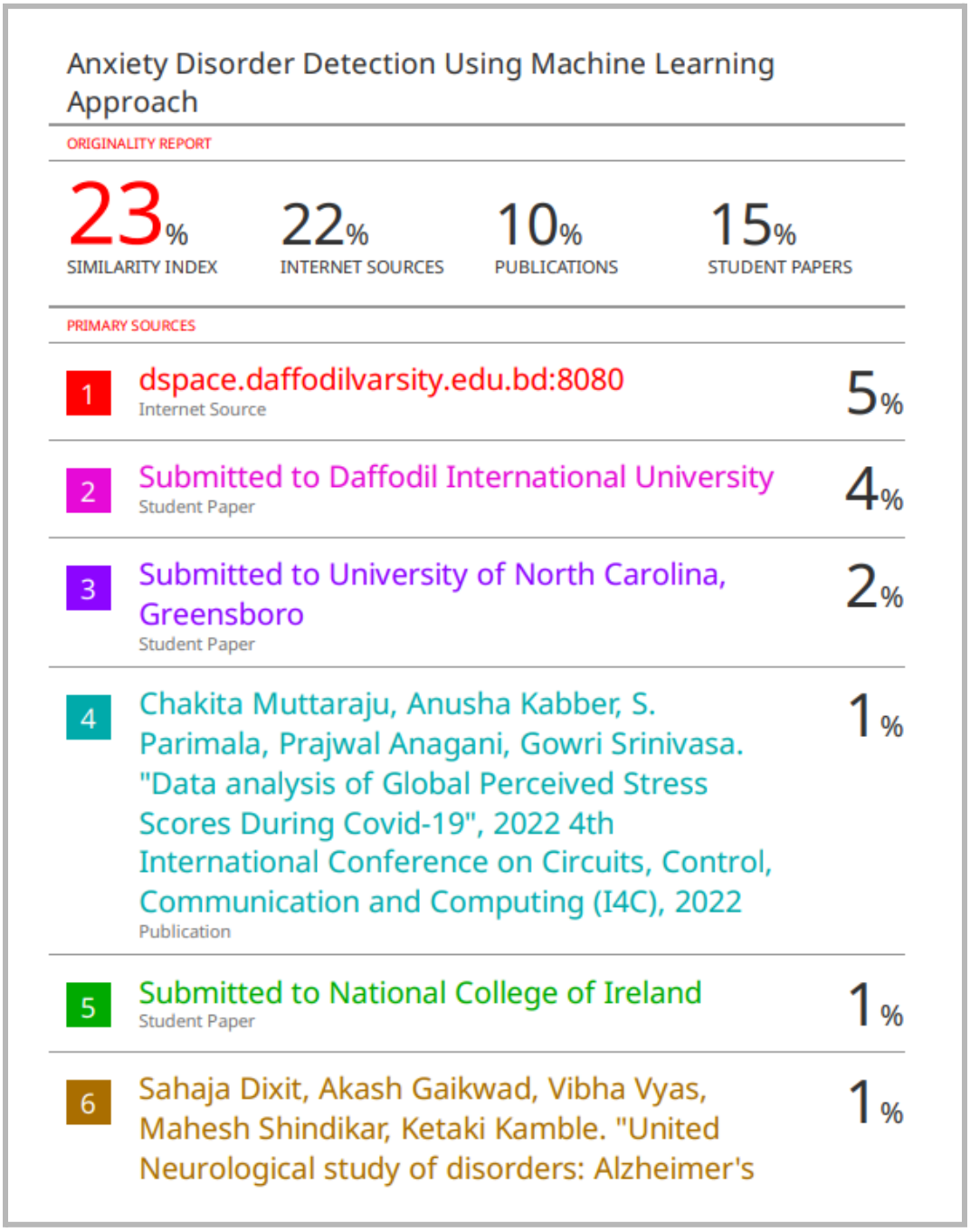
[17]. Srividya, M., Subramaniam Mohanavalli, and Natarajan Bhalaji. "Behavioral modeling for mental health using machine learning algorithms." *Journal of medical systems* 42 (2018): 1-12.

[18]. Bari Antor, Morshedul, A. H. M. Jamil, Maliha Mamtaz, Mohammad Monirujjaman Khan, Sultan Aljahdali, Manjit Kaur, Parminder Singh, and Mehedi Masud. "A comparative analysis of machine learning algorithms to predict alzheimer's disease." *Journal of Healthcare Engineering* 2021 (2021).

[19]. Richter, Thalia, Barak Fishbain, Andrey Markus, Gal Richter-Levin, and Hadas Okon-Singer. "Using machine learning-based analysis for behavioral differentiation between anxiety and depression." *Scientific reports* 10, no. 1 (2020): 16381.

[20]. Khanna, Shubh, and William Das. "A novel application for the efficient and accessible diagnosis of ADHD using machine learning." In *2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G)*, pp. 51-54. IEEE, 2020.

PLAGIARISM REPORT



Anxiety Disorder Detection Using Machine Learning Approach

ORIGINALITY REPORT

23%	22%	10%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	4%
3	Submitted to University of North Carolina, Greensboro Student Paper	2%
4	Chakita Muttaraju, Anusha Kabber, S. Parimala, Prajwal Anagani, Gowri Srinivasa. "Data analysis of Global Perceived Stress Scores During Covid-19", 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), 2022 Publication	1%
5	Submitted to National College of Ireland Student Paper	1%
6	Sahaja Dixit, Akash Gaikwad, Vibha Vyas, Mahesh Shindikar, Ketaki Kamble. "United Neurological study of disorders: Alzheimer's	1%