Arab American University

Faculty of Graduate Studies

Predicting Anxiety, Depression, and Stress Mental Health Disorders Using Machine Learning Algorithms

By

Mohannad Ahmad Yousef Amarneh

Supervisor

Prof. Dr. Mohammed Awad

This thesis was submitted in partial fulfillment of the requirements for the Master's degree in Data Science and Business Analytics

Feb / 2024

Thesis Approval

Predicting Anxiety, Depression, and Stress Mental Health Disorders Using Machine Learning Algorithms

By

Mohannad Ahmad Amarneh

This thesis was defended successfully on 20/02/2024 and approved by:

Committee members	Signature
1. Prof. Mohammed Awad: Supervisor	
2. Dr. Ahmad Ewais: Internal Examiner	
3. Dr. Yousef Daraghmeh: External Examiner	…………………………..

Declaration

This word provided in this thesis unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student Name: Mohannad Ahmad Yousef Amarneh

Student Number: 202113298

Signature: Mohanad Amarneh

Date: 10.11.2025

Dedication

I express my heartfelt dedication to my family, particularly my supportive parents, for their unwavering encouragement throughout my academic journey. A special acknowledgment goes to my wife and children for their constant support and patience, which played a crucial role in the completion of this thesis.

Acknowledgments

I want to thank Prof. Dr. Mohammed Awad a lot for his ongoing advice and support while I was writing this thesis. His great knowledge and commitment helped us overcome challenges and obstacles.

Abstract

As Artificial Intelligence rapidly continues to be improved and refined, it has become more possible to help mental health industry owners redefine the meaning of mental health in a more positive way than before. Thus, improving the way to diagnose mental illness in earlier stages, making the treatment more effective. This can be achieved by personalizing the diagnoses and treatment based on the individual's case characteristics. Mental health is a person's overall psychological well-being, including emotional, psychological, and social factors that affect how they think, feel, and behave. A balance and stability in mental health enable a person to handle normal life stresses, work efficiently, and contribute positively to their community. Since Depression, Stress, and Anxiety are considered the most common disorders, this research aims to employ machine learning algorithms to predict the diagnoses of stress using a dataset collected as part of this work. The dataset consists of around 700 records using an online survey, which was on top of DASS21 (Depression, Anxiety, and Stress international Scale). The data were collected from Palestinian participants, volunteers from the community, and university students. Then data preprocessing has been applied, cleansed from duplicates, and removed useless features like the hand and religion. Also converting textual values into numerical, and due to the imbalance in data, resampling technique has been used to resolve the imbalance to have more accurate results. Five different machine learning algorithms were utilized to analyze the data and achieve the early detection of mental health issues: Random Forest Model, SVM Model, K-Nearest Neighbors Model, XGBoost Model, and Multi-Layer Perceptron (3 Layers) Model. The results for Depression were SVM with the best model accuracy at 100%, followed by MLP with 98%, then XGBoost with 95%, Random Forest with 93%, and finally KNN with 77%. The results for Anxiety showed SVM gaining the highest score at 100%, MLP at 96%, both Random Forest and XGBoost at 95%, and KNN at 79%. Also, the results for Stress were 100% for SVM, then Random Forest and XGBoost at 97%, MLP at 96%, and KNN at 78%.

Keywords— Mental Health Disorders, Stress, Depression, Anxiety, Machine Learning, Classification, Random Forest, Support Vector Machine, KNN, Multi-Layer Perceptron, Xgboost

Table of Contents

List of Tables

List of Figures

List of Abbreviation

Abbreviation	Explanation
DASS	Depression Anxiety Stress Scale
AI	Artificial Intelligence
ML	Machine Learning
LSTM	Long Short-Term Memory
KNN	K-nearest-Neighbors
SVM	Support Vector Machine
DNN	Deep-Neural Network
RF	Random Forest
MLP	Multi-Layer Perceptron
GB	Gradient Boosting
WHO	World Health Organization
ANN	Artificial Neural Network
ROC	Receiver Operating Characteristic
AUC	Area Under Receiver
TP	True Positive
TN	True Negative
FP	False-Positive
FN	False-Negative
PC	Personal Computer

Chapter 1

1.1 INTRODUCTION

According to the World Health Organization (WHO), depression is the most widespread mental disorder and affects over 300 million individuals globally as show in figures (1.1,1.2)(WHO, 2017). This has led to a significant increase in focus from health researchers toward studying this issue. However, accurately distinguishing anxiety, depression, and stress is a challenge for machines and requires the use of an appropriate learning algorithm for proper diagnosis.

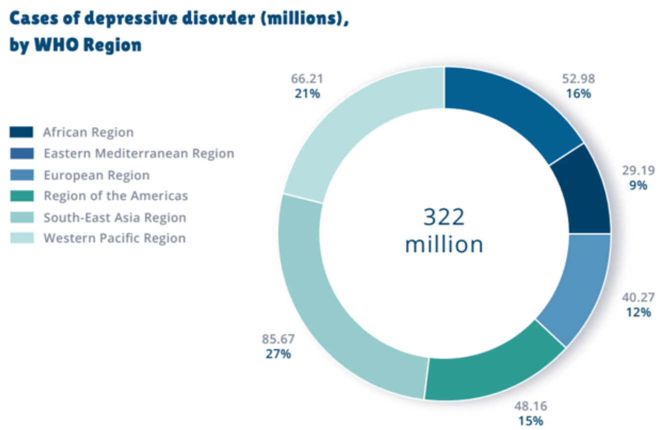

Figure 1.1.1 Cases of Depression by Region

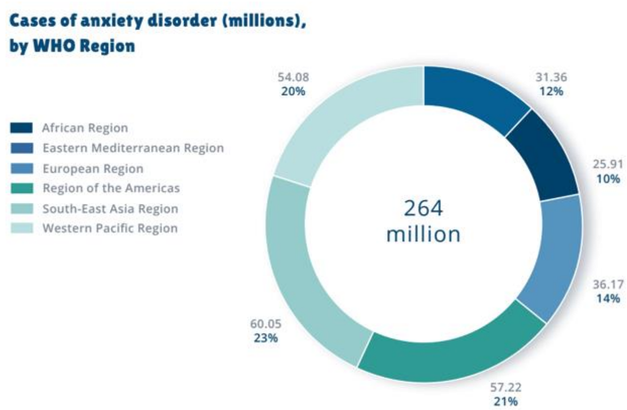

Figure 1.1.2 Cases of Anxiety by region

The WHO believes that a person's overall health, including both their physical and mental well-being, is essential for overall wellness (Sau & Bhakta, 2017), also the studies proved that a strong relationship between Stress and Depression and Anxiety which cause a suicide thought as shown in figure (1.3)(Samele et al., 2018).

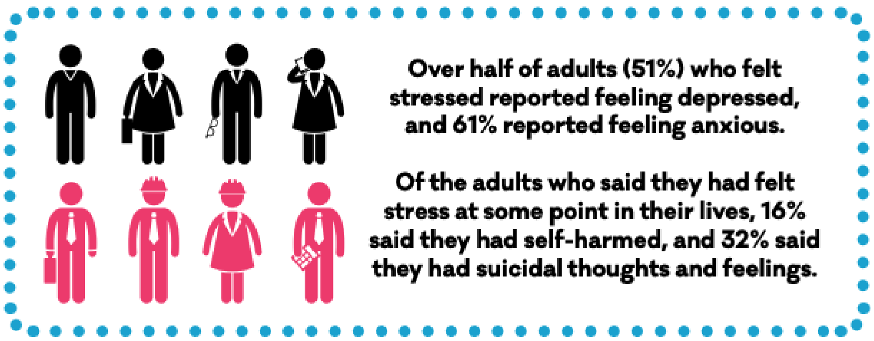

Figure 1.1.3 Stress Statistics

In Palestine, the political issues, and the Israeli occupation, and the surrounding circumstances, made it more vulnerable to have mental health illnesses, among 1254 patients 15.3 % reported depression,17.3 % anxiety disorder (other than PTSD or acute stress disorder), and 23.2 %post-traumatic stress disorder [PTSD]. Among children ≤15 years old, factors significantly associated with PTSD included being witness to physical abuse or murder, receiving threats, and property destruction or loss (Marie et al., 2016).

Increasing number of people who are looking for mental health aid and lack of number of specialists all around the world, late mental health disorder detection and the shortage in mental health governmental budget encourage researchers to find a solution for early detection of mental health issues and to employ the technologies for self-diagnoses which could lead to avoid development of the mental health issues which could lead into a disaster result.

Depression, Anxiety, and Stress are the most common disorders in the modern lifestyle, and many scales are used to self-identify the illness status of patients, one of the most common scales is DASS which will be used to collect data from local individuals, this scale has set of questions for Depression, Anxiety and Stress, the answers will be aggregated based on the scale equations, and the final score will be classified to give the final result, these results will be used for data labeling, and then the supervised machine learning will be used to learn from this data and generate a model that going to be used for future cases classification.

The Depression, Anxiety, and Stress Scale (DASS-21) is a method for gathering information that may be leveraged to create a best-fitting method for the detrimental consequences of despair, anxiety, and strain on mental health (Basha & Kaya, 2016a). Research shows that the DASS exhibits strong intrinsic coherence, and internal reliability, and effectively distinguishes between the characteristics of stress as well as despair more effectively than that of other assessments currently in use. Therefore, the DASS is a good tool for evaluating the symptoms of anxiety, distress, and stress in clinical as well as non-patient populations. Utilizing a large database and a learned system, the researchers hope to get a better understanding of which techniques perform best for visualizing anxiety, panic, and tension in real time (Ghorpade-Aher et al., 2023a).

This has led to a significant increase in focus from health researchers toward studying this issue. However, accurately distinguishing anxiety, depression, and stress is a challenge for machines and requires the use of an appropriate learning algorithm for proper diagnosis.

In this study, local data will be collected, analyzed, and then used five supervised machine learning. algorithms Random Forest, Support Vector Machines, KNN, XGBoost, and Multi-Layer Perceptron to early detection of mental health issues (Depression, Anxiety, and Stress) based on the data collected from several cases, where these algorithms will be trained against this data to be able to make decisions or future cases.

1.2 OBJECTIVES

The main objective of this study is to employ machine learning technology to find a model that can predict mental health issues specifically Depression, Anxiety, and Stress by responding to a few questions extracted from the DASS scale, then the model will be able to classify the patient whether he has one of the mentioned disorders and which level, to achieve this goal the below steps will be followed.

- Preparing an Arabic version of the questionnaire based on DASS.
- Collecting data from local individuals.
- Finding the best model using machine learning that can be used to automate the process of diagnoses of Depression, Anxiety, and Stress.

1.2 CONTRIBUTION

Depression, Anxiety, and Stress represent prevalent mental health disorders. However, a localized Arabic dataset specifically addressing these disorders for comprehensive study is lacking, alongside a dearth of prior experimental studies utilizing an Arabic dataset. The creation of such a local dataset and subsequent comparison with international studies will illuminate the distinct local features and characteristics pertaining to these three Mental Health disorders, delineating key indicators and symptoms. Moreover, the development of a robust predictive model capable of accurately gauging the severity of mental health conditions such as Depression, Anxiety, and Stress signifies a pioneering effort in Palestine's mental health landscape. This model, leveraging machine learning techniques on local data, can be deployed online for use by individuals and mental health specialists, aiding in early case detection and guiding treatment recommendations based on diagnostic outcomes.

1.3 OVERVIEW

Chapter (1) focuses on the problem of predicting mental health disorders especially Depression, Anxiety, and Stress using machine learning techniques. First, we discuss the common types of mental health disorders (Depression, Anxiety, and Stress) and how they affect human well-being. Furthermore, the objectives of this study have been discussed as what are the main goals that the research targeting to achieve. Finally, the main contribution of this study is mainly evolving about collecting local DASS data sets for mental health disorders.

Chapter Two provides a literature review on mental health disorders (Depression, Anxiety, and Stress) classification to give an overview of existing research. Various methods are discussed for classifying the different types of mental health disorders using the DASS scale and the techniques used for detecting and diagnosing the same. Moreover, a combination of machine learning and traditional statistical methods is introduced as a possible way to predict the risk of Mental health issues.

In chapter three, the machine learning methodology is presented through the development process to evaluate and deploy machine learning models. The method consists of data pre-processing, feature engineering, model selection, model evaluation, and model deployment.

Finally, in chapter (4), the results are presented and benchmarking between different models that were used. The presentation of the results contains all available means to reach the most appropriate model to achieve the optimal model through a fair comparison between the models used in the classification process.

In conclusion, Chapter 5 delves into the restrictions of existing approaches and potential developments in this area. It highlights that further research is needed to improve the accuracy and reliability of the models. Also, new techniques can be developed to handle large datasets and reduce computational complexity.

Chapter 2

2.1 BACKGROUND

In general, about 1 out of every 6 adults will have depression at some time in their life. Depression affects about 16 million American adults every year. Anyone can get depressed, and depression can happen at any age and in any type of person, many people who experience depression also have other mental health conditions. Anxiety disorders often go hand in hand with depression. People who have anxiety disorders struggle with intense and uncontrollable feelings of anxiety, fear, worry, and/or panic. These feelings can interfere with daily activities and may last for a long time (Kessler et al., 2010).

Machine Learning applications are increasing in the medicine and health care industry, in the mental health sector the need for Machine Learning is important to understand its current applications and future opportunities. And how Machine Learning can support mental health clinical practice while taking into consideration its current limitations. Machine Learning technology is becoming common in medicine for physical health applications, the adoption of Machine Learning in mental health has been slower. Mental health practitioners are more hands-on and patient-centered in their clinical practice than others in the Physical health sector, they depend more on soft skills and relationships with the patients and observing direct behaviors and emotions.

However, Machine Learning has great opportunities to enhance mental illness diagnosis and understanding of mental illnesses, collecting data of previously diagnosed cases and analyses these cases by learning the algorithms on it, will help to automatically predict new cases to make the diagnoses process more efficient and accurate.

2.2 LITERATURE REVIEW

Given the profound impact of mental health on human well-being, it stands as a pivotal area of research garnering substantial scholarly attention. This study specifically delves into the realms of Depression, Anxiety, and Stress, with an exhaustive literature review scrutinizing prior investigations within these domains. Notably, the review concentrates on research endeavors utilizing data derived from the Depression, Anxiety, and Stress Scale (DASS), providing a comprehensive foundation for the current research focus.

2.2.1 Using Machine Learning Algorithms with Mental Health dataset

Early and timely diagnoses of mental health disorders prevent possible harmful result of people, especially among elderly patients. (Sau & Bhakta, 2017) Sau, A., & Bhakta aim to develop an appropriate predictive model, to diagnose anxiety and depression among older patients from socio-demographic and health-related factors, using machine learning technology. The research was done on 510 geriatric persons using 10 machine learning classifiers. Random Forests obtained the best results with 89% of prediction accuracy. In another research (Sau & Bhakta, 2019) the same authors studied different populations of seafarers' vulnerability to suffering from mental health disorders mainly Anxiety and Depression. By doing aperiodic screening for the seafarers using automated procedures backed by machine learning algorithms to identify the risk of being diagnosed with these disorders for early referral to psychological counseling and treatment. A total of 470 seafarers were interviewed at Haldia Dock Complex, India. Various socio-demographic, occupational, and health-related information were collected. Five machine learning classifiers were used

(Catboost, Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine) implemented in Python, where Catboost gave the best results in terms of accuracy and precision (82.6% and 84.1%) respectively. The results were encouraging automation of the timely consuming screening procedure by using machine learning.

Another research by A. Priya, S. Garg, and N. P. Tigga (A. Priya et al., 2020) which focuses on the same set of disorders, but with different datasets for predictions of Anxiety, Depression, and Stress. This research has used Five different machine learning algorithms Decision Tree (DT); K-NN; Naïve Bayes (NB); Random Forest Tree (RFT); Support Vector Machine (SVM). Data were collected from 348 participants employed and unemployed from different communities and cultures. The participants aged between 20 and 60 years answered 21 questions on DASS 21 scale, to diagnose Stress, Anxiety, and Depression, with 7 questions for each. Due to Imbalanced classes f1 score the accuracy of naïve Bayes was found to be the highest, although Random Forest was identified as the best model.

While stress disorders are a common issue among people who are working in the IT field, (Srinivasulu et al., n.d.) applied machine learning algorithms to analyze stress patterns in working adults. OSMI mental health survey 2017 which includes responses from people who are working in Tech field, Various Machine Learning techniques were applied to train our model after due data cleaning and preprocessing. The accuracy of the above models was obtained and studied comparatively; Boosting gave the highest accuracy among other modules. Also, decision trees helped to identify the most influencing features on the expected results which were gender, family history, and availability of health benefits in the workplace.

2.2.2 DASS21-Based Related Studies

DASS is one of the common nonclinical scale for diagnosing the mental health disorders Depression, Anxiety and Stress. It has two versions, 42 and 21 versions, the 21 version consists of 21 related questions, divided into three groups, 7 questions for each mental health disorder, Depression, Anxiety and Stress, which has been used in this study. Also, many previous research has been conducted to analyze the results from these scales using machine learning algorithms. To gain insights about these results and to find out which are the most accurate questions that could predict the specific disorder. (Choudhury et al., 2019) research which was conducted in Bangladesh and due to the difficulties in the younger population of underdeveloped and developing countries. The research aiming to get insights that could explain why young especially university students suffer from depression, to check whether depression can be successfully predicted with the help of machine learning using related features among 935 Bangladesh undergraduates. They also used DASS 21 scale to collect responses and labelling the data. Random Forest was found to be the best algorithm, closely followed by Support Vector Machine with similar accuracy and f-measure of around 75% and 60% respectively. But Random Forest giving a better precision, recall and lower false negatives.in the other hand, (Ghorpade-Aher et al., 2023b) utilized the scale of the Depression, Anxiety, and Stress, 21 items (DASS-21), in analyzing mental health symptoms. The researchers collected the data using a cross-sectional study with 339,781 respondents using data from an online mental health resource center. Depression, anxiety, and stress were measured as both outcomes and predictors in machine learning models that included sociodemographic factors. A feed-forward artificial neural network was applied using IBM

SPSS version 23, the prediction accuracies were 76.4% for depression, 76.3% for anxiety, and 87.4% for stress. The study focused on the importance of stress and anxiety as significant contributors to the predictive models. The findings described the potential of machine learning in the early detection of mental health issues, providing a 76.4% to 87.4% accuracy in predicting depression, anxiety, and stress prevalence. However, the authors were concerned about the necessity of human judgment based on ethical principles in Islamic law (Maqasid al-Shariah) in considering and applying these machine-generated predictions.

2.2.3 DASS42 Based Related Studies

DASS42 is the second version of the DASSS scale and consists of 42 questions, divided into 3 groups of 14 questions for each disorder Depression, Anxiety, and Stress. The next three kinds of literature worked on an online available dataset for DASS42 survey responses(openpsychometrics.org, 2020). With more than 38K records from international participants, T. Singh (Tejveer Singh, 2022) worked on predicting the condition of Depression, Anxiety, and Stress for various people worldwide using various Machine learning methods. The models which were used are Naïve Bayes, SVM, Random Forest, Decision Tree, and Nearest Neighbors. Parameters were tuned using Randomized search or grid search for better selections. SVM was performing the best among all. The second research that used the same dataset was S. G. A. G. Prince Kumara (Kumar et al., 2020) which focused on the prediction of five severity levels of anxiety, depression, and stress using eight different machine learning models. These methods fall into four different categories: Bayes, neural network, lazy, and tree. Last is a hybrid technique of K-star and random forest method. The hybrid approach improved the accuracy of the single algorithm, but it took 30 to 45 minutes to execute, whereas single algorithms were executed in a maximum of five minutes. The results showed that neural networks performed better than all the others. Among the categories of neural networks, RBFN performed the best. The third one was (Sun et al., 2022) Y. H. Sun, H. Luo, and K. Lee proposed a novel Long to Short approach that uses machine learning to develop efficient and convenient short assessments or surveys to approximate an equivalent long one with acceptable accuracies for predicting anxiety disorders. They used several machine learning techniques which are Logistic Regression (LR), Gaussian Naive Bayes (GNB), Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP) neural network, Extreme Gradient Boosting decision trees (XGBoost) and Stacked Generalization Ensemble (Ensemble). As a result, accuracy achieved over 90% on a reduced DASS-42 scale for measuring low vs. high levels of anxiety that only has five questions plus demographics or seven questions without demographics, RF and Ensemble techniques performed better than others. The below table (2.1) explains the comparison between the three above-mentioned research on the same dataset.

Table 2.2.2.1 Comparison between the research on the same dataset

Research	Dataset	Participants	Objective	Methods Applied	Key Findings
T. Singh [11]	DASS42 Survey Responses (Online)	38K Worldwide	Predicting Depression, Anxiety, Stress	Naïve Bayes, SVM, Random Forest, Decision Tree, Nearest Neighbors	SVM performed the best among all methods after parameter tuning.
S. G. A. G. Prince Kumara [12]	DASS42 Survey Responses (Online)	38K Worldwide	Predicting severity levels of anxiety, depression, stress	Eight models: Bayes, Neural Network, Lazy, Tree, Hybrid of K-star and Random Forest	Neural networks, especially RBFN, outperformed others. The hybrid approach improved accuracy but had longer execution times.
Y. H. Sun, H. Luo, K. Lee [13]	DASS42 Survey Responses (Online)	38K Worldwide	Developing short assessments for anxiety disorders	Logistic Regression (LR), Gaussian Naive Bayes (GNB), SVM, Random Forest (RF), MLP Neural Network, XGBoost, Ensemble	Achieved over 90% accuracy on a reduced DASS-42 scale for measuring low vs. high levels of anxiety. RF and Ensemble techniques performed the best.

The previous three research were conducted against the DASS42 version, but it was the same online available dataset, added to that, this study data (Preethi & Mahalakshmi, 2023) also utilized DASS42. But with its own collected dataset to check how different personality affects influence depression, stress, and anxiety using machine learning by figuring out which characteristics accurately influence mental health disease. The data was collected using a self-report survey, three machine learning algorithms were used: Support Vector Machines, Logistic Regression, and Random Forest. Logistic Regression and Support Vector Machines resulted in being more accurate than Random Forest for all three outcomes. The results suggest that anxiety can be predicted more accurately than stress or depression.

2.2.4 Hybrid Methods

Machine learning with DASS scales is used to predict the early stages of mental health disorders, especially Depression, Anxiety, and Stress. To make the process more accurate the researchers merge third-party techniques to increase accuracy and try to gain more accurate insights. The following related studies which have been conducted using third party techniques, the first one (Likforman-Sulem et al., 2017). Which has merged the new database called EMOTHAW (EMOTion recognition from HAndWriting and draWing) that connects emotions with how people write and draw with to figure out when people feel down or stressed during everyday tasks like writing or drawing. They collected data from 129 people and used smart algorithms to figure out which factors are best at predicting someone's emotional state. They recorded seven different tasks using a digital tablet, including various writing, and drawing activities. The data collected includes things like where the pen is when it was used, how hard it was pressed, and other details. The results showed that the method worked better for recognizing anxiety and stress compared to identifying depression. The research highlights the importance of using devices like tablets to easily collect information about how people write and draw, with the potential to improve well-being by identifying negative emotions in daily activities, this research used the ensemble decision trees using Random Forest for ranking the best targeted emotional state, it revealed that anxiety and stress detection much better than depression, Whereas A novel approach of mixing the feature selection(FS) prior to classification process to eliminate the unwanted features and only select the significant ones this research (Arokkiya Mary & Jabasheela, n.d.), the author is addressing the same topic of classifying Depression, Anxiety and Stress using machine learning algorithms, on a dataset which was collected by his own consists of 267 students, the unique about it that he used the bio inspiring algorithms, and optimization techniques like specifically Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) and Feature Selection(FS) with the Ant-Miner-based classification model, The evaluation metrics include accuracy, sensitivity, specificity, Kappa coefficient, F-score, False Discovery Rate (FDR), and False Omission Rate (FOR). The results revealed that it is verified that the use of ACO in the FS for ant miner-based data classification is superior to the PSO-based FS, The third one (Fatima et al., 2021) used the DAS21 scale to look for Depression, Anxiety, or Stress symptoms using the sentiment meaning of a written text, it proposes a technique called DASentimental by splitting the text into 200 sequences of emotional words, then convert these sequenced into Embeddings which is a vector representation of the text including the semantic meaning, to do this, they collected a data of 142 suicide notes, they used a technique called cognitive network science, where they represented each set of recalled words as a bag of words (BOW) and as a walk over a network that represents how words are connected in our memories. This method considers the importance of words in our memory and the connections between them. The results of this model were very accurate in predicting depression (R = 0.7), anxiety (R = 0.44), and stress (R = 0.52). The model, powered by a neural network (Multi-Layer Perceptron), opens up possibilities for exploring how our brains organize emotions.

2.2.5 COVID-19 Related Studies

During the COVID-19 pandemic period, people around the world were affected by the circumstances, either on the level of their health or mental health or other fields of life, many kinds of research were conducted to study the prevalence of mental health issues among people during the lockdown, especially Depression, Stress and Anxiety, the following researches which were conducted in Bangladesh and Kingdom Of Saudi Arabia are related to this research, the first one (Sofia et al., 2023) focusses on Depression among these people as they are at higher risk of developing other health conditions. The authors conducted a study using a DAS21 survey with 1694 records and used machine learning algorithms such as Decision Tree KNN and Naïve Bayes, the study concluded that KNN was better in predicting than other techniques based on the accuracy matrix, and Decision Tree was more accurate in detecting the depression of a person, hence machine learning was suggested to replace the conventional methods being used. The second one which was also conducted in Bangladesh (Mahmud et al., 2023) used machine learning to develop a model that predicts which students are at risk of suicide. The model was trained on data from a survey of over 2,000 students and was able to predict suicidal risk with an accuracy of 79%. The most important predictors of suicidal behavior were found to be relationship status, family environment, family income, family type, sex, job loss, economic loss, loss of family/relatives due to COVID-19, depression, anxiety, stress, and insomnia. This information can be used to develop more effective suicide prevention interventions, such as offering additional support services, such as counseling or therapy, to students who are at high risk of suicide. And the third one (Al-Wesabi et al., 2022) was conducted in Saudi

Arabia during the COVID-19 pandemic using the same scale. This study proposed the development of an Intelligent Feature Subset Selection with Machine Learning-based DAS predictive (IFSSML-DAS) model to determine DAS (Depression, Anxiety, and Stress) among 938 people during the COVID-19 crisis. The model uses the Group Gray Wolf Optimization FSS (GGWO-FSS) technique to reduce the curse of dimensionality and the Beetle Swarm Optimization-based Least Square Support Vector Machine (BSO-LSSVM) model for classification. The performance of the proposed IFSSML-DAS model was tested using a benchmark DASS-21 dataset and achieved an ROC of 99%. The results suggest that the proposed IFSSML-DAS model is an effective tool for determining DAS among people during the COVID-19 crisis. Table (2.2) compares the three COVID-19-related papers:

Table 2.2 Comparison between the three COVID-19 related papers

Research	Location	Participants	Objective	Methods Applied	Key Findings
[17]	Bangladesh	1694 individuals	Focus on Depression during COVID-19	DAS21 survey, Machine Learning Algorithms (Decision Tree, KNN, Naïve Bayes)	KNN was identified as the best predictor of depression based on the accuracy matrix. Decision Tree was more accurate in detecting depression. Suggested the adoption of machine learning over conventional methods.
[18]	Bangladesh	Over 2,000 students	Predicting suicide risk among students during COVID-19	Machine Learning Model trained on survey data	The model achieved 79% accuracy in predicting suicidal risk. Key predictors included relationship status, family environment, income, job loss, COVID-19-related factors, and mental health indicators. Recommendations for targeted interventions, such as counseling, for high-risk students.

Research	Location	Participants	Objective	Methods Applied	Key Findings
[19]	Saudi Arabia	938 Instances	Developing an Intelligent Feature Subset Selection with a Machine Learning-based DAS predictive model	Group Gray Wolf Optimization based FSS (GGWO-FSS), Beetle Swarm Optimization based Least Square Support Vector Machine (BSO-LSSVM)	The proposed IFSSML-DAS model achieved a ROC of 99%. Effective in determining Depression, Anxiety, and Stress during the COVID-19 crisis. Suggested as a valuable tool for assessment in similar contexts.

2.2.6 Statistics Based Research

The previous research used machine learning models with different flavors, supervised, semi-supervised, and non-supervised algorithms, on the other hand, before machine learning, several kinds of research were conducted using statistics, to study the effectiveness of DASS scales and the correlation between its features and questions (Nada et al., 2022), (Ahmed et al., 2022) and (Shafiq et al., 2020) aimed to assess the psychometric characteristics of DASS21 scale using both classical test theory (CTT) and item response theory (IRT) techniques, the results showed that the scale was efficient in term of internal consistency in both techniques. (Das et al., 2021) tried to determine whether previous depression, anxiety, and stress diagnosis affect the postpartum depression among mothers, he collected data from 118 newborn mothers, and the screening happened thorough the DASS21, the results showed that there is as association between a history of previously diagnosed depression and current depression scores were statistically significant ($p=0.04$), and the associations between previous depression and current anxiety and stress scores ($p=0.02$ and $p=0.003$ respectively). In Malaysia, the prevalence of mental illness is rising each year (Zarkasi et al., n.d.) this research aims to study the levels of Stress, Anxiety, and Depressions and the associated factors in the rural village of Sabah, and to achieve that data was collected from 115 individuals based on DASS-21 and using the statistical approaches, the results shows that the rural village community at higher risk to depression and anxiety levels which requires the decision makers to take actions.

Based on the above literature it is obvious that DASS scales are widely used and trusted by many researchers and communities, also after the research we couldn't identify more than a couple of similar studies that used DASS with an Arab dataset, added to that, no local Palestinian data has been collected before based on the same scale or even in the mental health sector which was utilized in machine learning, which make our contribution unique in this field and emphasize us to employ the trend technologies like data science especially machine learning to find insights about the mental health disorders based on locally collected dataset from the Palestinian community.

Chapter 3

1.1 METHODOLOGY

1.1.1 Proposed Method

The purpose of this research is to use machine learning algorithms to predict mental health status, by preparing a model that can predict as much accurately as possible the mental health status. After training it on the collected data, the results of the model would be one of the five classes (Normal, Mild, Moderate, Sever and Extremely Sever), the data will be labeled using the DASS-21 mathematical equation. Most data science research is almost share common methods to get insights and knowledge from data, all the way from determining the goals to results discussion and future work recommendation, in this research the below steps were followed to achieve the goal as shown in Figure (3.1.1):

- Data collection where we have prepared an online form using google forms and distributed the form among individuals personally and online to fill out the survey.
- Exploratory Data Analysis (EDA) is to understand the data, how it is being constructed and what can we understand from it.
- Data preprocessing by polishing the data, detecting anomalies, balancing the distribution of the classed between the data and out layers dividing into groups, and making it ready to be used by machine learning to have accurate results.
- Feature selection by selecting the proper features that have a significant effect on the results.
- Model selection to determine which machine learning models to use that are suitable for the shape, type, and size of the data collected.

- Training the models by splitting the dataset into training and testing groups.
- Model Optimization by tweaking the models' parameters to hypertune and get better results.
- Results discussion, comparison, and evaluation.

The research is concerned with numbers and accuracy levels of ability for machine learning algorithms to predict Anxiety, Depression, and Stress, which is considered as quantitative kind of research and will use the experimental approach to test several machine learning algorithms to obtain the best results.

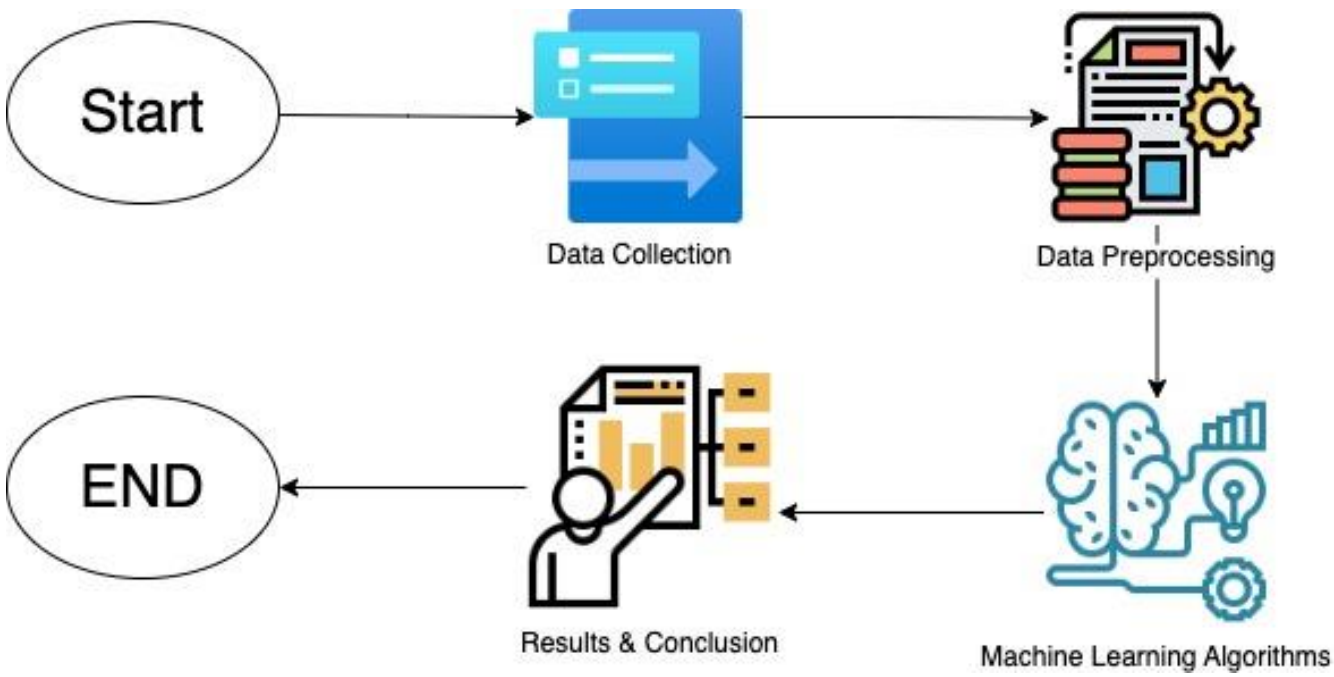

Figure 1.1.1: The steps of the process

1.2 DATASET DESCRIPTION

The Depression, Anxiety, and Stress Scale were applied to reveal the psychological status of the individuals. Also, this scale is used in determining the level of negative emotional states. The scale was developed by Lovibond and Lovibond (1995) and proposed by the Australian Psychological Society (Basha & Kaya, 2016b). After consulting the subject matter experts in the field, the dataset is divided into two main sections, the first one was the demographic data about the individuals like age, gender, City, type of living, academic level, financial status, family size, marital status and work type and the other section was collected based on the Arabic version of DASS-21 scale (Miriam Taouk Moussa, 2001), this scale consists of 21 questions shown in the figure (3.2.1) below, representing 3 disorders Depression, Anxiety and Stress, with 7 questions for each. Each subset of questions in the DASS21 scale belongs to a different disorder as below (Ali et al., 2021):

- Depression: ['Q3', 'Q5', 'Q10', 'Q13', 'Q16', 'Q17', 'Q21']
- Anxiety: ['Q2', 'Q4', 'Q7', 'Q9', 'Q15', 'Q19', 'Q20']
- Stress: ['Q1', 'Q6', 'Q8', 'Q11', 'Q12', 'Q14', 'Q18']

The total number of records collected was 731 records, using the Google Forms tool to collect the data, the questionnaire has been shared either personally or online using some channels targeting adults in Palestine for the sessility of the data and to preserve the privacy of the participant to personal data that can identify the participant identity was collected to maintain the ethical rules about the research.

استعمل التقديرات التالية:

٠ لا ينطبق عليّ بتاتاً

١ ينطبق عليّ بعض الشيء أو قليلاً من الأوقات

٢ ينطبق عليّ بدرجة ملحوظة أو بعض الأوقات

٣ ينطبق عليّ كثيراً جداً، أو معظم الأوقات

١	وجدت إنني مضطرب ومنزعج بسبب أمور تافهة جداً	٠	١	٢	٣
٢	شعرت بجفاف في حلقي	٠	١	٢	٣
٣	لم يبدو لي أن بإمكاني الإحساس بمشاعر إيجابية على الإطلاق	٠	١	٢	٣
٤	شعرت بصعوبة في التنفس (شدة التنفس السريع، اللهثان بدون القيام بمجهود جسدي مثلاً)	٠	١	٢	٣
٥	لم يبدو لي أن بإمكاني أن أبدأ في القيام بأعمالي	٠	١	٢	٣
٦	كنت أميل إلى ردة فعل مفرطة للظروف والأحداث	٠	١	٢	٣
٧	شعرت بالرجفة (إن رجلي لا تقوى على حملي مثلاً)	٠	١	٢	٣
٨	أجد صعوبة في الاسترخاء	٠	١	٢	٣
٩	وجدت نفسي في مواقف جعلتني قلقاً جداً، وكنت مرتاحاً للغاية بزوالها	٠	١	٢	٣
١٠	شعرت بأن ليس لدي أي شيء أتطلع إليه	٠	١	٢	٣
١١	وجدت نفسي أميل إلى الاضطراب والانزعاج بسهولة	٠	١	٢	٣
١٢	شعرت بأنني أستهلك الكثير في الطاقة العصبية (شعرت بأنني أستهلك الكثير من قدرتي على تحمل التوتر العصبي)	٠	١	٢	٣
١٣	شعرت بالحزن والإكتآب	٠	١	٢	٣
١٤	وجدت أنني قليل الصبر كلما أخرني شيء (عند انتظار المصعد، إشارات المرور، أو كلما طلب مني الانتظار، مثلاً)	٠	١	٢	٣
١٥	انتابني شعور بالإغماء	٠	١	٢	٣
١٦	شعرت بأنني فقدت الاهتمام بكل شيء تقريباً	٠	١	٢	٣
١٧	شعرت بأن قيمتي قليلة كشخص	٠	١	٢	٣
١٨	شعرت بأنني أميل إلى الغيظ بسرعة	٠	١	٢	٣
١٩	عرقت بشكل ملحوظ (عرق غزير من اليدين مثلاً) بدون أن يكون الطقس حاراً وبدون بذل مجهود جسدي	٠	١	٢	٣
٢٠	شعرت بالخوف بدون أي سبب وجيه	٠	١	٢	٣
٢١	شعرت بأن الحياة لا قيمة لها	٠	١	٢	٣

Figure 1.2.1 Dataset Description

1.3 DATA PREPROCESSING

There is a common says in the data science era “Rubbish in rubbish out”, which means, that to have mature and accurate models, clean data needs to be given to this model. This explains why around 60% - 80% of the effort is in the data science lifecycle (Klettke & Störl, 2022) Data preprocessing includes cleansing, normalization, validation, data encoding, imbalance handling, data quality checks, and other activities, in this research we have used python 3 and utilized different packages like Pandas, NumPy, and SkLearn.

The first step in data preprocessing is to check for missing values, we found that there are 252 missing in the religion column, so we had to exclude it, also there were 5 missing values in the gender, and forward filling was used to fill these missing data. The next step in data preprocessing is checking for duplicated records, and there were no duplicated records. After that, and since most of the data were collected as textual, and machine learning algorithms work better with numerical data, we have converted the data into numbers like gender, city, residence type, education level, finance level, marital status, work type using mapping these values from textual to numbers. Also, there are 21 questions, and the possible answers were mapped into integers to be able to aggregate the results and find the total of answers.

One more technique in the data preprocessing, is data grouping, like age, we have grouped the values into groups instead of having hundreds of distinct values, so the age was grouped into 6 groups 0 to 5 based on defined criteria like less than 10 is first group, 16-21 second group and so on. Scaling data is important for machine learning models is essential as most of the algorithms use the Euclidean distance to calculate the distance between two data points, which could be Standardization, Normalization, or Minmax scaler which is sensitive to the scale of the variable (Ahsan et al., 2021), so we have used the StandardScaler from sklearn. Preprocessing (scikit-learn.org, 2023c) to achieve this.

Data Labeling is one of the critical tasks in the supervised machine learning life cycle, in this research, the DASS21 uses an equation to calculate the results and based on these results it determines the level of infection of the individuals, so after responses were collected, the question's answers were used to calculate the results based on the scale's equation (K. B. Priya et al., 2022):

- **Depression**
    - 0-9 = Normal
    - 10-13 = Mild Depression
    - 14-20 = Moderate Depression
    - 21-27 = Severe Depression
    - 28+ = Extremely Severe Depression
- **Anxiety**
    - 0-7 = Normal
    - 8-9 = Mild Anxiety
    - 10-14 = Moderate Anxiety
    - 15-19 = Severe Anxiety
    - 20+ = Extremely Severe Anxiety

- **Stress**
    - 0-14 = Normal
    - 15-18 = Mild Stress
    - 19-25 = Moderate Stress
    - 26-33 = Severe Stress
    - 34+ = Extremely Severe Stress

Then Label encoders were used to convert the classes from textual to numeric values using sklearn. preprocessing library (scikit-learn.org, 2023b), Normal:0, Moderate:1, Mild:2, Severe:3 and Extremely Severe:4.

This research uses a dataset based on responses to the DASS21 scale which is used as mentioned earlier to identify Depression, Anxiety, and Stress, therefore the data was divided into three groups, one for each disorder, and from the first impression, it was oblivious and expected to have an imbalance in the data, figures (3.3.1,3.3.2,3.3.3) show the distribution of each disorder classes:

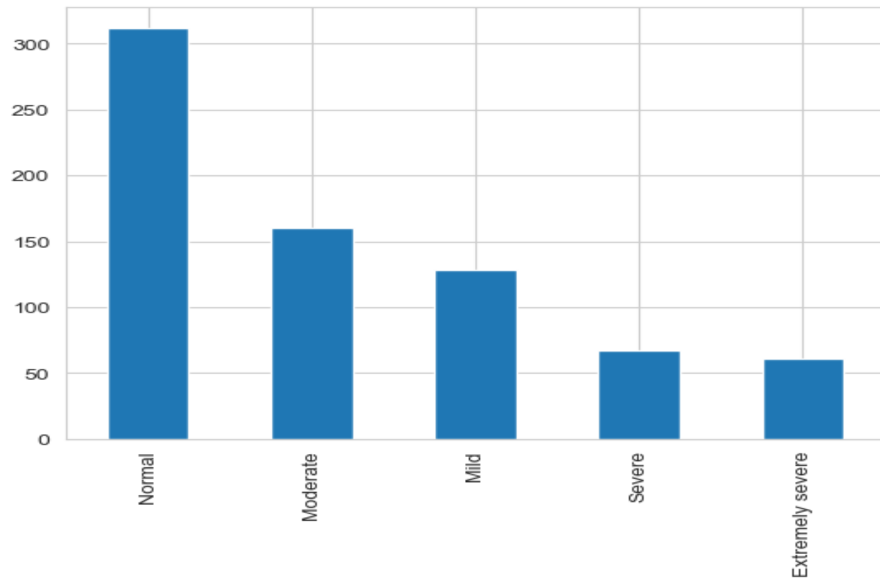

Figure 1.3.1: Depression class distribution

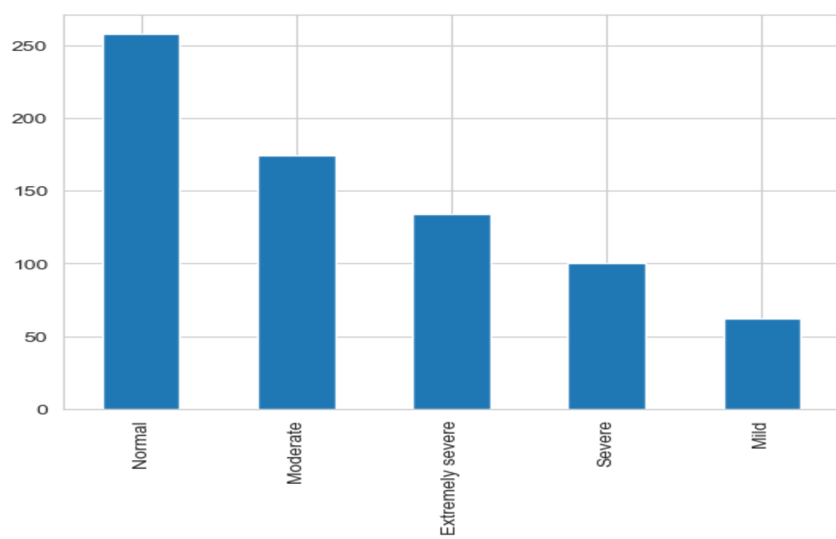

Figure 1.3.2 Anxiety class distribution

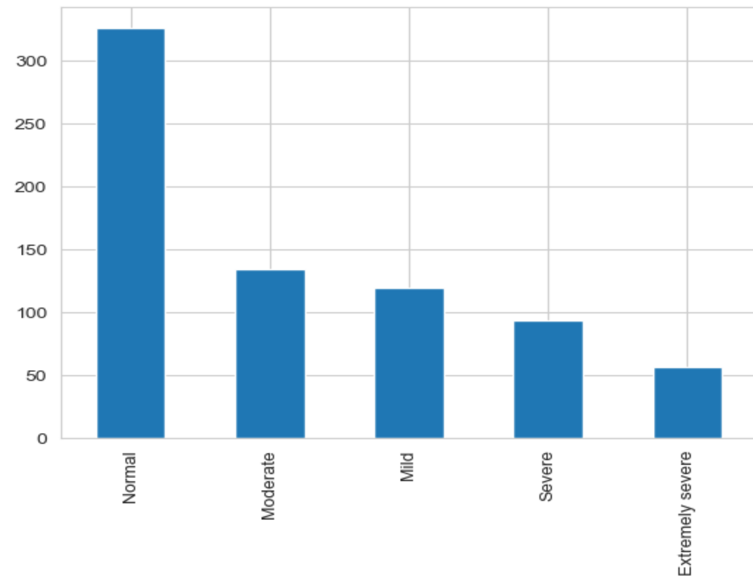

Figure 1.3.3: Stress Classes Distribution

To handle the imbalance in data, the data resampling technique is used to up-sample the low count classes, and the resampling technique used from sklearn.utils the function sklearn.resample, resamples arrays or sparse matrices in a consistent way, and the default strategy implements one step of the bootstrapping procedure. In simple terms, sklearn.resample doesn’t just generate extra data points to the datasets by magic, it creates a random resampling(with/without replacement) of your dataset. This equalization procedure prevents the Machine Learning model from inclining towards the majority class in the dataset. (scikit-learn.org, 2023a) to have a unified number of samples for each class of 200 samples for each class.

1.4 EXPLORATORY DATA ANALYSIS (EDA)

Exploration Data Analysis is essential to the scientific process and helps establish the validity of scientific conclusions (Jebb et al., 2017), in this stage data is being explored and analyzed to deeply understand the data and try to find insights and relations as well, in this research, each feature was analyzed individually, and the correlation between features also identified to understand the relation between them, figure (3.4.1) shows the description of the dataset:

skimpy summary

Data Summary

dataframe	Values
Number of rows	728
Number of columns	36

Data Types

Column Type	Count
string	32
float64	3
int64	1

number

column_name	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
age	0	0	2.7	0.74	0	2	3	3	5	
fam_size	0	0	6.3	2.2	1	5	6	8	20	
gender_encoded	0	0	0.58	0.49	0	0	1	1	1	
fam_size_new	0	0	6.2	2.2	1	5	6	8	20	

string

column_name	NA	NA %	words per row	total words
gender	0	0	1	730
governante	0	0	1.5	1092
address	0	0	1	731
edu_level	0	0	1.1	816
fin_level	0	0	1	740
hand	0	0	1	728
mstatus	0	0	1.1	833
work_type	0	0	2	1457
Q1	0	0	7.1	5168
Q2	0	0	6.1	4459
Q3	0	0	6.2	4492
Q4	0	0	6.3	4578
Q5	0	0	6.4	4681
Q6	0	0	6.6	4835
Q7	0	0	6.3	4583
Q8	0	0	6.8	4977
Q9	0	0	6.3	4584
Q10	0	0	5.4	3924
Q11	0	0	6.5	4754
Q12	0	0	6.7	4872
Q13	0	0	6.6	4793
Q14	0	0	7	5132
Q15	0	0	5.5	4018
Q16	0	0	6.6	4785
Q17	0	0	5.8	4231
Q18	0	0	6.7	4880
Q19	0	0	6	4368
Q20	0	0	5.8	4226
Q21	0	0	5.6	4069
Depression_total	0	0	72	52486
Anxiety_total	0	0	72	52168
Stress_total	0	0	82	59772

Figure 1.4.1: Dataset Description

As mentioned in the dataset description section, the data contains the 21 DASS scale questions and some demographic question answers, and as part of the data exploration, each feature will be discussed individually in the following section:

Gender: figure (3.4.2) shows the percentage of gender between the individuals who participated in the study, it shows 415 Females and 306 Males which means the data is almost distributed equally between genders.

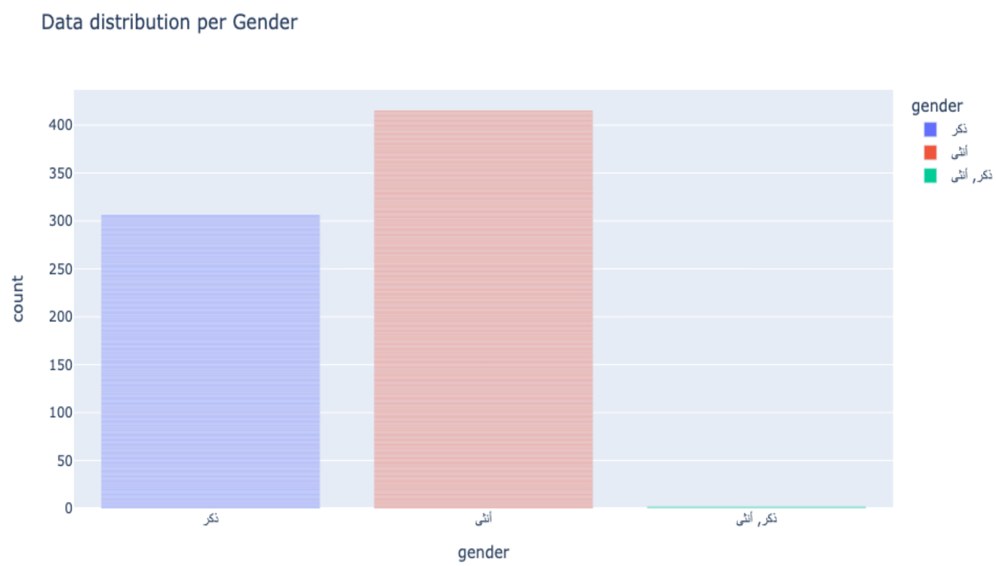

Figure 1.4.2: Gender Distribution

Also, the following figures (3.4.3,3.4.4,3.4.5) show the distribution of mental health disorders (Depression, Anxiety, and Stress) severity levels between genders which reveals that the infection status between females is a little bit higher than males in the three disorders:

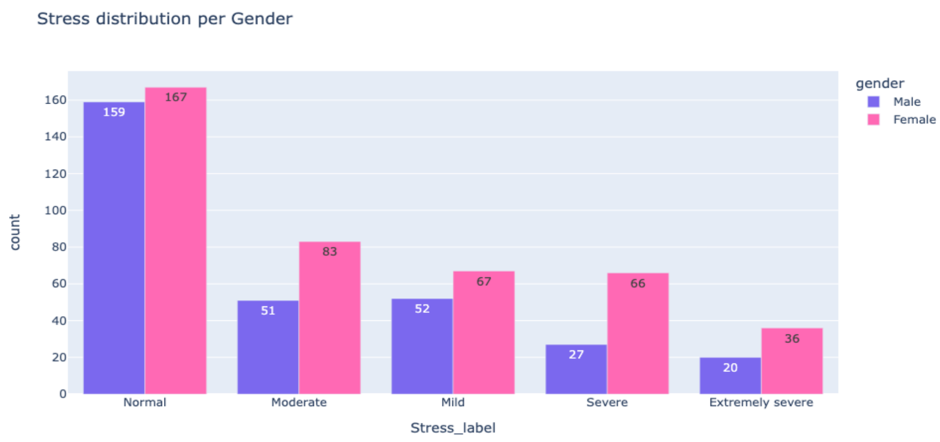

Figure 1.4.3: Stress Distribution Per Gender

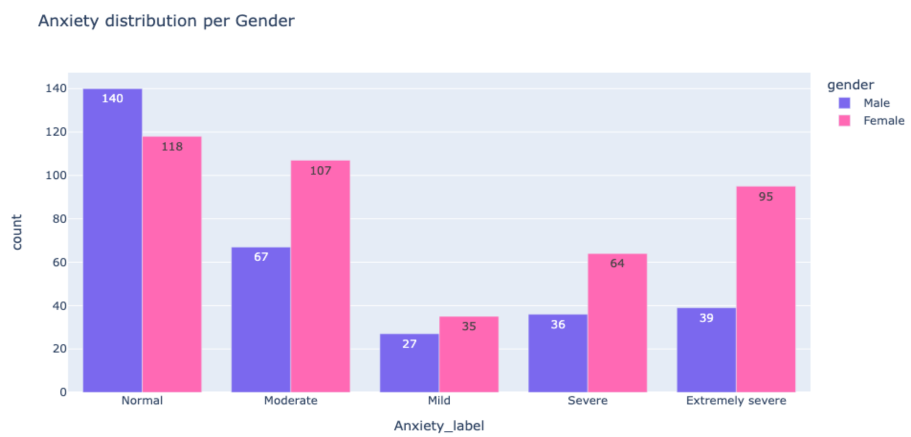

Figure 1.4.4: Anxiety Distribution Per Gender

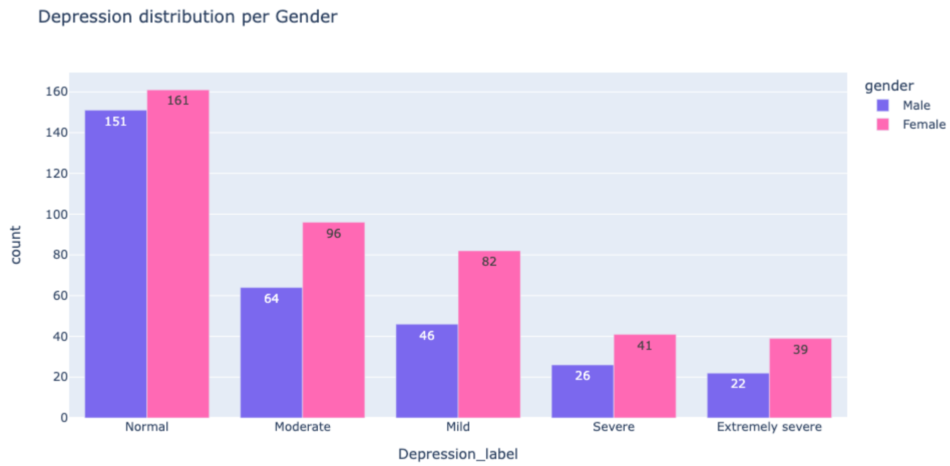

Figure 1.4.5: Depression Distribution Per Gender

City: this feature represents the city to which the individual belongs, the below figure (3.4.6) shows the distribution of responses between these cities which shows that 31.5% of the responses were from Jenin, 22.7% from Ramallah, 10.6 from Nablus, 8.52% from Gaza and the others are distributed between the rest of the cities.

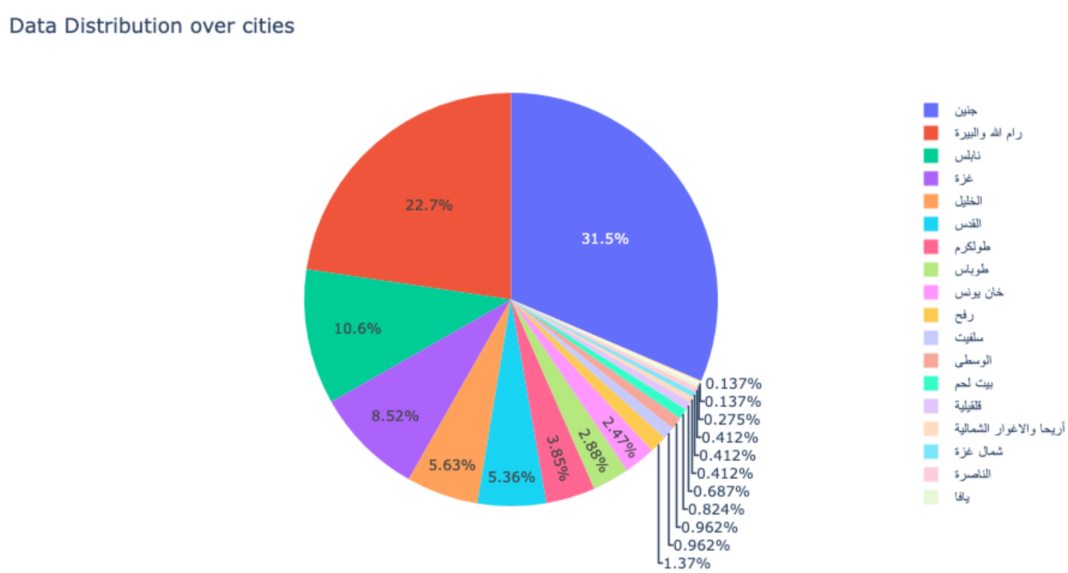

Figure 1.4.6: Distribution of the data between cities.

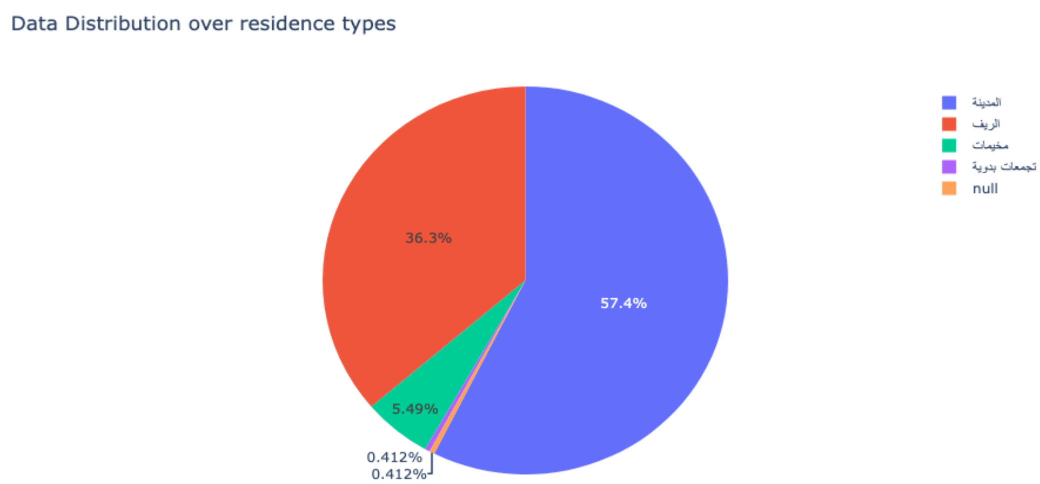

Figure 1.4.7: Data Distribution over residence type

Residence Type: which represents the type of place they are living in, figure (3.4.7) shows the percentages of the residence types among individuals, which shows that maturity of the participants was living in cities with 57.4%, 36.3% living in rural areas and 5.49% living in camps.

Education Level: which contains the education level of the participant, the following figure (3.4.8)

shows the percentages of each value of the education level, the figure reveals that the maturity of the participants has a bachelor's degree with 73.2%, then 16.1% with secondary school, on the other hand, 9.34% holding master's degree or above and 1.37 with low-level education.

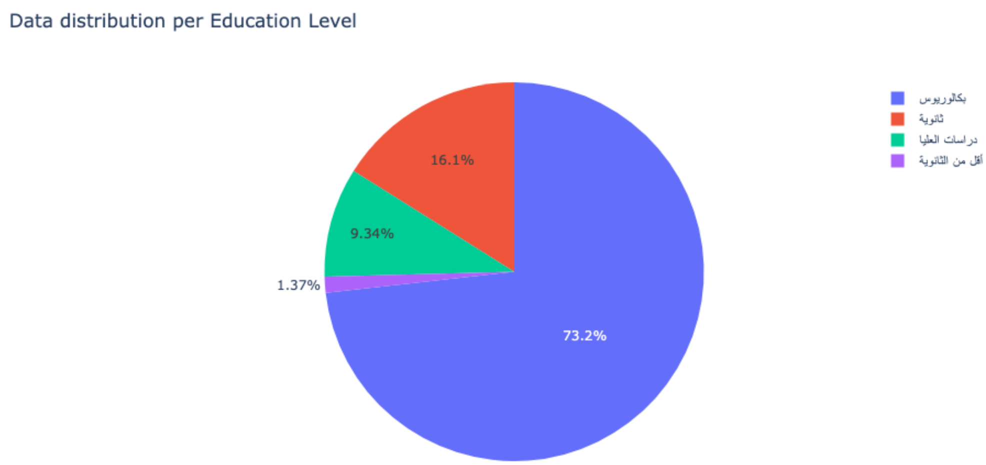

Figure 1.4.8: Distribution of data over education level

Financial Level: which includes the financial level of the participants, the figure below shows that 78.6% of the participants are from the mid-layer which represents the maturity of the data, then 12.4% were from high-income level, while 6.87% low-income level, 1.37% are nonexistent and .82% are extremely wealthy, this is explained visually in Figure (3.4.9).

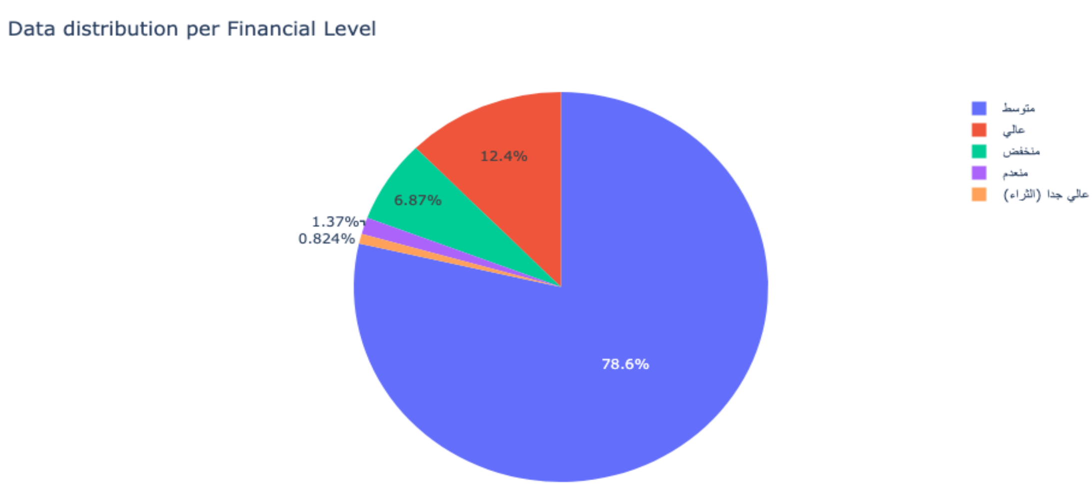

Figure 1.4.9: Data Distribution over financial level

Marital Status: represents the marital status of the participant, based on the data collected, figure (3.4.10) shows that 60.9% of the participants are single, and 33.4% are married.

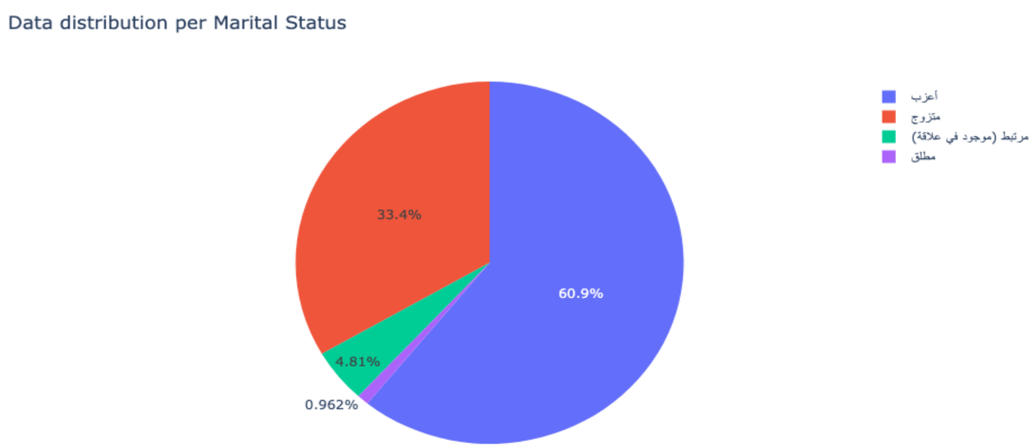

Figure 1.4.10: Data Distribution over marital status

Work Type: this feature represents the participant work type like private sector, public sector, or others, data analysis shows that 43% of the participants are students who are not working, 25% are private sector employees 12.4% are freelancers, also 10.4% are not currently working, for the government sector the analysis shows that 5.36% who are working in the government the rest of the participants 3.85% are from NGO sector, figure(3.4.11) represent these percentages.

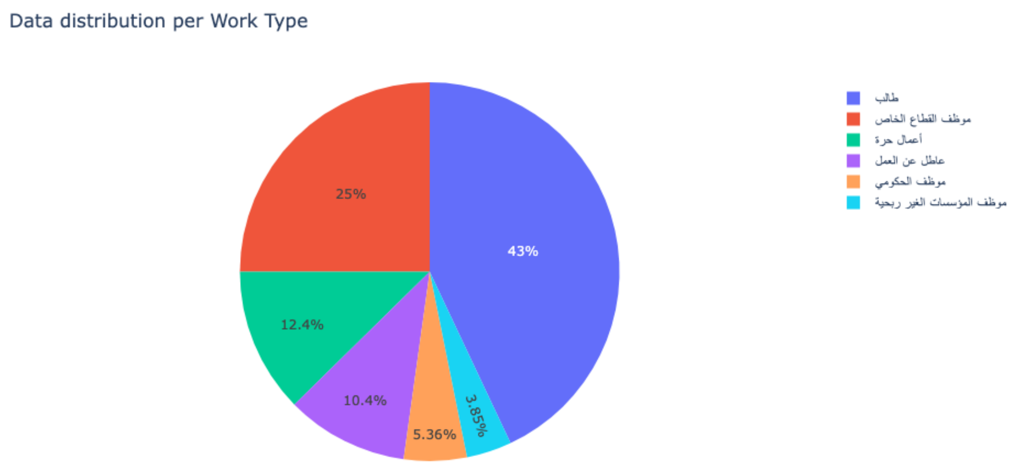

Figure 1.4.11: Data distribution over work type

Age: this represents the age of the participants, the below figure (3.4.12), explicitly shows that the maturity is between 20-30. Year which makes this dataset more youth based.

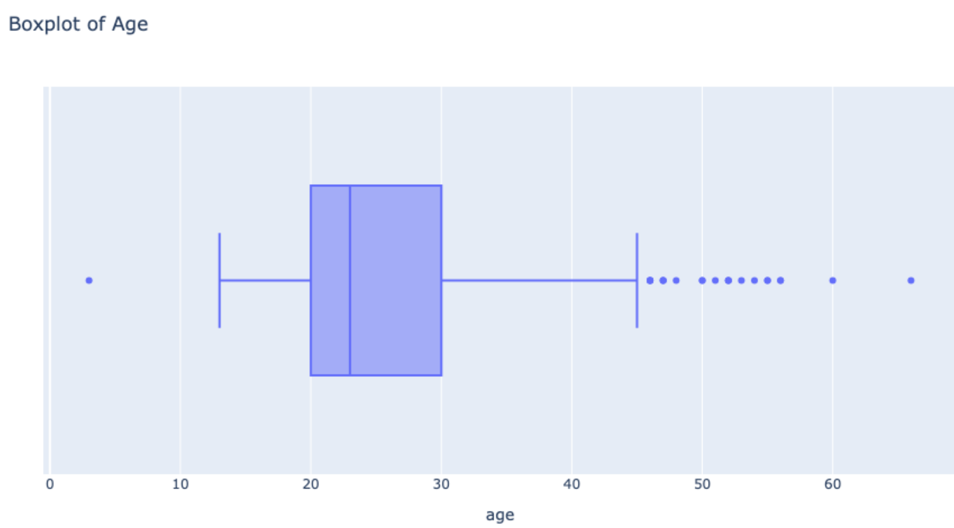

Figure 1.4.12: Boxplot of Age

Family Size: represents the family size of the participant, the analysis reveals that the data is symmetric and the maturity of data between 3 and 8 family members, figure (3.4.14) shows the distribution of the family size using the histogram chart:

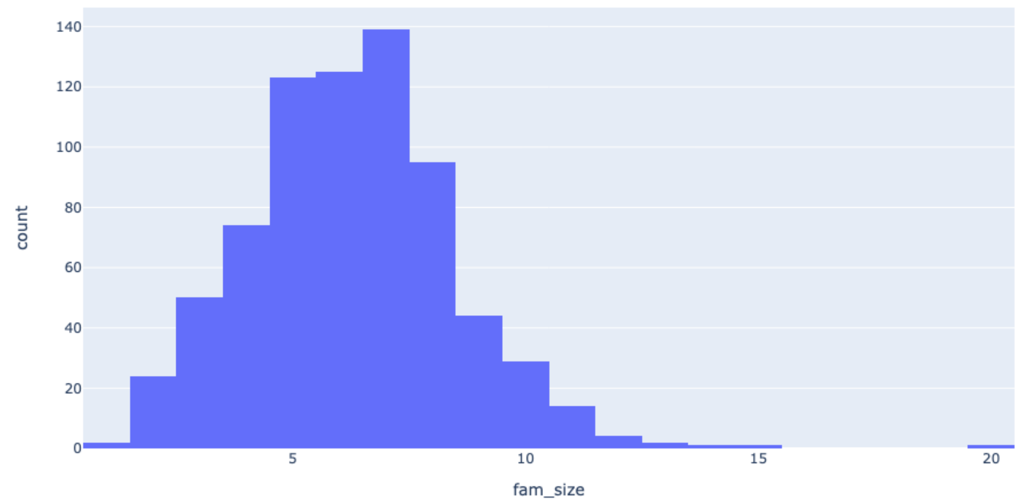

Figure 1.4.13: Family Size histogram

Questions Answers: As mentioned earlier, the dataset contains 21 questions used to diagnose the infection status by one of the mental health disorders (Depression, Anxiety, and Stress), and each question has 4 possible answers mapped as:

- 0 Did not apply to me at all.
- 1 Applied to me to some degree, or some of the time.
- Applied to me to a considerable degree or a good part of time.
- Applied to me very much most of the time.

The two figures (3.4.15,3.4.16) below show the percentages of each answer:

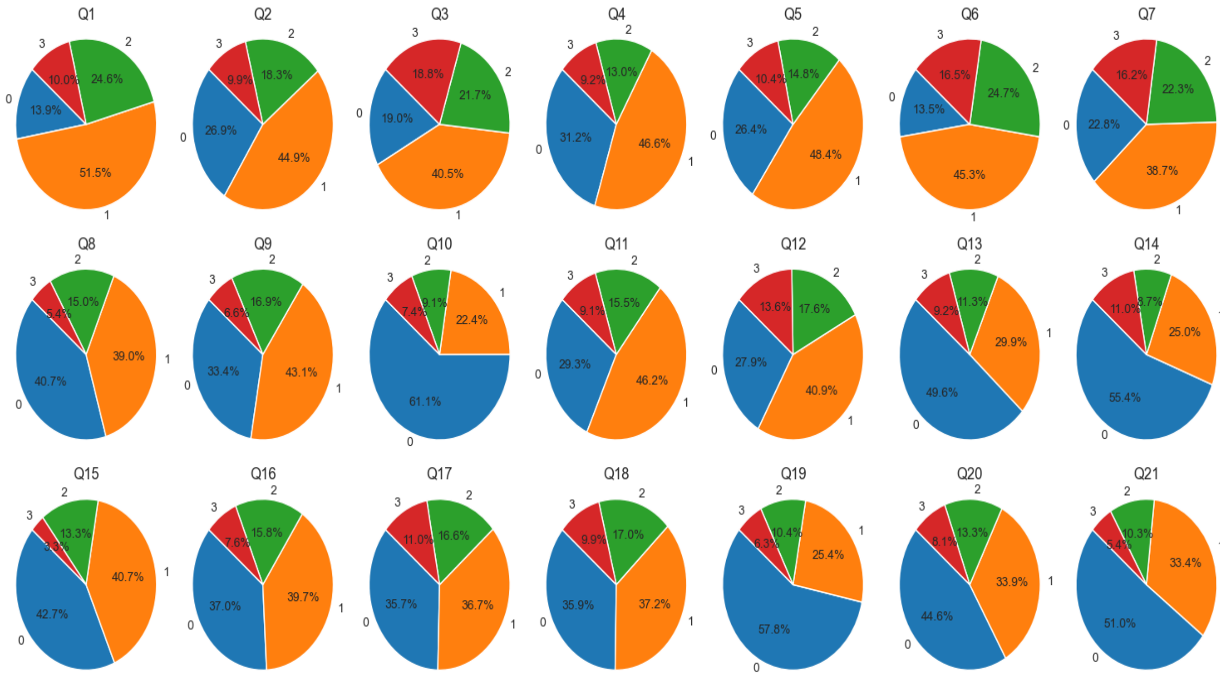

Figure 1.4.14: Percentages of each answer choices

1.5 APPLIED MODELS

As data science-based research, different machine learning algorithms need to be employed to be trained against the data, to find the best models that work better with the data and give accurate results in predicting the mental health status among the three disorders (Depression, Anxiety, and Stress). Therefore, these models need to be from different types and architectures. in this research five supervised machine learning algorithms have been used, Random Forrest, Support Vector Machines, K Nearest Neighbors, XGBoost, and Multi-layer Perceptron Neural Networks, in the following, each algorithm will be discussed individually.

Figure 1.5.1: Bar chart of the DASS question answers

1.5.1 Random Forest

Random forests are known for their robustness against overfitting and ability to handle large datasets with high dimensionality. They are also effective in capturing complex relationships within the data, which is beneficial for predicting mental health issues that may have multifaceted underlying factors. It is mainly used for classification and regression. (Belgiu & Drăgu, 2016; Edla et al., 2018; Kulkarni, n.d.) Random Forrest uses an ensemble learning technique, by having multiple decision trees as appears in figure (3.5.1), each tree classifies the input, then a voting technique is used to take the final decision and generate the output. Normal decision tree algorithms are rule-based and are solely based on some set of rules for prediction on data sets. In contrast to this, random forest classifiers instead of using the Gini index (Hegelich, 2016), or information gain for the calculation of the root node, find the root node and split the features randomly. it works with large datasets and maintains accuracy even when a significant proportion of the data is missing. It's widely used in various fields like finance, medicine, biology, and more for predictive modeling.

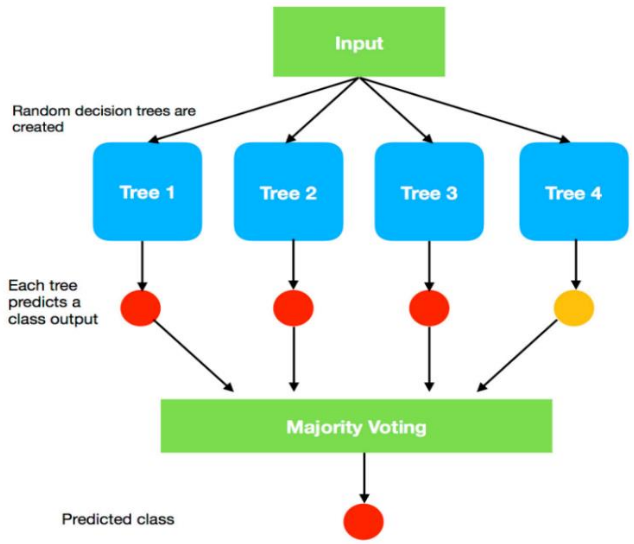

Figure 1.5.2: How Random Forest Works Internally (Edla et al., 2018)

Below is the mathematical representation of the Random Forrest Algorithm Where F(x) = The ensemble output of the Random Forest algorithm, N = The total number of decision trees in the forest, $f_i(x)$ = The result of the ith decision tree in the forest and x = The input data.

$F(x) = 1/N \sum_i (f_i(x))$(Madison Schott, 2019) (1)

1.5.2 Support Vector Machines

SVM is a supervised machine learning algorithm used for classification-based tasks, (Auria et al., 2008; Mavroforakis & Theodoridis, 2006) it creates a decision boundary between different classes, these boundaries enable the model to predict the datapoint class depending on the feature vectors. These decision boundaries, known as the hyperplane, are orientated in such a way that it is as far as possible from the closest data points from each of the classes. These closest points are called support vectors. SVMs are well-suited for binary classification tasks like detecting mental health disorders. They work by finding the optimal hyperplane that separates different classes in the data, making them particularly effective when there is a clear boundary between classes, as may be the case with mental health diagnoses.

Figure (3.5.2) represents how SVM works to separate data points into two classes and easily predict new data points based on the maximum marine hyperplane between support vectors that identify the best hyperplane. The types of hyperplane maximum margin classifier (hard margin), support vector classifier (soft margin), and support vector machines (non-linear) (Huang et al., 2018).

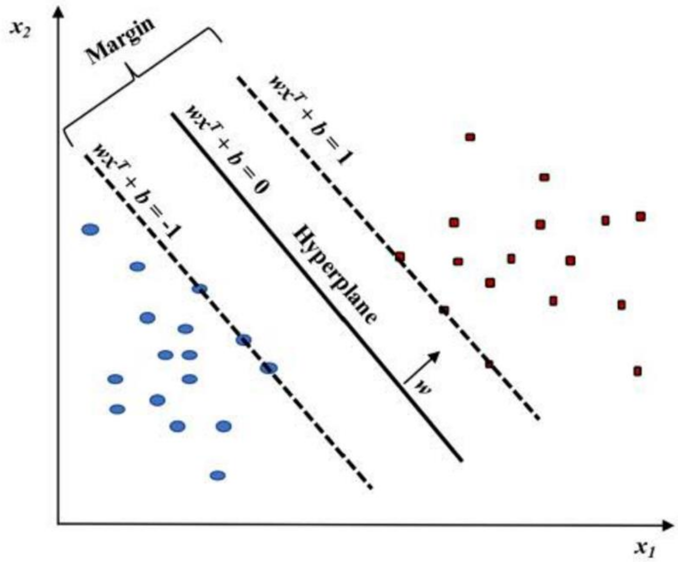

Figure 1.5.3: SVM Classification Criteria (Huang et al., 2018)

The SVM equation is a linear equation to make a classification for data points and try to find the optimal hyperplane the equation of SVM is below considering x as the input vector, W is the weight vector and b is the bias.

$w * x + b = 0$ (Auria et al., 2008) (2)

1.5.3 Gradient Boosting and Extreme Gradient Boosting

Gradient boosting is a machine learning technique that sequentially creates a predictive model, using an ensemble of many simple models, which are decision trees. It uses the concept of boosting, where each model corrects the errors made by previous models and minimizes the loss function using a technique called gradient boosting (Bentéjac et al., 2021). This enables the GB to achieve a wide range of regression and classification use cases. It has many variants like XGBoost, LightGBM, and CatBoost, which include enhanced performance and speed, especially in large and complex datasets, below is the mathematical representation of Gradient Boosting:

$F(x) = \Sigma_{i=1}{}^{m} (f_i(x))$ (3)

XGBoost is an alternative method for predicting a response variable given certain covariates (Pesantez-Narvaez et al., 2019). It was proposed by (Chen & Guestrin, 2016), The main concept behind this algorithm is that it builds D classification and regression trees, so that each subsequent model (tree) is trained using the residuals of the previous tree as shown in figure (3.5.3). In other words, the new model corrects the errors made by the previously trained tree and then predicts the outcome.

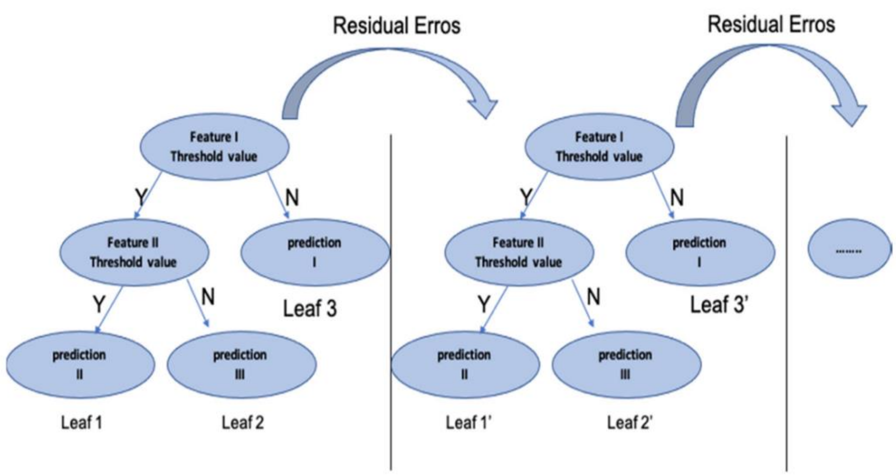

Figure 1.5.4: XGBoost Learning Technique (Ibrahem Ahmed Osman et al., 2021)

Xgboost is one of the implementations of gradient boosting machines (gbm) which is known as one of the best-performing algorithms utilized for supervised learning. It can be used for both regression and classification problems. Xgboost is preferred by data scientists because of its high execution speed out of core computation (Chen & Guestrin, 2016). Its ability to handle missing values and nonlinear relationships can be advantageous for predicting mental health issues accurately.

1.5.4 Multi-Layer Perceptron Neural Network

A Multilayer Perceptron (MLP) is one of the flavors of fully connected artificial neural network, considered as deep learning technique as well. Using feedforward artificial neural network that consists of multiple layers of nodes in a directed graph, it consists of multi layers, Input layer which receives the input data and pass it to the next layer. Usually, each feature in the dataset is mapped into one node in the input layer, Hidden Layers are the middle layers after the input layer and before output layer. The network should have at least one hidden layer, the main job of this kind of layers is to transform the values from the previous layer with a weighted linear summation followed by a non-linear activation function and Output layer. Which represents the final output of the network, and it could be one or more layer depends on the number of expected classes, in case of Binary classification only one output layer needed and in case of multiple, it has to be more (Delashmit & Manry, n.d.; Popescu & Balas, n.d.; Ruck et al., 1990).

The limitation of this algorithm could be the vulnerability for overfitting in the training phase which could make it hard to generalize, also it is considered computationally intensive if the hidden layers number is huge and it can’t be used for sequential kind of data. The inclusion of multiple layers allows for hierarchical feature learning, which can be beneficial for capturing nuanced features related to mental health.

Below is the equation of neurons (Perceptron) inside the hidden layer of MLP, and how it calculates the output, the input data are multiplied by the weights (Popescu & Balas, n.d.):

$$Y_{ij} = f\left(\sum_{i=1}^{n} Xi * w_{ij}\right) \tag{4}$$

where w_{ij} : is the connection weight between the ith node in the input layer and the jth node in the hidden layer, and x_i: is the ith input, and f is the activation function. The activation function f is calculated as in the following equation (Popescu & Balas, n.d.):

$$Y_j = \frac{1}{1 + e^{-y}} \tag{5}$$

Where Xk is the input to the next layer's node.

Figure (3.5.4) also explains the structure of the MLP network and how different layers (Input, Hidden, and Output) are connected and organized with each other.

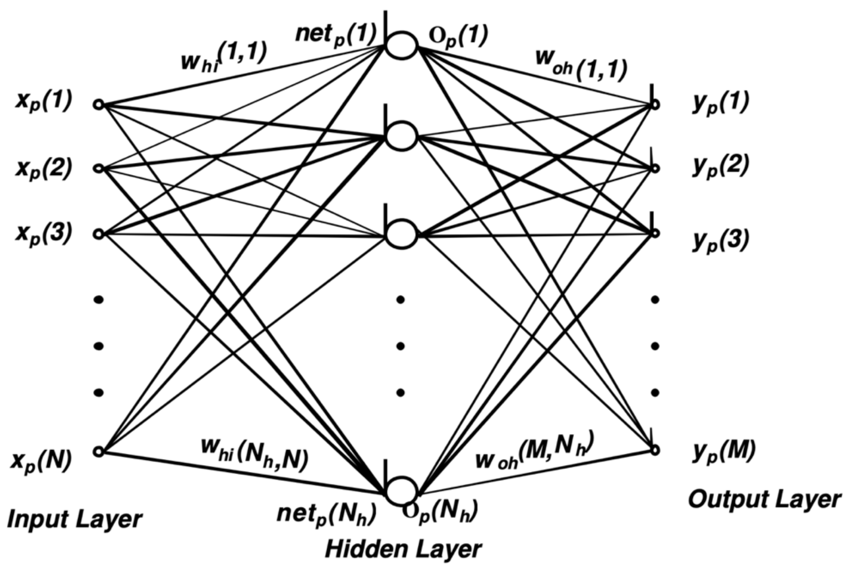

Figure 1.5.5: MLP Network Structure (Delashmit & Manry, n.d.)

1.5.5 K-Nearest Neighbor KNN

K-nearest neighbor (kNN) is one of the most popular machine learning algorithms and is a popular classification method in data mining and statistics because of its simple implementation and significant classification performance (Zhang et al., 2017, 2018) The k-Nearest Neighbors (KNN) algorithm is a simple, yet powerful machine learning technique used for classification and regression. It's part of a broader family of instance-based learning or non-generalizing learning methods, where the function is only approximated locally, and all computations are deferred until function evaluation.

KNN is a simple yet effective algorithm that works well for smaller datasets and can handle noisy data. It's based on the principle of similarity, making it suitable for tasks where similar cases tend to have similar outcomes, such as identifying patterns in mental health data.

Below are the steps of how the KNN algorithm works:

- Choose k: Choosing k significantly affects the prediction. A smaller k indicates that the algorithm is more sensitive to noise in the data, while a larger k increases the complexity and computation but could be more accurate.
- Distance Measure: Calculate the distance between the testing data instance and all the training data points. It used different distance-based techniques like Euclidean distance, Manhattan distance, and Minkowski distance.
- Identify Nearest Neighbors: compare the resulting distances and find the nearest k instances.

Aggregate Neighbors' Outputs: It uses voting to determine k neighbors to make a prediction. Figure (3.5.5) explains how KNN works by having a K value and tries to find the potential predicted classes based on the distance between different data points from the training stage:

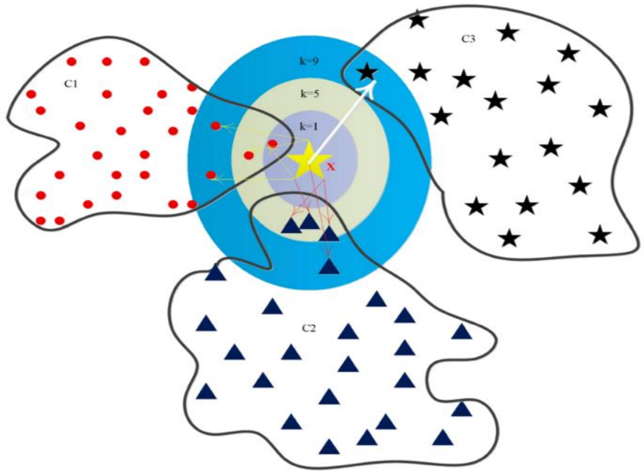

Figure 1.5.6: KNN Classification (Xing & Bei, 2020)

Also below is the Euclidean distance formula used by the KNN algorithm:

$$D = \{(x1, y1), (x2, y2), \ldots, (xn, yn)\}(\text{Virani et al.}, 2021) \quad (6)$$

1.6 TUNING PARAMETERS

Hyperparameter tuning in machine learning is the process of choosing the optimal set of parameters for an algorithm that improves its performance in a given case. This fine-tuning is necessary for enhancing accuracy, reducing overfitting, and speeding up the algorithm. Commonly tuned parameters include the learning rate, which controls the step size in the learning process; the regularization strength, which helps prevent the model from overfitting; and the structure parameters like the number of layers or nodes in neural networks. Specific techniques like grid search are employed to systematically test various combinations of these parameters to identify the best combination. Grid search, for instance, evaluates every possible combination of provided parameter values, searching for the optimal configuration based on performance metrics. Other approaches include random search, which samples random combinations, and more sophisticated methods like Bayesian optimization or genetic algorithms, which use probabilistic models or evolutionary concepts, respectively, to find suitable parameters.

In Support Vector Machines (SVM), parameters like the kernel function type, regularization parameter (C), and gamma value are fine-tuned. The kernel function allows the SVM to handle non-linear relationships, while C and gamma balance the trade-off between the model's complexity and its performance on the training data.

In ensemble methods like Random Forest or boosting algorithms like AdaBoost, Gradient Boosting, and XGBoost, parameters such as the number of trees (n_estimators), the depth of the trees (max_depth), and the learning rate are tweaked. These parameters control the complexity and learning capability of the models.

Neural networks, such as Multilayer Perceptron (MLP), have numerous tuning parameters including the architecture of the hidden layers, the activation functions, the learning rate, and the optimization algorithm. These affect the network's ability to learn from data and generalize to unseen data.

Tuning the K-Nearest Neighbors (KNN) algorithm includes adjusting different parameters to optimize the model's performance and accuracy. One of the main tuning parameters is the number of neighbors, which represents how many of the nearest points to a new data point are considered in the prediction. The distance metric is another important parameter; it determines how we calculate the distances between points. The most common metrics are Euclidean, Manhattan, and Minkowski distances, each measuring point proximity differently and impacting the classification outcome. Then there's the 'weighting scheme,' which is about how much influence each neighbor has in the voting process in a weighted KNN, closer points typically have more sway than more distant ones, potentially improving the model's performance by focusing on the most relevant instances.

For every machine learning algorithm, there's a set of parameters that, when appropriately tuned, can significantly enhance the model's predictive power. It's a process that requires expertise, experimentation, and sometimes automated techniques, to balance the bias-variance tradeoff and achieve the highest possible accuracy without overfitting or underfitting the data.

1.7 PERFORMANCE EVALUATION

In the machine learning ecosystem, the process for determining the success and accuracy of a model is known as performance evaluation. This is a major step in the model development lifecycle which gives insight into the model's capability to handle data it has never seen before in the training phase. There are different tools to measure the model performance, including the Confusion Matrix, which compares the correct and incorrect predictions, precision, and recall, which respectively measure the model's accuracy and completeness, as well as the F-1 score and AUC-ROC curves, which provide mature metrics for accuracy and the trade-off between true positive rate and false positive rate. Cross-validation methods like k-fold validation mix the data to prevent the model from learning from specific parts of the data, ensuring a more robust assessment of the model's true predictive power. Ultimately, these evaluations are pivotal in comparing models, pinpointing critical features, and identifying areas for refinement.

1.7.1 Confusion Matrix

Confusion matrix is a tool that represents the performance of a classification algorithm by displaying the correct and incorrect predictions in a tabular format. This matrix is a major component for evaluating the accuracy of classification algorithms. The matrix split the predictions into four sections: true positives, true negatives, false positives, and false negatives, as appears in Table (3.7.1) which finally provide a rich overview of the model's performance.(Heydarian et al., 2022)

Table 3.7.1: Confusion Matrix (Santra & Christy, 2012)

	Predicted	
Actual	TP	FN
	FP	TN

The entries in the confusion matrix have the following meanings:

- TP (true positive): The number of records classified as accurate while they were true.
- FP (false positive): The number of records classified as accurate while false.
- FN (false negative): The number of records classified as false while they were true.
- TN (true negative): The number of records classified as false while false.

The confusion matrix is an important tool for evaluating the performance of different classification models by using critical metrics like sensitivity (or true positive rate TPR) and specificity (or false positive rate FPR). Sensitivity measures the proportion of actual positives that are correctly identified by the classifier, essentially its ability to detect positives.

Specificity, on the other hand, assesses the classifier's precision in distinguishing between classes, reflecting its ability to identify negatives accurately(Markoulidakis et al., 2021). The confusion matrix provides valuable statistical measures such as Precision, which evaluates the model exactness; Recall, which indicates the model's capacity to correctly identify positive cases out of all actual cases; and the F1 score, which is the harmonic mean of Precision and Recall. This score is particularly useful for balancing the trade-off between Precision and Recall and is commonly used to measure the overall accuracy of a classifier. Below are the equations of each matric:

Accuracy is the proportion of the total number of correct predictions. It is determined using the equation:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{7}$$

Precision is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$\frac{TP}{TP+FP} \tag{8}$$

Recall The true positive rate (TPR) is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$\frac{TP}{TP+FN} \tag{9}$$

F1 score is a statistical measure used to evaluate the performance of binary classification models, which is especially useful when the class distribution is uneven.

$$2* \frac{precision * recall}{precision + recal} \tag{10}$$

1.7.2 Roc Curve

The Receiver Operating Characteristic (ROC) curve is a graphical tool used to measure the performance of a classification machine learning model without choosing a specific threshold. It plots sensitivity (true positive rate) against specificity (1 − false positive rate) at various threshold settings. The Area Under the Curve (AUC) shows the performance of the model, which represents the area under the ROC curve as shown in figure (3.7.1). The closer the AUC is to 1, the better the predictor's performance. ROC analysis takes into consideration the false positives and false negatives equally, and it can be used to select an optimal threshold based on different cost functions or objectives (Muschelli, 2020).

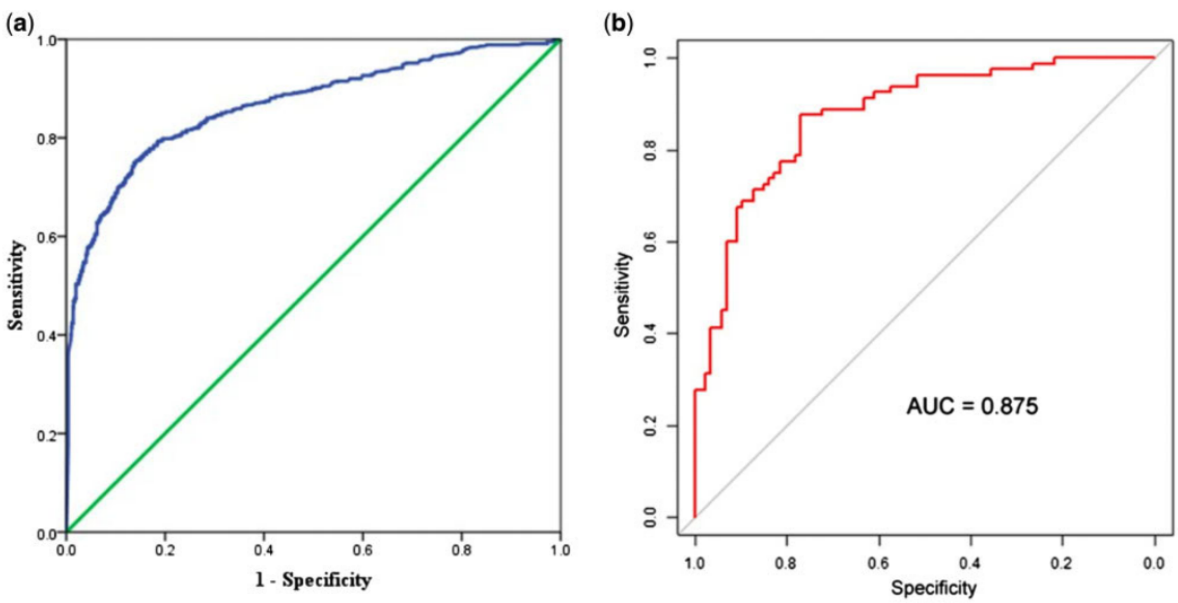

Figure 1.7.1: ROC& AUC Smooth Vs Hard learning (Janssens & Martens, 2020)

Chapter 4

Experiments and Results

In this chapter, the actual experiment results will be discussed in detail, where data is split into Depression, Anxiety, and Stress respectively, and five machine learning algorithms (SVM, Random Forest, KNN, XGBoost, and Multi-Layer Perceptron) were used to get the best results between them. The experiment was done on Macbook Pro 16, 32 RAM, 2.3 GHz 8-Core Intel Core i9 Processor, AMD Radeon Pro 5500M 4 GB Graphics as shown in figure (28), Anaconda used as a development environment with Pythons over Jupyter notebook.

4.1 RESULTS AND DISCUSSIONS OF PREDICTION

Since the dataset is imbalanced, we have used resampling techniques as mentioned in the EDA section, the experiment is divided into multiple phases for each disorder data set as follows:

1- Data splitting where data was split into 3 data sets one for each disorder one for Depression, one for Anxiety, and one for Stress.

2- Model building for each dataset is also divided into multiple steps as below:

   a. Model evaluation on the data as is.
   b. Model Evaluation on the data as is with Grid Search.
   c. Model Evaluation on the resampled data using the Resample technique by specifying the number of each class instance.
   d. Model Evaluation with resampled data using Grid Search.
   e. Model Evaluation using SMOTE resampling technique.

3- Comparing the results to get the best combination of the results.

Following the above strategy, after experimenting, the below tables show the results of each disorder in 5 phases with different techniques. Hence, we found that the best combination was using the resampling technique with grid search.

Table (4.1.1,4.1.2,4.1.3) show the results from all applied scenarios for the depression, Anxiety and Stress datasets when the data was as is, then when we resampled, the data using resampling and the SMOTE technique, also before and after applying the hypertuning using the grid search.

Depression: the results shown in table (4.1.1) represents the training and testing accuracy of the models in different stages, and from these results, the best combination was using the resampling after the grid search, also it is notable that there was an over fitting for Random Forest and XGBoost before the resampling and this has been resolved after the data became balanced.

Table 4.1.1: Depression Results

Technique / Model	Random Forest	SVM	K-Nearest Neighbors	XGBoost	Multi-Layer Perceptron
Data as is					
Training Accuracy	100%	99%	77%	100%	100%
Testing Accuracy	82%	89%	72%	81%	98%
Data as is with Grid Search					
Training Accuracy	100%	99%	76%	100%	100%
Testing Accuracy	84%	100%	71%	84%	98%
Data Resampled					
Training Accuracy	100%	99%	85%	100%	100%
Testing Accuracy	94%	95%	74%	95%	98%
Data Resampled with Grid Search					
Training Accuracy	100%	100%	91%	100%	100%
Testing Accuracy	94%	100%	77%	95%	97%

Technique / Model	Random Forest	SVM	K-Nearest Neighbors	XGBoost	Multi-Layer Perceptron
Data Resampled using SMOTE					
Training Accuracy	100%	94%	90%	100%	100%
Testing Accuracy	86%	83%	81%	87%	86%

Anxiety: Table (4.1.2) shows the Anxiety dataset accuracy classification results also, with the same scenarios for different machine learning algorithms, from these results, the accuracy of the models especially XGBoost and Random Forest was less than in Depression case, with 4% difference.

Table 4.1.2: Anxiety Results

Technique / Model	Random Forest	SVM	K-Nearest Neighbors	XGBoost	Multi-Layer Perceptron
Data as is					
Training Accuracy	100%	94%	75%	100%	100%
Testing Accuracy	74%	80%	60%	74%	96%
Data as is with Grid Search					
Training Accuracy	100%	100%	75%	100%	100%
Testing Accuracy	77%	100%	61%	76%	98%
Data Resampled					
Training Accuracy	100%	97%	80%	100%	100%
Testing Accuracy	94%	89%	69%	91%	97%
Data Resampled with Grid Search					
Training Accuracy	100%	100%	90%	100%	100%
Testing Accuracy	92%	100%	79%	91%	95%
Data Resampled using SMOTE					
Training Accuracy	100%	90%	86%	100%	100%
Testing Accuracy	83%	77%	68%	82%	76%

Stress: the last one which represents the Stress results in different stages for models' accuracy as well which is explained in Table (4.1.3) the results for Stress was the best for all models except KNN, but in general the results for the three disorders weren't far from each other and the gaps are not significant.

Table 4.1.3: Stress Results

Technique Model	Random Forest	SVM	K-Nearest Neighbors	XGBoost	Multi-Layer Perceptron
Data as is					
Training Accuracy	100%	99%	78%	100%	100%
Testing Accuracy	78%	87%	63%	80%	94%
Data as is with Grid Search					
Training Accuracy	100%	100%	79%	100%	100%
Testing Accuracy	80%	99%	65%	80%	94%
Data Resampled					
Training Accuracy	100%	99%	82%	100%	100%
Testing Accuracy	97%	97%	74%	96%	97%
Data Resampled with Grid Search					
Training Accuracy	99%	100%	90%	100%	100%
Testing Accuracy	96%	100%	78%	97%	97%
Data Resampled using SMOTE					
Training Accuracy	100%	93%	89%	100%	100%
Testing Accuracy	88%	86%	80%	88%	85%

After the comparison between the different combinations, we have followed the track of the resampling with grid search, and we will get into details of each disorder with the five models individually:

Depression:

The depression data set was split into 80/20 as training testing with 5 K-fold for cross-validation, the accuracy results after doing the hyper tuning using the grid search the results become (RF:94%, SVM: 100%, KNN:77%, XGBoost:95%, MLP:98%) from confusion matrix figures (4.1.1,4.1.2,4.1.3,4.1.4,4.1.5) it shows that the Random Forest as shown in figure (4.1.1), has classified 190 cases correctly out of 200 cases, while SVM in figure (4.1.2) classified 200 case out of 200, for KNN in figure (4.1.3), could classify correctly only 164 out of 200 in the other hand in figure (4.1.4), XGBoost has classified 191 out of 200 cases and finally in figure (4.1.5), MLP 195 of 200 cases has been classified correctly.

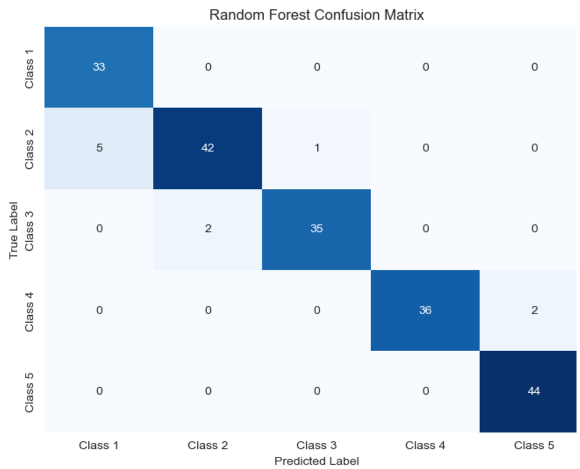

Figure 4.1.1: Random Forest Confusion Matrix for Depression

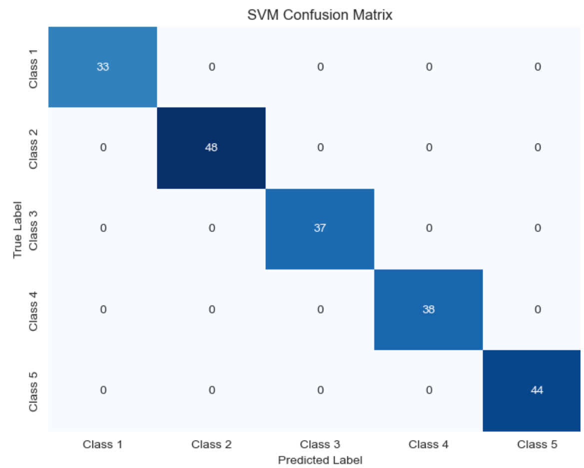

Figure 4.1.2: SVM Confusion Matrix for Depression

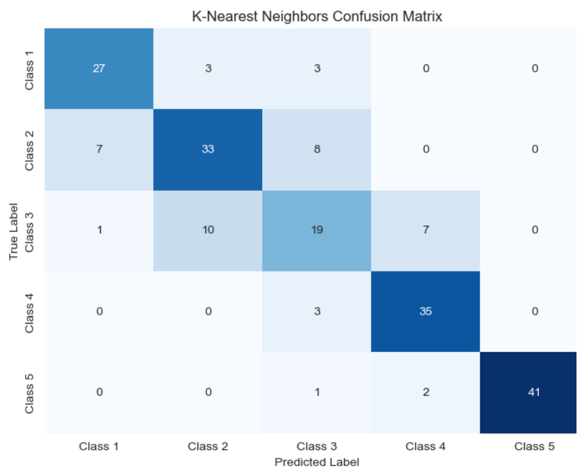

Figure 4.1.3: KNN Confusion Matrix for Depression

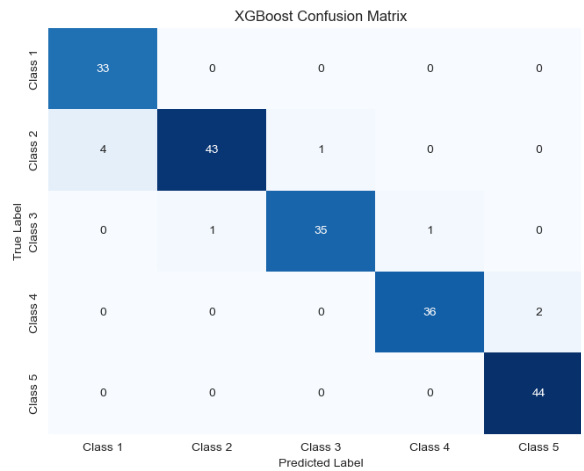

Figure 4.1.4: XGBoost Confusion Matrix for Depression

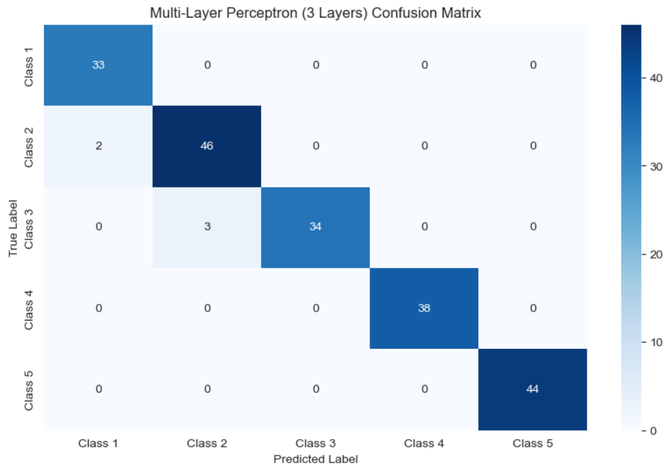

Figure 4.1.5: MLP Confusion Matrix for Depression

Additional metrics that evaluate the model performance including Precision, Recall, F-Score, and Support were used also to gain more insights about the models, tables (4.1.4, 4.1.5, 4.1.6, 4.1.7, 4.1.8) show the results measure the depression dataset results of 5 different classes, table (4.1.4) shows the results of Random Forest with average of precision 95%, Recall 96%and F-1 score 95%.

Table 4.1.4: Random Forest metrics Depression

Class	**Precision**	**Recall**	**F1-Score**
0	0.87	1.00	0.93
1	0.95	0.88	0.91
2	0.97	0.95	0.96
3	1.00	0.95	0.97
4	0.96	1.00	0.98
Macro Average	0.95	0.96	0.95
Weighted Average	0.95	0.95	0.95

In table (4.1.5) is shows the results of SVM, with an exceptional performance of 100% for Precision, Recall and F1-score.

Table 4.1.5: SVM Metrics Depression

Class	Precision	Recall	F1-Score
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
Macro Average	1.00	1.00	1.00
Weighted Average	1.00	1.00	1.00

While KNN as in Table (4.1.6) the results were 77% for the three measures Precision, Recall and F-1 score.

Table 4.1.6: KNN Metric Depression

Class	Precision	Recall	F1-Score
0	0.77	0.82	0.79
1	0.72	0.69	0.70
2	0.56	0.51	0.54
3	0.80	0.92	0.85
4	1.00	0.93	0.96
Macro Average	0.77	0.77	0.77
Weighted Average	0.77	0.78	0.77

In Table (4.1.7) is the results of XGBoost, with 95% average of Precision, 96% average of Recall and 95% average of F1-score.

Table 4.1.7: XGBoost Metric Depression

Class	Precision	Recall	F1-Score
0	0.89	1.00	0.94
1	0.98	0.90	0.93
2	0.97	0.95	0.96
3	0.97	0.95	0.96
4	0.96	1.00	0.98
Macro Average	0.95	0.96	0.95
Weighted Average	0.96	0.95	0.95

MLP metrics which is shown in table (4.1.8), it has 98% for Precision, Recall and F1-score.

Table 4.1.8: MLP Metric Depression

Class	Precision	Recall	F1-Score
0	0.94	1.00	0.97
1	0.94	0.96	0.95
2	1.00	0.92	0.96
3	1.00	1.00	1.00
4	1.00	1.00	1.00
Macro Average	0.98	0.98	0.98
Weighted Average	0.98	0.97	0.97

The ROC Curve (Receiver Operating Characteristic) and AUC (Area Under the Curve) are considered primary visualization tools to illustrate model performance. In this context, a unified ROC Curve is presented, incorporating data from all utilized models into a single diagram. This consolidated visualization facilitates a comparison of the models' predictive capabilities. The ROC Curve clearly shows the trade-off between true positive rates and false positive rates among different model thresholds, offering valuable insights into their predictive power, Figure (4.1.6) is the result for all models used for stress data which reflects the results of the accuracy.

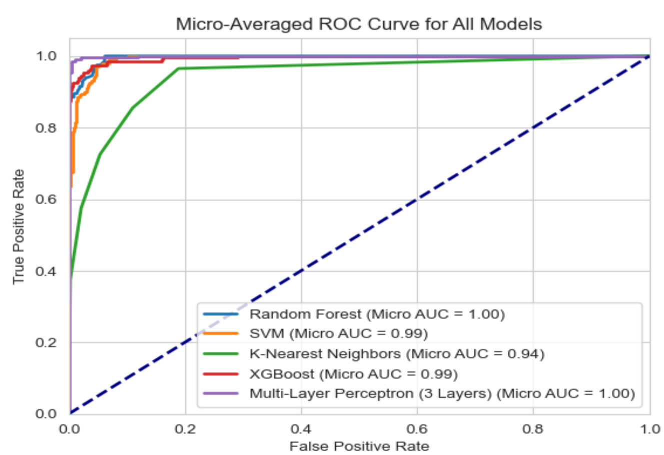

Figure 4.1.6: ROC Curve for Stress Models

Anxiety:

The anxiety dataset was divided into an 80/20 split for training and testing, with 5 K-fold cross-validation. The accuracy results after hyperparameter tuning using grid search, results became RF: 91%, SVM: 100%, KNN: 79%, XGBoost: 91%, and MLP: 96%. For a detailed understanding of the outcomes, the confusion matrix for each model after parameter tuning with the best parameters is provided in figures (4.1.7, 4.1.8, 4.1.9, 4.1.10,4.1.11) figure (4.1.7) shows that Random Forest was able to correctly classify 186 out of 200 cases, in figure(4.1.8) SVM classified correctly 200 out of 200 cases, KNN misclassified 42 out of 200 cases as shown in figure (4.1.9), XGBoost in figure (4.1.10), classified correctly 182 out of 200 and MLP only misclassified 7 out of 200 cases as shown in figure (4.1.11).

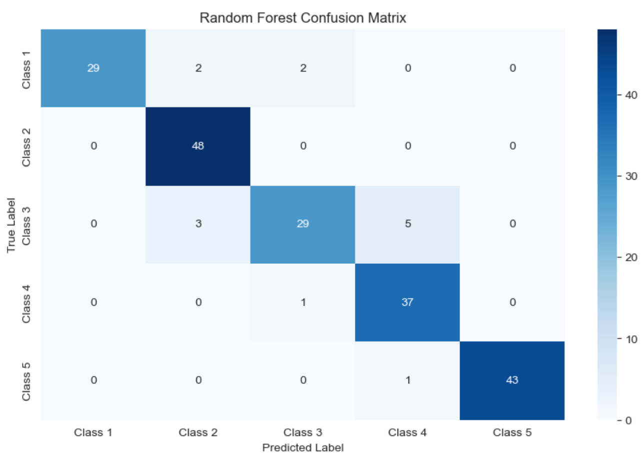

Figure 4.1.7: Random Forest confusion matrix Anxiety

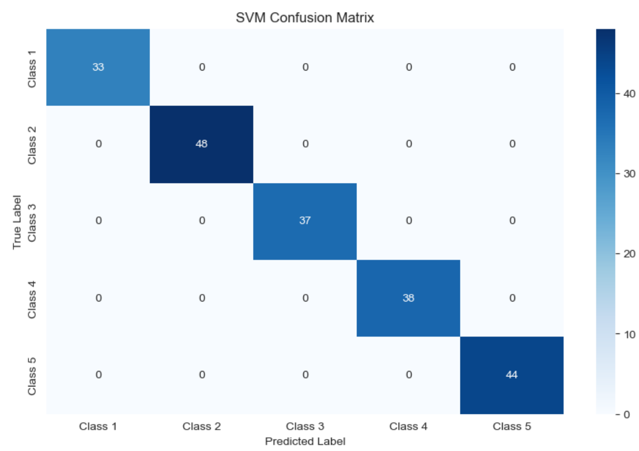

Figure 4.1.8: SVM Confusion Matrix Anxiety

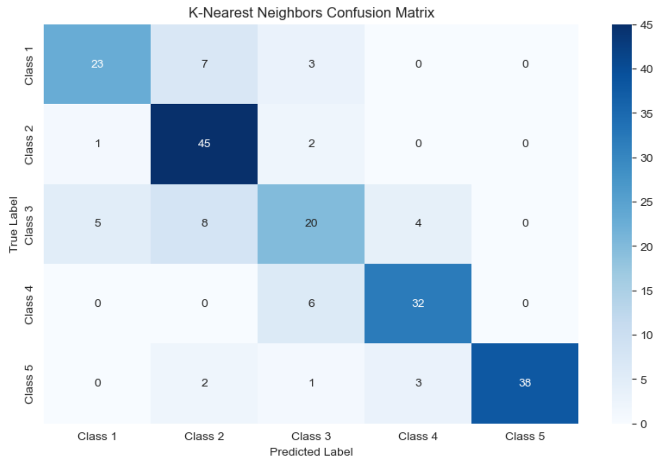

Figure 4.1.9: KNN confusion Matrix Anxiety

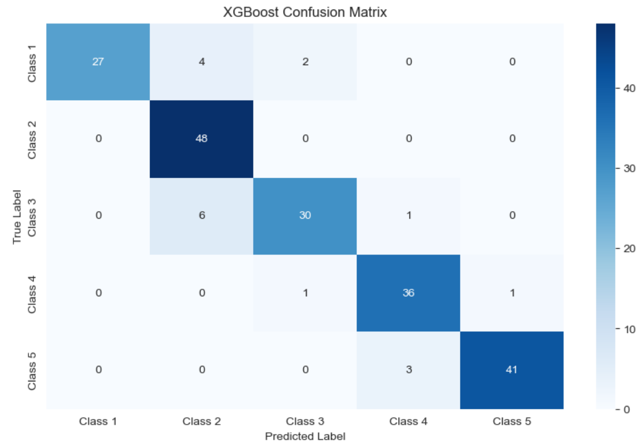

Figure 4.1.10: XGBoost Confusion Matrix Anxiety

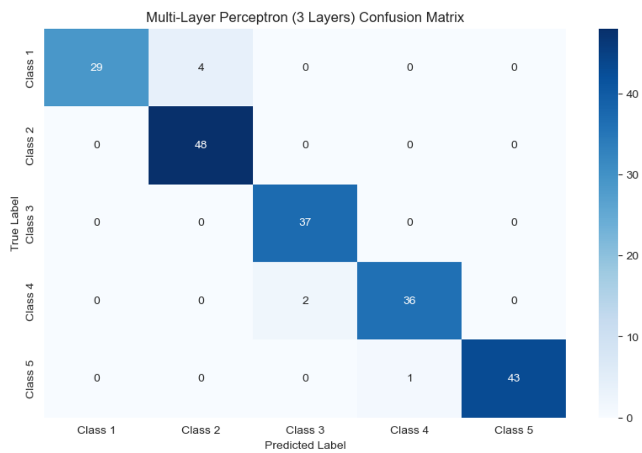

Figure 4.1.11: MLP confusion matrix Anxiety

Supplementary metrics assessing the model's performance, such as Precision, Recall, F-Score, and Support, were employed to provide additional insights into the models. The following tables (4.1.9, 4.1.10, 4.1.11, 4.1.12, 4.1.13) illustrate these performance measures. table (4.1.9) shows the results of Random Forest with average of precision 93%, Recall 92%and F-1 score 93%.

Table 4.1.9: Random Forest metrics Anxiety

Class	Precision	Recall	F1-Score
0	1.00	0.88	0.94
1	0.91	1.00	0.95
2	0.91	0.78	0.84
3	0.86	0.97	0.91
4	1.00	0.98	0.99
Macro Average	0.93	0.92	0.93
Weighted Average	0.93	0.93	0.93

In table (4.1.10) is shows the results of SVM, with an exceptional performance of 100% for Precision, Recall and F1-score.

Table 4.1.10: SVM Metrics Anxiety

Class	Precision	Recall	F1-Score
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
Macro Average	1.00	1.00	1.00
Weighted Average	1.00	1.00	1.00

While KNN as in Table (4.1.11) the results were 79% as an average Precision, 78% for average Recall and average F-1 score of 78%.

Table 4.1.11: KNN Metric Anxiety

Class	Precision	Recall	F1-Score
0	0.79	0.70	0.74
1	0.73	0.94	0.82
2	0.62	0.54	0.58
3	0.82	0.84	0.83
4	1.00	0.86	0.93
Macro Average	0.79	0.78	0.78
Weighted Average	0.80	0.79	0.79

In Table (4.1.12) is the results of XGBoost, with 92% average of Precision, 90% average of Recall and 91% average of F1-score.

Table 4.1.12: XGBoost Metric Anxiety

Class	**Precision**	**Recall**	**F1-Score**
0	1.00	0.82	0.90
1	0.83	1.00	0.91
2	0.91	0.81	0.86
3	0.90	0.95	0.92
4	0.98	0.93	0.95
Macro Average	0.92	0.90	0.91
Weighted Average	0.92	0.91	0.91

MLP metrics which is shown in table (4.1.13), it has 97% for average Precision, 96% for both average Recall and F1-score.

Table 4.1.13: MLP Metric Anxiety

Class	**Precision**	**Recall**	**F1-Score**
0	1.00	0.88	0.94
1	0.92	1.00	0.96
2	0.95	1.00	0.97
3	0.97	0.95	0.96
4	1.00	0.98	0.99
Macro Average	0.97	0.96	0.96
Weighted Average	0.97	0.96	0.96

ROC Curve achieves a superior balance between sensitivity and specificity. The AUC further measures this performance by measuring the area under the ROC Curve. A higher AUC implies a model's enhanced ability to distinguish between positive and negative instances. Therefore, the presented ROC Curve and AUC analysis provide an accurate view of the predictive power of each model, which helps in the identification of the most effective model during the prediction phase, below figure (4.1.12) shows the ROC for Anxiety models which shows that MLP and Random Forest have near results while MLP gained the best result.

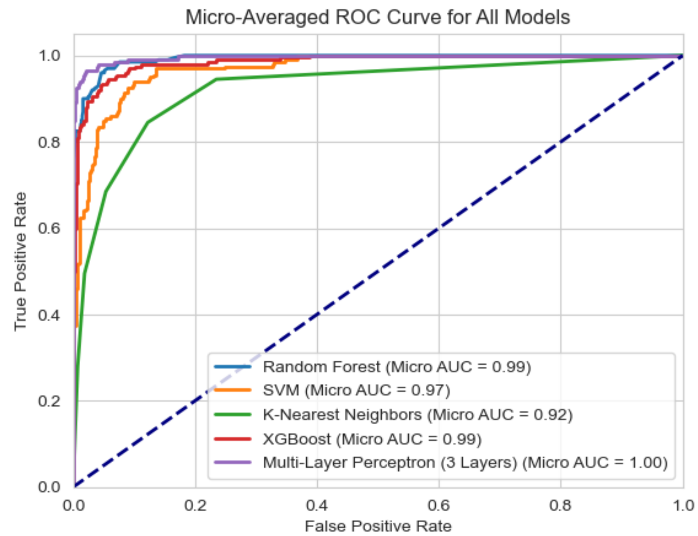

Figure 4.1.12: ROC Curve for Anxiety models

Stress:

The stress dataset was divided into an 80/20 split for training and testing, with 5-fold cross-validation. After hyperparameter tuning using grid search, the accuracy results became RF: 97%, SVM: 100%, KNN: 78%, XGBoost: 97%, and MLP: 96%, figures (4.1.13, 4.1.14, 4.1.15, 4.1.16, 4.1.17) show how each class gets predicted versus the actual results which give more insights. Figure (4.1.13) shows that Random Forest succeed to classify 193 cases out of 200, SVM could classify all of cases 200 out of 200 as shown in figure (4.1.14), KNN in figure (4.1.15) failed to classify 34 cases out of 200, XGBoost explained 194 and failed with 6 cases as figure (4.1.16) shows, and MLP could successfully classify 196 cases out of 200 as in figure (4.1.17).

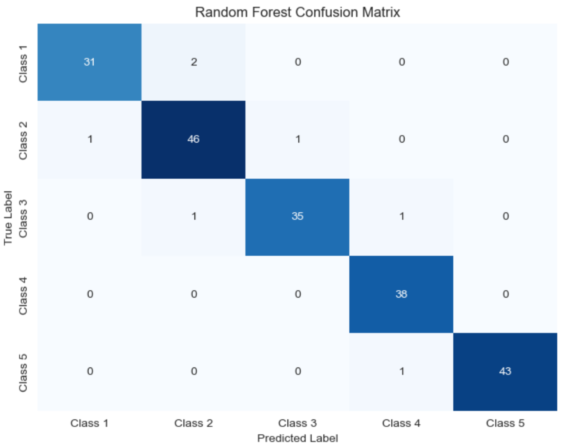

Figure 4.1.13: Random Forest confusion matrix Stress

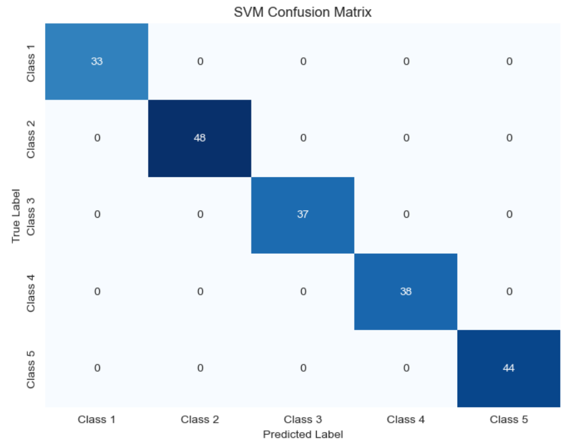

Figure 4.1.14: SVM confusion matrix Stress

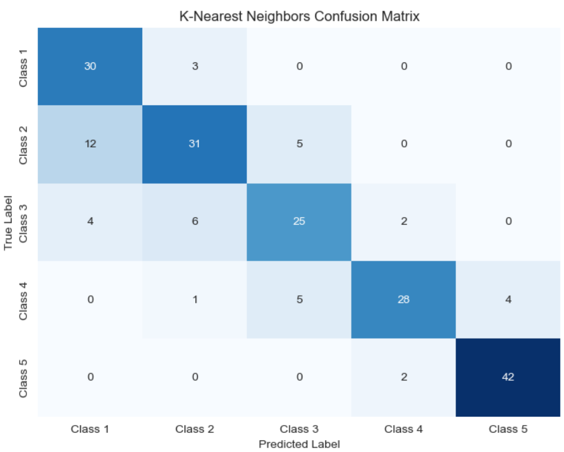

Figure 4.1.15: KNN confusion matrix Stress

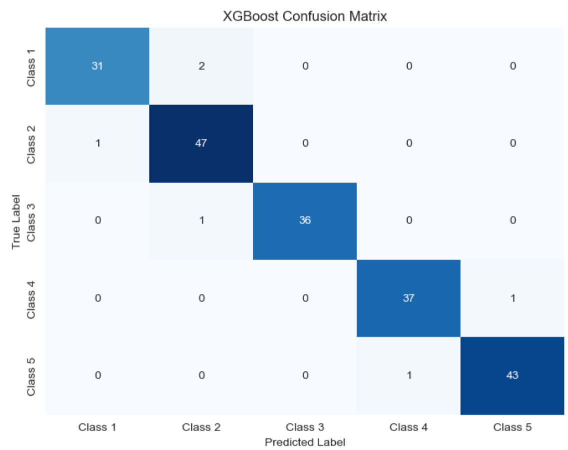

Figure 4.1.16: XGBoost confusion matrix Stress

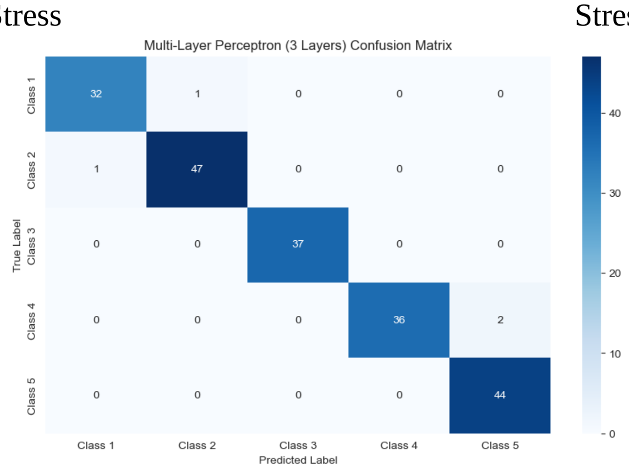

Figure 4.1.17: MLP confusion matrix Stress

Additional metrics evaluating the model's performance, including Precision, Recall, F-Score, and Support, were utilized to offer further insights into the models. The following tables (4.1.14, 4.1.15, 4.1.16, 4.1.17, 4.1.18) illustrate these performance measures. table (4.1.14) shows the results of Random Forest with average of precision 97%, Recall 96%and F-1 score 96%.

Table 4.1.14: Random Forest metrics Stress

Class	Precision	Recall	F1-Score
0	0.97	0.94	0.95
1	0.94	0.96	0.95
2	0.97	0.95	0.96
3	0.95	1.00	0.97
4	1.00	0.98	0.99
Macro Average	0.97	0.96	0.96
Weighted Average	0.97	0.96	0.97

In table (4.1.15) is shows the results of SVM, with an exceptional performance of 100% for Precision, Recall and F1-score.

Table 4.1.15: SVM Metrics Stress

Class	Precision	Recall	F1-Score
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
Macro Average	1.00	1.00	1.00
Weighted Average	1.00	1.00	1.00

While KNN as in Table (4.1.16) the results were 78% for average of Precision, Recall and F-1 score.

Table 4.1.16: KNN Metric Stress

Class	Precision	Recall	F1-Score
0	0.65	0.91	0.76
1	0.76	0.65	0.70
2	0.71	0.68	0.69
3	0.88	0.74	0.80
4	0.91	0.95	0.93
Macro Average	0.78	0.78	0.78
Weighted Average	0.78	0.78	0.78

In Table (4.1.17) is the results of XGBoost, with 97% for average of Precision, Recall and F1-score.

Table 4.1.17: XGBoost Metric Stress

Class	Precision	Recall	F1-Score
0	0.97	0.94	0.95
1	0.94	0.98	0.96
2	1.00	0.97	0.99
3	0.97	0.97	0.97
4	0.98	0.98	0.98
Macro Average	0.97	0.97	0.97
Weighted Average	0.97	0.97	0.97

MLP metrics which is shown in table (4.1.18), it has 98% for average Precision, Recall and F1-score.

Table 4.1.18: MLP Metric Stress

Class	Precision	Recall	F1-Score
0	0.97	0.97	0.97
1	0.98	0.98	0.98
2	1.00	1.00	1.00
3	1.00	0.95	0.97
4	0.96	1.00	0.98
Macro Average	0.98	0.98	0.98
Weighted Average	0.98	0.98	0.98

The ROC Curve shows balance sensitivity and specificity, while the AUC measures this performance by assessing the area under the curve. A higher AUC implies better discrimination between positive and negative instances. This combined ROC Curve and AUC analysis provide a significant evaluation of each model's predictive power. Figure (4.1.18) illustrates the ROC for Stress data models, which shows that MLP, Random Forest, and XGBoost are in almost the same area but MLP was the best.

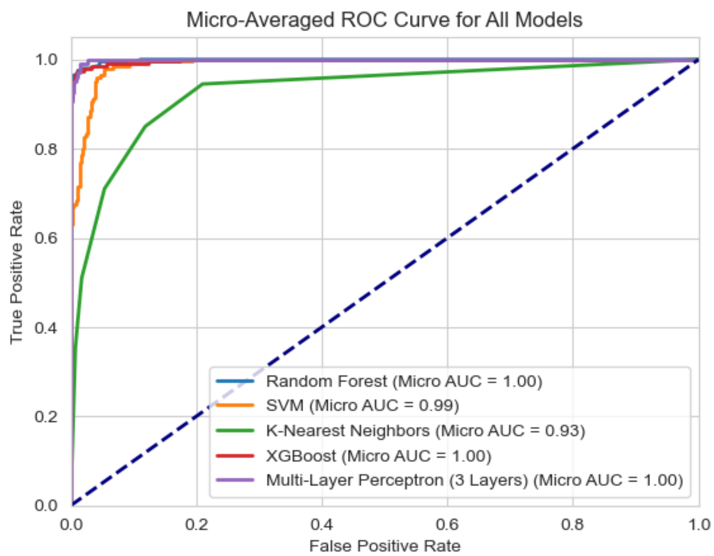

Figure 4.1.18: ROC Curve for Stress Data

4.2 DISCUSSION OF RESULTS AND COMPARISON

The dataset used in this thesis is split into 3 main sets, Depression, Anxiety, and Stress, each dataset contains the same number of features and instances, but each one represents responses to its question, and for each dataset, five machine learning algorithms were utilized to complete the experiment and investigate which algorithm gives the best result, when we started the experiment, we have started it was oblivious that there is an imbalance in the data, so we had to try the most common balancing techniques like resampling and SMOTE (Naseriparsa, 2013), to get the most accurate performance and results, in this section we will discuss the results from model point of view and how the data balance affected the performance using the accuracy, recall, precision, F-1 score and ROC techniques.

For Random Forest, the results of Random Forest for the 3 disorders with imbalanced data which can be found in the below table (4.2.1) show an overfitting as there are huge gaps between the training and testing accuracy:

Table 4.2.1: Random Forest Results before resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	84%	76%	79%

During the hypertuning of the parameters using the grid search we have tweaked the "n_estimators" which represent the number of decision trees that going to be used in the forest, and "max_depth" is the number of splits that each decision tree is allowed to make, the number of K-folds for cross-validation, the best values we have obtained from the grid search was {'max_depth': 20, 'n_estimators': 50} and applying the cross-validation during the training phase, this has increased the testing accuracy from 84% to 95% for Depression, 76% to 91% from Anxiety and from 79% to 97% for Stress as shown in table (4.2.2) which could be considered as huge improvement, we can say then that Random Forrest could learn well and also worked efficiently in the unseen data and hence the overfitting is no longer exists.

Besides the accuracy, The average precision for Random Forrest has improved also after handling the imbalance issue, for Depression, as shown in table (4.1.4) it increased from 82% to 95%, with average recall from 77% to 96% and average F1-score from 79% to 95%, for Anxiety, as shown in table (4.1.9), The average precision has increased from 70% to 92%, average recall also increased from 67% to 91% and average F-1 score changed from 67% to 91% and finally for Stress, the average precision as shown in figure (4.1.14) enhanced from

78% to 97%, the average recall increased from 75% to 98% and the average F1-score changed from 76% to 97%.

Table 4.2.2: Random Forest Results after resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	99%	100%
Testing Accuracy	95%	91%	97%

Regarding Support Vector Machines(SVM), tables (4.2.3,4.2.4) represents the results before and after the sampling, it is clear that SVM performance didn’t change in both cases, and it could deal with imbalance perfectly with no over fitting and with the best accuracy, and that’s is due to the nature of SVM where it tries to draw the hyper plane that divide the data between multiple classes and doesn’t have issues with the minor number of instances as it help the model to focus on the majority of the instances.

Table 4.2.3: SVM before resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	100%	100%	99%

We have done a grid search for SVM, to get the best performance by tweaking the parameters “C” which is a regularization parameter that controls the trade-off between achieving a low training error and a low testing error, and “Kernel” which is a crucial parameter that defines the type of decision boundary used by the classifier common types used are linear which implies that the data has the shape of linear data, polynomial, RBF, and sigmoid, and the best values were {'C': 10, 'kernel': 'linear'} as we can see how much the accuracy has increased for stress from 99% to 100%.

Since there is no change in the results before and after the data resampling to resolve the imbalance, and the classification accuracy is almost 100% for all cases this means that the model didn’t misclassify that much of cases, and hence, the precision, recall and the F1-score results that can be found in tables (4.1.5,4.1.10,4.1.15) didn’t change as well and also the same.

Table 4.2.4: SVM after resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	100%	100%	100%

After that K-Nearest Neighbors (KNN), instance-based learning, or lazy learning algorithm. Unlike traditional machine learning algorithms that involve a training phase where a model is trained on a dataset, KNN does not explicitly "learn" a model during a training phase. Instead, it memorizes the entire training dataset (Wang et al., n.d.), this model didn’t do well with this dataset it has been worst results among the other algorithms In all disorders, which could be due to the Curse of dimensionality since we have a big number of features which make the task harder for KNN as the distances between data become uniform which makes it harder for KNN to learn, the below tables (4.2.5,4.2.6) show how the accuracy changed before and after the data resampling, and these tables show a small improvement in the accuracy but still not competitor to the other models.

Table 4.2.5: KNN before resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	76%	75%	79%
Testing Accuracy	71%	61%	65%

The only parameter that can be tuned using the grid search in KNN is “n_neighbors” which represents the number of neighbors, and the best value in this case was {'n_neighbors': 3}.

Table 4.2.6: KNN after resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	91%	90%	90%
Testing Accuracy	77%	79%	78%

The fourth algorithm in this study was the XGBoost, which considered the powerful ensemble algorithms, the below tables (4.2.7,4.2.8) show the difference between the performance of the model before and after resampling to resolve the imbalance issue, for depression, the testing accuracy enhanced from 84% to 95%, while for Anxiety it improved from 76% to 91% and also for Stress, there was an improvement from 80% to 96% from these results shows that XGBoost improved after the resampling and the results was impressive as in some cases overfitting was there and big gap between the training accuracy and the testing accuracy was there, so that this was disappeared after the data resampling.

Table 4.2.7: XGBoost before resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	84%	76%	80%

XGBoost as an ensemble algorithm, which is based on the decision trees, the hypertuning parameters are the same as for Random forest, where we tuned the “max_depth” and the “n_estimators” and in this case, the best values for these parameters that gave the highest accuracy were {'max_depth': 3, 'n_estimators': 200}which is different than the results in Random forest as values due to the nature of XGBoost and the working mechanism.

For the other metrics that can also evaluate the model performance, tables (4.1.7,4.1.12,4.1.17) show how the improvement of the precision, recall, and F-score have improved after the data balance has been applied, for Depression the average precision improved from 81% to 95%, the average recall improved from 82% to 96% and the average F-1 score also improved from 82% to 95%, for Anxiety, the average of precision improved from 67% to 92%, the average recall improved from 68% to 90% and the average F-1 score also improved from 67% to 91% for Stress, the average of precision improved from 77% to 97%, the average recall improved from 75% to 97% and the average F-1 score also improved from 76% to 97%.

Table 4.2.8: XGBoost after resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	95%	91%	96%

The last algorithm is the Multi-Layer Perceptron(MLP) which is a neural network based model, MLP obtained stable results before and after the resampling the difference wasn’t significant, in some cases it did better before like in Depression it improved from 97% to 98% for Stress there was significant improve from 93% to 96% and in another case it did a little worse like in the Anxiety it decreased from 89% to 96%, which implies that MLP can handle the imbalance data, in general, it was one of the best models in both training and testing accuracy, there was no overfitting nor underfitting the results can be shown in table (4.2.9,4.2.10).

Table 4.2.9: MLP before resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	97%	98%	93%

In MLP we have used the model with 3 layers, Input layer, Hidden layer, and Output layer, a grid search has been done with the size of the hidden layer parameter “hidden_layer_sizes”, the activation function “activation” and the “max_iter” parameters, the best values obtained for these parameters were “Best Parameters: {'activation': 'relu', 'hidden_layer_sizes': (64,),’max_iter’:’1000’}.

Table 4.2.10: MLP after resampling

Accuracy/Disorder	Depression	Anxiety	Stress
Training Accuracy	100%	100%	100%
Testing Accuracy	98%	96%	96%

Additionally, if we check other metrics rather than accuracies like precision, recall, and F1-score, the results in tables (4.1.8,4.1.13,4.1.18) show that for Depression, the average precision, recall, and F1-score (98%,98%,98%) respectively, didn’t change which explains the stability of the model and ability to handle imbalanced data. For Anxiety the average precision before data balance was 97% and didn’t change after the resampling, the average recall was also 97% and has been changed to 96% which means a slight change also the average F1-score was 97% and has been changed to 96% which also implies slight change, and this slight change happens because of some misclassification in some classes compared to other classes.

The ROC curve is a common indicator that is being used to compare and measure the model's performance, and since we have a data set for each mental health disorder Depression, Anxiety, and Stress, the figures (4.1.6,4.1.12,4.1.18) that show the ROC curve for the models in each disorder respectively (Random Forest, SVM, KNN, XGBoost and MLP) before data resampling and the figures (4.2.1,4.2.2,4.2.3) explains the ROC curve before data resampling for the three disorders as well. For Depression, as shown in figure (4.2.2) before sampling we can see that the MLP has 100% AUC which means the model has achieved a true positive rate of 1 (sensitivity) and a false positive rate of 0, then Random Forest, XGBoost and SVM has the AUC of 97% and last one is KNN with 94%, but in figure (4.1.6) which represent the ROC after resampling, we can see that MLP still has the 100% AUC, then Random Forest has improved to 100%, SVM and XGBoost has improved to 99% and KNN didn’t improve.

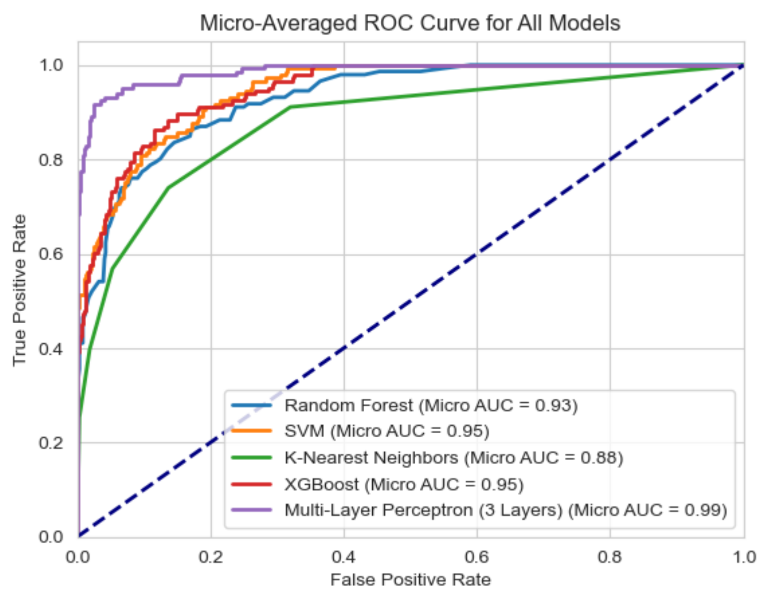

Figure 4.2.1: Anxiety ROC Curve before resampling

For Anxiety, as shown in figure (4.2.1) which shows the ROC curve for all models before resampling, the MLP had the highest AUC with 99%, then XGBoost and SVM achieved

95% after that the Random Forest gained 93% and the last one is KNN with 88%, if we compare these results to figure (4.1.12) which represent the ROC after resampling, we can see that MLP improved to 100%, Random Forest and XGBoost becomes 99% then SVM achieved 97% and lastly KNN with 92%.

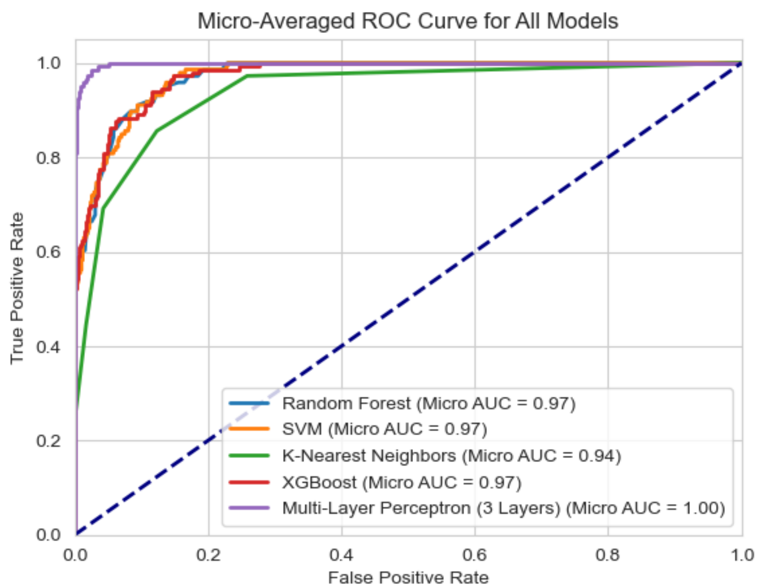

Figure 4.2.2: Depression ROC curve before resampling

The third disorder is Stress, as shown in Figure (4.2.3), before resampling, as in previous cases MLP was the best in terms of the area under the curve (AUC) with 99%, then SVM, Random Forest, and XGBoost with equal results of 95% and the least value for KNN with 91%, but after data resampling to balance the distribution of the classes among the data, figure (4.1.18) shows that MLP improved slightly to become 100% area under the curve, Random Forest also improved to become 100%, XGBoost also enhanced and resulted to 100% then SVM achieved 99% and last one was KNN with 93%.

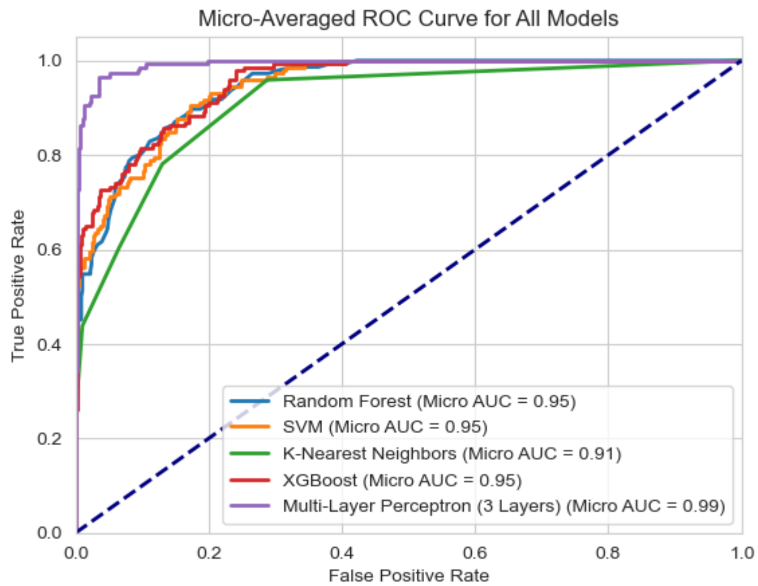

Figure 4.2.3: Stress ROC Curve before resampling

4.3 CHALLENGES AND LIMITATION

The main challenge in this study was the data collection, as mental health is sensitive topic and people usually hesitate to frankly talk about their mental issues, also the local mental health clinics don't have digital records for their patients and if they have, they are very careful to talk about it or share it, that's why it was not easy to have much of real infected cases in our data set, so in order to collect data from clinics is would take months to have hundreds of cases.

The other challenge is lack of studies in integration mental health issues with machine learning algorithms in the Arab world and especially in Palestine, which could be used to build on top of it and compare the results.

Chapter 5

Conclusion and Future Work

5.1 CONCLUSION

In recent years, the evolution of Artificial Intelligence and machine learning has significantly enhanced the healthcare practices, which helps in automation of mental health diagnostics. This study is about the application of machine learning algorithms for the early detection of mental health disorders, specifically Depression, Anxiety, and Stress. The importance of employing machine learning in mental health becomes evident as it facilitates personalized diagnoses, enabling timely interventions for improved patient outcomes. Through the utilization of sophisticated algorithms, this research aims to propose a new way of mental health assessments, emphasizing the need for accurate predictions and classifications, also collecting local mental health disorders (Depression, Anxiety and Stress) dataset in Palestine which was used with machine learning algorithms, considered the first initiative in Palestine. The experiment of different machine learning models, including Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), XGBoost, and Multi-Layer Perceptron (MLP), prove their effectiveness in analyzing datasets and achieving early detection of mental health issues. The initial imbalance in the data makes it important to implement resampling techniques, such as SMOTE, and resampling to enhance model performance. The comparative analysis of these models proved their adaptability and responsiveness to the complexities of mental health data.

Each algorithm's performance metrics, including accuracy, precision, recall, and F1-score, were evaluated and compared. It was obvious that SVM demonstrated the highest and exceptional performance with almost 100% accuracy in all disorders before and after resampling, and effectively handling imbalanced data with no overfitting. XGBoost and Random Forest, as an ensemble algorithm, showed a significant improvement in accuracy after resampling with (RF:95%, 91%, 97%) (XGBoost: 95%, 91%, 96%) for Depression, Anxiety, and Stress respectively, aligning training and testing accuracy and showcasing its robustness in handling mental health data analysis. Both algorithms initially exhibited signs of overfitting with a significant gap between training and testing accuracy. However, after hyperparameter tuning and resampling, its accuracy improved. MLP, a neural network-based model, demonstrated stability and effectiveness before and after resampling with slight improvement after resampling with no overfitting, with accuracy scores (98%, 96%, 96%) for Depression, Anxiety, and Stress respectively. KNN showed the worst improvements in accuracy after and before resampling with accuracy scores (77%, 79%, 78%) for Depression, Anxiety, and Stress respectively.

In conclusion, the study conducted an in-depth analysis of machine learning algorithms applied to datasets related to Depression, Anxiety, and Stress. The results are promising and showed high accuracy which could be generalized and used as a tool for mental health early detection for self-use.

5.2 RECOMMENDATIONS

The results show an exceptional performance of Support Vector Machines (SVM), with almost 100% accuracy in all disorders both before and after resampling, and its ability to handle imbalanced data without overfitting showcases its robustness, it is recommended to consider the using SVM in mental health diagnostics frameworks and Building an interactive web and mobile application that call this model which helps individuals to use the machine learning models to help them in self-diagnoses or enabling the therapists from using these application which helps them in their diagnoses process.

On the contrary, the study revealed that K-Nearest Neighbors (KNN) result was the least improvement in accuracy, both before and after resampling. It is advised to avoid considering KNN for mental health diagnostics and explore alternative algorithms that could resulted in better performance on imbalanced datasets. While machine learning models show promising results, collaboration with mental health professionals is recommended, to have input from their experience to make sure these results are aligned with the actual clinical perspective and results. Their integration into mental health practices should be done in collaboration with mental health professionals.

5.3 FUTURE WORK

This research considered the first Palestinian study in the field of using machine learning in early detection of Mental health disorders especially Depression, Anxiety and Stress, using local dataset, even though, there is still ways for enhancement and build on top of this, the first thing that can be done is enhanced data collection, where it could be expanded to include more diversity and wider range of demographic and cultural background instance, which can lead to more robustness in generalization of these models, also more finetuning for the used machine learning models or exploring more models which will lead to better accuracy or new trends in the data itself, the other thing that could be done, is more collaboration with mental health experts in order to have better version of the data and corporation in the result analysis and interpretation which could be lead to more insights that experts can judge .

References

Ahmed, O., Faisal, R. A., Alim, S. M. A. H. M., Sharker, T., & Hiramoni, F. A. (2022). The psychometric properties of the Depression Anxiety Stress Scale-21 (DASS-21) Bangla version. Acta Psychologica, 223. https://doi.org/10.1016/j.actpsy.2022.103509

Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies, 9(3). https://doi.org/10.3390/technologies9030052

Ali, A. M., Alkhamees, A. A., Hori, H., Kim, Y., & Kunugi, H. (2021). The depression anxiety stress scale 21: Development and validation of the depression anxiety stress scale 8-item in psychiatric patients and the general public for easier mental health measurement in a post COVID-19 world. International Journal of Environmental Research and Public Health, 18(19). https://doi.org/10.3390/ijerph181910142

Al-Wesabi, F. N., Alsolai, H., Hilal, A. M., Hamza, M. A., Duhayyim, M. Al, & Negm, N. (2022). Machine learning based depression, anxiety, and stress predictive model during COVID-19 crisis. Computers, Materials and Continua, 70(3), 5803–5820. https://doi.org/10.32604/cmc.2022.021195

Arokkiya Mary, S. T., & Jabasheela, L. (n.d.). ANT COLONY OPTIMIZATION BASED FEATURE SELECTION AND DATA CLASSIFICATION FOR DEPRESSION ANXIETY AND STRESS (Vol. 9). www.iioab.org

Auria, L., Berlin, R. A. M., & Moro, R. A. (2008). Support Vector Machines (SVM) as a Technique for Solvency Analysis. DIW Berlin German Institute for Economic Research. https://dx.doi.org/10.2139/ssrn.1424949

Basha, E., & Kaya, M. (2016a). Depression, Anxiety and Stress Scale (DASS): The Study of Validity and Reliability. Universal Journal of Educational Research, 4(12), 2701–2705. https://doi.org/10.13189/ujer.2016.041202

Basha, E., & Kaya, M. (2016b). Depression, Anxiety and Stress Scale (DASS): The Study of Validity and Reliability. Universal Journal of Educational Research, 4(12), 2701–2705. https://doi.org/10.13189/ujer.2016.041202

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785

Choudhury, A. A., Khan, M. R. H., Nahim, N. Z., Tulon, S. R., Islam, S., & Chakrabarty, A. (2019). Predicting Depression in Bangladeshi Undergraduates using Machine Learning. Proceedings of 2019 IEEE Region 10 Symposium, TENSYMP 2019, 789–794. https://doi.org/10.1109/TENSYMP46218.2019.8971369

Das, A., Gordon-Ocejo, G., Kumar, M., Kumar, N., & Needlman, R. (2021). Association of the previous history of maternal depression with post-partum depression, anxiety, and stress in the neonatal intensive care unit. Journal of Maternal-Fetal and Neonatal Medicine, 34(11), 1741–1746. https://doi.org/10.1080/14767058.2019.1647162

Delashmit, W. H., & Manry, M. T. (n.d.). Recent Developments in Multilayer Perceptron Neural Networks.

Edla, D. R., Mangalorekar, K., Dhavalikar, G., & Dodia, S. (2018). Classification of EEG data for human mental state analysis using Random Forest Classifier. Procedia Computer Science, 132, 1523–1532. https://doi.org/10.1016/j.procs.2018.05.116

Fatima, A., Li, Y., Hills, T. T., & Stella, M. (2021). Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. Big Data and Cognitive Computing, 5(4). https://doi.org/10.3390/bdcc5040077

Ghorpade-Aher, J., Memon, A., Chugh, S., Chebolu, A., Chaudhari, P., & Chavan, J. (2023a). DASS-21 Based Psychometric Prediction Using Advanced Machine Learning Techniques. Journal of Advances in Information Technology, 14(3), 571–580. https://doi.org/10.12720/jait.14.3.571-580

Ghorpade-Aher, J., Memon, A., Chugh, S., Chebolu, A., Chaudhari, P., & Chavan, J. (2023b). DASS-21 Based Psychometric Prediction Using Advanced Machine Learning Techniques. Journal of Advances in Information Technology, 14(3), 571–580. https://doi.org/10.12720/jait.14.3.571-580

Hegelich, S. (2016). Decision Trees and Random Forests: Machine Learning Techniques to Classify Rare Events. European Policy Analysis, 2(1), 98–120. https://doi.org/10.18278/epa.2.1.7

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. IEEE Access, 10, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048

Huang, S., Nianguang, C. A. I., Penzuti Pacheco, P., Narandes, S., Wang, Y., & Wayne, X. U. (2018). Applications of support vector machine (SVM) learning in cancer genomics. In Cancer Genomics and Proteomics, 15(1), 41–51. International Institute of Anticancer Research. https://doi.org/10.21873/cgp.20063

Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. Human Resource Management Review, 27(2), 265–276. https://doi.org/10.1016/j.hrmr.2016.08.003

Kessler, R. C., Birnbaum, H. G., Shahly, V., Bromet, E., Hwang, I., McLaughlin, K. A., Sampson, N., Andrade, L. H., De Girolamo, G., Demyttenaere, K., Haro, J. M., Karam, A. N., Kostyuchenko, S., Kovess, V., Lara, C., Levinson, D., Matschinger, H., Nakane, Y., Browne, M. O., … Stein, D. J. (2010). Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: Results from the WHO world mental health survey initiative. Depression and Anxiety, 27(4), 351–364. https://doi.org/10.1002/da.20634

Klettke, M., & Störl, U. (2022). Four Generations in Data Engineering for Data Science. Datenbank-Spektrum, 22(1), 59–66. https://doi.org/10.1007/s13222-021-00399-3

Kulkarni, V. Y. (n.d.). Random Forest Classifiers: A Survey and Future Research Directions. In International Journal of Advanced Computing, 36(1), 1144-1153. https://adiwijaya.staff.telkomuniversity.ac.id/files/2014/02/Random-Forest-Classifiers_A-Survey-and-Future.pdf

Kumar, P., Garg, S., & Garg, A. (2020). Assessment of Anxiety, Depression and Stress using Machine Learning Models. Procedia Computer Science, 171, 1989–1998. https://doi.org/10.1016/j.procs.2020.04.213

Likforman-Sulem, L., Esposito, A., Faundez-Zanuy, M., Clemencon, S., & Cordasco, G. (2017). EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing. IEEE Transactions on Human-Machine Systems, 47(2), 273–284. https://doi.org/10.1109/THMS.2016.2635441

Madison Schott. (2019). Random Forest Algorithm for Machine Learning. Part 4 of a Series on Introductory Machine Learning Algorithms.

Mahmud, S., Mohsin, M., Muyeed, A., Nazneen, S., Abu Sayed, M., Murshed, N., Tahrin Tonmon, T., & Islam, A. (2023). Medicine ® Machine learning approaches for predicting suicidal behaviors among university students in Bangladesh during the COVID-19 pandemic A cross-sectional study. https://doi.org/10.21203/rs.3.rs-2069873/v1

Marie, M., Hannigan, B., & Jones, A. (2016). Mental health needs and services in the West Bank, Palestine. International Journal of Mental Health Systems, 10(1), 1–8. https://doi.org/10.1186/s13033-016-0056-8

Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. ACM International Conference Proceeding Series, 9(4), 412–419. https://doi.org/10.1145/3453892.3461323

Mavroforakis, M. E., & Theodoridis, S. (2006). A geometric approach to support vector machine (SVM) classification. IEEE Transactions on Neural Networks, 17(3), 671–682. https://doi.org/10.1109/TNN.2006.873281

Miriam Taouk Moussa, P. L. and R. L. (2001). Taouk Moussa et al. Arabic DASS translation. https://Www2.Psy.Unsw.Edu.Au/Groups/Dass/Arabic/Arabic.Htm.

Muschelli, J. (2020). ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. Journal of Classification, 37(3), 696–708. https://doi.org/10.1007/s00357-019-09345-1

Nada, Q., Herdiana, I., & Andriani, F. (2022). Testing the validity and reliability of the Depression Anxiety Stress Scale (DASS)-21 instrument for individuals with Psychodermatology. Psikohumaniora, 7(2), 153–168. https://doi.org/10.21580/pjpp.v7i2.11802

Naseriparsa, M. (2013). Combination of PCA with SMOTE Resampling to Boost the Prediction Rate in Lung Cancer Dataset. In International Journal of Computer Applications, 77(3), 33-38, https://arxiv.org/pdf/1403.1949

openpsychometrics.org. (2020). DASS Dataset. https://openpsychometrics.org/tests/DASS

Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGboost versus logistic regression. Risks, 7(2). https://doi.org/10.3390/risks7020070

Popescu, M.-C., & Balas, V. E. (n.d.). Multilayer Perceptron and Neural Networks.

Preethi, P., & Mahalakshmi, R. (2023). EasyChair Preprint The Impact of Different Dispositions and Gender on Depression, Anxiety, and Stress-DASS The impact of different dispositions and gender on depression, anxiety, and stress-DASS.

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. https://doi.org/10.1016/j.procs.2020.03.442

Priya, K. B., Anilkumar, A., Alexander, S. E. E., Dinesh, T., & Akash, S. (2022). Psychological Assessment using Modified DASS-21 Scale during COVID-19 Pandemic. Journal of Pharmaceutical Research International, 34(47A), 16–26. https://doi.org/10.9734/jpri/2022/v34i47a36394

Ruck, D. W., Rogers, S. K., & Kabrisky, M. (1990). Feature Selection Using a Multilayer Perceptron. In Journal of Neural Network Computing, 2(2), 40-48, https://www.researchgate.net/publication/2430790_Feature_Selection_Using_a_Multilayer_Perceptron

Samele, C., Lees-Manning, H., Zamperoni, V., Goldie, I., Thorpe, L., Wooster, E., Giugliano, T., O'sullivan, C., Kousoulis, A., Breedvelt, J., Rowland, M., & Grange, R. (2018). Stress: Are we coping? London: Mental Health Foundation. In Mental Health Foundation.

Sau, A., & Bhakta, I. (2017). Predicting anxiety and depression in elderly patients using machine learning technology. Healthcare Technology Letters, 4(6), 238–243. https://doi.org/10.1049/htl.2016.0096

Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among the seafarers using machine learning technology. Informatics in Medicine Unlocked, 16. https://doi.org/10.1016/j.imu.2018.12.004

scikit-learn.org. (2023a). Data Resampling Sklearn. Https://Scikit-Learn.Org/Stable/Modules/Generated/Sklearn.Utils.Resample.Html.

scikit-learn.org. (2023b). Label Encoder Sklearn. Https://Scikit-Learn.Org/Stable/Modules/Generated/Sklearn.Preprocessing.LabelEncoder.Html#sklearn.Preprocessing.LabelEncoder.

scikit-learn.org. (2023c). Standard Scaler Sklearn. Https://Scikit-Learn.Org/Stable/Modules/Generated/Sklearn.Preprocessing.StandardScaler.Html.

Shafiq, F., Haider, S. I., & Ijaz, S. (2020). Anxiety, depression, stress, and decision-making among orphans and non-orphans in Pakistan. Psychology Research and Behavior Management, 13, 313–318. https://doi.org/10.2147/PRBM.S245154

Sofia, Malik, A., Shabaz, M., & Asenso, E. (2023). Machine learning based model for detecting depression during Covid-19 crisis. Scientific African, 20. https://doi.org/10.1016/j.sciaf.2023.e01716

Srinivasulu, U., Vivek, A., Tech, T. B., Dharun, A., & Tech, B. (n.d.). Machine Learning Techniques for Stress Prediction in Working Employees.

Sun, Y. H., Luo, H., & Lee, K. (2022). A Novel Approach for Developing Efficient and Convenient Short Assessments to Approximate a Long Assessment. Behavior Research Methods, 54(6), 2802–2828. https://doi.org/10.3758/s13428-021-01771-7

Tejveer Singh. (2022). Prediction of Depression, Stress and Anxiety. International Journal of Research in Engineering, Science and Management. https://journal.ijresm.com/index.php/ijresm/article/view/2003/1942

Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Cheng, S., Delling, F. N., Elkind, M. S. V., Evenson, K. R., Ferguson, J. F., Gupta, D. K., Khan, S. S., Kissela, B. M., Knutson, K. L., Lee, C. D., Lewis, T. T., … Tsao, C. W. (2021). Heart Disease and Stroke Statistics - 2021 Update: A Report From the American Heart Association. In Circulation, 143(8), 254–743, https://doi.org/10.1161/CIR.0000000000000950

Wang, J., Zucker, J.-D., & Fr, J.-D. Z. (n.d.). Solving the Multiple-Instance Problem: A Lazy Learning Approach.

WHO. (2017). Depression and Other Common Mental Disorders Global Health Estimates.

Zarkasi, K. A., Danker, T. J., Julani, I. M., Ilya, N., Rasid, M., Kumaran, J., Panirselvan, P., Ramli, N. Z., Hijaz, M., & Sani, M. (n.d.). Levels of Depression, Anxiety, and Stress Measured by DASS-21 and their Associated Factors in a Rural Village of Ranau District, Sabah, Malaysia. http://www.openepi.com

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for kNN Classification. ACM Transactions on Intelligent Systems and Technology, 8(3). https://doi.org/10.1145/2990508

Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN classification with different numbers of nearest neighbors. IEEE Transactions on Neural Networks and Learning Systems, 29(5), 1774–1785. https://doi.org/10.1109/TNNLS.2017.2673241

الملخص

في خضم ثورة الذكاء الصناعي وعلوم البيانات، أصبح من الممكن أكثر مساعدة أخصائي الصحة في إعادة تعريف معنى الصحة النفسية بطريقة أكثر إيجابية مما كانت عليه في السابق، مما يعزز امكانية تشخيص الأمراض النفسية في مراحل مبكرة لتكون العلاجات أكثر فعالية. يمكن تحقيق ذلك من خلال تخصيص التشخيص والعلاج بناءً على خصائص حالات الأفراد. الصحة النفسية هي الرفاهية النفسية الشاملة للفرد، بما في ذلك العوامل العاطفية والنفسية والاجتماعية للفرد التي تؤثر في كيفية تفكيره وشعوره وسلوكه. التوازن والاستقرار في الصحة النفسية يمكنان الشخص من التعامل مع ضغوط الحياة الطبيعية بفعالية والعمل بكفاءة والمساهمة بشكل إيجابي في مجتمعه. نظرًا لأن الاكتئاب والتوتر والقلق يعتبرون الاضطرابات النفسية الأكثر شيوعًا، تهدف هذه البحث إلى استخدام خوارزميات تعلم الآلة لتوقع تشخيص التوتر والقلق والاكتئاب باستخدام مجموعة البيانات التي تم جمعها كجزء من هذا العمل. تتألف مجموعة البيانات من حوالي 700 سجل باستخدام استطلاع عبر الإنترنت، والذي كان على أساس مقياس الاكتئاب والقلق والتوتر الدولي (DASS21). تم جمع البيانات من مشاركين فلسطينيين من المجتمع وطلاب الجامعات. ولضمان فعالية تطبيق خوارزميات الذكاء الصناعي، تم العمل على معالجة البيانات قبل تحليلها حيث تم تنقيحها والبحث عن بيانات مكررة، وحذف بعض العناصر الغير لازمة مثل اليد المستخدمة في الكتابة والدين، وتم تحويل النصوص الى ارقام لسهولة استخدامها مع الخوارزميات، ونظرا لعدم توازن البيانات حيث ان بعض التصنيفات تحتوي على عدد اكبر بكثير من غيرها تم عمل إعادة التوازن للحصول على نتائج ادق وافضل.

تم استخدام خمس خوارزميات مختلفة لتحليل البيانات وتحقيق الكشف المبكر عن اضطرابات الصحة النفسية: Random Forest، SVM، KNN، نموذج XGBoost ونموذج Multi-Layer

Perceptron(3layers). كانت النتائج بالنسبة للاكتئاب SVM الأفضل من حيث دقة النموذج بنسبة 100٪، ثم MLP بنسبة 98٪، بعد ذلك XGBoost بنسبة 95٪، ثم Random Forest بنسبة 93٪، وحصل KNNعلى اقل نتيجة بنسبة 77٪. أما بالنسبة لنتائج القلق، حققت SVM النقطة الأعلى بنسبة 100٪، يليه MLP بنسبة ٪96 ، كل من Random Forest و XGBoost حققا نتيجة ٪95 وحصل KNN على٪79 ، كما كانت نتائج التوتر 100٪ لـSVM ، يليه ال MLP بنسبة 97% وحصل ال Random Forest and XGBoost على نتيجة متماثلة بنسبة 96% وحصل ال KNN على اقل نتيجة بنسبة 78%