Identification of Mental Disorder based on Changes in Personal Behaviour using Machine Learning Algorithms.

MSc Research Project
Data Analytics Sep 2020-2021

Harshita Singh
Student ID: x19196725

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Harshita Singh
Student ID:	x19196725
Programme:	Msc in Data Analytics
Year:	2020-2021
Module:	MSc Research Project
Supervisor:	DR. Rashmi Gupta
Submission Due Date:	16-Aug-2021
Project Title:	Identification of Mental Disorder based on Changes in Personal Behaviour using Machine Learning Algorithms.
Word Count:	8724
Page Count:	26

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Harshita Singh
Date:	16th August 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Identification of Mental Disorder based on Changes in Personal Behaviour using Machine Learning Algorithms.

Harshita Singh
x19196725@student.ncirl.ie

16th August 2021

Abstract

People are having behavioral issues and mental disorders as a result of increased strain and stress in their daily lives. Anxiety, Depression, Stress, Schizophrenia, Bipolar Disorder, and many more types of Mental Disorders exist. There are various types of physical and emotional symptoms in mental disorder. This research will identify mental illness based on the occurrences and feelings that a person is experiencing. Panic attacks, sweating, palpitations, sorrow, anxiety, overthinking, hallucinations, and illusions are all signs of mental disease, and each symptom reveals something about the kind of mental illness. The (Feed Forward Neural Network)FFNN, XGBoost, Support Vector Machine, Logistic Regression, and Decision Tree are five machine learning algorithms used in this study. We used an additional tree classifier as a feature selection approach in this study, along with other preprocessing procedures. Following the feature selection approach, a machine learning algorithm was used to diagnose a mental illness based on the person's symptoms. The effectiveness of machine learning models was assessed using the Recall, Accuracy, Precision, and F1-score parameters. Logistic regression produced the greatest results, with an accuracy of 79.9%, while the FFNN model had the lowest accuracy of 40.9%.

1 Introduction

In today's world, mental illness has wreaked havoc on society and has risen to the fore as a major issue. People suffering from anxiety, depression, schizophrenia, bipolar disorder, panic disorder, and other mental disorders are mainly ignoring the reality that mental disorders are the world's most severe problem. [Silvana et al., 2018]Because mental illnesses are significant issues, they may have an impact on the environment and people's stability and security. It has a negative influence and effect on a person's life since it causes disruption in their behavior, thoughts, and emotions. Mental illness causes a wide range of changes and negative consequences in a person's life, including a loss of faith in himself and his family.

Mental illness has societal consequences as well as a global economic cost. [Silva et al., 2019]For the diagnosis of mental illness, no biological tests are performed or accessible.

Biological tests are not conducted or available for the diagnosis of mental disease. The expert's judgment, based on the numerous symptoms, is used to make the diagnosis. Based on the person's bodily symptoms, there might be a clear correlation and influence.

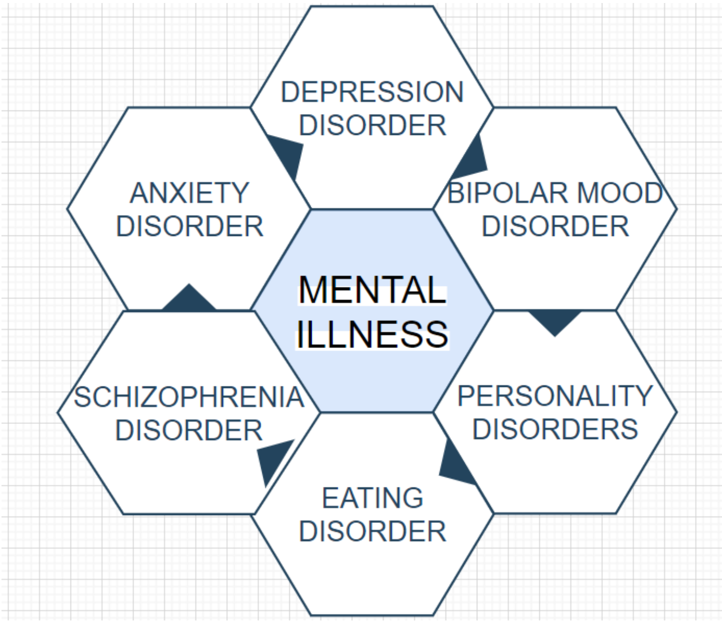

Figure 1: Types of Mental Disorders

By confirming the symptoms of the patients, the study fundamentally identifies the categories of disease, such as anxiety, depression, schizophrenia, bipolar disorder, stress, and panic disorder. Mental health and a healthier state of mind are critical to one's overall well-being. According to WHO estimates, 800,000 individuals attempt suicide each year, with a significantly greater number of people committing themselvesGore and Rathi [2019]. The bulk of suicide situations are caused by fairly common mental disorders Gore and Rathi [2019]. One of the main causes of mortality among teenagers is suicide. The gathering of data in the field of mental health treatment has risen dramatically in recent years.

As a result of the depressed state, hearing may be relatively poor. In general, people react in odd ways. They might include signs like rage over insignificant things, overeating or undereating, pondering, sleep problems, and even suicide thoughts. Psychologists suggest antidepressants, anxiolytics, and other medications. Patients can also share their difficulties with their doctor in a single personal interaction.

Nowadays, mental illness is on the rise. Everyone else suffers from some form of mental illness. However, some people throughout the world have not been cured. In high-income countries, only 35–50% of individuals in need receive basic therapy. Low and middle-income nations have an alarming gap of 76-85% untreated people. Because diagnostic therapy is limited and the problem is not first and foremost diagnosed, this imbalance varies. Only a few psychiatrists are accessible in Chad, Liberia and Rwanda, for example.

In this research, the main objective is to evaluate the state-of-the-art methodologies for mental illness classification. We have used five machine learning technique selected according to the previous researches which are FFNN, XG Boost, Logistic Regression, SVM and Decision tree is used among which FFNN and XGBoost are novel approaches in the state-of-the-art method [Kavitha et al., 2019]. A state-of-the-art technique is used to assess the type of psychological problem a person is experiencing based on the symptoms of the person. This research aimed at applying the Machine learning models using feature selection technology to the information set for mental disease and testing the model with different performance measurements. The findings exceeded the techniques described in the article and examined the success of state-of-the-art approaches. It is still necessary to establish which approaches are more effective in the identification of mental illness and which algorithm provides superior results. The difficulty in classifying various symptoms such as stress, breathing problems, overthinking, lack of interest in today's events, lifestyle, sweating, anger, and so on with various types of mental disorders such as depression, anxiety, schizophrenia, bipolar disorder, loneliness, and so on as distinct disorders with different symptoms depending on the person is the major challenge. As a result, we use a supervised machine learning method called multiple instance learning to identify mental disorders based on the person's symptoms.

We present our contributions as follows: (i) pre-process the supervised mental health data; (ii) Implementation and design of ML algorithms; (iii) implementation of algorithms and proposed new ways to provide a valid study comparison.

In section 2 we describe the prior study work done on this subject and explain the methodology used in section 3. We explain the architecture of the study model in Section 4. Section 5 lays forth the state-of-the-art and new method for identification of mental disorders by the person's symptoms and we give the assessment of the models in section 6. This research is concluded in Section 7.

2 Related Work

The most pressing issue in today's world is mental illness. We will look at some of the datasets and approaches that have been utilized in various sorts of mental health in this research review. The journals provide a brief overview of the approach, field, and assessment process for forecasting and analyzing the outcome. The review is divided into categories based on the technique used.

2.1 Machine Learning models used for Analysis of Mental Disorder

Social networks usage to obtain different information in decision-making and automation, such as assisting in software development activities, the usage of crypto-currencies, and the identification of network communities, is discussed in general terms in this paper. The authors describe how to utilize social media and big data analytics to identify individuals with mental diseases including depression, schizophrenia, eating disorders, anxiety, or addictive behaviors in their paper [Tariq et al., 2019]. The use of social media and big data analytics to anticipate patients' mental illnesses has become a popular issue. It also emphasizes that typical techniques of identifying a patient with a mental illness or condition require either a significant quantity of historical data or constant monitoring of patient behaviors. The author has developed a technique for categorizing individuals with persistent mental illness illnesses to solve this issue (such as anxiety, depression, bipolar disorder, and attention deficit hyperactivity disorder). The suggested method employs the Co-training methodology, which combines the discriminative power of frequently used classifiers including Random Forest, Support Vector Machine, and Nave Bayes. The Reddit API is used to extract posts and the top five associated comments in order to create a feature space. In line with every state-of-the-art approach, the experimental findings demonstrate that Co-training-based classification beats state-of-the-art classifiers by an average margin of 3%.

The journal [Elmunsyah et al., 2019] highlights how psychological well-being concerns are becoming increasingly important in the workplace. These difficulties have an impact on worker efficiency and, as a result, on the productivity of the business. Organizations should understand elements associated with representatives' emotional well being in order to reduce psychological well-being concerns among them. As a result, an order approach to determine whether or not a representative demands psychological well-being therapy is very actually required. The phases below have been completed:

- Collecting data from the Open Sourcing Mental Illness database (OSMI).
- Pre-processing of data measurements (information cleaning, highlight choice, information change).
- Executing the KNN computation while sorting the data.
- Using a disordered grid to generate exactness, review, and grouping using the KNN computation, the exactness, accuracy, and review were 87.27%, 84.21%, and 66.7%, respectively.

In summary, psychological well-being information for categorization of KNN with a substantial degree of accuracy is important.

The research [Laijawala et al., 2020] illustrates how a person's emotional, psychological, and social health indicate their emotional goodness. The reaction of the individual is influenced by thoughts and feelings. Mental health is vital throughout all phases of life from birth to maturity. Stress, social sadness, employment problems, fear, drug addiction and obsessive compulsive disorder are only few variables that contribute to psychiatric diseases. In this work, the author has encoded the data in order to enhance prediction. The data are submitted to a number of machine learning algorithms in order to obtain labels. The accuracy of the algorithm is evaluated prior to construction of the model. In this study, classification methods like Decision Tree, Random Forest and Nave Bayes were utilized. The target audience is those aged 18 or over who belong to the working class. Following the model creation based on the user's information, it was incorporated and utilized to forecast results in the websites. The article finds that out of 315 data instances, 258 data instance properly categorized, the decision tree obtained an accuracy of 82%. Whenever a person answers the inquiry at these sites, a probability of the status of their mental health will be provided to him or her and a recommendation will also be issued based on the condition of the person. The article also indicates that the proper output is presented and that the probability of a disease being rejected is minimal.

The research [Saleque et al., 2020] discusses depressed illnesses. Global Disease Burden Study claims the depression one of the biggest and major reason of the mental disorder. Unfortunately, many in our culture dismiss depression, refusing to seek medical assistance and to acknowledge it as a mental disorder. They are being reduced because of the limited biological markers for the diagnosis of depression. This paper's major objective is to provide a nonintrusive technique for identifying and differentiating depression through the use of brain signals from individuals with MDD and healthy people. In this investigation, a Raw EEG dataset was applied. Some of the key characteristics of depression detection are retrieved in order to recognize depression. Classification methods utilized in the study to input those depression detection features include Vector Machine, Naive-Bayes and Logistic Regression support. Additional ten-fold cross-validation is conducted to assure clarity and accuracy of findings that stimulate and encourage persons with mental health condition to obtain appropriate therapy.

AI is perhaps the most important part of a human being's consciousness, according to the journal [Kumar et al., 2021]. The health area is one of those fields in which the application of AI has produced surprising outcomes. AI has also proven quite productive in the field of medical services in conjunction with Internet of Things (IoT). In all cases, a few locations remain without the development of innovation. Dysfunctional behavior is one of the areas where no optimal therapy has occurred. With one-to-one effective cooperation, therapists evaluate and treat their clients. Although the drugs approve many drugs for their clients, such as anti-depressants, rest pills etc, the drug has not been able to repair or kill the infection. However, it does not have a solution. Reason may be anything for a persons to get in a certain situation such like society, job, family etc. The research on this topic will also be confined to predicting such a plight in the human body, identifying how the person uses a data set recently recorded. In this work the author used calculations for the production of troupe model and subsequent consideration of models for Calculated Regressions, Support Vector Machines (SVM), Choice Tree, the K-Nearest neighbor and Naivé-Bayes. The study utilized the suggested Kaggle data collection of 334 examples in 31 diverse domains of unemployment and psychiatric illness. Finally, the post-effect test of this application can be a genuine medical IOT demonstration.

The study shows that the genesis of genuine mental illnesses, such Schizophrenia and Bipolar, has substantial impact on the administration of social and well-being.The author [Zebin et al., 2019] talks about the Locomotor action adjusting most of the time of the SMI, which in turn allows early guidance of wearable action trackers in the earlier position of the SMI background. The generated model has an accuracy of 91.3% using a thorough computation of the neural organizations and the typical results acquired from wearable movement data.Disarray the DNN model and group disarray networks.It is generally apparent that the DNN ensemble beats the single DNN substantially with a total precision of 91.3% and 83.8%. Actually, the number of data correctly categorized into the control state across the two DNN setups is fairly comparable.

Journal [Yadav et al., 2020] reports Depression, a major reason for disability, affects 264 million individuals globally. A poor workplace can create a number of physical, health and efficiency problems. Studies have shown that people are reluctant to seek treatment from specialists of mental health. This is mostly due to the stigma of mental health concerns. The family history of mental illness, the working environment and their house have been questioned and a short routine survey data was provided.K-Nearest

Neighbours, Multinomial Logistic Regression, Decision Tree, stacking, boosting, Random Forest Classifier, and bagging have been used for analyzes.This article then ends with the algorithm resulting in the best results, with a precision score of 81.75%, followed by the Random Forest Classifier at 81.22% and other. Depression is a frequent condition, and the results of mental disease in several areas help to recognize it early.

During the future imagination, the study article has discussed thoroughly the mental state evaluation by change of mood using ECG.The article [Kitagawa and Kato, 2019] suggested a short research on the mental state estimate. The major focus is on the research of conduct and mood shifts, whereas the ECG signal is altered by a person picturing the next future. In this examination the algorithm utilized is support vector machine learning.During the future imagining, ECG is measured. The ECG function extraction is done by cardiac variability analysis and it is possible that mood changes might impact the mental state utilizing supporting characteristics of HF. It shows the efficiency using ECG influenced by mood change in envisioning mental status estimate in the future. The estimate is performed by the specified functions in advance. The end result of the MS TMS estimated based on chosen characteristics indicates the best F-measurement of 0.48, showing that HF and ccvHF are essential for the mental health of the person.

[Gore and Rathi, 2019] general, explain that technology has enabled the detection and access to many sorts of data to persons of mental illness. Mental health assessments are carried out by psychiatrists. This study aims to give an overview of data formats utilized for the control of the mental health by diverse authors. Examples of data utilized in research for evaluation of mental health problems are speech/audio clips, email chats, wrist sensors, f-MRI scans, EOG data, structural-MRI scans, EEG state-rest, questionnaire/survey information, twitter data and computer-based wireless monitoring.This article gives an overview of the many data kinds and the algorithms that various writers study and apply.The ratios of psychiatrists to patients, the source of the mental disease and the percentage of mental health problems globally are to be found generally. The criteria used to measure the performance of the models are F-score and accuracy. Since there are comparable forms of mental disorder, the study does not focus on similar data superior to others. Similar data types are also included.The inputs for mental health evaluations can be illustrated through textual data, image data, audio and signal data.

2.2 Statistical approach used in Analysis of Mental Disorder

This article [Silva et al., 2019] provides a completely new paradigm to be used by mental fitness practitioners to tackle their problems using statistical technologies. Although many study articles were published on public mental fitness, few focused on the use of statistical technology in public intellectual health, the way in which we handle, review and exploit statistics in healthcare enterprises has changed the record science.Record science differs from traditional workplaces in the examination of facts, in particular since the medical procedures used in data management motivates health specialists to use 'statistics' to tackle the difficult conditions of mental health.

This research provides a novel analysis framework for data collecting, fusion records, storage statistics, processing, analysis, visualization and modeling for public IT initiatives. One of the main reasons why this new framework was developed is to help medical practitioners in addressing public mental health concerns.

A completely new framework that mental health practitioners may use to meet the obstacles of using the information technology is the major aim of researchers. This structure depends greatly on advance planning. This is really important. This waterfall mode structure demands a purposeful undertaking from beginning to end, without a portion of the task beginning until the previous one is over.One of the reasons why a new framework has been put in place is to motivate health professionals to use "statistical science" to cope with difficult mental fitness issues.

In this research [Park et al., 2020] the description for the manipulation of the stress of understanding at work and to introduce an intellectual health management device, in particular whether or not harassment is perpetrated and recommend strain relief optimized for individuals by gathering and studying statistics on physiological sensors and other statistics. Stress mode must be further developed. Current models include SVN and ANN are typically used by researchers.However, in this study, the author employed an extra sophisticated computational technique to replace the green strain version.

2.3 Artificial Intelligence used in Analysis of Mental Disorder

The study [Niederriter et al., 2020] shows proposals as to how sensing technologies (for example, eye monitoring) may be integrated with a VR simulation of healthcare facilities to improve the science's decision support to diagnose and assess intellectual impairments. As the basis for the advancement of VR modules the scenario-based person simulations are utilized.Data collected by the sensing technology are used to create analytical ways to prevent the risk of intellectual contamination.In addition, VR-based medical school artificial intelligence (AI) technology helps students in medical college research more quickly and makes more wise judgments.They construct a VR simulation environment for the evaluation of intellectual difficulties and utilize sensing technologies to gather and analyze data and the behavior of participants.The goal is to improve medical evaluation and the handling of intellectual problems (for example, stress, desperation and schizophrenia) by increasing cognitive meta skills.

Stress is defined as the body's response to environmental changes via mental, physical, or emotional responses. This can have a detrimental impact on work performance and, in the long run, increase the risk of physiological issues such as hypertension and mental illnesses such as anxiety.A trustworthy, cost-effective, intense pressure identification framework might help customers better monitor and manage their stress, reducing the long-term negative repercussions. The author [Abaei and Osman, 2020] examines and audits the written documents using AI stress recognition methods.The study seemed as wise as the current arrangements have been written using the notion of the edge to constantly monitor pressure in a probable arrangement.

2.4 Natural Language processing used for Analysis of Mental Disorder

The author [Mulyana et al., 2019] discusses in this study how mental illness is a significant concern in Indonesia. The whole check is carried out and the patients are examined on the basis of their different symptoms.In Indonesia serious mental illnesses reached 1.7%, according to the 2013 fundamental health research statistics. The number of patients with mental problem is 0.47 per 100,000 professionals, meaning that the WHO standards remain lower.As a result, professionals cannot treat all patients with mental disorders directly.The creation of a case-based thinking system might be an alternative to resolve these gaps and help to diagnose different forms of mental illnesses. In most areas of health care non-expert medical practitioners undertake an initial assessment of mental patients. In the event that patients can't communicate their complaints, the examiner analyzes patients according to their symptoms and develops the narration text on the state of the patient.Text papers, however, cannot be used directly since their structures are not suited for processing in CBR systems.In this work, natural language processing is explored utilizing a text medical recording paradigm for producing mental health symptoms. This approach is intended to help the diagnostic and management of many kinds of mental illnesses by developing a case-based computer reasoning system.

The research [Hemmatirad et al., 2020] of mental disorders by people's demand for expressions in two prominent social media platforms, Reddit and Twitter. Their objective is to enhance the empirical model to hit and identify people's most significant psychological issues. The key characteristics of this article are the following in increasing the precise nature of an early detection device for sadness in social media:

- Use unique measures intended primarily for our early detectors.

- Extraction by using one-level feature extraction and class, text functions that are more accurate than traditional techniques (most of them rely on tf-idf vectorisation).

According to studies [Jain et al., 2020], mental health is a vital element of a character. Several variables, notably tension, worry and many more damage mental fitness.When someone with an intellectual disease does not receive the appropriate care, the results can be awful. Today, with the age high, the power of a laptop may be used to await the beginning of such intellectual diseases. Researchers propose specific procedures for important modules such as facial emotion and pulse load detection of these gadgets. He also proposes comparison research for a range of facial emotion detection methods. The author has produced results with the help of OpenCV and emotional analysis. OpenCV modules are utilized. The heart pulse loading module measures an individual's heart charge. The command behind it is basically to extract statistics on heartbeat from the shade of the facial skin owing to movement of the blood.

The alterations in detachment, on the other hand, represent psychological instability in the person's thinking, emotion and articulation. If an individual considers psychological instability like any other clinical problem, it may be addressed within the limitations of clinical research. A person's mental health and identification have become essential since his/her behaviour affects the social environment directly. In case of the emotional well-being state of an individual and the quantification of methods there is an exceptional advantage. The study [Sadasivuni and Zhang, 2019] aimed to differentiate the behaviour and the link between these diseases of individuals through the posted tweets. Moreover, Word Frequency technology is used to customize tweeting words. The test was well aligned with the results of the SVD method.

2.5 Deep learning models used for Analysis of Mental Disorder

The journal [Kwon et al., 2019] states that Gloom is a psychological disturbance endured worldwide by more than 3 million persons. In severe instances misery can quickly kill itself, although hardly exactly 50% of them are treated properly. In order to prevent the most striking decision patients are making, the study shows that it is essential for them to notify the patients that the person is miserable and has to go to a medical clinic. The study, which utilized the CNN and VGG16 models, revealed both a steady improvement in accuracy and a decline in the loss rate as the era increased. The results obtained with the CNN model demonstrate the accuracy of 75% while the results obtained with the VGG16 model are 87,5% accurate.

The identification of mental condition depends significantly on the autonomous symptoms of the existing technique and on a clinical diagnosis. To instruct the biological indicators of illnesses, to build a system that automatically identifies the mental illness. It is based on several methods such as speech contents, gestures, auditory characteristics and facial expressions. It is therefore a challenging process since indicators need to be considered. To solve this challenge, the author [Zhang et al., 2020] suggested a multimodal Deep Learning architecture that comprises many modalities including textual characteristics and the visual acostic, which may be independently analyzed while taking into consideration inter-modality correlations. A Multimodal Deep denoising autoencoder that is subsequently encoded to create session-level descriptors using a Fisher Vector coding is utilized to obtain a multimodal presentation of audio-visual information.

In the textual format, the records of interview sessions that are a paragraph vector suggested by integrated transcripts are provided to capture questions relating to mental illnesses.In a final classification, Multitask Deep Neural Network is integrated prior to an early fusion method, including audio-visual and textual characteristics.Two mental illnesses are evaluated for automatic diagnosis.The method uses two separate datasets to assess conditions such as depression and bipolar illness.

- The Bipolar Disorder Corpus.
- The Extended Distress Analysis Interview Corpus, respectively.

Study results showed similar successes to the state of the art to generate effective multi-modal interpretive learning and potential in various mental disorders, according to a research on bipolar disease and depression detection.

After studying the latest literature in this field, various methods were examined and this subject has become a topic of worldwide interest for scholars. Based on a literature study, it can be anticipated that most standard models, such as: KNN, SVM, Naive Bayes, Linear Regression, Logistic Regression, Random Forest and Decision Tree, would be utilized. The profound learning algorithms have also been developed and analyzed. The issue also included examination of feelings, statistical methods and artificial intelligence. Different techniques are already employed with sufficient precision in the investigation and diagnosis of mental illness. Deep learning was also seldom utilized for the same reason. We present a deep learning technique based on FFNN in diagnosing a mental disease. FFNN would also be compared with other models like Decision Tree, Logistic Regression, XGBoost, and SVM.

3 Methodology

The Knowledge Discovery in Databases approach is being used in this project, first suggested by [Fayyad et al., 1996]. This methodology is the basic way for gaining information on the area of mental illness, processing, experimentation, evaluation, and eventually insights. The difficulty of KDD is largely because of the type and process nature of the analyzed data. Summarizing this technique involves the selection and assessment of the data, preparation of the data, data transformation and application. This methodology is best suited for this dataset, starting from data collection, extraction and execution.

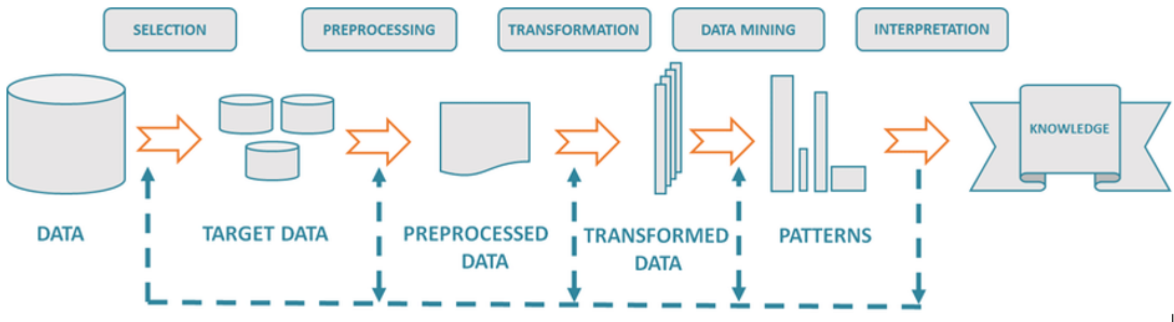

Figure 2: KDD Methodology

3.1 Business Understanding

The main aim before any project begins is to reach the entire goal of the project. This research aims primarily at identifying the mental illness in terms of the person's symptoms. It will require a thorough knowledge of the company objectives, enabling the complete conversion of data collected into information by means of data mining tools. The first stage is to read and identify the variable and the purpose of the KDD technique, then use it to analyze the factors that help to forecast the target variable. There are various types of symptoms a person go through dealing with mental disorder such as anxiety, depression, schizophrenia, bipolar disorder etc have symptoms such as sleeping trouble, stress, panic attacks, palpitation, hallucination. Our aim in this research is the identify the what kind of a mental disorder a persons deals with on the basis of their symptoms We are using five machine learning algorithms after the study of the previous researches. By following the KDD methodology the implementation of the machine learning algorithms are done which will be helpful in the topic of mental disorder and its identification on the basis of the symptoms. For this research we have done comparison of five algorithms among which ffnn and xgboost are the novel approaches.

3.2 Dataset Understanding

It is essential to develop a dependable model before executing the model with a correct knowledge of its components. The data requested may be accessible from several sources yet it might be unethical to collect it from a few. Therefore, data from a reputable source might also not be dependable on a few sources, and this demands substantial effort in terms of ethics. The dataset utilized for this investigation is provided without demographic information about the patient on Kaggle, therefore it is ethical to conduct the research with this data set.

To proceed with this research, we require data which contains various symptoms related to mental disorder. According to research, the identification of mental disorders is done with respect to symptoms the person experiences. The dataset has various mental disorders and symptoms, and each disorder has various different symptoms which are indicated in the form of Boolean. The dataset consists of 25 columns. The attributes are feeling nervous, panic, breathing rapidly, sweating, trouble in concentration, having trouble in sleeping, having trouble with work, hopelessness, anger, over react, change in eating, suicidal thought, feeling tired, close friend, social media addiction, weight gain, material possessions, introvert popping up stressful memory, having nightmares, avoids people or activities, feeling negative, trouble concentrating blaming yourself and Disorder. These symptoms are based on the disorder of the types of mental illness. These disorder names and symptoms are gathered on the basis of the actual indicators during mental disorder by various websites.

3.2.1 Data Exploration

The use of visualizations is one of the easiest ways to study and comprehend data. Exploring data visually may help understand the structure of the data, how the numbers are distributed, and whether there is any link within the dataset. Therefore below are the findings of the Mental disorder dataset based on the symptoms.

- **Identifying the datatype of all the variables**: As part of the data pre-processing, the majority of the variables are in string datatype and must be transformed to boolean datatype.
- **Checking the presence of null values in the dataset**: The null values of the dataset were checked, and the finding explains that there were no null values found in the dataset as the data was in the form of a boolean format with one string column to select the variables that are significant to the prediction.
- **The correlation within the variables is observed using below figure**: The multi-collinearity and correlation between our dependent variable and the independent variables are depicted in this graph. In this example, the 'disorder' variable is our dependent variable which shows the types of mental illness are there based on the symptoms. Fig 3 shows the Correlation graph between the columns of the dataset which is showing the linkage between different columns of dataset.

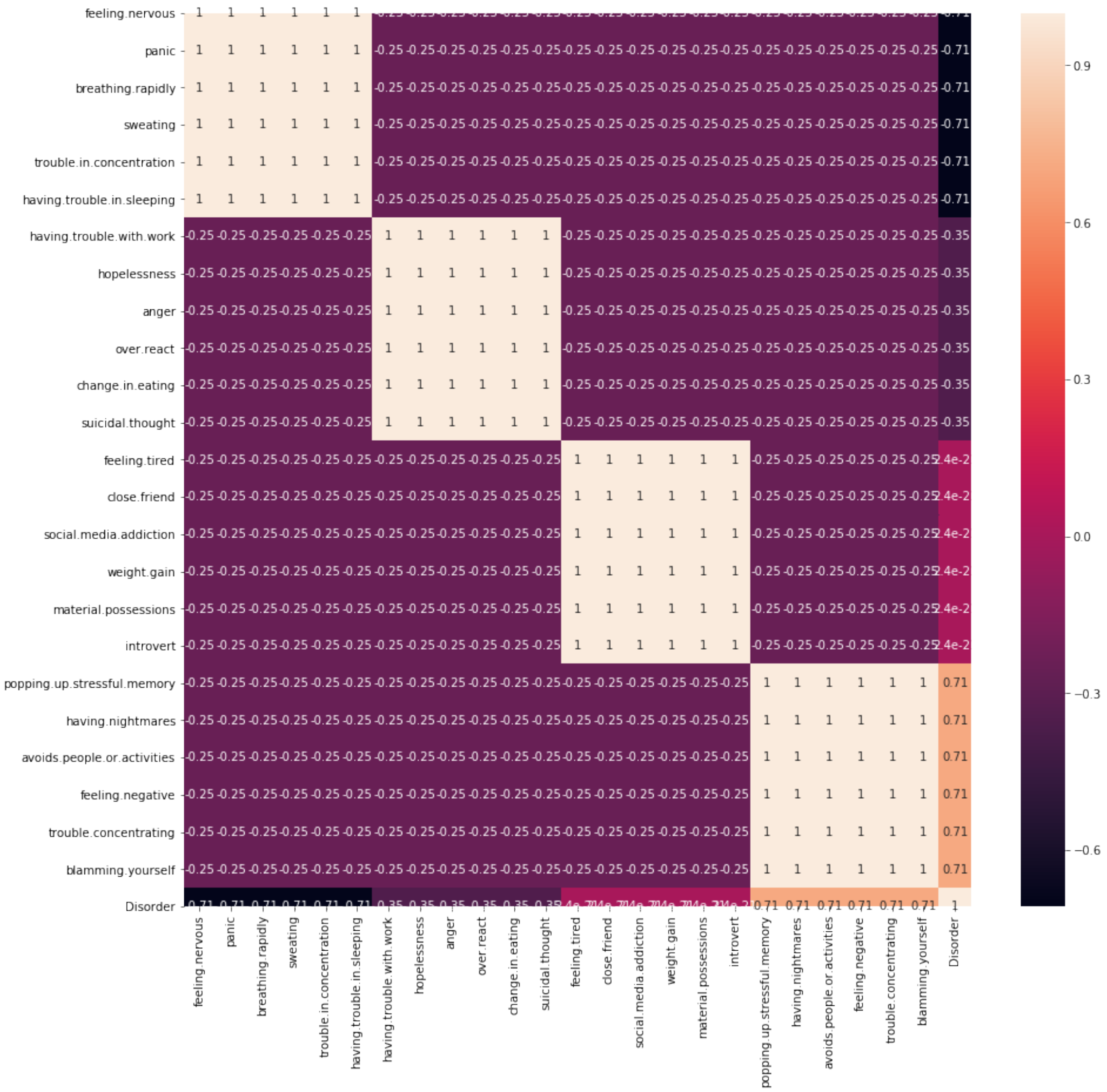

Figure 3: Correlation graph between the columns of the dataset which is showing the linkage between different columns of dataset.

3.3 Data Pre-processing and Transformation

Splitting the data into train and test, missing data, categorical values, and normalization are the most essential processes in data preparation. Large data often contains noise and the existence of special characters, missing values and blank spaces can impact the performance of the model and should be treated with great care in the interests of preserving and not manipulating the essential information during the data preparation. Even if the more information we have are expected, the better forecasts may be predicted, but this is not always the case. When examining the data it is seen that the data are not filled in consistently and need to be cleaned. The entire dataset is divided into data sets of training and validation. Training data is used to prepare the model, while validation data is used to evaluate model performance. The figure 4 highlights some key attribute that characterize the fundamental demographic information of the mental illness and its symptoms, its shows that there are no outlier in the dataset.

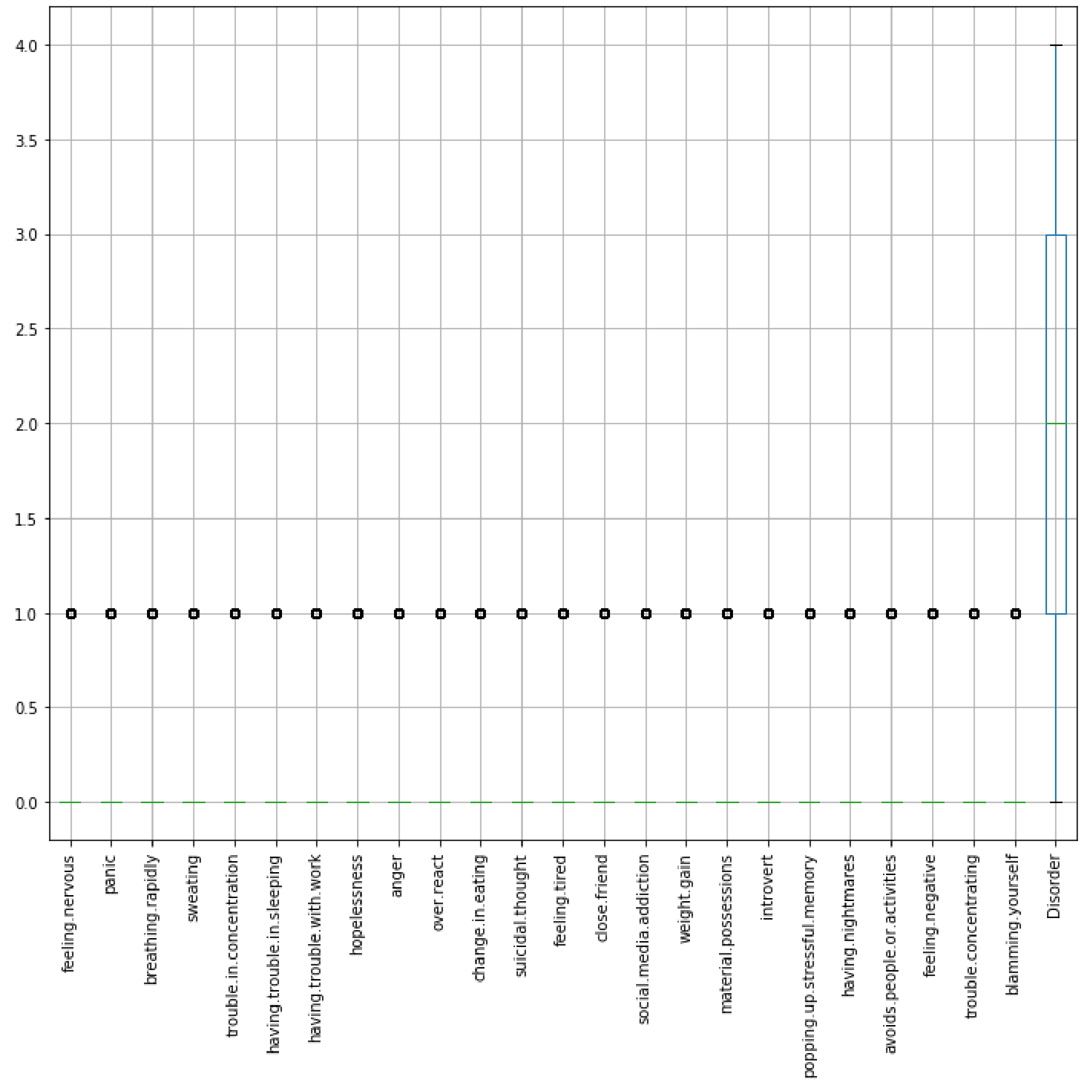

Figure 4: Box plot of dataset showing there is no outlier in the dataset and dataset is perfectly normalized

With redundant and unnecessary variables, computational time and costs are increased and hence the preprocessing phase of the data is typically seen in the KDD technique as a particularly difficult and time-consuming step. The procedures performed to prepare the necessary data are mentioned in the next paragraph:

- **File type**: The format of the file of the dataset is given in csv format which was not required to do any changes on it. So the same file was used for the implementation of the models.
- **Changing the datatype of all the variables**: The data types of all variables are usually noticed after the data collection is originated. The datatype of the file was in string form which was changed into Boolean form.
- **Handling the NA's or missing values**: The presence of missing values is problematic for the model and may impede performance of the models. The dataset has

a special character (" ?") instead of missing values. In order to deal further with these numbers the proportion of missing values are changed to NA's in each column. But for this dataset there were no missing values found hence in the pre-processing part the string values which were in the form of yes and no were changed in boolean form which is 0 and 1.

- **Multicolinearity**: The correlation test shows how these factors impact the target variables and also the correlation between them. The strongly linked variables (p value of 0.90) may be described as redundant, infact data, which will equally contribute to the prediction of the target variable. It is recommended that one variable should be removed and the other kept.

3.3.1 Feature Selection

Feature selection is the technique to reduce the number of elements for the promotion of an advanced model. The primary objective of the methods of feature selection is to discover the subset of characteristics to increase model correctness. Feature selection methods are used to enhance precision and minimize the calculation cost of a model and raise the understanding of models and approaches so as to eliminate non-relevant features and make them more compatible with a dataset. It helps also to reduce trainings duration, data complexity and overfitting of the model. The feature selection is a pre-processing phase utilized before a classification model is established and deals with the curse of the dimensional problem, which affects the algorithm negatively.

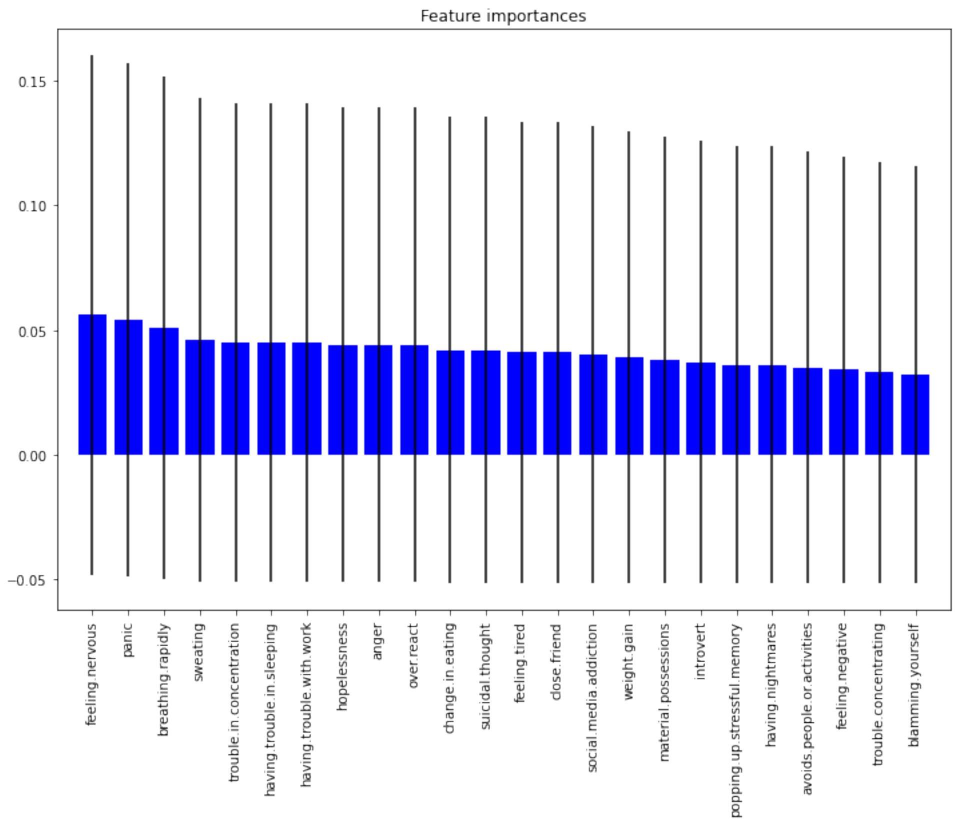

Figure 5: Output of extra tree classifier

For the selection function in the mental disorder dataset, we have used extra tree classifier. The extra tree classification is the sort of ensemble learning technology used to create the aggregate results in the forest of different or numerous decoration for the classification of the output. The data set utilized in this study comprised of 24 characteristics which could not be beneficial for the dataset on mental disorders. The tree-based techniques utilized in the Extra Tree Classification are based on the extent to which they can increase the node purity. Near the beginning of the tree, the biggest decline in the pure node occurs while the lowest decline occured at the end of the tree. This creates a subgroup of key characteristics by enclosing a certain node and then by pruning below it.

Important characteristics are noted, which will be utilized for further preprocessing and modeling.

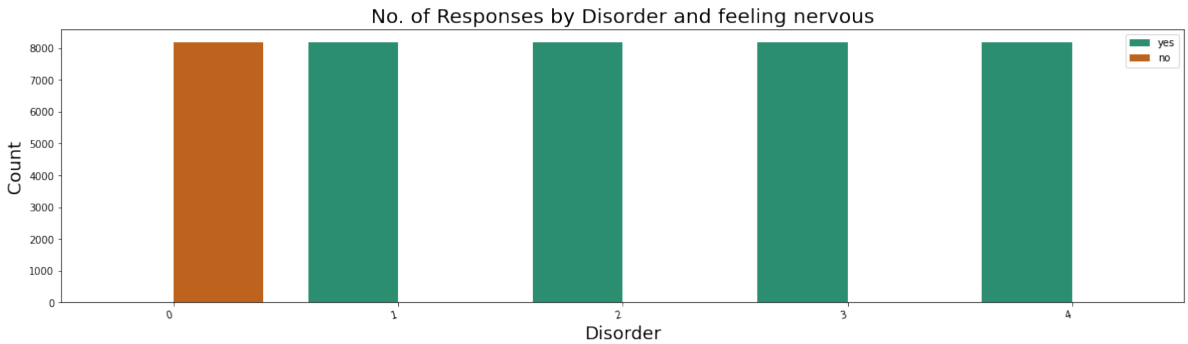

Figure 6: Graph between number of responses and Disorder

3.4 Modelling approach

This component is regarded vital and crucial for the process of machine learning. The recommended models must be applied in addition to the data preparation procedure, which includes feature selection and resampling. It is therefore possible to assess and compare the influence of preprocessing technology on different models. This section discusses the intricacies of the model utilized and the workings of the model.

- **Logistic Regression**: Logistic Regression is used in this case since we have approximate binary values like 1 or 0, no or yes, false or true based on independent variables like discrete values. When the dependent variable is dichotomous, this sort of model is suited for regression analysis. They may also describe the data and explain how one or more nominals relate to the dependent variable (binary). In high multi-colinearity, it does not work well and does not outline the variables. As the output of the Decision tree Classifyer, the results variable is determined hierarchically or sequentially and depends on the forecast data.

$$\Pi(\mathrm{x}) = 1/1 + e^{-\mathrm{y}} \quad (1)$$

for logistic regression

- **Decision Tree Classifier**: The decision tree classification model for the data set has been implemented. When compared to other processes, a decision tree is unquestionably incredibly fast. The only stumbling block is the state of overfitting, which occurs as trees grow and become unexpected or dense. We should use the irregular timberland in conquering the problem of overfitting, namely just the collection of trees that operate dynamically on a subset of the dataset, so that the impact of overfitting remains rapid. Decision-trees use several computations to determine the major components, division and the optimum value that provides a new subpopulation set thereafter.

  Decision trees need more attention to pre-processing of information, since there are many attributions in the dataset which we cannot require and each feature adds to the calculating dynamic, then flawless and schemed information, in calculation like a decision tree, so that there is no chance of an undesirable result..

- **Support Vector Machine**: In the two-dimensional analysis, the main objective of the vector holder is to identify the optimal line to distinguish the data points. These focuses are called support vectors. The computation then determines the distance between the reference vectors and the isolating plane at that location. The distance is referred to as the hole. The basic purpose of the computation is to increase the leeway distance. A hyperplane for which this hole is quite as extensive as predicted is the optimum hyperplane.

$$S(x) = \mathrm{sign}(\sum_k \alpha_k y_k k(x_k, x) + b) \tag{2}$$

  for svm

- **Boosting Algorithms**: Boosting is a computation which helps to reduce differences and inclinations in an AI collection. The measurement contributes to the transformation of powerless students into sound students by consolidating the N number of pupils.Additionally, boosting can operate on model learning assumptions. The fragile pupils are corrected successively by their archetype and transformed into solid students at all times.

$$F(x) = sign\left(\sum_{m=1}^{M} \theta_m f_m(x)\right) \tag{3}$$

- **Gradient Boosting**: Gradient boosters provide indicators to the group and follow the system to check before indications show up at a precise indicator for the system's finish, truly as with some other equipment AI technique. Ada boosts its prior errors by adjusting the burdens for each misperception of each cycle. All the considerations are that slope boosting goals fit an indication in the remaining blinders of the preceding indicator. Prompting inclination employs a decline in angle to identify the problems in students' expectations.

- **Ada Boosting**: Adaboost aims to build a number of fragile pupils in order to form a strong student in solitude. Adaboost concentrates on powerless pupils who are often chosen trees with only one split and are usually referred to as stumps. The primary Adaboost stump has equally weighted perceptions. Previous errors are corrected and views which were wrongly categorized are designated as having greater weight than distinct perceptions that were not incorrect in the scheme. Adaboost calculations are notably used in the approach of relapse and characterisation. A bug noticed in previous models is modified by weighting until an exact indication is produced.

- **Feed Forward Neural Network**: Complex neural organizations comprise an information layer and a yield layer, which are essentially like simple neural structures, but are also packed into numerous secret layers. As a result, they're known as profound neural organizations, and they're useful for deep learning. A deep learning framework emerges and becomes more "proficient" as time goes on, dividing data through several hidden layers, much like the human cerebrum with all of its complexities. The specific purpose to include repetition in a network is to provide a dynamic behavior, especially in connection with time series issues like pattern identification, which require internal memory to support the process of learning.

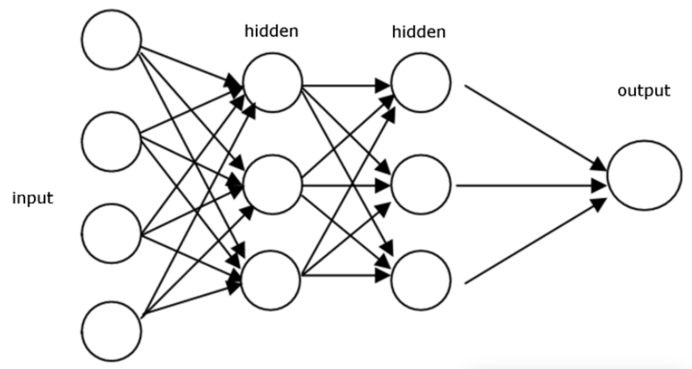

Figure 7: Neural Network

4 Design Specification

The architecture/diagram approach followed by our study is shown in Figure 7. We initially collected data from the source and then pre processed the NA values, which included the deletion of unnecessary columns. We also utilized the feature selection method for identifying the most significant features and eliminating the rest. Finally, the processed data is classified to four separate classes and the results are assessed on the test data.

1. Data Extraction: Obtaining and downloading reliable data.
2. Pre-processing: The collected data was cleaned to eliminate any Null values and the category variable was transformed to a boolean value. To obtain a better understanding of the data variable significance, an exploratory data analysis (EDA) was performed.
3. Model Building: This is the process of normalizing the target variable and dividing the data 70-30 for training and testing. Testing data is used to test the model, whereas training data is used to train it.
4. Evaluation: This is a method of determining how well a model performed.

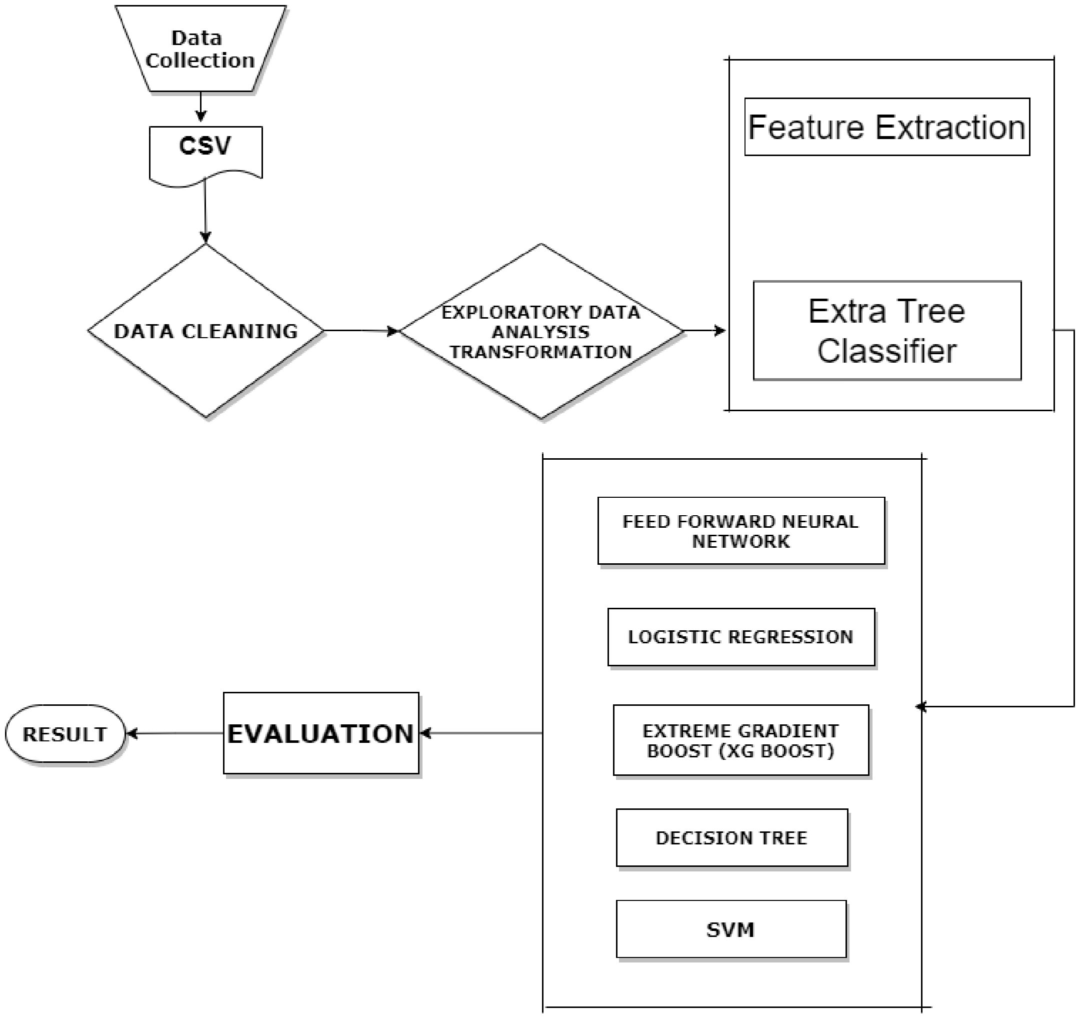

Figure 8: Design Flow process

5 Implementation

This section discusses the application of the suggested models to identify the person's symptoms for mental illness. It also discusses how most essential characteristics are selected and the dataset is tested again. The entire implementation portion in python v.3.7,2 was carried out and the integrated development environment was selected for the Jupyter notebook (v.6.0.2) (IDE). For the implementation portion Python was chosen since it is easy to use, has a large online support forum and one of the best languages in terms of code readability. Because of the active python community, there are several tools for data preparation and managing unbalanced data, making it a popular choice for machine learning applications.

The data was available on the website named Kaggle. Data source in this study consists of one single file with string datatype consists of types of disorders and symptoms which was in CSV format. The data was in the form of yes or no based on the symptoms of various disorders which was then converted in Boolean form. After that, the dataset was imported as a Dataframe into Python and the missing values were verified. After exploring the missing columns using PandasProfilingPackage there were no missing values found in the dataset. In the pre-processing part the string datatype was converted into Boolean type which is 0 and 1. After completing the basic clea ning task, we used a feature selection approach to reduce the characteristics and choose only the finest. The extra tree selection approach for classifying functions was utilized and this was done using the sklearn.feature selection library SelectFromModel 7 package. For filtering the best characteristics, a 0.03 threshold value of gini was employed. The filtered data set was then used for additional training and testing processes. The sklearn library's StratifiedK-Fold8 package is used to execute a stratified k fold split with a k value of 5.

After importing the dataset and completing the preprocessing and feature selection part we used five alternative models. The various models that were implemented Decision tree, SVM, Feed Forward Neural Network, Logistic Regression and XG Boost. These models may be found in the Sklearn Python library in the form of various packages. The metrics like true positives, true negatives, false positives, and false negatives were appended to a list for each model during the iteration process that occurs a number of times based on the value of K, and then the mean of these values was treated as the model outcome and was used to calculate the individual metrics. The accuracy, specificity, sensitivity, and geometric mean of specificity and sensitivity were utilized as assessment measures in this study. These metrics were computed for each model and compared using bar graphs. It was observed that Decision Tree gave the best result among comparison of models. Further a detailed evaluation and comparison of the results is carried in section 6 in the below section.

6 Evaluation

The primary goal of our research is to analyze the performance of our model and determine whether the approaches we employed are appropriate for this problem. The performance of the models is compared to measurements like precision, reminder, specificity and geometrical mean of reminder and speciality. These metrics are chosen because they are appropriate for issues of binary classification in previous study. The values obtained from the confusion matrix, which include true positives, true negatives, false positives, and false negatives, may be used to calculate these metrics. Machine learning model performance is assessed based on predictive accuracy, this predictive accuracy is correctly used when the data set is balanced, data set is divided into two sections: A model is trained at 70% and a model is tested using 30% testing data. The outcome of the confusion matrix determines the performance of models. The following parameter is used to create the confusion matrix.

True Positive (TP): This is a numerical number for instances that are successfully classified for mental disorder identification, and it is calculated using the formula TP/TP+FN

False Negative(FN): This a (type 1) error this is a numeric value of instance that are incorrectly classified for mental disorder identification is measure by FN/TP+FN

False Positive(FP): This (type 1) error is assessed by FP/TN+FP, the numerical instance values in-correctly classified for mental disorder identification.

True Negative(TN): This is a numeric value of instance that are correctly classified for mental disorder identification is measured by TN/TN+FP

6.1 Logistic Regression Evaluation / Case Study 1

Figure below illustrates the results of the regression following training and adopts 70/30% data for the training and testing, 79.9% accuracy, 74.7% precision, and F1-score of 70.3%, respectively. The model was not as good as previous models.

Table 1: Evaluation score for Logistic Regression

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	**0.799**	**0.747**	**0.798**	**0.703**

When using a logistic regression model, it is likely to add to the model's fair performance, with 79.8% True positive and 73.5% True negative predictions, 78.4% and 69.8% False Positive and False Negative predictions, respectively.

6.2 Decision Tree Evaluation / Case Study 2

A Decision tree is without a doubt extremely quick when contrasted with different procedures. The model is simple to use and performs well even when there are outliers. This model has an accuracy of 79.6%, precision 70.0% a recall of 80.1%, and a f1-score of 73.0%.

Table 2: Evaluation score for Decision Tree

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	**0.796**	**0.700**	**0.801**	**0.730**

It demonstrates that when using the Decision tree model, it got 88.6% right True positive predictions and 70.4% correct True negative predictions, with just 55.6% and 67.8% erroneous False-positive and False Negative predictions, respectively.

6.3 SVM Evaluation / Case Study 3

SVM was used in identification of mental disorder, the dataset wasdivided into two sections Training and Testing sets 75% and 25% respectively. Confusion matrix was employed to for predicting a good result. The result is 59.7% accuracy, precision 56.5%, recall 59.8% and f1-score 52.4% value.

Table 3: Evaluation score for SVM

Model	Accuracy	Precision	Recall	F1-score
SVM	**0.597**	**0.565**	**0.595**	**0.524**

6.4 Boosting Algorithm Evaluation / Case Study 4

As a result of the calculation, it can be deduced that it is a weighted average of multiple week learners. XGboost is one of the and most effective boosting algorithms for binary classification tasks, and it improves decision tree performance. In terms of recall, the XGboost classifier beat the competition with a score of 79.6%. This indicates that the model can properly identify mental illness symptoms with a good accuracy rate, which is a appropriate score.

Table 4: Evaluation score for XGBoost

Model	Accuracy	Precision	Recall	F1-score
XGBoost	**0.796**	**0.775**	**0.799**	**0.724**

Precision foe Xgboost classifier is 77.5%, recall is 79.9% and F1 score is 72.4% which gave the better accuracy results among various models for diagnosis of mental illness classification based on the symptoms of the person.

6.5 Feed Forward Neural Network Model Evaluation / Case Study 5

This approach used to build and implement the model was also a novel approach used to identify the mental disorder according to the symptoms of the person.

- **Create a Baseline NN Model** We have established a basic neural network to attempt to get our best outcomes using logistic regression and decision-making. We utilized a single hidden layer for our task to learn important characteristics and then a final layer for classification.
- **Beat the Baseline** Beat the above average accuracy of the tests. In the next cell, we have monitor the network and learn hyper-parameters. After training of the model the result was obtained.
- **Non-Determinacy** The initial network parameters are selected at random. Every time we begin on a random portion of the loss surface. This lead to a different result. We have trained our network 20 times and took the average. That was still giving some variance but was much more reliable. The more often we average the more dependable the outcome. It is certainly much better than taking a single run as an indicator of how good a model is.

Table 5: Evaluation score for FFNN

Model	Accuracy	Precision	Recall	F1-score
FFNN	**0.409**	**0.397**	**0.409**	**0.384**

After splitting the data into train, test and validate the accuracy obtained for FFNN model which was a novel approch is 40.9%, precision score is 39.7%, recall 40.9% and f1 score is 38.4% with the lowest accuracy among other four models used to determine the mental disorder.

In the above fig.6 comparison of Neural Network classifier Accuracy vs Ephocs it shows the validation accuracy and training accuracy is constant because the dataset used for the identification of mental disorder based on the symptoms of the person was a balanced dataset.

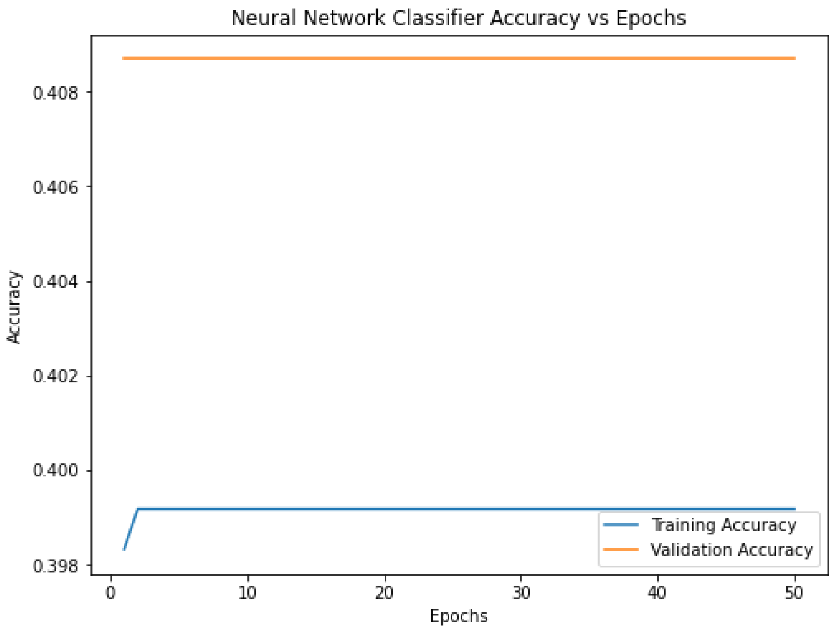

Figure 9: Neural Network Classifier Accuracy vs Epochs

6.6 Model Comparison

All of the models, including Logistic Regression, Decision Tree, SVM, XGboost, and Feed Forward Neural Network, were applied, with some of the models yielding good results and some yielding bad results. XGboost and FFNN models are innovative approaches, according to a prior study review. Performance can be deduced from the following table that the results of Logistic Regression, Decision Tree, and XGBoost were very close to each other, with Logistic Regression providing the best result. Feed forward neural network model gave the poor accuracy result for the mental disorder identification according to the symptoms. Any model's performance is determined by the type of the dataset used, and the dataset's dimensionality influences the model's performance. As the dataset was not too large and in boolean form might be the reason for the poor performance of some of the model. Because this study subject is in the medical field, the models' precision, recall, F1score as well as their accuracy, determine their dependability. FFNN and XGBoost were the novel approaches introduced in the mental disorder iden tification among which XGBoost showed the better results and FFNN failed with the poor results.

From the bar chart below its shows that the accuracy of three models which are Logistic Regression, Decision Tree, and XGBoost gave the almost similar accuracy and SVM and FFNN performed poor among which FFNN has the lowest performance with lowest accuracy for mental illness identification. They achieved a comparable result in all of the measurement parameters, which took a long time. Among all five models one of the novel approach which was xgboost algorithm was succesful and other which was FFNN failed to give the good result with achieving the minimum accuracy.

Table 6: Comparison of Models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.799	0.747	0.798	0.703
Decision Tree	0.796	0.700	0.801	0.730
SVM	0.597	0.565	0.598	0.524
XGBoost	0.796	0.775	0.799	0.724
Feed Forward Neural Network	0.409	0.397	0.409	0.384

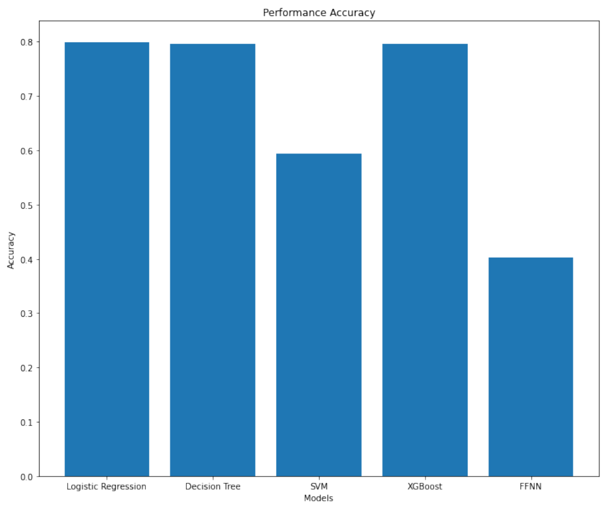

Figure 10: Comparison of the models based on accuracy

6.7 Discussion

This project was executed in stages; finding an appropriate dataset and the procedure of extraction, cleaning, and transformations were difficult. After pre-processing the datasets, the project objective was archived to balance unbalanced data utilizing the data balance feature selection approach. Preprocessing the information was the biggest problem as the variables in the dataset were understood and significant variables were identified.

The metric recall tells us how precise our model has been for the identification of the mental illness based on the symptoms of the person. This is a crucial metric for such classification of issues as we do not want to identify the wrong diagnosis due to variations of the mental condition. Though this metric is also important, recalls are more important than on a broader scenario and have a model which can classify the identification of a mental disorder according to their symptoms to give more accurate results as the correct disorder for each mental disorder should be diagnosed with different symptoms. As a result, the goal is to enhance recall while maintaining the model's credibility.

In table 7, individual performance is compared. The model initially provided varied recall values for the methods employed. Even the Decision tree method had the best recall of 80.0%. Table 7 shows that logistic regression had the maximum accuracy of value of 79.9% while having the lowest accuracy of 49.9%. Our innovative approach to FFNN underperformed. For this dataset of mental disorders, SVM achieved the lowest accuracy with unsatisfactory results. XGBoost performed well with accuracy of 79.6% also it was the novel approach. Following a review of prior research, it was discovered that these approaches, with variations in feature selection strategy methodology, have been utilized sparingly in this issue area, with XGBoost, Decision Tree, and Logistic Regression achieving substantial success in the implementation. Although our model's performance does not differ much from that of the prior research, a few measures, such as XGBoost's Recall are promising.

7 Conclusion and Future Work

As seen in the earlier literature, the subject of identification and diagnosis of mental disorders has been quite interesting in the last few decades and various approaches have been used to seek optimal outcomes and results in this field. Psychological health is one of the current medical concerns, and it may affect people of all ages. Medicines are also being manufactured, although full medicines are still necessary. With the use of Logistic Regression, SVM, Decision Tree, XGBoost, and FFNN Models, I have concluded that the application of machine learning technique to detect the mental disease on the individuals behavior according to their symptoms has been effectively achieved. The accuracy and forecasts preserved have met the goals of the study inquiry.

However from the list of list of contributing factors it is observed that the major various types of disorders and dependent on the symptoms for every mental disorders have different symptoms of the person. It also demands that creative work initiatives recognize and carefully respond to the broader practical and moral implications of the usage of ML frameworks for individuals, medical care, and society. When using machine learning to deal with the capture and evaluation of rich human needs and experiences, analysts must be careful not to discern and extract too much from the unique individual and their unique context in data analysis, translation, and depiction. To shape the future work large amount data with different attributes can be used for the diagnosis also different architectures can be used to perform the neural network for better results including Dimensional reduction techniques.

References

N. Abaei and H. A. Osman. A hybrid model for bipolar disorder classification from visual information. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4107–4111, 2020. doi:10.1109/ICASSP40776.2020.9054648.

H. Elmunsyah, R. Mu'awanah, T. Widiyaningtyas, I. A. E. Zaeni, and F. A. Dwiyanto. Classification of employee mental health disorder treatment with k-nearest neighbor algorithm. In *2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, volume 6, pages 211–215, 2019. doi:10.1109/ICEEIE47180.2019.8981418.

Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37, Mar. 1996. doi:10.1609/aimag.v17i3.1230.

E. Gore and S. Rathi. Surveying machine learning algorithms on eeg signals data for mental health assessment. In *2019 IEEE Pune Section International Conference (PuneCon)*, pages 1–6, 2019. doi:10.1109/PuneCon46936.2019.9105749.

K. Hemmatirad, H. Bagherzadeh, E. Fazl-Ersi, and A. Vahedian. Detection of mental illness risk on social media through multi-level svms. In *2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS)*, pages 116–120, 2020. doi:10.1109/CFIS49607.2020.9238692.

M. P. Jain, S. Sribash Dasmohapatra, and S. Correia. Mental health state detection using open cv and sentimental analysis. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 465–470, 2020. doi:10.1109/ICISS49785.2020.9315984.

Muthusubash Kavitha, Novanto Yudistira, and Takio Kurita. Multi instance learning via deep cnn for multi-class recognition of alzheimer's disease. In *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)*, pages 89–94, 2019. doi:10.1109/IWCIA47330.2019.8955006.

A. Kitagawa and S. Kato. Mental states estimation using ecg affected by mood change during imagining the near future. In *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech)*, pages 91–92, 2019. doi:10.1109/LifeTech.2019.8883951.

P. Kumar, R. Chauhan, T. Stephan, A. Shankar, and S. Thakur. A machine learning implementation for mental health care. application: Smart watch for depression detection. In *2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 568–574, 2021. doi:10.1109/Confluence51648.2021.9377199.

H. Kwon, S. Kang, W. Park, J. Park, and Y. Lee. Deep learning based pre-screening method for depression with imagery frontal eeg channels. In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 378–380, 2019. doi:10.1109/ICTC46691.2019.8939919.

V. Laijawala, A. Aachaliya, H. Jatta, and V. Pinjarkar. Classification algorithms based mental health prediction using data mining. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 1174–1178, 2020. doi:10.1109/ICCES48766.2020.9137856.

S. Mulyana, S. Hartati, R. Wardoyo, and Subandi. A processing model using natural language processing (nlp) for narrative text of medical record for producing symptoms of mental disorders. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–6, 2019. doi:10.1109/ICIC47613.2019.8985862.

B. Niederriter, A. Rong, F. Aqlan, and H. Yang. Sensor-based virtual reality for clinical decision support in the assessment of mental disorders. In *2020 IEEE Conference on Games (CoG)*, pages 666–669, 2020. doi:10.1109/CoG47356.2020.9231896.

K. Park, M. Jung Kim, J. Kim, O. Cheon Kwon, D. Yoon, and H. Kim. Requirements and design of mental health system for stress management of knowledge workers. In *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1829–1832, 2020. doi:10.1109/ICTC49870.2020.9289464.

S. T. Sadasivuni and Y. Zhang. Analyzing tweets to discover twitter users' mental health status by a word-frequency method. In *2019 IEEE International Conference on Intelligent Systems and Green Technology (ICISGT)*, pages 5–53, 2019. doi:10.1109/ICISGT44072.2019.00016.

S. Saleque, G. A. Z. Spriha, R. Ishraq Kamal, R. Tabassum Khan, A. Chakrabarty, and M. Z. Parvez. Detection of major depressive disorder using signal processing and machine learning approaches. In *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1032–1037, 2020. doi:10.1109/ICIEA48937.2020.9248419.

C. Silva, M. Saraee, and M. Saraee. Data science in public mental health: A new analytic framework. In *2019 IEEE Symposium on Computers and Communications (ISCC)*, pages 1123–1128, 2019. doi:10.1109/ISCC47284.2019.8969723.

Meza Silvana, Ricky Akbar, Derisma, Mia Audina, and Firdaus. Development of classification features of mental disorder characteristics using the fuzzy logic mamdani method. In *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, pages 410–414, 2018. doi:10.1109/ICITSI.2018.8696043.

S. Tariq, N. Akhtar, H. Afzal, S. Khalid, M. R. Mufti, S. Hussain, A. Habib, and G. Ahmad. A novel co-training-based approach for the classification of mental illnesses using social media posts. *IEEE Access*, 7:166165–166172, 2019. doi:10.1109/ACCESS.2019.2953087.

S. Yadav, T. Kaim, S. Gupta, U. Bharti, and P. Priyadarshi. Predicting depression from routine survey data using machine learning. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pages 163–168, 2020. doi:10.1109/ICACCCN51052.2020.9362738.

T. Zebin, N. Peek, and A. J. Casson. Physical activity based classification of serious mental illness group participants in the uk biobank using ensemble dense neural networks. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1251–1254, 2019. doi:10.1109/EMBC.2019.8857532.

Z. Zhang, W. Lin, M. Liu, and M. Mahmoud. Multimodal deep learning framework for mental disorder recognition. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 344–350, 2020. doi:10.1109/FG47880.2020.00033.