A

Preliminary Project Report

on

PSYCHOLOGICAL WELL BEING PREDICTION USING MACHINE LEARNING ALGORITHMS

SUBMITTED TOWARDS THE
FULFILLMENT OF THE REQUIREMENTS OF

Bachelor of Engineering (Computer Engineering)

BY

Student Name : Annu	Exam No:72021469G
Student Name : Priya	Exam No:72021608H
Student Name : Puja Nayak	Exam No:72021610K
Student Name : Mahima Chauhan	Exam No:72021545F

Under The Guidance of

Prof. Yogita Hambir

Onward to Glory

Department of Computer Engineering

Army Institute of Technology, Pune - 411015.

SAVITRIBAI PHULE PUNE UNIVERSITY

2022-23

ARMY INSTITUTE OF TECHNOLOGY,
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

PSYCHOLOGICAL WELL BEING PREDICTION USING MACHINE LEARNING ALGORITHMS

Submitted by

Student Name : Annu	Exam No:72021469G
Student Name : Priya	Exam No:72021608H
Student Name : Puja Nayak	Exam No:72021610K
Student Name : Mahima Chauhan	Exam No:72021545F

is a bonafide work carried out by students under the supervision of Prof. Yogita Hambir and it is submitted towards the fulfillment of the requirement of bachelor of engineering (Computer Engineering) Project.

Prof. P R Sonawane
Internal Guide

Prof. Sunil Dhore
H.O.D

____
External Examiner

Dr. B P Patil
Principal

Place : AIT, Pune

Date : 05/06/21

PROJECT APPROVAL SHEET

A

Project Report

on

Psychological Well Being Prediction Using Machine Learning Algorithms

is successfully completed by

Annu (Exam No: 72021469G)

Priya (Exam No: 72021608H)

Puja Nayak (Exam No: 72021610K)

Mahima Chauhan (Exam No: 72021545F)

at

Onward to Glory

Department Of Computer Engineering

Army Institute of Technology, Pune-411 015.

SAVITRIBAI PHULE PUNE UNIVERSITY

2022-23

Prof. Yogita Hambir Department of Computer Engg.	Prof. Sunil Dhore Head

Abstract

These days, mental health is a big problem. It is crucial to find any problems and fix them before they have a significant effect. We strive to do this with the Mental Health Tracker App. Because users may be coping with mental illness, we will need to design the app to be extremely friendly and inviting. The goal of this project is to develop a simple piece of machine learning software that tracks its users' progress while suggesting actions to help them improve their mental health. The user is posed with a number of questions, the application analyses their answers, proposes tasks, monitors their mental health, and displays the results on a dashboard. Machine learning was employed throughout this procedure.

Keywords: Random Forest, Naïve Bayes, Support Vector Machine, Recurrent Neural Network, Sentiment Analysis, Machine Learning Algorithms, Convolutional Neural Network, Logistic Regression, Linear Regression, Mental Health, Deep Learning, Binary Classification.

Acknowledgments

Presentation inspiration and motivation have always played a key role in the success of any venture.
It gives us great pleasure in presenting the preliminary project report on Psychological Well Being Prediction Using Machine Learning Algorithms. '

We express our sincere thanks to ***Prof. B P Patil****, Principal, Army Institute of Technology.*
We pay our deep sense of gratitude to ***Prof. Sunil Dhore*** *(HOD) of Computer Department, to encourage us to the highest peak and to provide us the opportunity to prepare the project.*
We feel to acknowledge our indebtedness and deep sense of gratitude to our guide ***Prof. Yogita Hambir*** *and our project coordinator* ***Prof. Anup Kadam*** *whose valuable guidance and kind supervision given to us throughout the course which shaped the present work as its show.*

Annu
Priya
Puja Nayak
Mahima Chauhan
(B.E. Computer Engg.)

INDEX

List of Figures

Chapter 1

INTRODUCTION

1.1 DETAILS OF PROJECT WORK

- ### 1.1.1 Machine Learning Algorithms

  Through the use of sample data or prior knowledge, machine learning allows computers to optimise performance criteria. A model has been defined up to a certain point, and learning is the process of running a computer programme to adjust the model's parameters based on training data or prior knowledge. The model could be descriptive or predictive, allowing users to learn from data and make future predictions. The challenge of creating computer programmes that automatically get better with use is central to the branch of research known as machine learning. Machine learning is a subfield of artificial intelligence that studies statistical models and algorithms to help computers improve their performance. These models and algorithms are designed to produce predictions or conclusions based on data without being explicitly instructed to do so. There are several different types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, models are taught using labelled data, whereas in unsupervised learning, models are trained using unlabeled data. The foundation of reinforcement learning is the training of a model through trial and error. Machine learning has a wide range of uses, including natural language processing, audio and image recognition, and recommender systems.

1.1.2 Natural Language Processing

The process of accurately translating voice data into text is known as speech recognition, commonly referred to as speech-to-text. Our programme will responds to voice commands or inquiries using speech recognition. The way individuals speak- quickly, slurring words together, with varied emphasis and intonation, in various dialects, and frequently using improper grammar- makes speech recognition particularly difficult. We are using natural language processing to convert users audio input into texts which will be then given to machine learning model for prediction. Building machines that can control written, spoken, and organised human language is the field of natural language processing (NLP), which aims to create data that is similar to or entirely composed of human language. It developed from computational linguistics, which employs computer science to comprehend the fundamentals of language; nevertheless, NLP is an engineering subject that focuses on creating technology to carry out practical functions rather than constructing theoretical frameworks. Natural language understanding (NLU), which concentrates on semantic analysis or figuring out the intended meaning of text, and natural language generation (NLG), which concentrates on text synthesis by a machine, are two overlapping subfields of NLP. NLP is distinct from speech recognition, which aims to turn sound into words by parsing spoken language, but it is frequently used in conjunction with speech recognition.

1.1.3 Language Translation API

Translation apps make communication easier by converting written material between two languages. The basic goal of a translation app is to make it possible for readers to read a translation as though it were the original. In our application we are using Libre language Translation API . Libre is an open source language translation software . we are using it API to take user input in different languages. We will convert different language inputs into English inputs and feed it into machine learning models.

1.1.4 Python Libraries

Python's Matplotlib toolkit provides a complete tool for building static, animated, and interactive visualization. Matplotlib makes difficult things possible and simple things easy. Produce plots fit for publishing. Make interactive charts with zoom, pan, and update capabilities. Visual style and layout can be changed. Export to a variety of file types. Integrate with Graphical User Interfaces and JupyterLab. Utilize the extensive selection of third-party programmes based on Matplotlib. We are using Matplotlib to graphically represent users weekly mental progress.

1.1.5 ReactJS

One of the most widely used JavaScript libraries for creating mobile and online applications is ReactJS. React, developed by Facebook, consists of components, which are reusable JavaScript code fragments used to design user interfaces (UI). Create self-contained, encapsulated components and combine them to create sophisticated user interfaces. Reactjs is being used for front end development for our app. By generating reusable components in React, which you may imagine as separate Lego blocks, you can build applications. These parts make up a final interface piece by piece, which when put together creates the full application's user interface.

React's main function in an application is to manage the view layer by offering the best and most effective rendering execution, much like the V in a model-view-controller (MVC) paradigm. Instead of treating the entire user interface as a single entity, React.js enables developers to break these intricate UIs down into distinct reusable components that serve as the foundation for the entire UI. Thus, the ReactJS framework combines the effectiveness and speed of JavaScript with a more effective way to manipulate the DOM.

1.1.6 NodeJS

An open-source, cross-platform run time environment for JavaScript called Node.js allows JavaScript code to be executed outside of a web browser. Node.js is a well-liked, easy-to-learn web framework that is utilised by many well-known businesses, including Netflix and Uber. Usually, when we think about JavaScript, we think of web development. There truly wasn't a method to run JavaScript outside of a browser before Node.js came around. Because we can run our code as a standalone programme rather than something that can only be assessed in a browser context, Node.js is a common choice when we develop a back end server and database.

1.2 OBJECTIVES

- 1.2.1 To create an app for early prediction of psychological disorder and proving solution according to the stages of disorder.

  1.2.2 To create awareness about mental health among common people through our application

  1.2.3 To encourage people think positive and be conscious of their psychological well being.

1.3 SCOPE OF THE PROJECT

- 1.3.1 According to World Health Organization (WHO) by 2030 depression will be the chef source worldwide disease burden. Our app has a huge market base in providing services to these people.

  1.3.2 This app will help people to improve their mental condition and encourage to enjoy their life and stay always positive.

  1.3.3 This app can be used by literate and illiterate people and people from all region without any language barrier.

1.4 MOTIVATION OF THE PROJECT

- Mental health disorders are now among the top leading causes of health burden worldwide.197.3 million people require care for mental health condition. The problem is huge for current hospital management system for costly prediction and providing solutions to mental health issues. People in India lacks awareness and interest in mental wellbeing. Our application will provide solution to this problem. We are using machine learning algorithms to predict current psychological state of a person. On the basis of our prediction we are providing solution to the user to improve his mental health being.

Chapter 2

LITERATURE SURVEY

2.1 A RANDOM FOREST MODEL FOR MENTAL DISORDERS DIAGNOSTIC SYSTEMS

The advantages of the random forest classifier are its resistance to missing and noisy data, and its ability to achieve the maximum accuracy in tests employing various classifiers.

Decision trees, neural networks, and support vector machines trailed random forest in accuracy by 92.10%, 92.01%,91.04% respectively. The relevance of these findings is to confirm the potential use of a data mining strategy in diagnostic procedures.

The trained models could be useful in developing mental disorder diagnosis models. Future work should focus on making more tries with more patients.

2.2 PREDICTING GENERALIZED ANXIETY DISORDER AMONG WOMEN USING RANDOM FOREST APPROACH

It is revealed that all prediction accuracy scores are over 0.9, indicating that the random forest technique can successfully predict generalized anxiety disorder.

The results show variability between balanced and unbalanced data sets in terms of sensitivity. They suggest that the classifier's performance may be impacted by the data set's imbalance. In terms of assessing specificity, the random forest technique consistently predicts those who do not have generalized anxiety disorder.

Comparing the prototype's results to the baseline that is implemented in the R tool, they are fairly consistent. Overall, the findings show that the random forest technique has high predictive performance.

2.3 DETECTION OF CHILD DEPRESSION USING MACHINE LEARNING METHODS

Early depression identification in children and adolescents is crucial for their academic, social, and learning development. In this work, Random Forest (RF), a large-dimensional data set of mental health of children and adolescents in Australia, has shown to be an effective and accurate classifier to identify depression.

The capabilities of the TPOT classifier in this model have been shown by the performances of all four algorithms (XGB, RF, DT, and Gaussian NB) in terms of confusion matrix parameters, results of K-fold cross validation, and scores of area under curve (AUC) and receiver operating characteristic curve area (ROC).

The sad people in the sample have been more correctly identified thanks to the features extraction technique that was implemented using correlation and weighted classifiers.

2.4 APPLICATION OF MACHINE LEARNING METHODS IN MENTAL HEALTH DETECTION

This systematic review's goal is to evaluate mental health problem identification critically. The data source, feature extraction approach, and better performance in machine learning or deep learning techniques make up this study.

This study also examines the utility of pre-mental health detection by outlining its data analysis technique, difficulties, and constraints.

The study's findings suggest that OSNs have great promise as sources of information for identifying mental health issues but cannot ever fully replace more conventional techniques for doing so that rely on in-person interviews, self-reporting, or the distribution of questionnaires. However, OSNs can offer further information, and combining the two methods for identifying mental health can enhance future study.

To improve the accuracy and precision of mental health problem diagnosis in the future, this study calls for widespread adoption, creative algorithms, and computational linguistics.

2.5 CLASSIFICATION ALGORITHMS BASED MENTAL HEALTH PREDICTION USING DATA MINING

According to a WHO report, depression will be a leading contributor to mental disease worldwide, and individuals need to care more about their mental health in order to manage their personal and professional lives.

Online outcome predictors can be used by those who are apprehensive to seek a diagnosis from a person. We initially encoded the data before doing the prediction. After that, we built a model with the decision tree method and applied it to our website.

Decision trees achieved an accuracy of 82%, with 258 out of 315 instances of data being properly identified. The customer receives advice and a likelihood of their mental health state after answering the questions on our website.

2.6 MACHINE LEARNING IN MENTAL HEALTH: A SCOPING REVIEW OF METHODS AND APPLICATIONS

Remarkable developments have been made in the field of ML for mental health, especially in recent years. Overall, it is evident that ML may considerably enhance mental health issue identification and diagnosis.

Research into various ML applications, such as public health, therapy and support, as well as research and clinical management, has so far shown promising results.

It is anticipated that the discipline will expand and that new applications for mental health will follow as ML techniques become more available to academics and doctors.

2.7 MACHINE LEARNING FOR MENTAL HEALTH DETECTION

We were successful in predicting the degrees of depression on our dataset using the EMU mobile Android application in combination with several machine learning methods. The data set included SMS messages as well as open- and closed-ended audio prompt replies from real people who were working on Amazon Mechanical Turk (MTurk).

We tested several tool and machine learning algorithm combinations to determine which ones worked best, and we were able to get results with appreciably high accuracy.

However, given that our results are not flawless, we do expect that further research in this area will increase overall accuracy. This procedure makes it simpler for a person to get a kind and estimate of their depression's severity.

2.8 SUICIDAL IDEATION DETECTION: A REVIEW OF MACHINE LEARNING METHODS AND APPLICATIONS

The majority of this field's research has been done by psychologists using statistical analysis, and computer scientists using feature engineering- and deep learning-based machine learning.

We outlined present activities and additionally suggested new potential tasks based on recent research. Finally, we explore some of the limits of the present literature and suggest a number of new lines of inquiry, including the use of cutting-edge learning strategies, interpret able intention understanding, temporal detection, and proactive conversational intervention.

The primary medium for SID in the future is almost certainly going to be social media information online. In order to identify online messages containing suicidal ideation and to prevent suicide, it is crucial to create new techniques that may bridge the gap between clinical mental health diagnosis and automatic machine detection.

2.9 DEEP LEARNING FOR DETECTION OF DEPRESSION USING TEXTUAL DATA

By putting the suggested method into practise and lowering the false positive rate, an accuracy of 99% was achieved.

The evaluation findings demonstrated that, in comparison to Naive Bayes, SVM, CNN, and Decision Trees, this framework delivers good accuracy, precision, recall, and F1-measures. The study was conducted in 2022 and methods used are RNN and LSTM.

To characterise depressive symptoms from the text data, the RNN approach using the LSTM technique was used, together with a depression prognosis strategy.

2.10 DETECTION OF CHILD DEPRESSION USING MACHINE LEARNING METHODS

When it came to accurately and precisely predicting depressed classes in just 315 milliseconds of training time, RF outperformed all other systems by 99 %.

The techniques utilised here are Gaussian Naive Bayes, Decision Tree (DT), XG

Boost (XGB), and RF.

In order to make an early and correct diagnosis and prevent serious effects in the future, it is essential to recognise the right indications to anticipate mental disease, such as depression, in children and adolescents.

There hasn’t been any research using machine learning (ML) techniques to identify depression in kids and teenagers aged 4 to 17 in a carefully crafted high prediction dataset like Young Minds Matter.

Chapter 3

REQUIREMENT ANALYSIS

3.1 INTRODUCTION

3.1.1 purpose

The project “Psychological well-being prediction and solutions using machine learning” aims towards developing a software product which shows predicts the psychological wellbeing of a person using machine learning for this purpose we are collecting data from user in audio format. Our application will show weekly progress of user output.

3.1.2 Intended Audience

Intended audience for this software will be all the people dealing with mental illness or who are susceptible of any mental disorder. People from all regions whether literate or illiterate can use this software for checking their psychological wellbeing and in case of mild illness can follow suggestions given by the application to improve his/her mental state.

3.2 OVERALL DESCRIPTION

3.2.1 Product Perspective

Providing the common people an easy to use and interactive app to check their psychological wellbeing and find solutions to their illness. This application is based on machine learning model and natural language processing. Our application is using simple question based inputs.

3.2.2 Product Feature

Following are the features of this product:

a) Easy and interactive UI for convenience of user.

b) Storing the user data in organised manner and using it for detection of mental illness.

c) Providing user's progress in graphical representation.

3.2.3 Constraints

Following are the main constraints of the project:

a) The data being provided here is static, so in future we will have to take care of the scenario where data is provided dynamically.

b) We have not worked too much on the UI since the main aim of this project was to predict psychological wellbeing. So, in the future, work on UI will also be done.

3.2.4 Assumptions

Following are the main assumptions of this project:

a) The data is static.

b) Customers have Microsoft Edge, Google Chrome, and Firefox browsers. c) Users don't mind not having the best UI and only want the efficient

prediction of their mental wellbeing and best solutions for improvement.

3.3 SPECIFIC REQUIREMENTS

3.3.1 Functional Requirement:

a) Real time prediction: The user can enter data in near real time. The data has been stored in database which will help in prediction using machine learning model in real time.

b) Login: If a user is already registered, then they will have to enter user id and password. If the details entered are correct, they will be logged in otherwise error message will be shown to them. Forget id password feature will also be shown in case of wrong password.

c) Register: A person will have to give basic details to register. If the details entered are valid then the person will be able to register and create an user id and password for login.

d) View: User will be able to see weekly progress on their screen in a tabular format.

e) Update: User will be able to update different attributes like password, address etc.

3.3.2 Hardware Requirement

a) CPU Processor: Intel i3 1153HE or Above

b) RAM: 4 Gigabytes or above

c) Router: Cisco RV300 series

d) Available Space: 30 GB (including OS)

3.3.3 Software Requirement

a) Operating system: Any modern version of MacOS, Windows, or Linux.

b) A code editor or IDE such as Atom, Visual Studio Code or WebStorm.

c) Browser: Any modern browser such as Chrome, Firefox, Safari, Opera, etc.

d) Node.js, version ¿= 6.0.0 and a package manager such as npm or yarn.

e) React version 17.0.0.

f) MongoDB and MySQL installed.

3.4 USE CASE MODEL

3.4.1 Use Case 1: Registration

Primary Actor: User

Pre-Conditions: Internet should be available.

Main Scenario:

a) Terms and conditions are digitally signed by the user.

b) Valid personal information is filled into the registration form.

c) The registration form is submitted for verification, and once it gets verified, the user receives a notification in his/her email ID regarding the same.

d) User account is made.

Alternate Scenario:

Network failure. Now, the user needs to try again when a secure internet connection is re-established.

3.4.2 Use Case 2: Login

Primary Actor: User

Pre-Conditions: The user should have an existing account.

Main Scenario:

a) Open the website.

b) Fill in your username and password.

c) Authentication is done

d) User is directed to the main screen (or dashboard)

Alternate Scenario:

Re-enter your username and password in case they were incorrect.Click on forgot password (in case the user has forgotten their password or wants to set a new password) and set a new password for your account by following some verification steps.

3.4.3 Use Case 3: Reset password

Primary Actor: User

Pre-Conditions: The user has an existing account in the system which has an email ID associated with it.

Main Scenario:

a) The user clicks on forgot password.

b) He/she receives and email on their registered email ID which will guide them to change their password.

Alternate Scenario:

The new password is not long enough or doesn't contain a special character or a number, etc.

3.4.4 Use Case 4: Weekly Progress

Primary Actor: User

Pre-Conditions: progress presented to the user once he/she log in

Main Scenario:

a) He/she can check weekly progress though graphical representation.

b) He/she can select the week dates to show the progress.

Alternate Scenario:

A dialog box appears in case of a network error and table is not displayed to the user

3.4.5 Use Case 5: Show Details

Primary Actor: User

Pre-Conditions: Internet should be available. Database should have details that user wants to see.

Main Scenario:

a) User choses the details he wants to use.

b) App gives the response according to the server.

Alternate Scenario:

The internet connection gets broken during this process prompting the user to re-establish an internet connection and repeat the Use Case 5.

3.4.6 Use Case 6: Provide Inputs

Primary Actor: User

Pre-Conditions: Internet should be available. Database should have questions that user wants to see.

Main Scenario:

a) The user can provide input in audio format.

b) Questions will be fetched from the database.

c) They will then be shown to the user.

Alternate Scenario:

The internet stops working while the user selects the choose question option, the user is then prompted to re-establish an internet connection and then repeat use case 6.

3.4.7 Use Case 7: Insert/Delete user

Primary Actor: Admin

Pre-Conditions: The user should be logged into the website as an admin.

Main Scenario:

a) The admin has the power to add a user.

b) He also has the power to delete the user.

Alternate Scenario:

The internet stops working while the admin clicks on the add/delete user option, prompting the user to re-establish an internet connection and then repeat use case 7.

3.4.8 Use Case 8: Logout

Primary Actor: User

Pre-Conditions: The user should be logged into the website.

Main Scenario:

a) The user clicks on the logout button.

b) He/she is directed to the login page.

Alternate Scenario:

The internet stops working while the user clicks on the logout button prompting the set to re-establish an internet connection and then repeat use case 8.

Chapter 4

DETAILED DESIGN

4.1 ARCHITECTURAL DESIGN

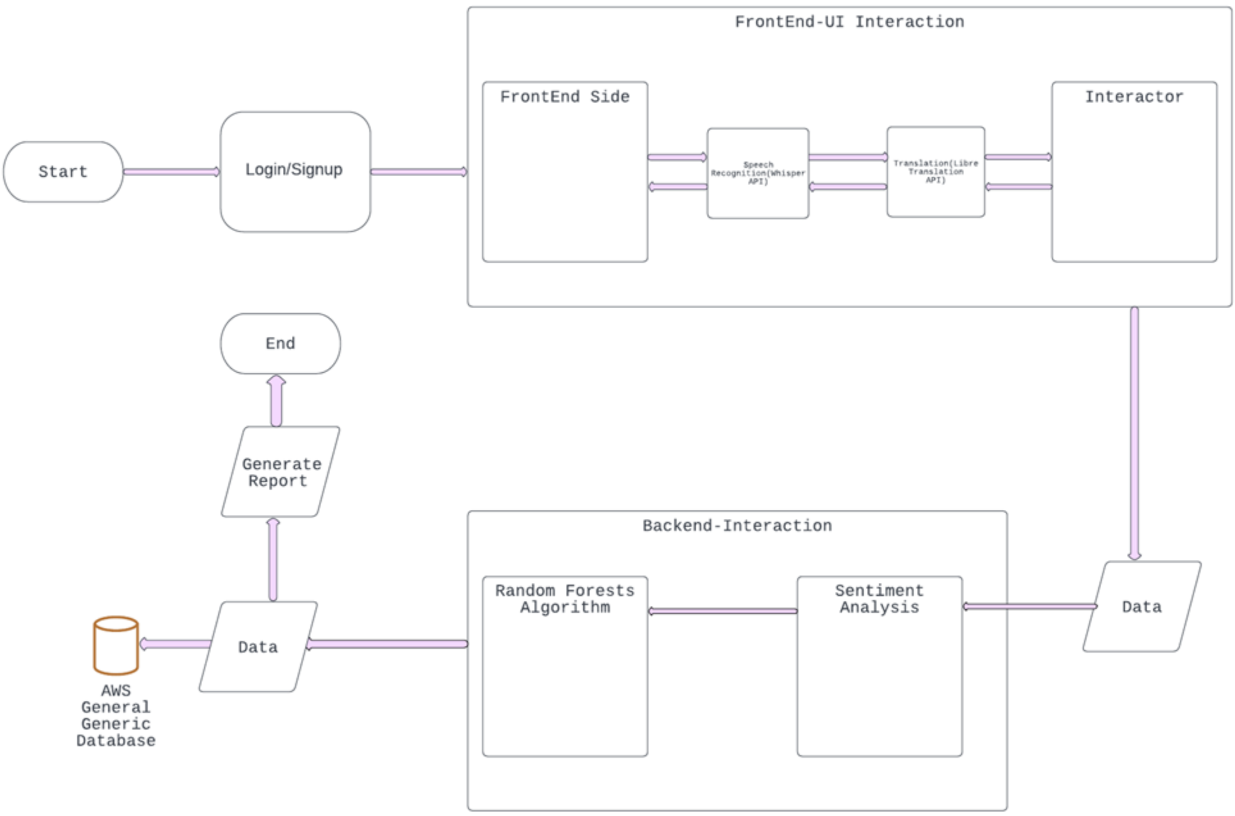


Figure 4.1: Architectural Diagram

4.2 LOGIN AUTHENTICATION DIAGRAM

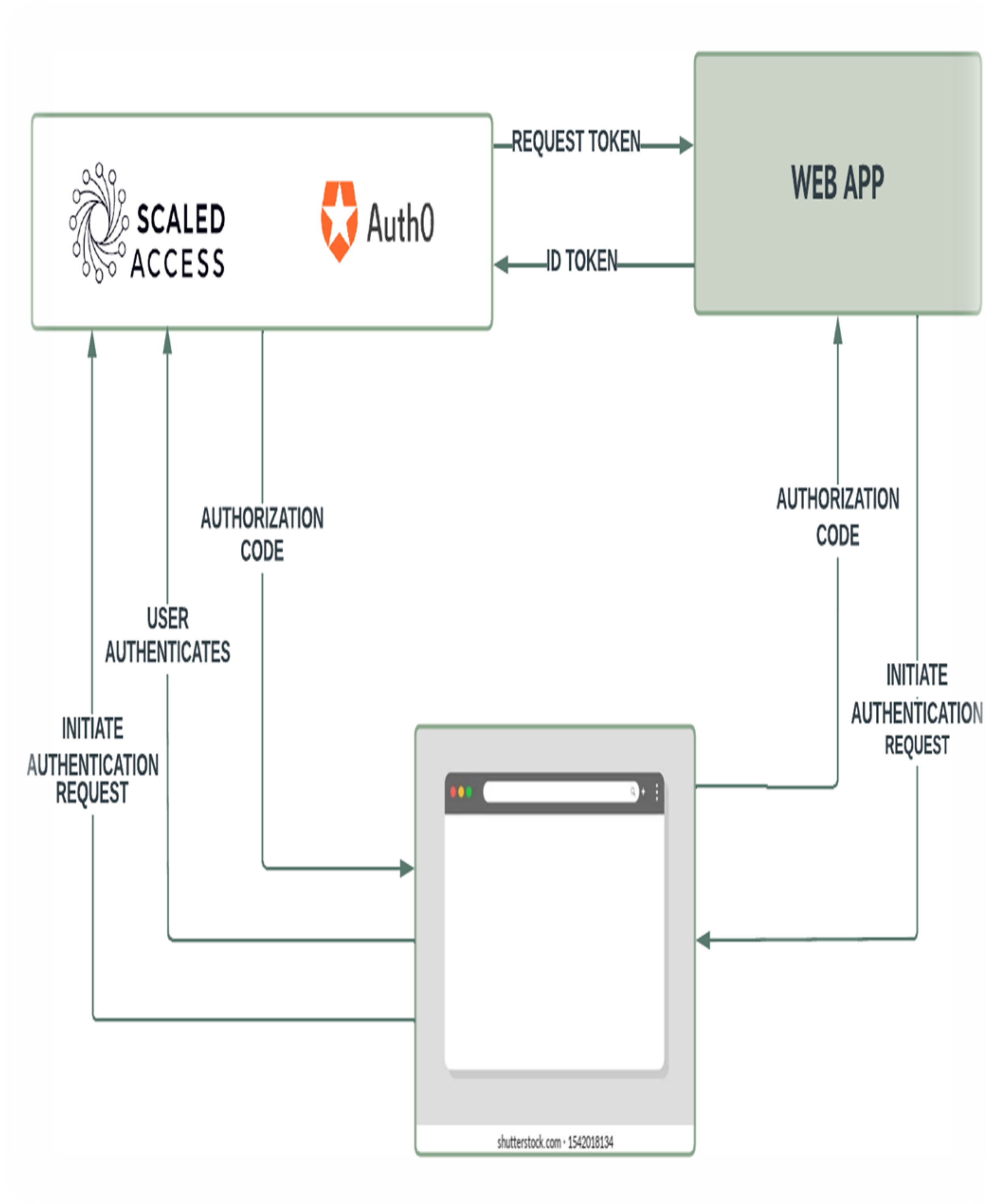


Figure 4.2: Login Authentication Diagram

4.3 USE CASE DIAGRAM

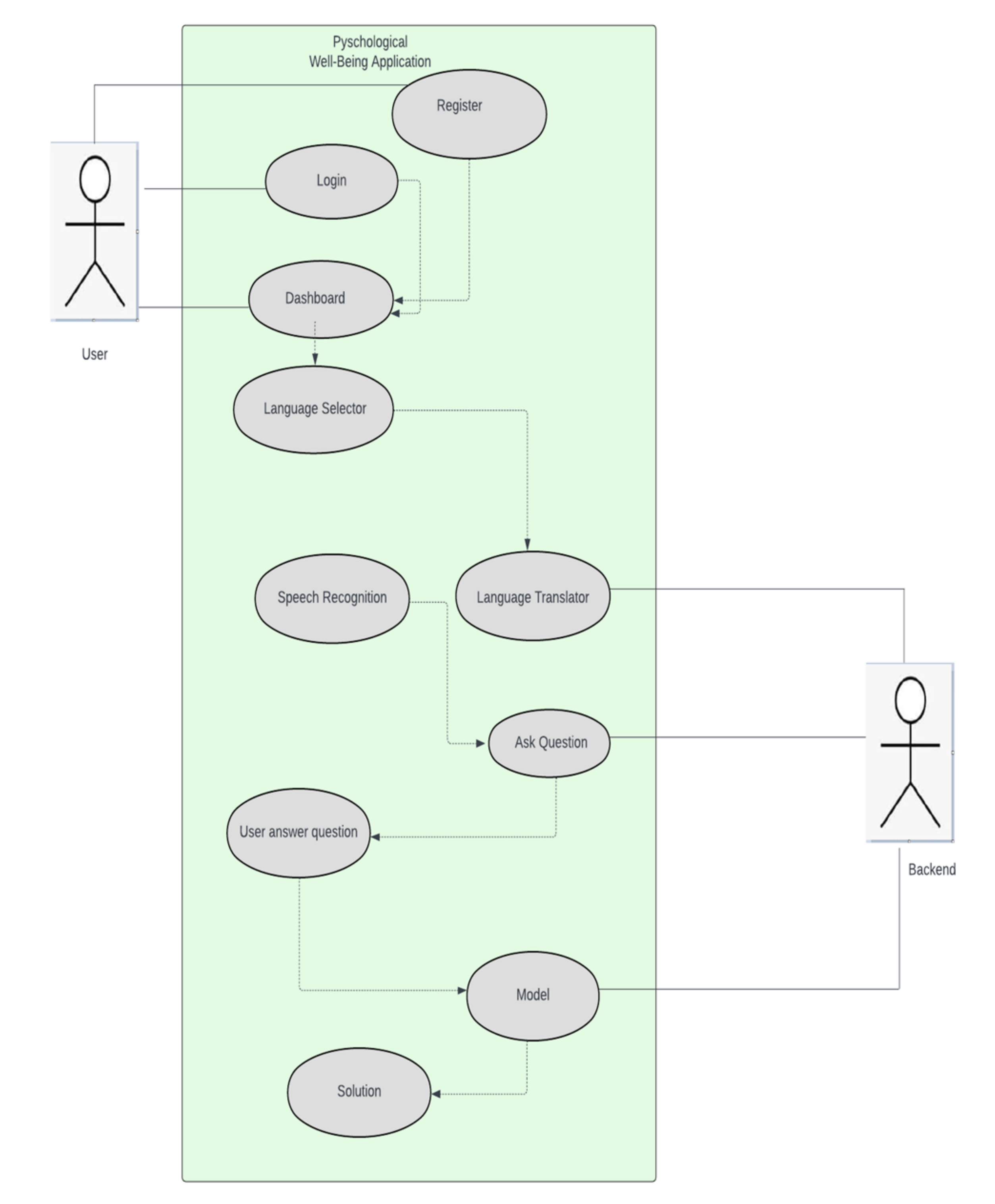


Figure 4.3: Use Case Diagram

4.4 CLASS DIAGRAM

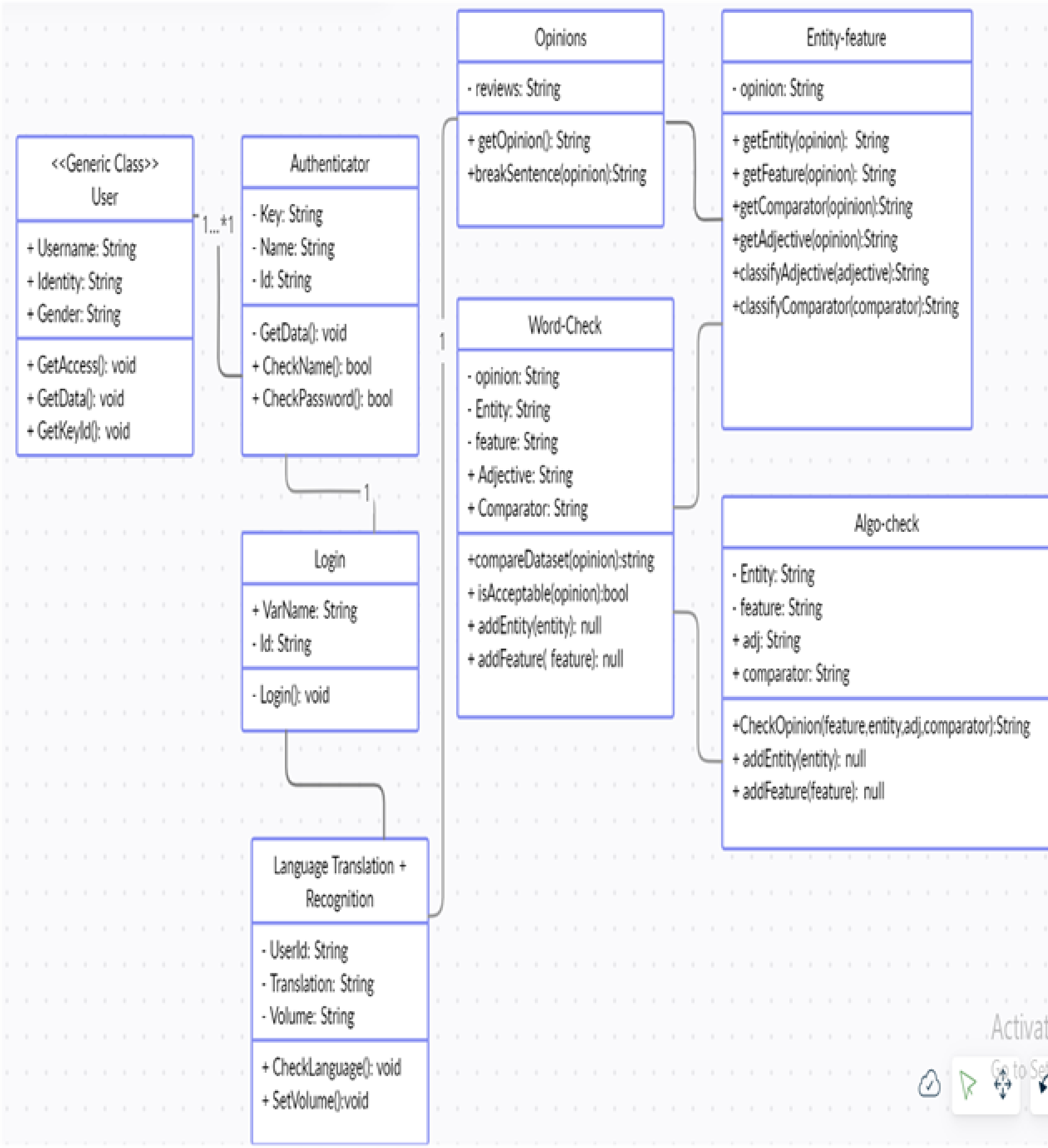


Figure 4.4: Class Diagram

4.5 ACTIVITY DIAGRAM

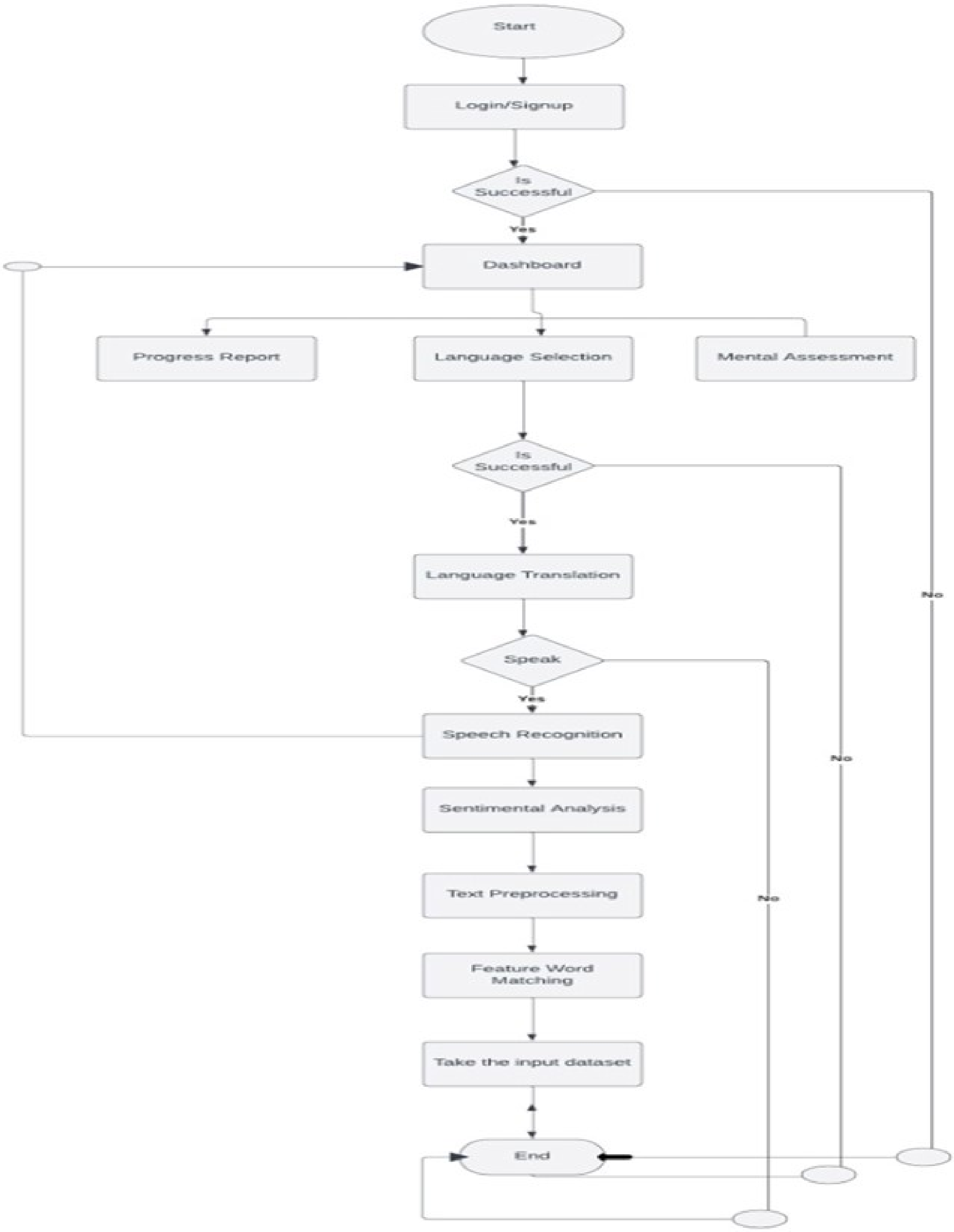


Figure 4.5: Activity Diagram

Figure 4.6: Activity Diagram 2

4.6 SEQUENCE DIAGRAM

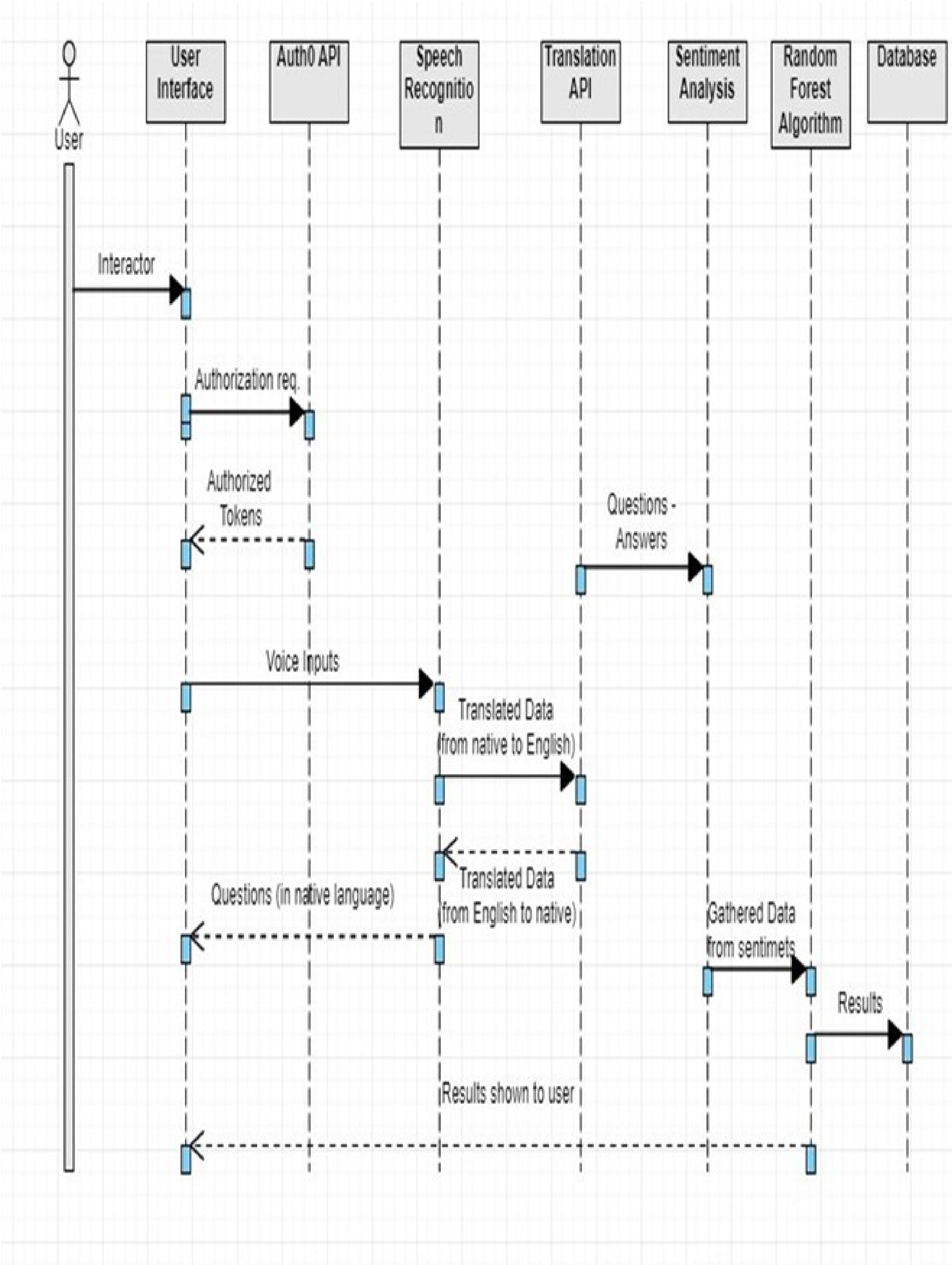


Figure 4.7: Sequence Diagram

4.7 DATA FLOW DIAGRAM

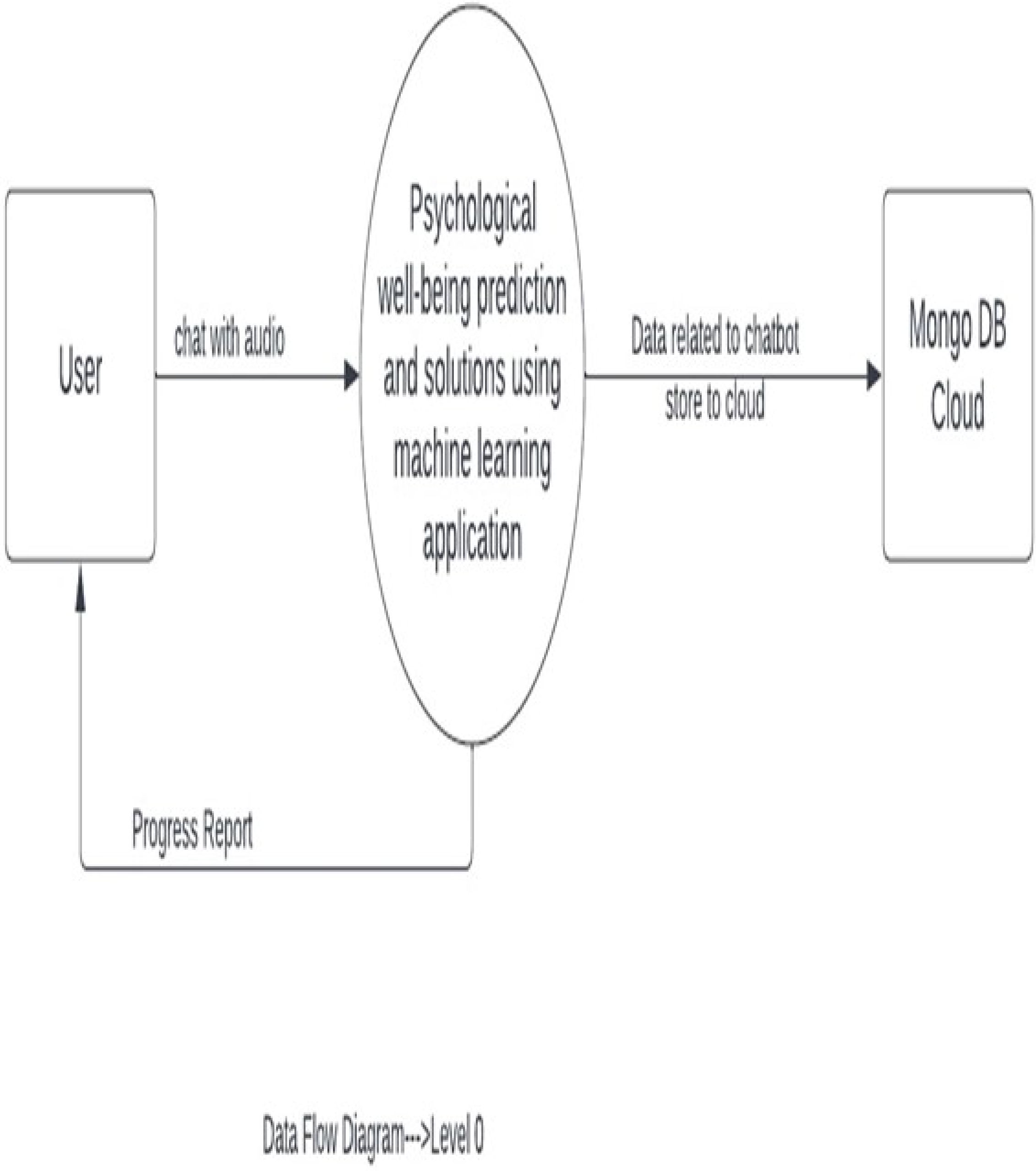


Figure 4.8: Data Flow Diagram Level 0

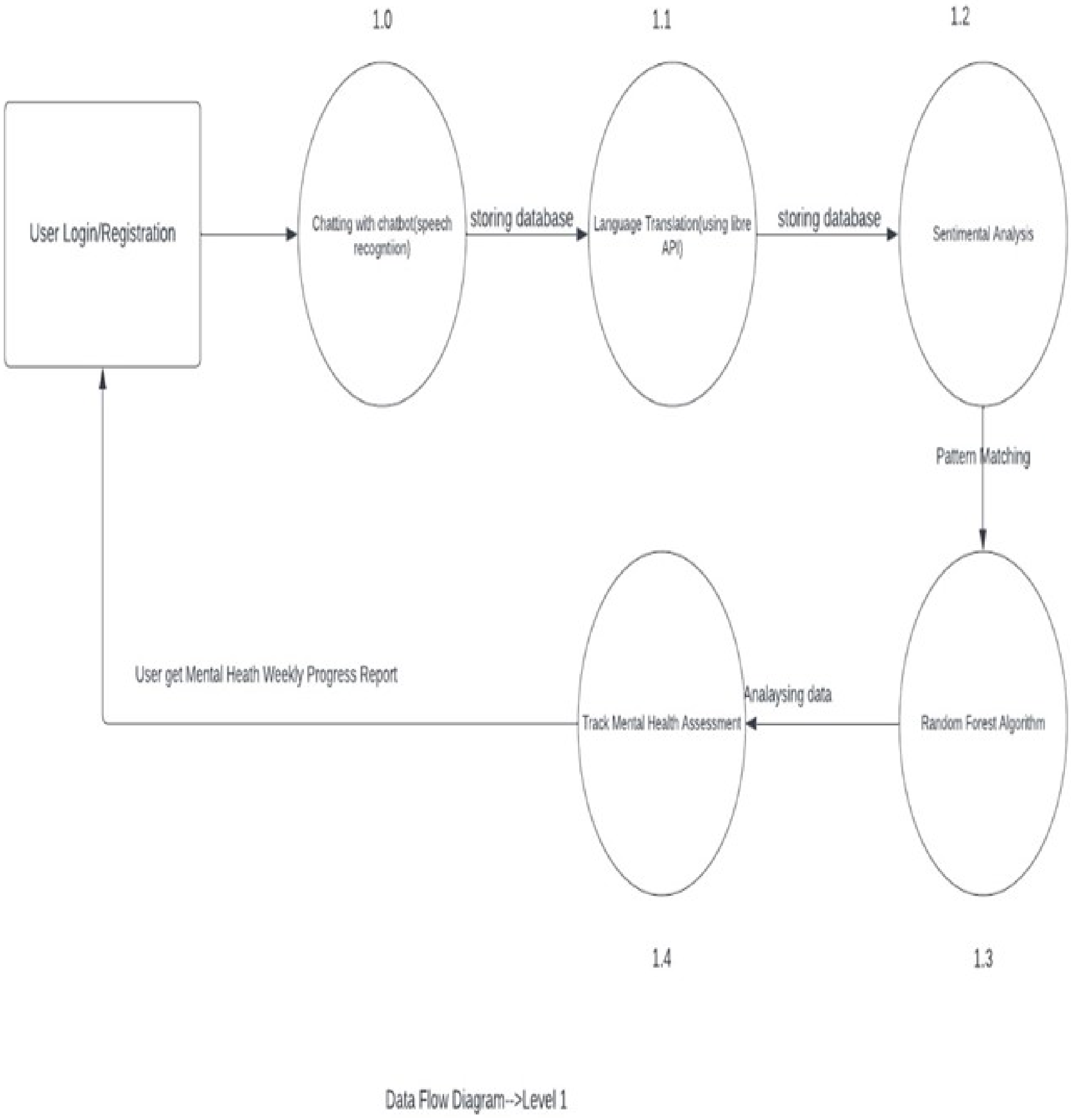


Figure 4.9: Data Flow Diagram Level 1

4.8 STATE DIAGRAM

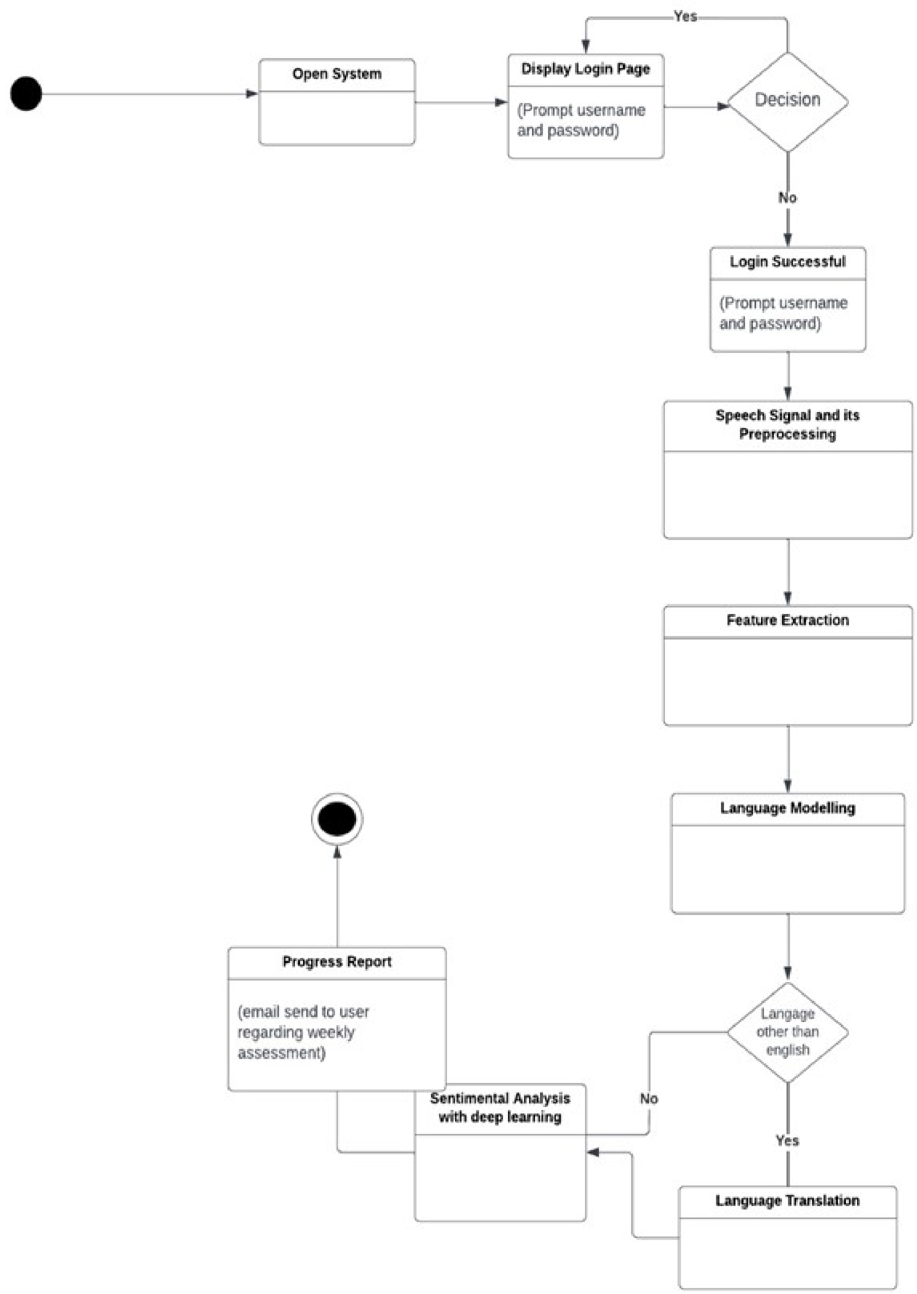


Figure 4.10: State Diagram

4.9 DEPLOYMENT DIAGRAM

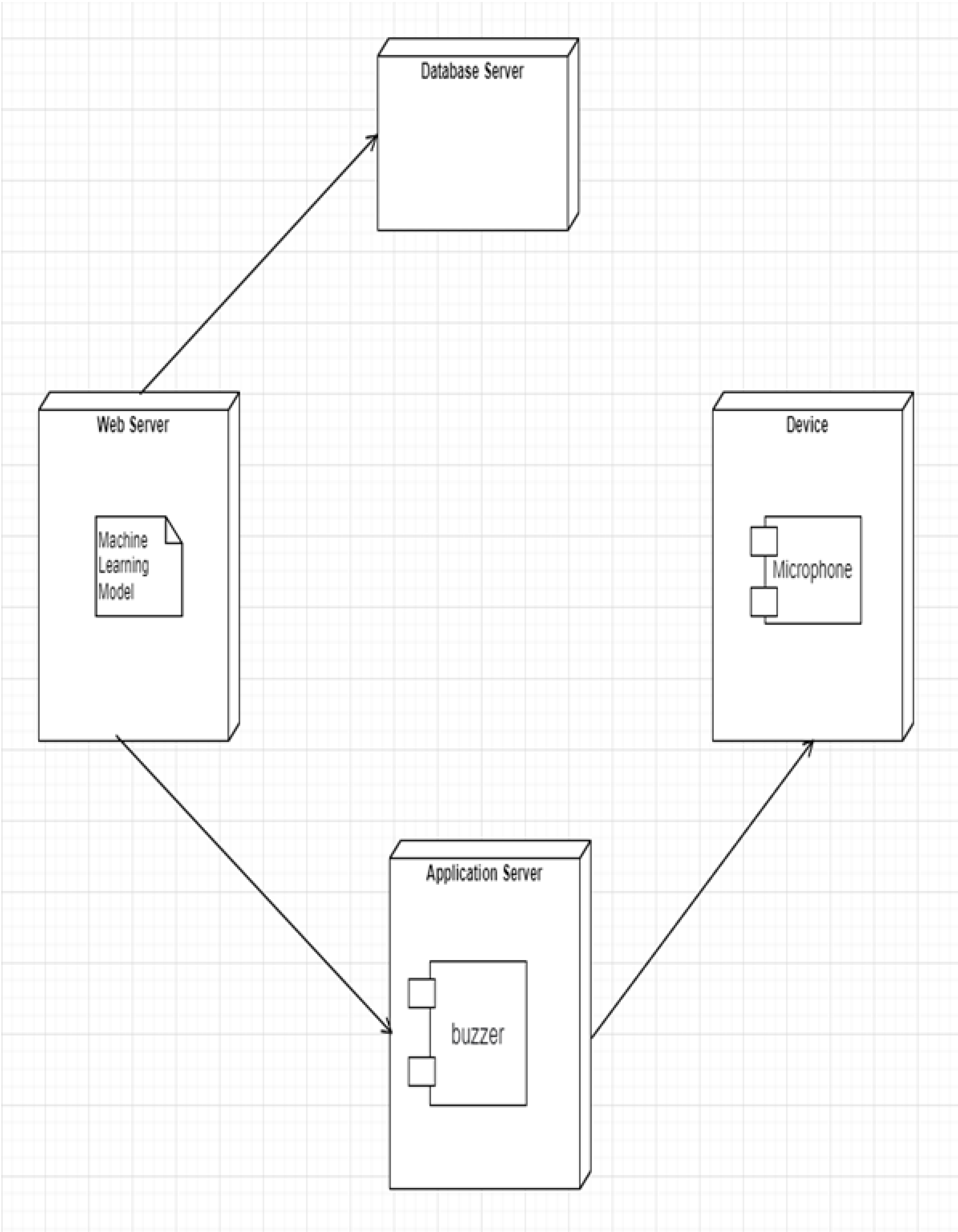


Figure 4.11: Deployment Diagram

4.10 SYSTEM FLOW CHART

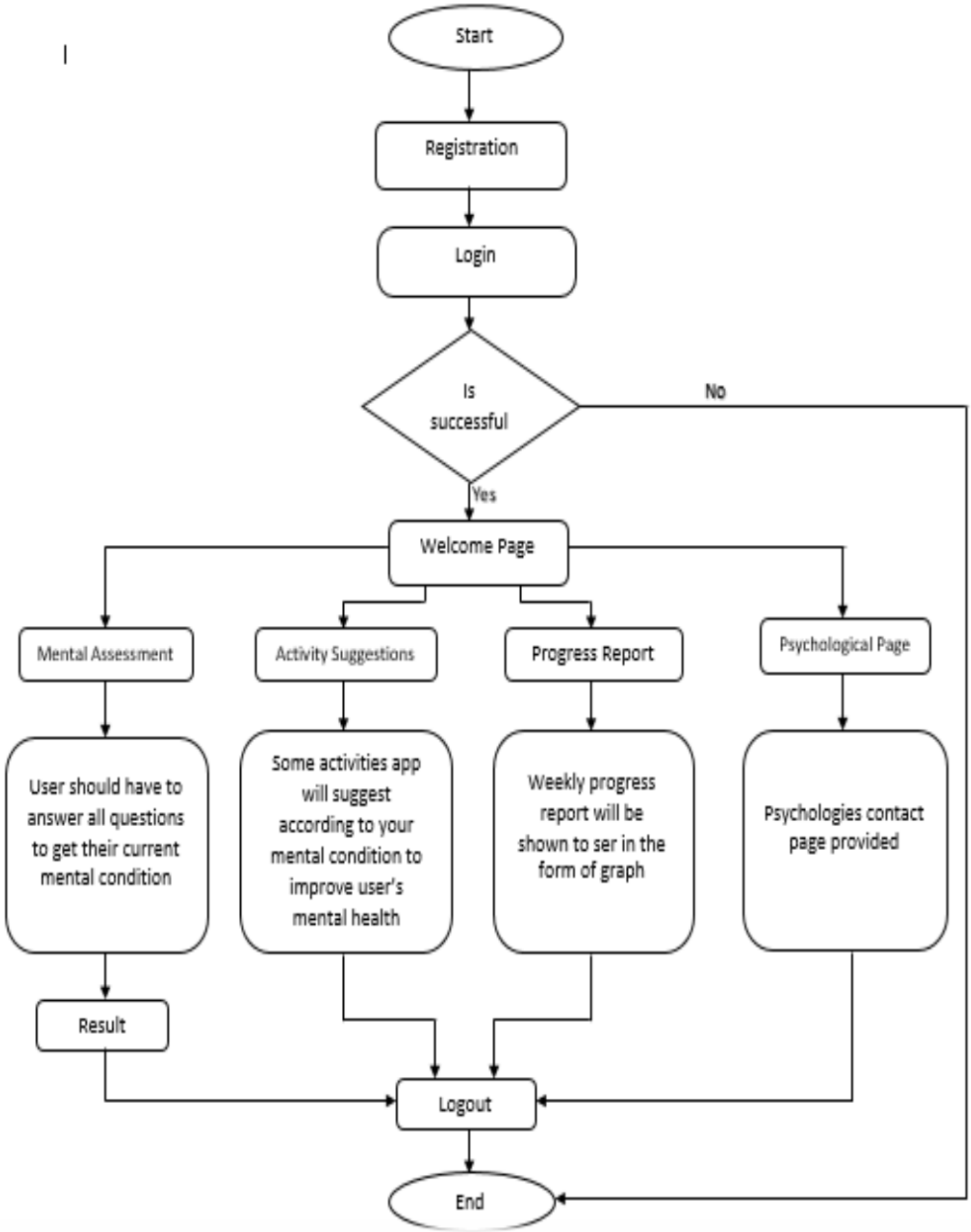


Figure 4.12: System Flow Chart

4.11 BLOCK DIAGRAM OF MODEL

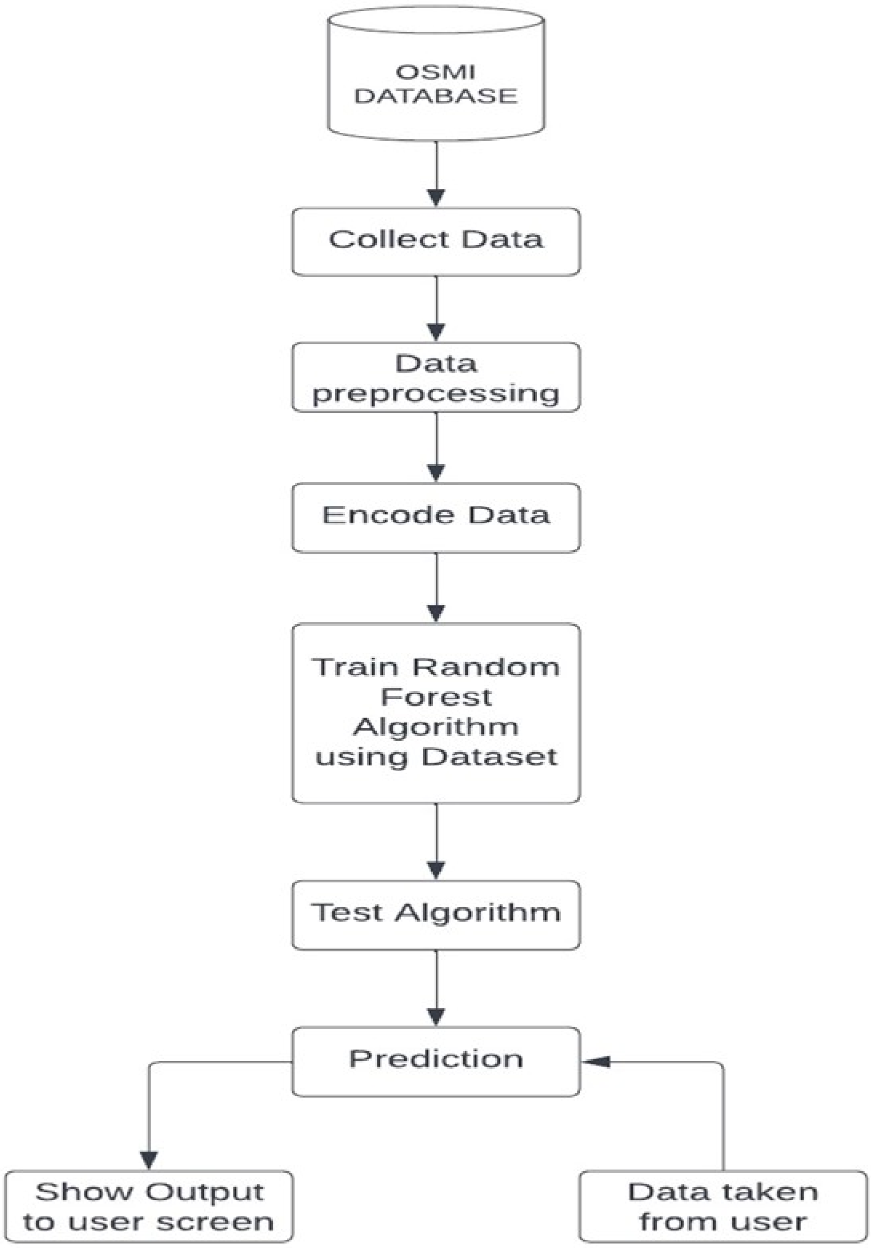


Figure 4.13: Model Block Diagram

Chapter 5

PROJECT PLANNING

5.1 PROJECT TASKS

Task 1: Create machine learning model for prediction.

Task 2: To train the machine learning model.

Task 3: To provide a simplistic USER-INTERFACE.

Task 4: To test the entire system.

5.2 RESPONSIBILITIES

a) Stay up to date on technology.

b) Creating the required software based on the latest technology.

c) Testing the software regularly for identifying bugs.

d) Training the model based on online data set

5.3 TECH-STACK COST

Software Cost Estimation is a very difficult thing to do, and no approach is going to give highly reliable results, but we have used COCOMO model to estimate the cost for this project.

Estimated lines of code 7K

Values of constant a = 3, b = 1.12, c = 2.5, d = 0.35

Effort = 26.626

Time = 7.8 months

Person Required = 3

Since, we are using a team of 4 members, so the delivery of final product is expected earlier.

Chapter 6

MATHEMATICAL MODELING

6.1 COUNT VECTORISER

With the help of this feature extraction method, text data is converted into a token count matrix. The number of words in a document is determined using the following formula:

Count (w, d) is the number of times the word ”w” appears in the text ”d.”

Features used in total: 16

Data entries total: 46234.

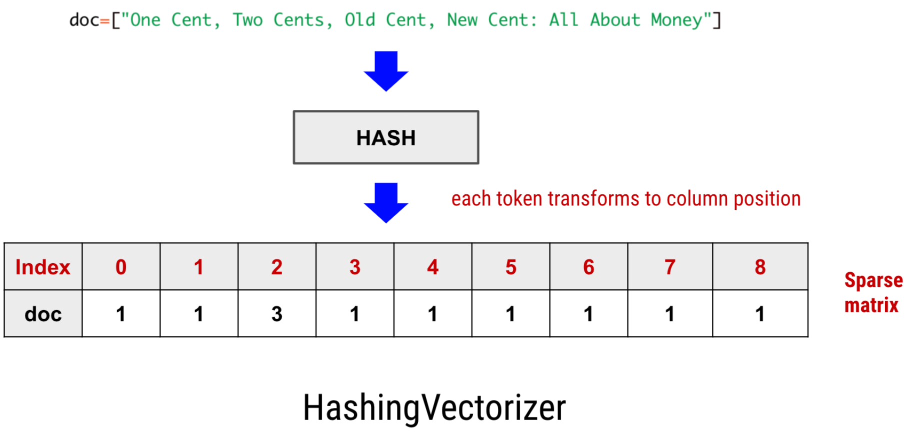


Figure 6.1: Vectoriser function5

6.2 MULTIDIMENSIONAL NAIVE BAYES

This categorization model is probabilistic. It is assumed that given the class, the features—in this example, the words—are conditionally independent.

The Naive Bayes equation is as follows:

code:

P(c—x) equals P(x—c) * P(c) / P(x).

where:

class (in this instance, Positive or Negative). (In this example, characteristics are words)

P(c—x): Probability of class c given characteristics x (what we want to forecast).

Count Vectorizer is used to calculate P(x—c), which stands for probability of features x given class c.

The training set was used to calculate the prior probability of class c, or P(c).

P(x): probability of a feature with a constant value for the given sample

Classify any new date instance $x = (f_1, f_2, \ldots, f_n)$ as:

$$\underset{Naive\,Bayes}{C} = \underset{C}{argmax}\, P(C)P(x|C) = \underset{C}{argmax}\, P(C) \prod_{i=1}^{n} P(f_i|C)$$

To do this based on training examples, estimate the parameters from training examples:

- For each target class (hypothesis) C

  $\hat{P}(C) = estimate\, P(C)$

- For each feature value f_i of each date instance

  $\hat{P}(f_i|C) = estimate\, P(f_i|C)$

Figure 6.2: Naive Bayes[11]

6.3 SUPPORT VECTOR MACHINE

This linear classifier seeks out the hyperplane that most effectively divides the classes. SVM's formula is code: wT * x + b = 0, where:

w: weight vector x: input vector (in this example, features)

biased term

transposition of the weight vector

The formula for the decision boundary is wT * x + b = 0.

Positive samples are those with wT * x + b ¿ 0, whereas negative samples are those with wT * x + b 0.

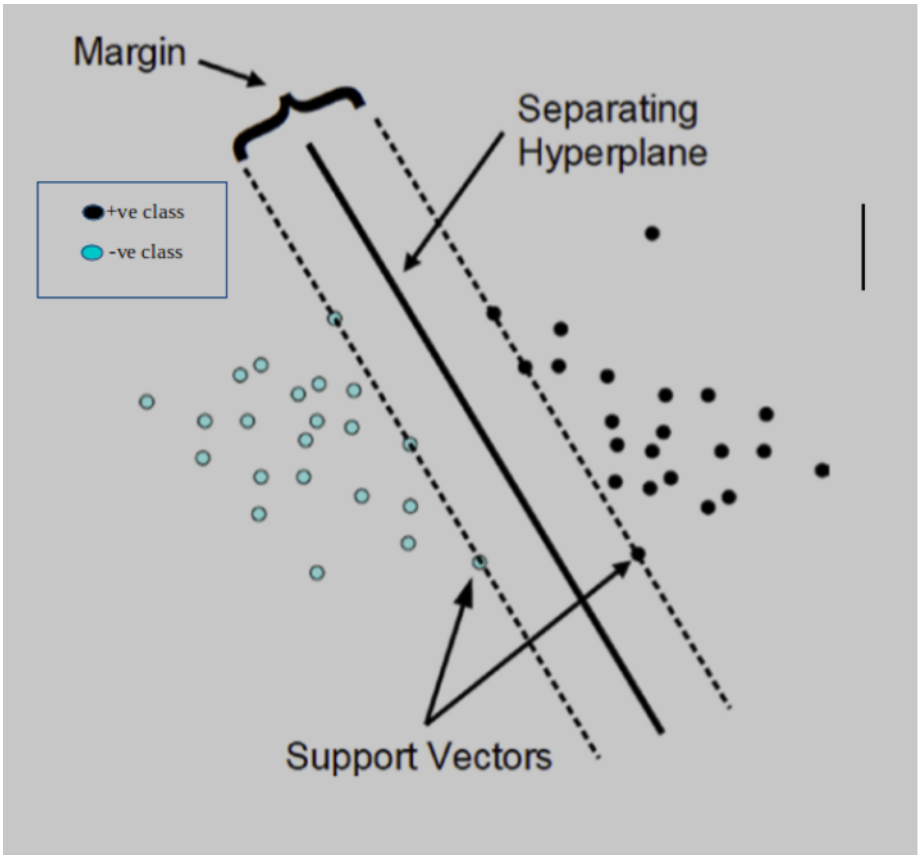


Figure 6.3: Support Vector Machine15

6.4 RANDOM FOREST ALGORITHM

This is an ensemble method where the predictions from different decision trees are combined after training. A decision tree's formula is:

code:

When a feature's value is greater than or equal to a threshold's value,

go_left()

otherwise: go_right() where:

feature_value: The value of the currently being taken into consideration

feature. The data were divided into two subsets (left and right) using the value _value.

Recursively use the same procedure on the left subset with go_left().

Apply the same procedure iteratively to the right subset using go_right().

The sum of several decisions determines the decision boundary.

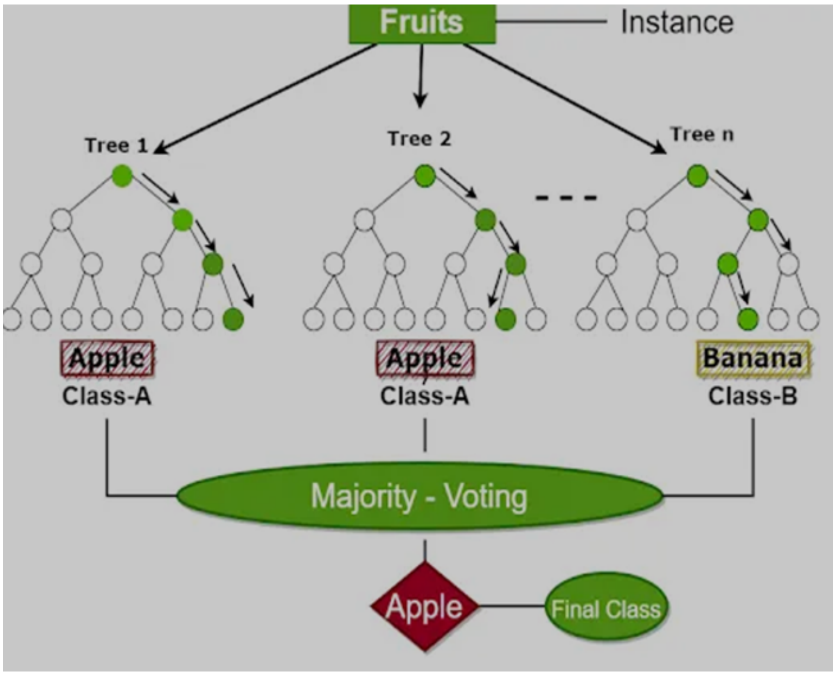


Figure 6.4: Random Forest18

6.5 RECURRENT NEURAL NETWORK

This neural network architecture is effective at handling data sequences (in this example, word sequences). An LSTM cell's formula is as follows:

code:

i_t = sigmoid(W_i * [h_t-1, x_t] + b_i)

f_t = sigmoid(W_f * [h_t-1, x_t] + b_f)

o_t = sigmoid(W_o * [h_t-1, x_t] + b_o)

g_t = tanh(W_g * [h_t-1, x_t] + b_g)

c_t = f_t * c_t-1 + i_t * G_t h_t = o_t * tanh(c_t)

where:

x_t: input vector at time step t (in this instance, the current word is embedded in the vector).

h_t-1: hidden state at time step t-1 (result of the LSTM cell before it)

i_t: input gate (decides how much of the input should be allowed through).

f_t: forget gate (chooses how much of the preceding cell state should be ignored).

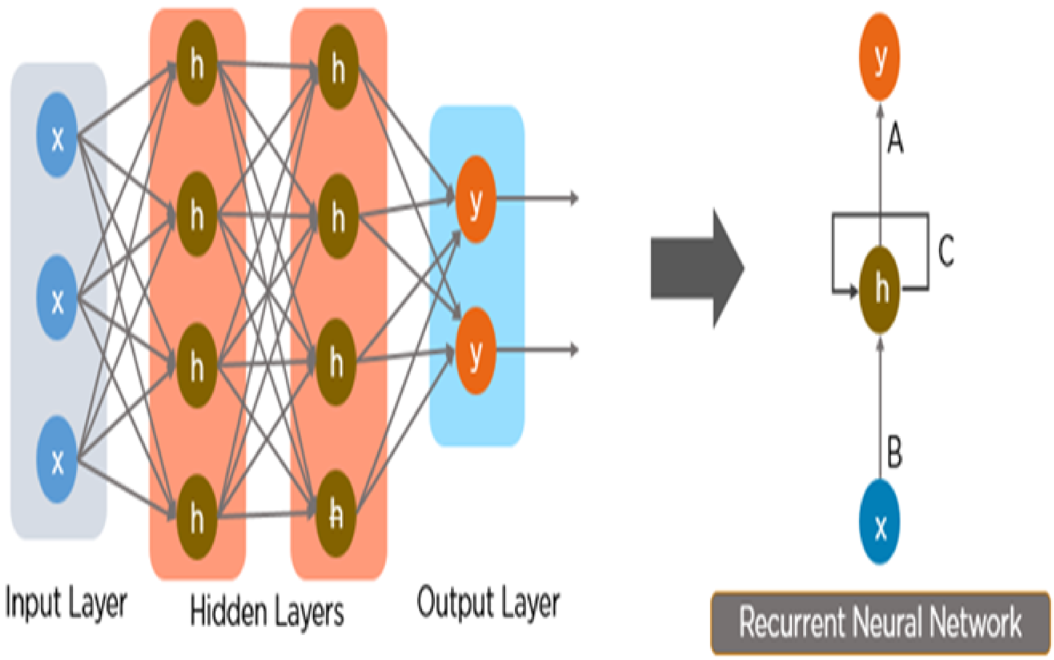


Figure 6.5: RNN30

Chapter 7

IMPLEMENTATION

7.1 DATA COLLECTION AND PREPROCESSING

we have collected dataset from Kaggle "OSMI Tech Survey 2016". The dataset contain 46234 rows and 64 columns. The data has record of past mental health of employees and their current and previous working conditions. This dataset is valuable for our project as we can train our machine learning models on this dataset to make accurate predictions for any mental health disorder. After collecting the dataset, it was cleaned and preprocessed to give better results. The null values present in dataset were replaced with the mode value of that column values.

mental_df[col].mode() function was used to calculate the mode value.

mental_df[col].fillna(mode, inplace=True) function fills the missing values in the dataset with the calculated mode values.

Total 15 features were manually selected to train model as they are the most relevant ones for our study.

s.no	features
1	Are you self-employed
2	Do you work remotely
3	Have you had a mental health disorder in the past
4	Do you believe your productivity is ever affected by a mental health issue
5	Do you have a family history of mental illness
6	Do you feel comfortable in your working environment
7	Do you feel comfortable working with your direct supervisor
8	Do you feel that your organisation takes mental health as seriously as physical health
9	Have you observed or experienced an unsupportive or badly handled response to an issue i
10	Are you stressed about your career
11	How willing would you be to share with friends and family about your work stress
12	Do you currently have a mental health disorder
13	Have you observed or experienced an unsupportive or badly handled response to an issue i
14	Did you feel that your previous employers took mental health as seriously as physical healt
15	Have your previous employers provided mental health benefits

In the dataset, few columns contain employee name, age and department information which is not relevant to our study so those columns are not used for prediction and are simply dropped.

The mapping function given below is used to convert answers given by user into numerical values -1,0,1 which will later be used to train the model. mapping_dict = 'yes': 1, 'no': -1,'Yes': 1, 'No': -1, "I don't know": 0, 'not eligible': 0, I am not sure': 0, 'Maybe': 0, 'Yes, I know several': 1,'Yes, I think it would':1,'Unknown':0,'I know some': 1, ' no': -1, "I don't know any": 0, 'Not applicable to me': 0, 'No, because it doesn't matter': -1, 'Sometimes, if it comes up': 1, "I'm not sure": 0, 'No, none did': 0, 'Yes, they all did': 1, 'Some did': 1,'N/A (not currently aware)': 0, 'I was aware of some': 1,'None did': -1, 'Yes, always': 1, 'Some of them': 1, 'Yes, all of them': 1, 'Some of my previous employers': 1,'Somewhat difficult':-1,'No, at none of my previous employers': -1, "No, I don't know any":0,"Unsure":0,'Very easy':1,'Somewhat easy':1,'Neither easy nor difficult':0, 'Yes, at all of my previous employers': 1, 'No, I don´t think they would': -1, 'Yes, they do': 1, 'Yes, I think they would': 1, 'Not applicable to me (I do not have a mental illness)': 0, 'Somewhat open': 1, 'Neutral': 0, 'Very open': 1,

”Never”:-1,”Often”:1,”Yes, it has”:1,”Not open at all”:-1, Somewhat not open’: -1, ’Maybe/Not sure’: 0 ’Yes, I experienced’: 1, ’Yes, I observed’: 1, ’Very difficult’:-1, ”No, they do not”:-1, ’No, it has not’:-1,’No, they have not’:-1,’Often’: 1, ’Not applicable to me’: -1, ”No, I don’t think it would”:-1, ”No, I only became aware later”:-1, ”No, I only became aware later”:-1,”No, because it would impact me negatively”:-1, ’Rarely’: -1, ’Sometimes’: 1, ’Always’: 1,”i don’t know”:0,’Not eligible for coverage /N/A’:0,’None of them’:-1,’Yes, I was aware of all of them’:1

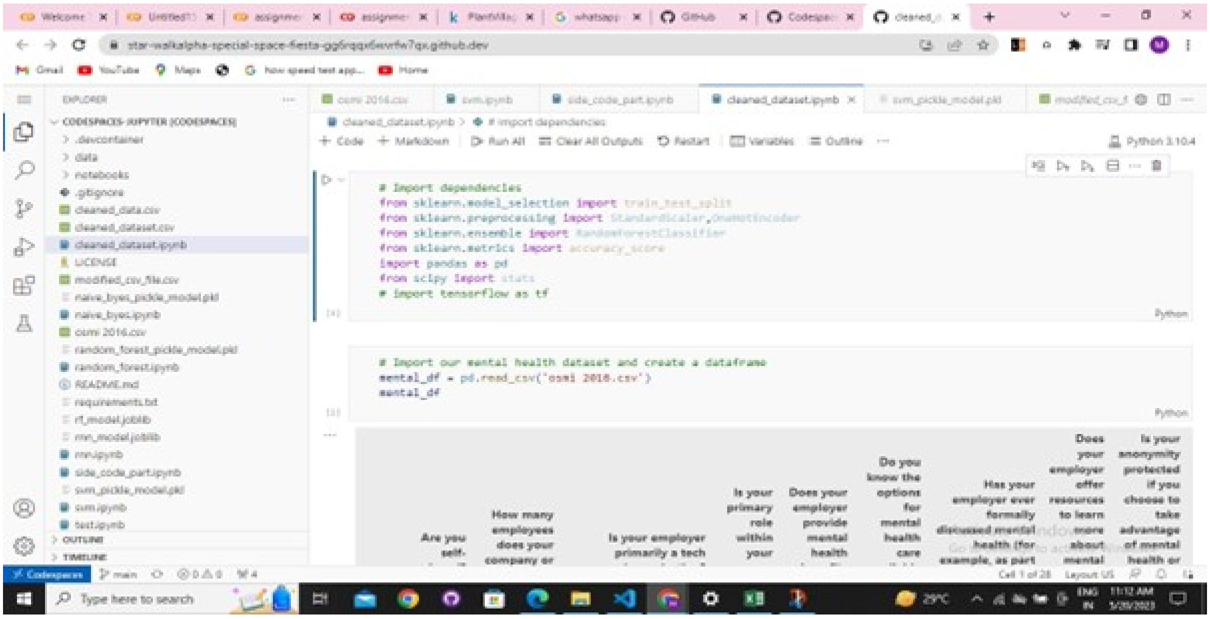


Figure 7.1: Dataset

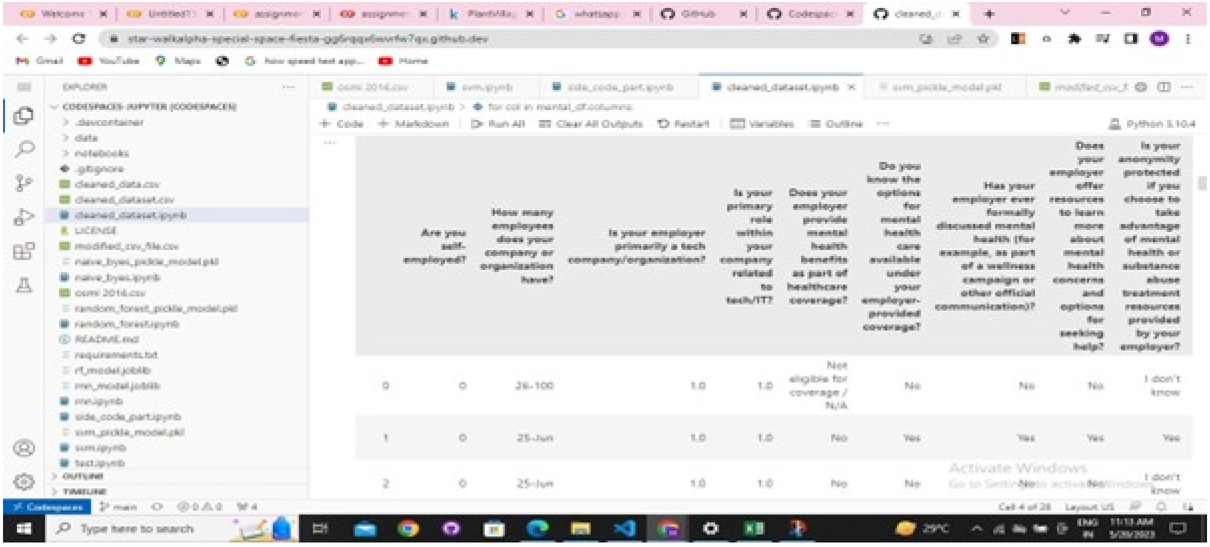


Figure 7.2: Mapping Of Dataset

7.2 TRAINING OF MODEL

We have trained four machine learning models: Random Forest, Support Vector Machine, Naïve Bayes, Recurrent Neural Network.

X: This variable often denotes our dataset's independent variables or input features.

y: This variable often represents our dataset's target variable, or dependent variable, which we are attempting to predict or categorise.

The dataset is split into training and testing set. Here we are using 75% dataset as training set and remaining 25% dataset as testing set.

7.3 RANDOM FOREST

Using supervised classification, the random forest technique can be used. This algorithm builds a forest with many trees, as the name would imply. The look of the forest is often stronger the more trees there are in the forest. The random forest classifier works in a similar fashion, with higher accuracy results being produced by more trees in the forest. If you're familiar with the decision-tree algorithm. The question of whether or not we should be developing more decision trees, as well as how to do so, may be on your mind. due to the fact that the same dataset will be used for all nodes selection calculations. Yes. Truly, you are. You are not going to develop a forest by modelling more decision trees.

Random Forest pseudocode:

1. Pick ”k” features at random from a total of ”m” features

2. Use the optimum split point to determine the node ”d” among the ”k” features.

3. Use the optimum split to divide the node into daughter nodes.

4. Up until the ”l” number of nodes, repeat steps 1 through 3 as necessary.

5. To produce a ”n” number of trees, repeat steps 1 through 4 a ”n” number of times.

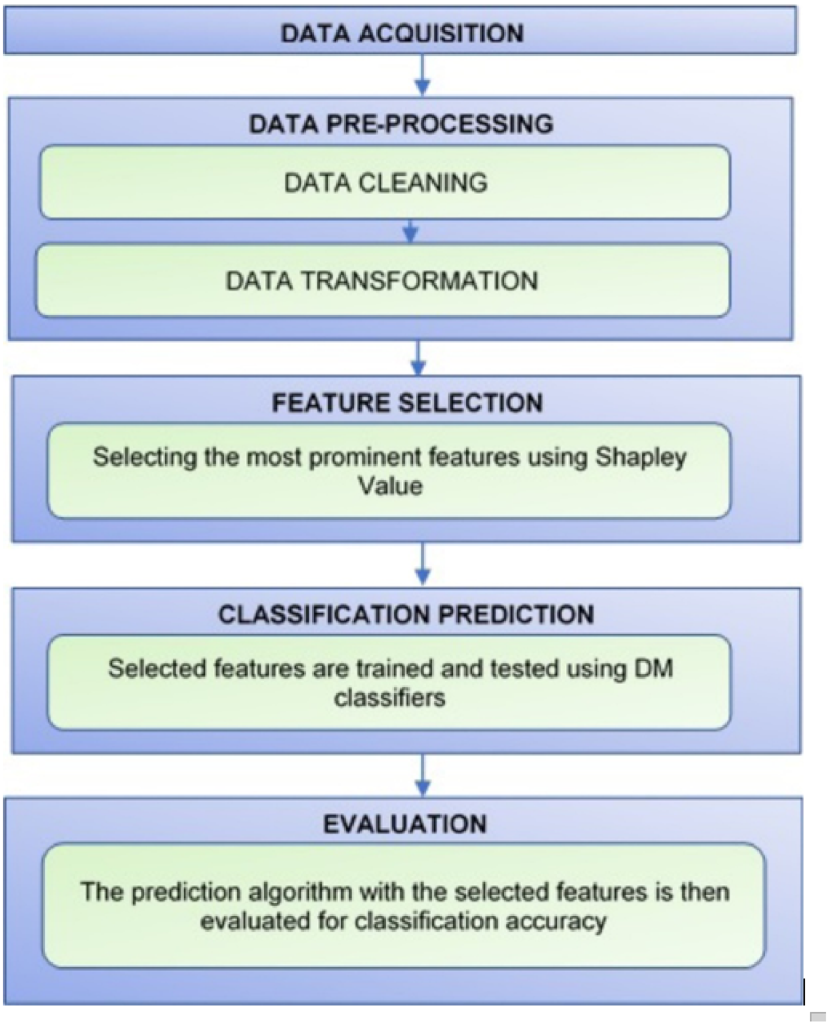


Figure 7.3: Flow Of ML Model

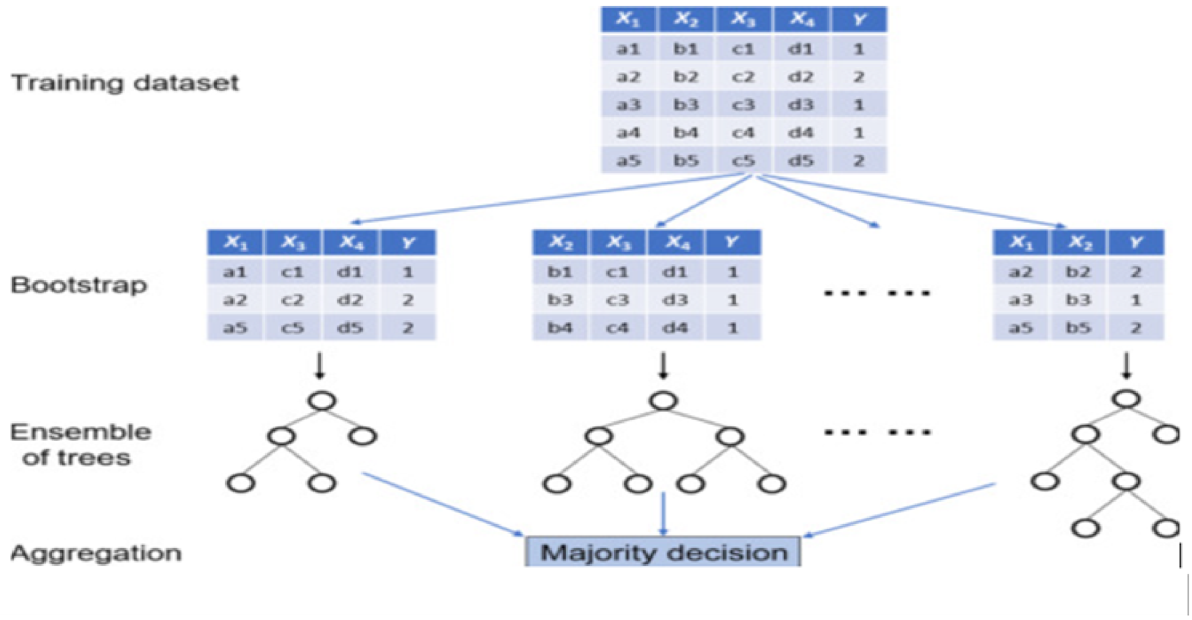


Figure 7.4: Working Of RF17

The next step is to use the "k" features that were chosen at random to discover the root node using the best split method.

We will calculate the daughter nodes in the following stage using the same optimal split methodology. Will the first three stages up until we create a tree with the target as the leaf node and the root node.

Then, to build "n" randomly generated trees, we repeat steps 1 through 4. The random forest is made up of these arbitrary trees.

Because it was trained on a random subset of the data, the random forest classifier is less likely to overfit. By creating several decision trees, each based on a random subset of the data, the random forest classifier is less prone to overfit. Because it averages the results of various decision trees, the random Forest classifier has a lower likelihood of being overfit.

7.4 SUPPORT VECTOR MACHINE

One of the most well-liked algorithms for supervised learning is called the Support Vector Machine(SVM) and it is used to solve both classification and regression issues. It is largely utilised in Machine Learning Classification issues, though.

Data : Dataset with p^* variables and binary outcome.
Output: Ranked list of variables according to their relevance.

Find the optimal values for the tuning parameters of the SVM model;
Train the SVM model;
$p \leftarrow p^*$;
while $p \geq 2$ **do**
  $SVM_p \leftarrow$ SVM with the optimized tuning parameters for the p variables and observations in **Data**;
  $w_p \leftarrow$ calculate weight vector of the SVM_p $(w_{p1}, \ldots, w_{pp})$;
  $rank.criteria \leftarrow (w_{p1}^2, \ldots, w_{pp}^2)$;
  $min.rank.criteria \leftarrow$ variable with lowest value in $rank.criteria$ vector;
  Remove $min.rank.criteria$ from **Data**;
  $Rank_p \leftarrow min.rank.criteria$;
  p $\leftarrow$ p - 1 ;
end
$Rank_1 \leftarrow$ variable in **Data** $\notin (Rank_2, \ldots, Rank_{p^*})$;
return $(Rank_1, \ldots, Rank_{p^*})$

Figure 7.5: SVM_{23}

The SVM algorithm's objective is to establish the best decision boundary or line that can divide n-dimensional space into classes so that subsequent data points can be quickly assigned to the appropriate category. The term "hyperplane" refers to this optimal decision boundary. In order to create the hyperplane, SVM selects the extreme points and vectors. Support vectors, which are used to represent these extreme instances, are what give the Support Vector Machine method its name.

Linear SVM: Linear SVM is used for linearly separable data, which is defined as data that can be divided into two classes by means of a single straight line. Linear SVM classifiers are used for such data.

Non-linear SVM: Non-Linear SVM is used for not linear separated data, which implies that if a dataset cannot be classified using a straight line, it is considered to be non-linear data, and the classifier employed is referred to as a Non-linear SVM classifier.

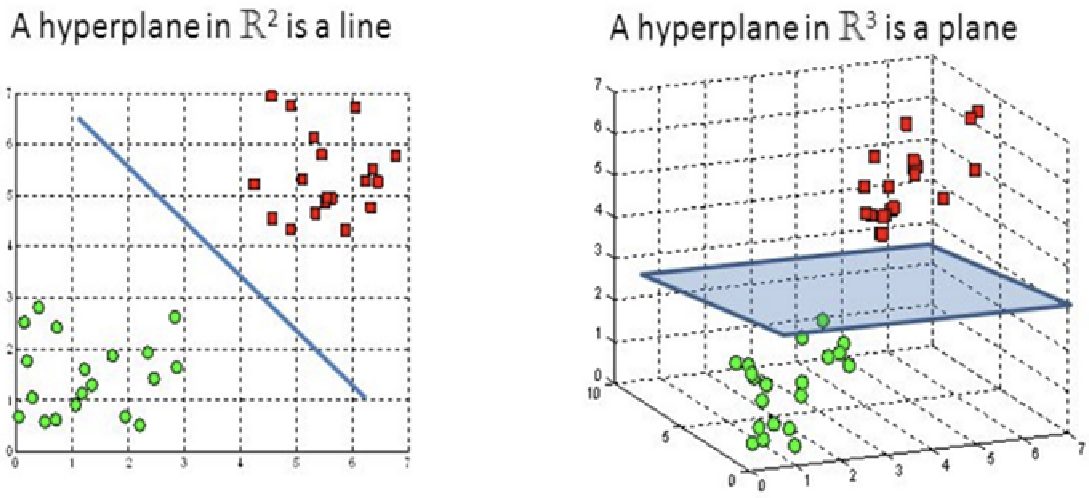


Figure 7.6: SVM Hyperplane$_{23}$

7.5 NAÏVE BAYES

By using sequential events, where later knowledge is learned and affects the original probability, Bayes' Theorem stands out. The terms "prior probability" and "posterior probability" are used to describe these probabilities. The initial likelihood of an event before it is contextualised in a particular scenario is known as the prior probability, also known as the marginal probability. The chance of an event following the observation of a piece of data is known as the posterior probability.

In contrast to other classifiers, nave Bayes classifiers operate under a few fundamental presumptions, hence the term "nave." It makes the supposition that a Naive Bayes model's predictors are conditionally independent of one another or unconnected to any of the other features. Additionally, it is predicated on the idea that each element influences the result equally. Although these presumptions are frequently broken in real-world situations, they make categorization problems more manageable from a computational perspective. Meaning that each variable will now only need one probability, making the computation of the model simpler. The classification algorithm performs well, especially with small sample sizes, despite this false independence assumption.

Input:

Training dataset T,

F= (f_1, f_2, f_3,.., f_n) // value of the predictor variable in testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

   Until the probability of all predictor variables (f_1, f_2, f_3,.., f_n) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

Figure 7.7: NB Pseudocode[15]

7.6 RECURRENT NEURAL NETWORK

For time series analysis, supervised deep learning is employed using RNN. One of the most sophisticated algorithms in the field of supervised deep learning is the recurrent neural network.

RNN and the frontal lobe:

Similar to short-term memory are RNNs. We will discover that they are able to recall recent events from the prior few observations and use that information moving ahead. One of the Frontal Lobe's roles in humans is short-term memory.

The concept of RNN

Weights are said to have Long Term Memory, often known as LTM. For instance, the weights in a classical ANN are known, thus no matter what input we give it, it will process in the same way as it did yesterday. Given that the human brain's temporal lobe controls long-term memory (LTM), the weights can be found there.

Bidirectional Recurrent Neural Network (BRNN)Algorithm
1. *Ino= Number of input layers* 2. *Hno= Number of hidden layers* 3. *Ono= Number of output layers* 4. *S= number of data set instances* ***Forward pass*** 5. *for i=1 to Hno* 6. *for j=1 to S* *calculating the forward pass for the forward hidden layer's activation function* h_t^f *using eq. (4)* 7. *end for* 8. *for j=S to 1* 9. *calculating the backward pass for the backward hidden layer's activation function* h_t^b *using eq. (5)* 10. *end for* 11. *end for* 12. *for i=1 to Ono* 13. *calculating the forward pass for the output layer using the previous stored activations using eq. (3)* 14. *end for* ***Backward pass*** 15. *for i= Ono to 1* 16. *calculating the backward pass for the output layer using the previous stored activations using eq. (3)* 17. *end for* 18. *for i=1 to Hno* 19. *for j=1 to S* *calculating the backward pass for the forward hidden layer's activation function* h_t^f *using eq. (5)* 20. *end for* 21. *for j=S to 1* 22. *calculating the forward pass for the backward hidden layer's activation function* h_t^b *using eq. (4)* 23. *end for* 24. *end for*

Figure 7.8: RNN Pseudocode$_{27}$

RNN is similar to short-term memory in that it can recall events from the prior couple of observations. To achieve the best result, we must locate the global minimum throughout the gradient descent. When information is transmitted through an RNN, there are several errors to assess, and we must keep in mind that the blue nodes are not only neurons but also an entire hidden layer. Now, the weights of every single neuron involved in the output computation should be changed in order to reduce error. The Weight Recurring in the temporal loop must be updated frequently, which causes them to shrink.

Recurrent neural networks are adjusted to create Long Short-Term Memory (LSTM) networks, which facilitate better memory retention for prior information. Here, the RNN's vanishing gradient issue is fixed. The classification, processing, and prediction of time series with uncertain time lags are all excellent applications for LSTM. The model is trained via back-propagation. There are three gates in an LSTM.

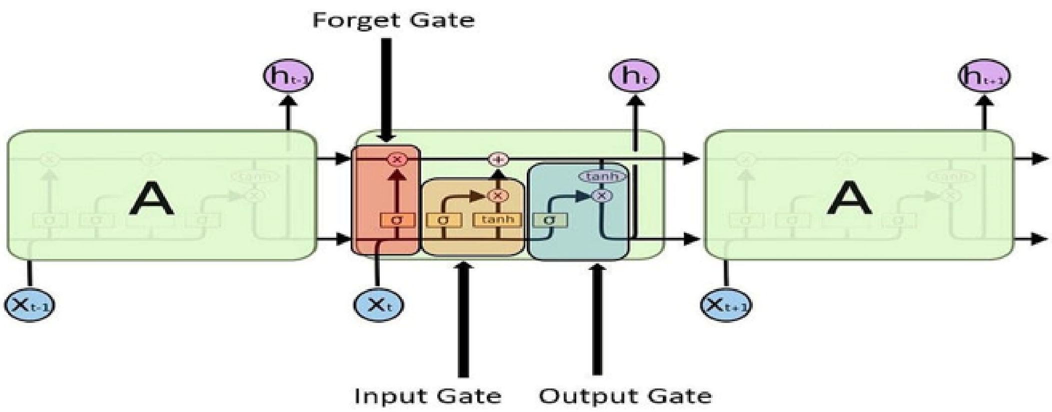


Figure 7.9: RNN With LSTM13

7.7 TESTING THE MODEL

Once the training of ML model is completed, the model is then tested on remaining 25% dataset. We calculate the accuracy, precision, recall, F score for the algorithms and find the optimum algorithm for our project.

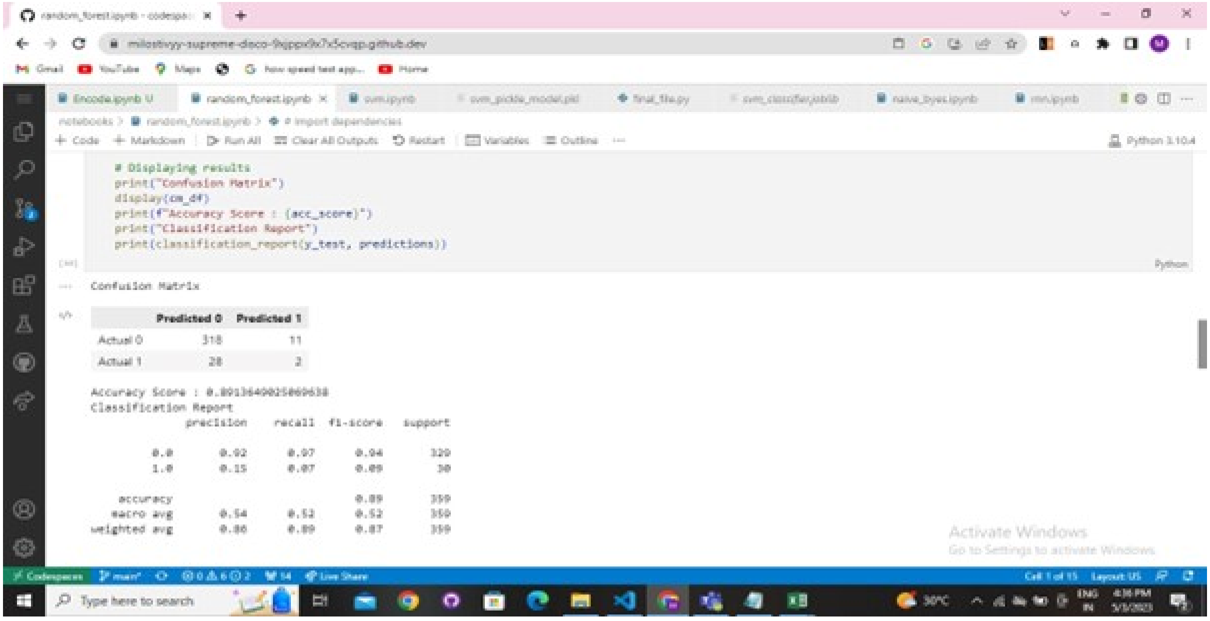


Figure 7.10: Screenshot of RF

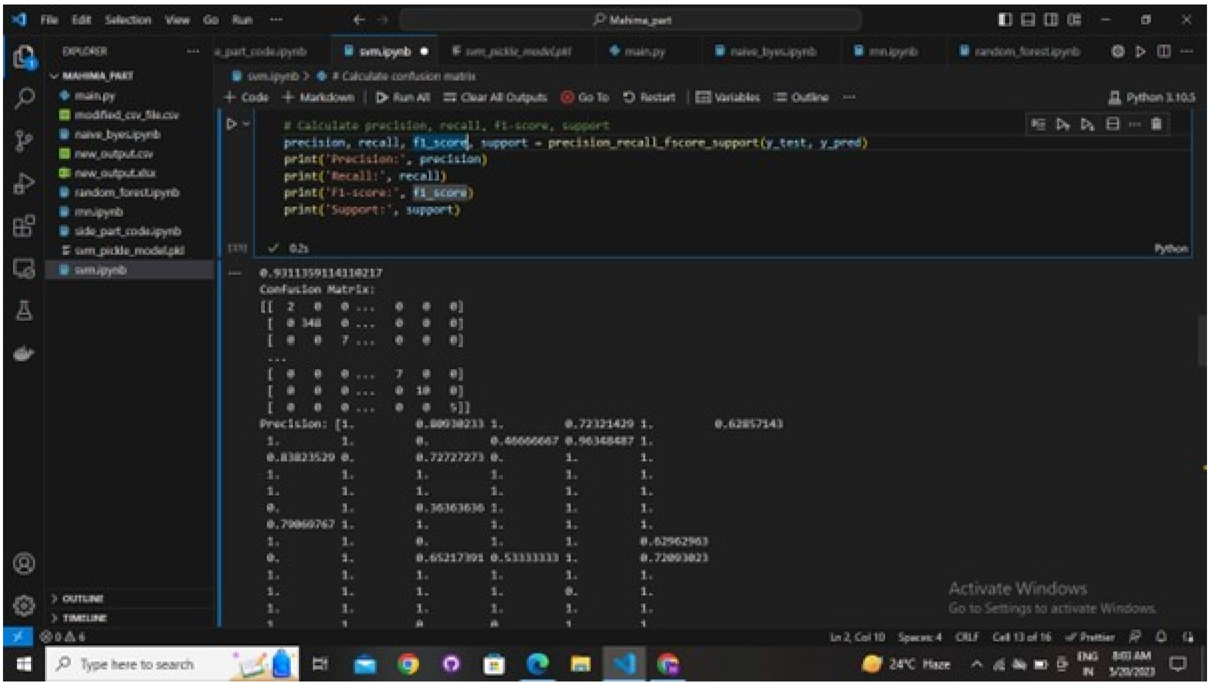


Figure 7.11: Screenshot of SVM

```
2023-05-18 01:47:04.570785: I tensorflow/core/common_runtime/executor.cc:1197] [/devic
         [[{{node gradients/split_1_grad/concat/split_1/split_dim}}]]
29/29 [==============================] - 4s 33ms/step - loss: 0.5084 - accuracy: 0.854
Epoch 2/10
29/29 [==============================] - 0s 17ms/step - loss: 0.3311 - accuracy: 0.898
Epoch 3/10
29/29 [==============================] - 0s 16ms/step - loss: 0.3297 - accuracy: 0.898
Epoch 4/10
29/29 [==============================] - 0s 15ms/step - loss: 0.3299 - accuracy: 0.898
Epoch 5/10
29/29 [==============================] - 0s 17ms/step - loss: 0.3306 - accuracy: 0.898
Epoch 6/10
29/29 [==============================] - 1s 19ms/step - loss: 0.3316 - accuracy: 0.898
Epoch 7/10
29/29 [==============================] - 0s 15ms/step - loss: 0.3311 - accuracy: 0.898
Epoch 8/10
29/29 [==============================] - 0s 15ms/step - loss: 0.3319 - accuracy: 0.898
Epoch 9/10
29/29 [==============================] - 0s 16ms/step - loss: 0.3303 - accuracy: 0.898
Epoch 10/10
29/29 [==============================] - 0s 15ms/step - loss: 0.3295 - accuracy: 0.898
9/9 [==============================] - 0s 5ms/step - loss: 0.3581 - accuracy: 0.8850
Test loss: 0.3580659329891205
Test accuracy: 0.8850173950195312
```

Figure 7.12: Screenshot of RNN

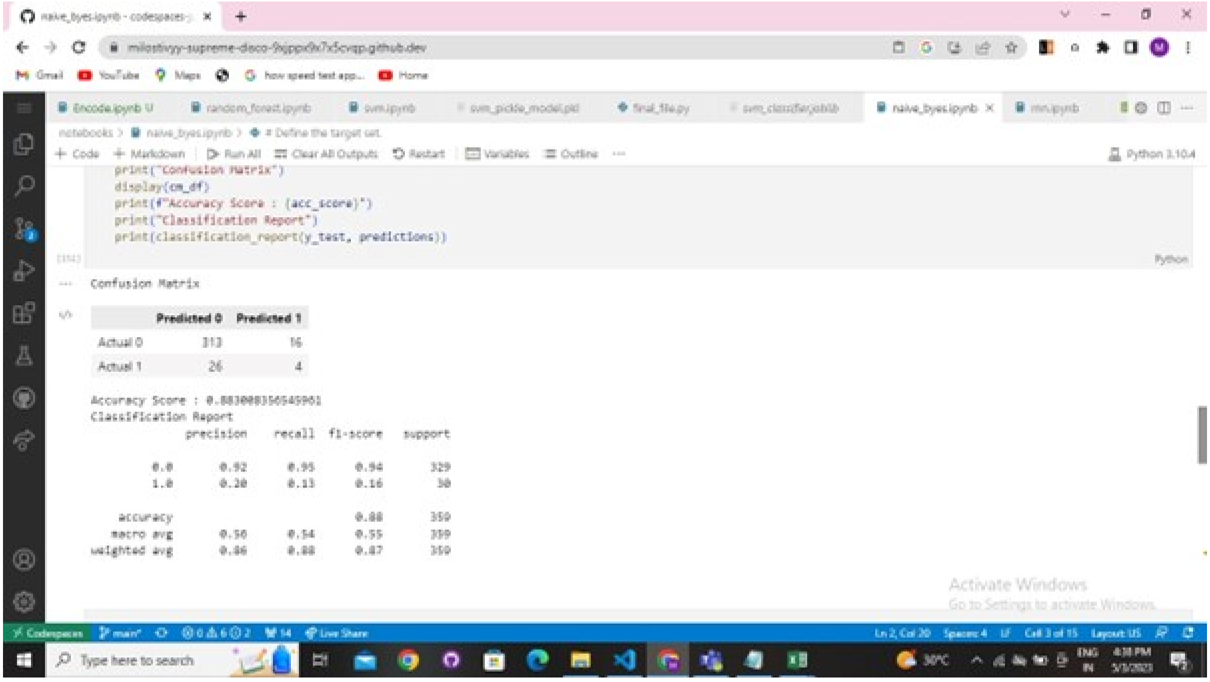


Figure 7.13: Screenshot of NB

After testing all four algorithms, it was found that support vector machine algorithm performed better than all other algorithms. Support Vector Machine was able to predict mental health disorders with an accuracy of 93.11%. Random forest proved to be the best after SVM in predicting psychological wellbeing of a person.

7.8 RESULTS

ML Algorithm	Accuracy
Support Vector Machine	93.11%
Recurrent Neural Network	88.50%
Naïve Bayes	88.30%
Random Forest	89.13%

ML Algorithm	Precision	Recall	Fscore
Support Vector Machine	92.5	93.11	92.434
Naïve Bayes	86	88	87
Random Forest	86	89	87

RNN Loss value – 35.80

7.9 USER INPUT AND OUTPUT

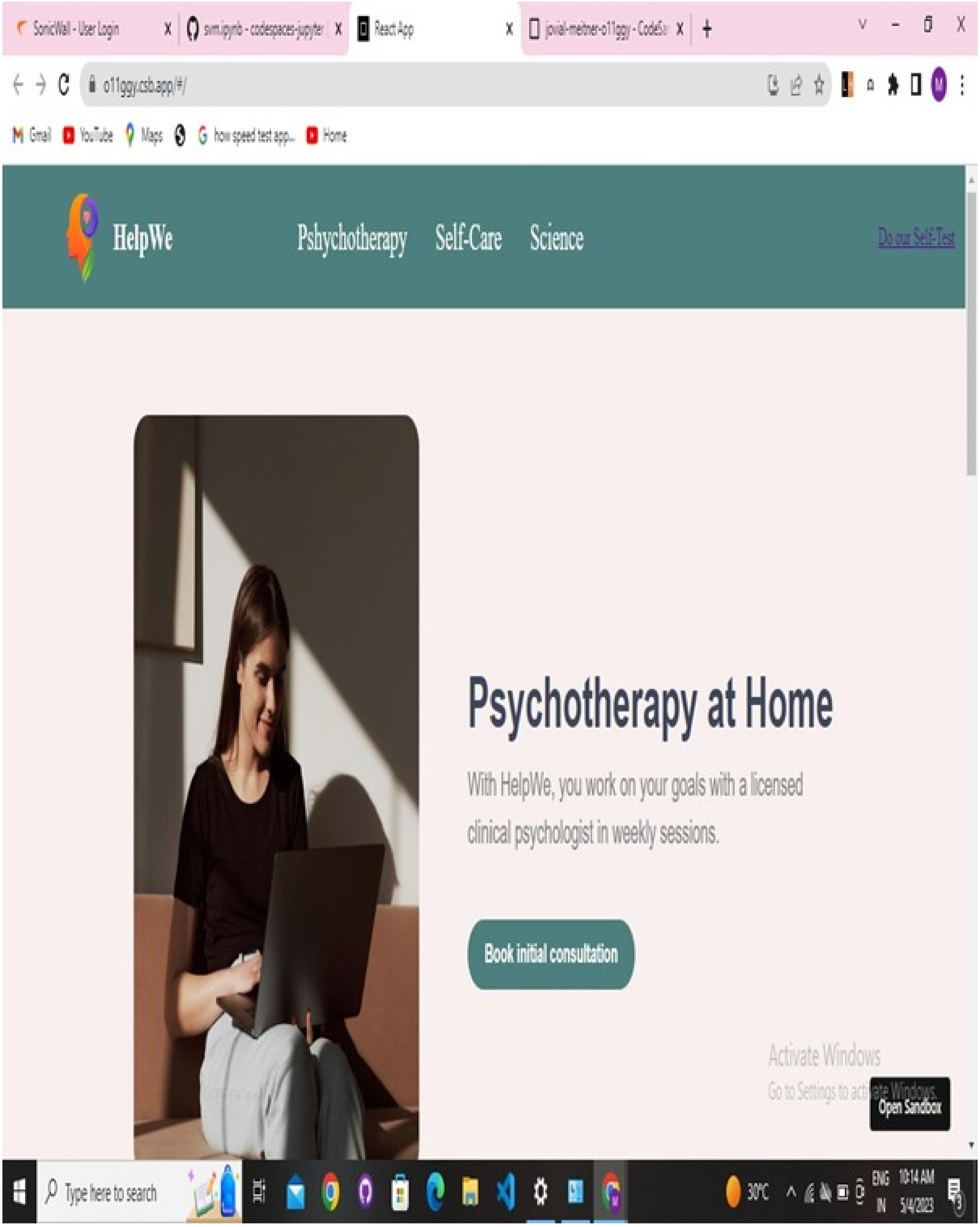


Figure 7.14: User Interface

Figure 7.15: User Input

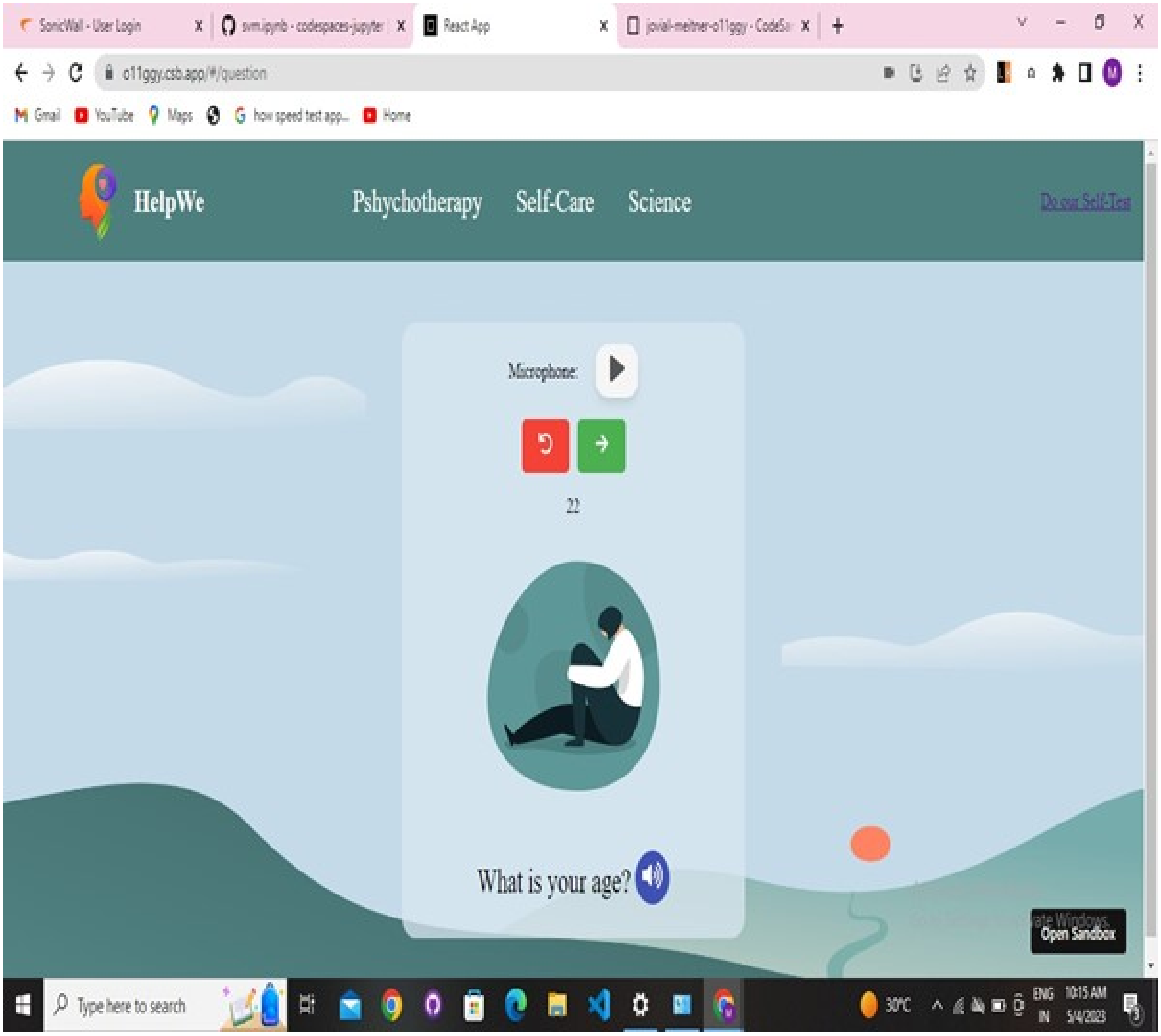


Figure 7.16: User Output

Chapter 8

CONCLUSION

In today's world, for the smooth functioning of society it has become important that an individual physically and mentally healthy. Numerous people are dealing with pathological problems worldwide. With the current hospital management system in country, it is difficult to provide health care to all the people suffering from mental illness. We are implementing an app which can detect mental disorder at an early stage using machine learning algorithms. For this purpose, we are taking input for user in the form of questionnaire. We are using language translation API and natural language processing to take input in any language and in audio format. Our application will solve the problem detection of mental disorder at early stage.

8.1 FUTURE SCOPE

1) In future the mental health prediction can be done using personal data of the user like call duration, number of missed calls, screen time, walking and sleeping time, heart rate and facial expressions.

2) We can add a system that will take dynamic data instead of static data.

3) We can add more features like consulting directly with psychiatrist in this application

Chapter 9

GLOSSARY

ML- Machine Learning

AI- Artificial Intelligence

RF- Random Forest

RNN- Recurrent Neural Network

NB- Naïve Bayes

SVM- Support Vector Machine

DL- Deep Learning

NLP- Natural Language Processing

LSTM- Long Short-Term Memory

UI- User Interface

GNB- Gaussian Naïve Bayes

DT- Decision Trees

GB- Gradient Boost

REFERENCES

[1] Anu Priya, Shruti Garg, Neha Prerna Tigga: Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms, IEEE,2019.

[2] Georgia Tech, ERIC P.S BAUMER: Human-Centered Machine Learning:The Case of Predicting Mental Health from Social Media,2019

[3] MS. Purude Vaishali Narayanrao, Dr. P. Lalitha Surya Kumari: Analysis of Machine Learning Algorithms for Predicting Depression, IEEE ,2023.

[4] Tianhua Chen, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis, Pan Su:Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning , IEEE 2021.

[5] Jolly Ehiabhi and HaifengWang: Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi Channel Multi-Modal Biometric Signals, IEEE 2023

[6] Shuo Yu, Member, Qing Qing, Chen Zhang, Ahsan Shehzad, Giles Oatley, and Feng Xia : Data-Driven Decision-Making in COVID-19 Response:, IEEE2021

[7] Shreya Ghosh: A Deep Learning Approach Depression, Intensity Estimation via Social Media, IEEE 2021

[8] Ayan Seal, Rishabh Bajpai, Jagriti Agnihotri, Anis Yazidi, Enrique Herrera-Viedma: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG, IEEE 2021

[9] Jianlong Zhou, Senior Member, IEEE, Hamad Zogan, Shuiqiao Yang, Shoaib Jameel, Guandong Xu, and Fang Chen: Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia, IEEE 2021

[10] Mario Ezra Arag , Adrian Pastor Lopez-Monroy, Luis Carlos Gonzalez-Gurrola , and Manuel Montes-y-Gomez : Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression , IEEE 2023

[11] Kresimir osi, Sinisa Popovi, Marko Sarlija Ivan Kesedzi: Impact of human disasters and covid 19 pandemic on mental health: potential of digital psychiatry ,2020

[12] Amna Amanat, Muhammad Rizwan, Abdul Rehman Javed, Maha Abdelhaq, Raed Alsaqour,Sharnil Pandya : Deep Learning for Depression Detection from Textual Data, 2022

[13] Ashley E. TateID, Ryan C. McCabeID, Henrik Larsson1, Sebastian Lundstro, Paul Lichtenstein Predicting mental health problems in adolescence using machine learning Techniques, April 6, 2020

[14] Ngumimi Karen Iyortsuun , Soo-Hyung Kim ,Min Jhon , Hyung-Jeong Yang and Sudarshan Pant : A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis , 17 January 2023

[15] Siriporn Sawangarreerak , Putthiporn Thanathamathee : Random Forest with Sampling Techniques for Handling Imbalanced Prediction of University Student Depression , 5 November 2020

[16] Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam : Detection of child depression using machine learning methods , December 16, 2021

[17] Masanori Nagamine, Szu-Nian Yang, Yueh-Ming Tai, Chin Lin : Machine Learning Based Suicide Ideation7, JULY 2020

[18] Safwan Wshah, Christian Skalka, Matthew Price : Prediction for Military Personnel ,2019

[19] Predicting Posttraumatic Stress Disorder Risk - A Machine Learning Approach , 2019

[20] Alina Trifan, Maryse Oliveira , José Luís Oliveira Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitation , 2019

[21] Chul-Hyun Cho1,Taek Lee, Min-Gwan Kim, Hoh Peter In, Leen Kim ,Heon-Jeong LeeMood : Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study, 2019

[22] Jonah Meyerhoff Tony Liu, Konrad P Kording Lyle H Unga Susan M Kaiser Chris J Karr : Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study , 2021

[23] Andreas K Triantafyllidis : Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature , 2019

[24] Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, X Patricia B. Munroe,Bina Joe Predicting Next Call Duration : A Future Direction to Promote Mental Health in the Age of Lockdown, March 5, 2023.

[25] Adrian B. R. Shatte, Delyse M. Hutchinson and Samantha J. Teague : Machine learning in mental health: a scoping review of methods and applications ,10 January 2019

[26] R. A. Rahman :Application of Machine Learning Methods in Mental Health Detection: A Systematic Review, 8, 2020

[27] Vidit Laijawala ,Aadesh Aachaliya :Classification Algorithms based Mental Health Prediction using Data Mining , 2019

[28] Shaoxiong Ji , Shirui Pan , Member, Xue Li, Erik Cambria , Guodong Long , and Zi Huang : Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications ,2021

[29] Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam :Detection of child depression using machine learning methods , December 16, 2021

[30] Predicting Generalized Anxiety Disorder Among Women Using Random Forest Approach

[31] Horeya Abou-Warda1, Nahla A. Belal : A Random Forest Model for Mental Disorders Diagnostic Systems ,2017

[32] M. Srividya1 S. Mohanavalli1 N. Bhalaj:Behavioral Modeling for Mental Health using Machine Learning Algorithms ,2018

[33] Ms. Sumathi M.R: Prediction of Mental Health Problems Among Children Using Machine Learning Techniques , 2016

[34] Shailesh Hinduja Mahbuba Afrin Sajib Mistry, Aneesh Krishna : Machine learning-based proactive social-sensor service for mental health monitoring using twitter data , 2022

BE Project Report Stage 2

ORIGINALITY REPORT

19%	16%	12%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journals.plos.org Internet Source	2%
2	www.mdpi.com Internet Source	1%
3	www.irjmets.com Internet Source	1%
4	ijircce.com Internet Source	1%
5	www.leewayhertz.com Internet Source	1%
6	Rohizah Abd Rahman, Khairuddin Omar, Shahrul Azman Mohd Noah, Mohd Shahrul Nizam Mohd Danuri, Mohammed Ali Al-Garadi. "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review", IEEE Access, 2020 Publication	1%
7	ijarcce.com Internet Source	1%

8	shodhgangotri.inflibnet.ac.in Internet Source	1 %
9	www.researchgate.net Internet Source	1 %
10	pdfcoffee.com Internet Source	1 %
11	www.repository.smuc.edu.et Internet Source	1 %
12	vitalflux.com Internet Source	1 %
13	"Machine Learning and Information Processing", Springer Science and Business Media LLC, 2020 Publication	<1 %
14	link.springer.com Internet Source	<1 %
15	pt.scribd.com Internet Source	<1 %
16	github.com Internet Source	<1 %
17	Adrian B. R. Shatte, Delyse M. Hutchinson, Samantha J. Teague. "Machine learning in mental health: a scoping review of methods and applications", Psychological Medicine, 2019	<1 %

Publication

18	media.springernature.com Internet Source	<1 %
19	www.ijraset.com Internet Source	<1 %
20	www.coursehero.com Internet Source	<1 %
21	www.turcomat.org Internet Source	<1 %
22	Arti Sharma, Niyati Aggarwal, Harsh Khatter, Saurabh, Abhinav Tripathi, Shashank Awasthi. "Comparative Analysis of Different Algorithms in Link Prediction on Social Networks", 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), 2023 Publication	<1 %
23	happydaycards.org Internet Source	<1 %
24	Sushama A Shirke, Tushar Gosavi, Anurag Kumar Mishra, Vishal Vir Singh, Akshay Srivastava. "Performance enhancement of apache APEX", 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017 Publication	<1 %

25	aclanthology.org Internet Source	<1 %
26	codeinstitute.net Internet Source	<1 %
27	journals.fedpolyado.edu.ng Internet Source	<1 %
28	K Nayana, S Vinay, S P Ashwini. "Anomaly Detection in Cardiac Related Datasets", 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2022 Publication	<1 %
29	human-resources-health.biomedcentral.com Internet Source	<1 %
30	cajotas.centralasianstudies.org Internet Source	<1 %
31	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
32	espace.library.uq.edu.au Internet Source	<1 %
33	fdocuments.net Internet Source	<1 %
34	Bhawna Dey, Parthh Dikshit, Samridhi Sehgal, Vasuda Trehan, Vivek Kumar Sehgal. "Intelligent Solutions for Earthquake Data	<1 %

Analysis and Prediction for Future Smart Cities", Computers & Industrial Engineering, 2022
Publication

35 air.ashesi.edu.gh <1 %
Internet Source

36 repository.untag-sby.ac.id <1 %
Internet Source

37 Amna Amanat, Muhammad Rizwan, Abdul Rehman Javed, Maha Abdelhaq, Raed Alsaqour, Sharnil Pandya, Mueen Uddin. "Deep Learning for Depression Detection from Textual Data", Electronics, 2022 <1 %
Publication

38 www.slideshare.net <1 %
Internet Source

39 123docz.net <1 %
Internet Source

40 nova.newcastle.edu.au <1 %
Internet Source

41 www.edtechnology.in <1 %
Internet Source

42 Shweta S. Aladakatti, S. Senthil Kumar. "PIRAP: A study on Optimized Multi-language classification and Text Categorization using Supervised Hybrid Machine Learning <1 %

Psychological well -being prediction and solutions using machine learning

[1] Yogita Hambir
Assitant Professor, Department of Computer Engineering
Army Institute of Technology, Pune
[2]Annu
Student, Department of Computer Engineering
Army Institute of Technology, Pune
[3]Priya
Student, Department of Computer Engineering
Army Institute of Technology, Pune
[4]Puja Nayak
Student, Department of Computer Engineering
Army Institute of Technology, Pune
[5]Mahima Chauhan
Student, Department of Computer Engineering
Army Institute of Technology, Pune

ABSTRACT: *Mental health is a major issue today. Any issues must be identified and fixed very away to avoid any negative consequences. With the Mental Health Tracker App, we try to achieve this. We must make the app incredibly warm and welcoming because users may be dealing with mental illness.*
Our project's objective is to create a straightforward piece of machine learning software that monitors users' advancement while recommending steps to help them better their mental health. The application asks the user a series of questions, assesses their responses, suggests tasks, keeps tabs on their mental health, and displays the results on a dashboard. This process was completed using machine learning.

KEYWORDS: Random Forest, Naïve Bayes, Support Vector Machine, Recurrent Neural Network, Sentiment Analysis, Machine Learning Algorithms, Convolutional Neural Network, Logistic Regression, Linear Regression, Mental Health, Deep Learning, Binary Classification.

INTRODUCTION

An individual's overall perspective and general nature can both be inferred from their level of mental health. Irregularities in cerebrum science lead to dysfunctional behaviours. It's really simple to understand how to assess mental health and provide remedies for patients who behave mentally erratically. Because of many factors, most people are susceptible to pressure and are affected by discouragement.
In 2011, a management board of the World Health Organisation (WHO) conducted a survey that found that by 2030, destitution will be the primary cause of global infection problems.
A person's mental well-being can be used to assess both their mental state and general personality. Chemical abnormalities in the brain cause mental disease. In order to comprehend and cure individuals who exhibit abnormal mental behaviour, it is imperative to examine mental health. The majority of people are susceptible to stress, whereas other people experience depression for a variety of reasons.
By 2030, depression would, in line with a 2011 prediction from a World Health Organisation (WHO) administrative panel, be the leading cause of illness worldwide. Mental health disorders must be recognised and treated as soon as possible. Those who are experiencing mental health issues can find relief through early discovery, precise diagnosis, and efficient treatment. Mental illness can have major effects on those who are afflicted, their families, and society as a whole.

Commonly utilised in general methods of mental health diagnosis are interviews, observations, or questionnaires. Although labour- and time-intensive, traditional methods are frequently employed. The medical industry has already employed mobile and wearable sensors to identify mental disorder. However, most users of these technologies have a mental illness and are subject to severe supervision. A person's mental health reflects all of their physical, psychological, emotional, and social well-being. It is established what one is thinking, experiencing, and how one reacts to situations. With healthy mental health, one may perform at their best.

A person's health depends on all phases of life, including childhood, adolescence, and adulthood. Men can develop drug addiction, personality disorders, and OCD for a variety of reasons, including stress, social anxiety, depression, obsessive compulsive disorder, substance abuse, and others. Understanding the early warning symptoms of mental illness is increasingly important to maintaining a healthy work-life balance. In order to accurately forecast the onset of mental illness, it is possible to fully leverage AI and machine learning. These apps will benefit from being implemented in real time.

RELATED WORK

Predictions in [1] decision trees have been made using the RF, SVM, NB, and K closest neighbour algorithms. The random forest classifier outperformed all other algorithms at predicting anxiety and depression.

Deep learning techniques have been applied to [2] binary classification. The article was released in 2019. The conclusion of this article is that human-focused strategies will be more effective for engagement and will help to develop a good mode of representation for mental health prediction. [6] studies the covid 19 decision-making reactions. The common policies utilised for COVID 19 include data-driven policies, financial aid policies, and counselling policies. In COVID 19, data-driven policy proved to be the most effective response strategy. [14] asserts that ML works well as a stand-alone treatment for mental health problems. However, the introduction of the DL approach raises the possibility of accurately identifying and predicting a single sickness. In the group of patients with many comorbidities, changes in the frequency of active applications, the length of exercise, and GPS features occurred before depression. Every three weeks, online surveys were sent out in [21]. Participants in the study were signed up for 16 weeks. All techniques received approval from the Institutional Review Board at North Western University, and each participant's informed permission was obtained before to participation. Due to its speedy execution and excellent accuracy, Decision Tree was found to be the most efficient algorithm in [27] when employing data mining to predict mental health. For prediction, we used decision trees, RF, and Naive Bayes.

[28] predicts that in the future, online social media would dominate as the primary SID channel. It is essential to create new techniques that could link clinical mental health with automatic machine detection. On the substance misuse dataset, the random forest classification accuracy values were 87.72% and 92.15%, respectively, with and without an impute missing values learner. The effectiveness of classification was improved by 4.43% when the learner was added to impute missing data. [31] uses random forest as its only algorithm.

The SVM, KNN, ensemble classifiers provide the psychologically unwell class label an overall score of 0.95, according to [32]. The score, which is quite near to 1, indicated that the data sample was as mental distress actually belongs into that group.

Multilayer Perceptron, Multiclass Classifier, LAD Tree were found to provide outcomes for mental health difficulties in children that are more accurate, as reported in a 2016 [33] paper.

DATA COLLECTION AND PREPROCESSING

we have collected dataset from Kaggle “OSMI Tech Survey 2016”. The dataset contain 46234 rows and 64 columns. The data has record of past mental health of employees and their current and previous working conditions. This dataset is valuable for our project as we can train our machine learning models on this dataset to make accurate predictions for any mental health disorder. After collecting the dataset, it was cleaned and preprocessed to give better results. The null values present in dataset were replaced with the mode value of that column values.

mental_df[col].mode() function was used to calculate the mode value.
mental_df[col].fillna(mode, inplace=True) function fills the missing values in the dataset with the calculated mode values.

Total 15 features were manually selected to train model as they are the most relevant ones for our study.

Table 1: list of features used to train ML model

S.No	Features
1	Are you self-employed
2	Do you work remotely
3	Have you had a mental health disorder in the past
4	Do you believe your productivity is ever affected by a mental health issue
5	Do you have a family history of mental illness
6	Do you feel comfortable in your working environment
7	Do you feel comfortable working with your direct supervisor
8	Do you feel that your organisation takes mental health as seriously as physical health
9	Have you observed or experienced an unsupportive or badly handled response to an issue in your current workplace
10	Are you stressed about your career
11	How willing would you be to share with friends and family about your work stress
12	Do you currently have a mental health disorder
13	Have you observed or experienced an unsupportive or badly handled response to an issue in your current workplace
14	Did you feel that your previous employers took mental health as seriously as physical health
15	Have your previous employers provided mental health benefits

In the dataset, few columns contain employee name, age and department information which is not relevant to our study so those columns are not used for prediction and are simply dropped.

The mapping function is used to convert answers given by user into numerical values {-1,0,1} which will later be used to train the model.

TRAINING THE MODEL

We have trained four machine learning models: Random Forest, Support Vector Machine, Naïve Bayes, Recurrent Neural Network.
X: This variable often denotes our dataset's independent variables or input features.

y: This variable often represents our dataset's target variable, or dependent variable, which we are attempting to predict or categorise.
The dataset is split into training and testing set. Here we are using 75% dataset as training set and remaining 25% dataset as testing set.

RANDOM FOREST

Using supervised classification, the random forest technique can be used. This algorithm builds a forest with many trees, as the name would imply. The look of the forest is often stronger the more trees there are in the forest. The random forest classifier works in a similar fashion, with higher accuracy results being produced by more trees in the forest. If you're familiar with the decision-tree algorithm. The question of whether or not we should be developing more decision trees, as well as how to do so, may be on your mind. due to the fact that the same dataset will be used for all nodes selection calculations. Yes. Truly, you are. You are not going to develop a forest by modelling more decision trees.
Random Forest pseudocode:
Pick "k" features at random from a total of "m" features
Use the optimum split point to determine the node "d" among the "k" features.
Use the optimum split to divide the node into daughter nodes.
Till the "l" number of nodes, repeat steps 1 to 3 as required.
To produce a "n" number of trees, repeat steps 1 to 4 a "n" number of times.

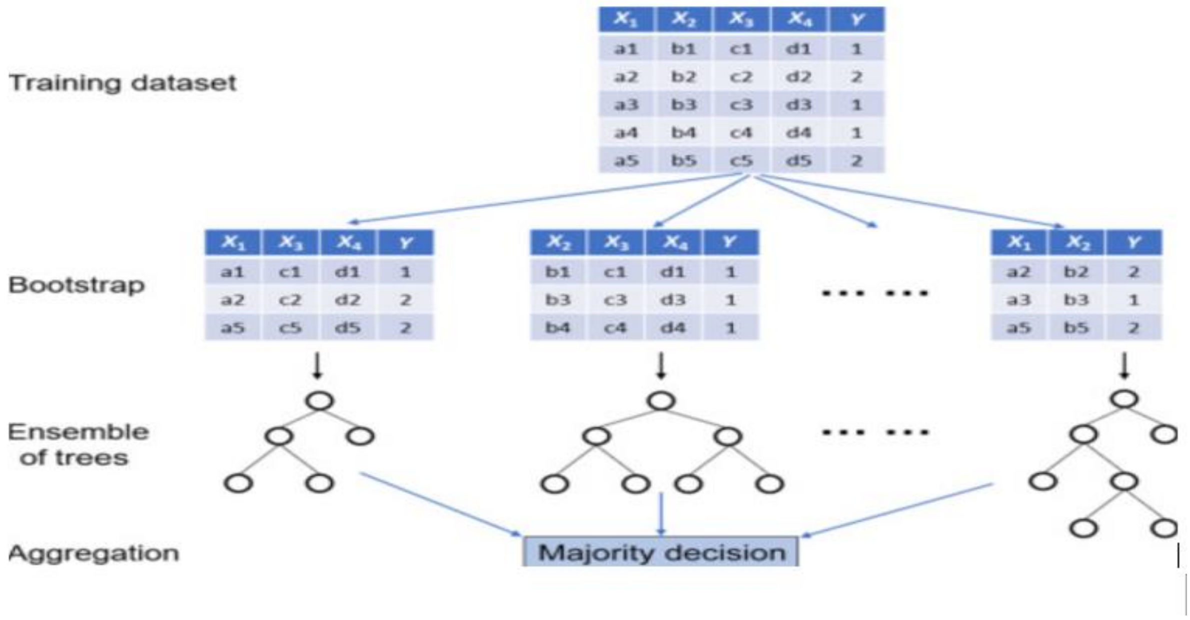


Figure 1: Random Forest model training [18]

SUPPORT VECTOR MACHINE

Support Vector Machine is one of the most popular methods for supervised learning, is used to address classification and regression problems. However, Machine Learning Classification problems are where it is most frequently used.
The SVM algorithm's goal is to find the optimal decision boundary or line that can categorise following data points in n-dimensional space and enable rapid assignment. This best-case decision boundary is referred to as a "hyperplane". SVM chooses the extreme points and vectors in order to construct the hyperplane. The term "Support Vector Machine" refers to a technique that uses support vectors to represent these extreme occurrences.

Data that can be separated into two groups by a single straight line are said to be linearly separable, which is a term used to describe linear SVM. Such data are classified using linear SVM classifiers.

Non-linear SVM is used for not linear separated data, which means that if dataset can't be distinguished using a straight line, it is regarded to be non-linear data, and the classifier utilised is referred to as a Non-linear SVM classifier.

NAÏVE BAYES

By using sequential events, where later knowledge is learned and affects the original probability, Bayes' Theorem stands out. The terms "prior probability" and "posterior probability" are used to describe these probabilities. The initial likelihood of an event before it is contextualised in a particular scenario is known as the prior probability, also known as the marginal probability. The chance of an event following the observation of a piece of data is known as the posterior probability.
In contrast to other classifiers, nave Bayes classifiers operate under a few fundamental presumptions, hence the term "nave." It makes the supposition that a Naive Bayes model's predictors are conditionally independent of one another or unconnected to any of the other features. Additionally, it is predicated on the idea that each element influences the result equally. Although these presumptions are frequently broken in real-world situations, they make categorization problems more manageable from a computational perspective. Meaning that each variable will now only need one probability, making the computation of the model simpler. The classification algorithm performs well, especially with small sample sizes, despite this false independence assumption.

RECURRENT NEURAL NETWORK

For time series analysis, supervised deep learning is employed using RNN. One of the most sophisticated algorithms in the field of supervised deep learning is the recurrent neural network.
RNN and the frontal lobe:
Similar to short-term memory are RNNs. We will discover that they are able to recall recent events from the prior few observations and use that information moving ahead. One of the Frontal Lobe's roles in humans is short-term memory.
The concept of RNN
Weights are said to have Long Term Memory, often known as LTM. For instance, the weights in a classical ANN are known, thus no matter what input we give it, it will process in the same way as it did yesterday. Given that the human brain's temporal lobe controls long-term memory (LTM), the weights can be found there.
Recurrent neural networks are adjusted to create LSTM networks, which facilitate better memory retention for information. Here, the RNN's vanishing gradient issue is fixed. The classification, processing, and prediction of time series with uncertain time lags are all excellent applications for LSTM. The model is trained via back-propagation. There are three gates in an LSTM.

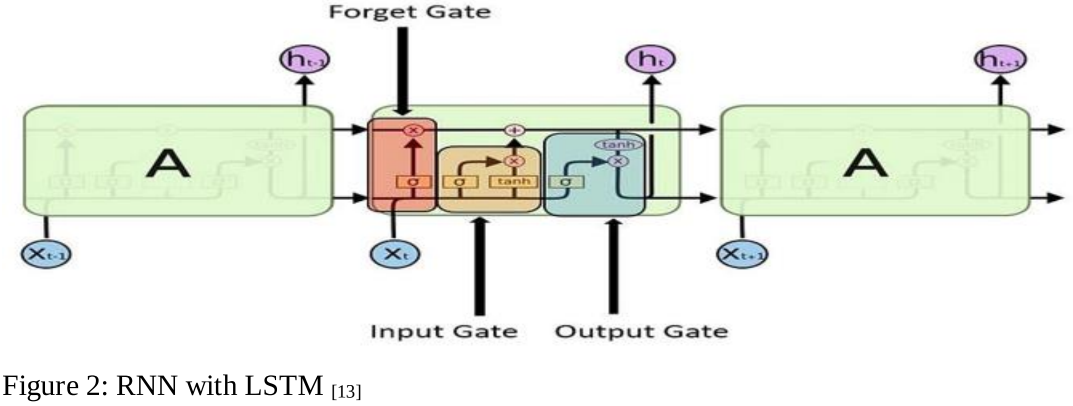


Figure 2: RNN with LSTM [13]

TESTING THE MODEL

Once the training of ML model is completed, the model is then tested on remaining 25% dataset. We calculate the accuracy, precision, recall, F score for the algorithms and find the optimum algorithm for our project.
After testing all four algorithms, it was found that support vector machine algorithm performed better than all other algorithms. Support Vector Machine was able to predict mental health disorders with an accuracy of 93.11%. Random forest proved to be the best after SVM in predicting psychological wellbeing of a person.

RESULTS:

Table 2: Accuracy comparison of algorithms

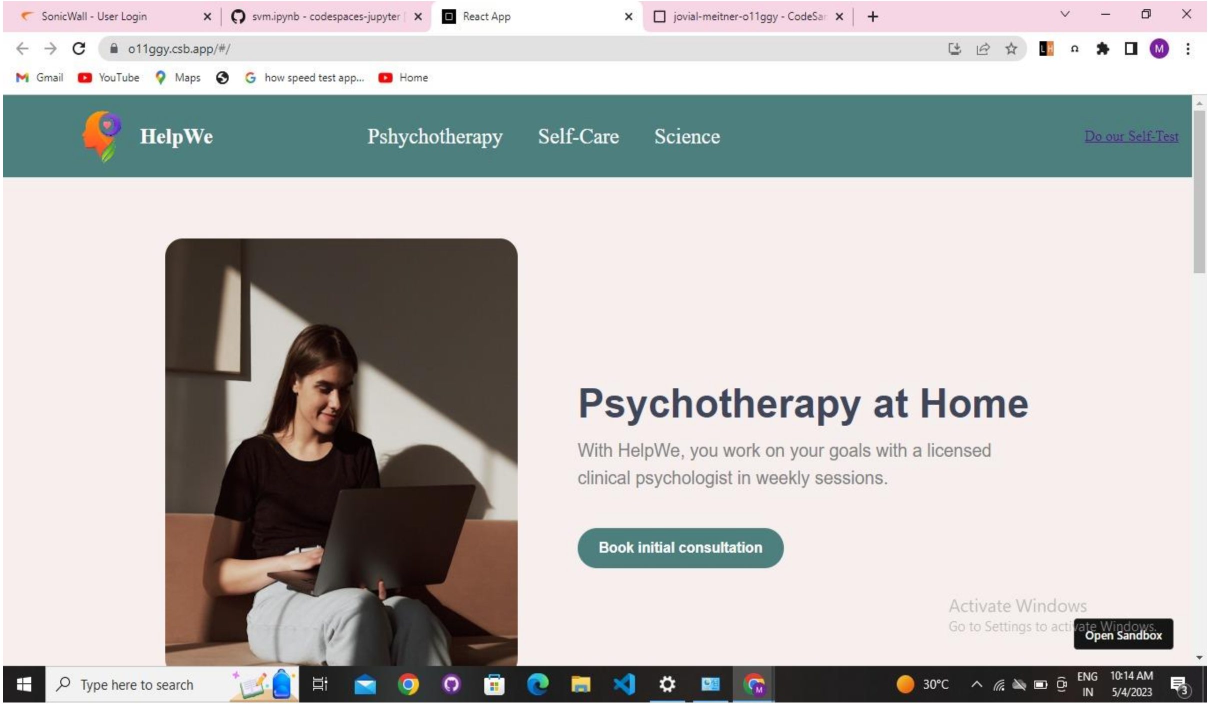
Machine Learning Algorithm	Accuracy
Support vector machine	93.11%
Recurrent neural network	88.50%
Naïve bayes	88.30%
Random forest	89.13%

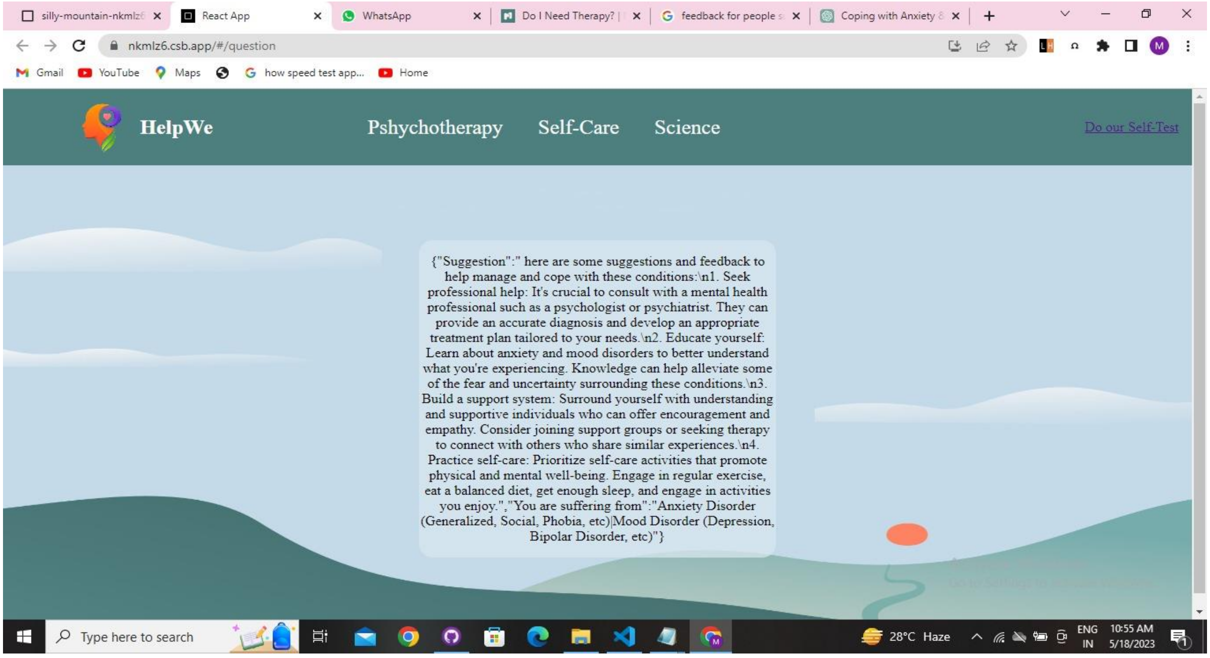
Table 3: other comparisons

Machine Learning Algorithms	Precision	Recall	F-score
Support vector machine	92.5	93.11	92.434
Naïve bayes	86	88	87
Random forest	86	89	87

RNN Loss value – 35.80

APPLICATION IMAGES:





CONCLUSION

In today’s world, for the smooth functioning of society it has become important that an individual physically and mentally healthy. Numerous people are dealing with pathological problems worldwide. With the current hospital management system in country, it is difficult to provide health care to all the people suffering from mental illness. We are implementing an app which can detect mental disorder at an early stage using machine learning algorithms. For this purpose, we are taking input for user in the form of questionnaire. We are using language translation API and natural language processing to take input in any language and in audio format. Our application will solve the problem detection of mental disorder at early stage.

REFERENCES

1. Anu Priya , Shruti Garg , Neha Prerna Tigga : Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms, IEEE,2019.

2. Georgia Tech, Eric P.S Baumer : Human-Centered Machine Learning:The Case of Predicting Mental Health from Social Media,2019

3. MS. Purude Vaishali Narayanrao, Dr. P. Lalitha Surya Kumari : Analysis of Machine Learning Algorithms for Predicting Depression , IEEE ,2023.

4. Tianhua Chen, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis, Pan Su :Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning , IEEE 2021.

5. Jolly Ehiabhi and HaifengWang: Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi Channel Multi-Modal Biometric Signals, IEEE 2023

6. Shuo Yu , Member, Qing Qing, Chen Zhang , Ahsan Shehzad , Giles Oatley, and Feng Xia : Data-Driven Decision-Making in COVID-19 Response:, IEEE2021

7. Shreya Ghosh: A Deep Learning Approach Depression, Intensity Estimation via Social Media , IEEE 2021

8. Ayan Seal, Rishabh Bajpai , Jagriti Agnihotri , Anis Yazidi ,Enrique Herrera-Viedma : A Deep Convolution Neural Network Framework for Detecting Depression Using EEG , IEEE 2021

9. Jianlong Zhou , Senior Member, IEEE, Hamad Zogan, Shuiqiao Yang, Shoaib Jameel, Guandong Xu , and Fang Chen : Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia ,IEEE 2021

10. Mario Ezra Arag , Adrian Pastor Lopez-Monroy, Luis Carlos Gonzalez-Gurrola , and Manuel Montes-y-Gomez : Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression , IEEE 2023

11. Kresimir osi , Sinisa Popovi , Marko Sarlija & Ivan Kesedzi : Impact of human disasters and covid 19 pandemic on mental health :potential of digital psychiatry ,2020

12. Amna Amanat , Muhammad Rizwan , Abdul Rehman Javed , Maha Abdelhaq , Raed Alsaqour,Sharnil Pandya : Deep Learning for Depression Detection from Textual Data, 2022

13. Ashley E. TateID, Ryan C. McCabeID, Henrik Larsson1, Sebastian Lundstro , Paul Lichtenstein Predicting mental health problems in adolescence using machine learning Techniques , April 6, 2020

14. Ngumimi Karen Iyortsuun , Soo-Hyung Kim ,Min Jhon , Hyung-Jeong Yang and Sudarshan Pant : A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis , 17 January 2023

15. Siriporn Sawangarreerak , Putthiporn Thanathamathee : Random Forest with Sampling Techniques for Handling Imbalanced Prediction of University Student Depression , 5 November 2020

16. Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam : Detection of child depression using machine learning methods , December 16, 2021

17. Masanori Nagamine, Szu-Nian Yang, Yueh-Ming Tai, Chin Lin : Machine Learning Based Suicide Ideation7, JULY 2020

18. Safwan Wshah, Christian Skalka, Matthew Price : Prediction for Military Personnel ,2019

19. Predicting Posttraumatic Stress Disorder Risk - A Machine Learning Approach , 2019

20. Alina Trifan, Maryse Oliveira , José Luís Oliveira Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitation , 2019

21. Chul-Hyun Cho1,Taek Lee, Min-Gwan Kim, Hoh Peter In, Leen Kim ,Heon-Jeong LeeMood : Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study, 2019

22. Jonah Meyerhoff Tony Liu, Konrad P Kording Lyle H Unga Susan M Kaiser Chris J Karr : Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study , 2021

23. Andreas K Triantafyllidis : Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature , 2019

24. Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, X Patricia B. Munroe,Bina Joe Predicting Next Call Duration : A Future Direction to Promote Mental Health in the Age of Lockdown, March 5, 2023.

25. Adrian B. R. Shatte, Delyse M. Hutchinson and Samantha J. Teague : Machine learning in mental health: a scoping review of methods and applications ,10 January 2019

26. R. A. Rahman :Application of Machine Learning Methods in Mental Health Detection: A Systematic Review, 8, 2020

27. Vidit Laijawala ,Aadesh Aachaliya :Classification Algorithms based Mental Health Prediction using Data Mining , 2019

28. Shaoxiong Ji , Shirui Pan , Member, Xue Li, Erik Cambria , Guodong Long , and Zi Huang : Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications ,2021

29. Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam :Detection of child depression using machine learning methods , December 16, 2021

30. Predicting Generalized Anxiety Disorder Among Women Using Random Forest Approach

31. Horeya Abou-Warda1, Nahla A. Belal : A Random Forest Model for Mental Disorders Diagnostic Systems ,2017

32. M. Srividya1 & S. Mohanavalli1 & N. Bhalaj :Behavioral Modeling for Mental Health using Machine Learning Algorithms ,2018

33. Ms. Sumathi M.R : Prediction of Mental Health Problems Among Children Using Machine Learning Techniques , 2016

34. Shailesh Hinduja Mahbuba Afrin Sajib Mistry, Aneesh Krishna : Machine learning-based proactive social-sensor service for mental health monitoring using twitter data , 2022