



COLLEGE OF ENGINEERING AND TECHNOLOGY

MULAVOOR P.O, MUVATTUPUZHA, KERALA – 686673



**PUBLIC MENTAL HEALTH PREDICTION
USING MACHINE LEARNING TECHNIQUES
INNERHOUR: MENTAL HEALTH FRIEND**

PROJECT REPORT

FATHIMA K S (ICE19CS048)

NANDANA P R (ICE19CS085)

NEERAJA SUBHASH (ICE19CS087)

PAVITHRA C A (ICE19CS092)

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SEPTEMBER 2022



COLLEGE OF ENGINEERING AND TECHNOLOGY

Mulavoor P.O, Muvattupuzha, Kerala – 686673



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that the project report entitled “**PUBLIC MENTAL HEALTH PREDICTION USING MACHINE LEARNING TECHNIQUES**” is a bonafide record of the project report presented by FATHIMA K S (ICE19CS048), NANDANA P R (ICE19CS085), NEERAJA SUBHASH (ICE19CS087), PAVITHRA C A (ICE19CSE092) during the academic year 2021- 2022 towards the partial fulfilment of the requirement of the award of B. Tech Degree in Computer Science and Engineering of APJ Abdul Kalam Technological University, Thiruvananthapuram.

Dr. Vadhana Kumari S

Professor
CSE Department
Project Guide

Ms. Chitra Rani P R

Assistant Professor
CSE Department
Project Coordinator

Dr. Sujith Kumar P.S

Professor & HOD
CSE Department
Head of the Department

DECLARATION

We Fathima K S, Nandana P R, Neeraja subhash, Pavithra C A, hereby declare that, this project report entitled “Public mental health prediction using machine learning techniques” is the bonafide work carried out under the supervision of Ms. Chitra Rani. We declare that to the best of our knowledge, the work report here in does not form part of any project report or dissertation the basis of which a degree or award was conferred on an earlier occasion on any other candidate the content of this report is not being presented by any other student to this or any other university for the award of degree.

Signature:

Name of the Student: FATHIMA K S

Uni. Register No: ICE19CS048

Signature:

Name of the Student: NANDANA P R

Uni. Register No: ICE19CS085

Signature:

Name of the Student: NEERAJA SUBHASH

Uni. Register No: ICE19CS087

Signature:

Name of the Student: PAVITHRA C A

Uni. Register No: ICE19CS092

Signature:

Name of the guide: Dr. VADHANA KUMARI S

Professor, CSE, ICET

Signature:

Name of the Coordinator: Ms. CHITRA RANI PR

Assistant Professor, CSE, ICET

Countersigned with Name: Dr. SUJITH KUMAR

HOD, Computer Science & Engineering

Ilahia College of Engineering and Technology

Date:

Place:

ACKNOWLEDGMENT

Apart from the efforts, the success of this project report depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We would like to show our heartfelt gratitude towards **Prof. Dr. ABDUL GAFUR M**, Principal, Ilahia College of Engineering and Technology for granting us the permission to work this project. Also, we would like to show our greatest gratitude towards our head of the department of computer science & Engineering, **Dr. Sujith Kumar P.S** and project guide **Dr. Vadhana Kumari S** and our project Coordinator **Ms. Chitra Rani**, for their valuable advice and guidance.

Finally, we express our gratitude and thanks to all our teachers and other faculty members of the Department of Computer Science & Engineering, for their sincere and friendly cooperation in completing this project.

Date :

Place :

FATHIMA K S

NANDANA P R

NEERAJA SUBHASH

PAVITHRA C A

ABSTRACT

In a fast-growing society, it is a hectic task to keep track on our mental health and wellness. As mental health is as important as someone's physical health it's is each one's responsibility to take care of it. Interpretation of public emotional wellness concerns utilizing data science, deep learning and machine learning techniques by observing life style, schedules, emotions and sentimental analysis will pave way to give the user an insight of their mental health. Our project focus on this issue and give prediction about user's mental health and thus suggest them a review about their mental health issue and lend a helping hand nearby who can be a therapist or counselor of their taste. High prevalence of mental health and the need for effective mental fitness care, blended with current advances in AI, has led to a growth in explorations of ways the sphere of system getting to know Machine Learning can assist inside the detection, prognosis and treatment of mental health issues.

CONTENTS

Chapter	Title	Page No
	LIST OF FIGURES	8
	LIST OF ABBREVIATIONS	9
1	INTRODUCTION	10
2	LITERATURE SURVEY	11
3	PROPOSED SYSTEM	12
	3.1 INTRODUCTION	12
	3.2 ARCHITECTURAL DESIGN	13
4	SYSTEM DESIGN	16
	4.1 REGISTRATION	16
	4.2 LOGIN	16
	4.3 MANAGE THERAPIST	16
	4.4 MANAGE PATIENT	16
	4.5 UPDATE PROFILE	16
	4.6 ATTEND SURVEY	16
	4.7 BOOK THERAPIST	17
	4.8 VIEW BOOKING	17
	4.9 VIEW HISTORY	17
5	SYSTEM REQUIREMENTS	18
	5.1 HARDWARE REQUIREMENTS	18
	5.2 SOFTWARE REQUIREMENTS	18
	5.2.1 FRONT END	19
	5.2.2 BACK END	19
6	SOLUTION METHODOLOGY	20

	6.1 USE CASE DIAGRAM	20
	6.2 FLOW CHART	20
	6.3 ACTIVITY DIAGRAM OF PATIENT	21
	6.4 ACTIVITY DIAGRAM OF THERAPIST	21
	6.5 ACTIVITY DIAGRAM OF ADMIN	22
	6.6 SEQUENCE DIAGRAM OF ADMIN	22
	6.7 SEQUENCE DIAGRAM OF PATIENT	23
	6.8 SEQUENCE DIAGRAM OF THERAPIST	23
7	SYSTEM IMPLIMENTATION	
	7.1 DATA PREPROCESSING	24
	7.2 DATASET EXTRACTION	25
	7.3 STATISTICAL ANALYSIS	25
	7.4 PERFORMANCE OF PREDICTION	26
	7.5 MACHINE LEARNING ALGORITHM	27
	7.6 EXPERIMENTAL SETTINGS	28
	7.7 PERFORMANCE EVALUATION	29
8	RESULTS AND DISCUSSION	30
9	CONCLUSION AND FUTURE HOPE	
	9.1 CONCLUSION	39
	9.2 FUTURE SCOPE	39
	REFERENCES	41
	APPENDIX1	

LIST OF FIGURES

Fig No	Title	Page No
3.2	Architectural design	
6.1	Use case Diagram	20
6.2	Flow Chart	20
6.3	Activity diagram of admin	21
6.4	Activity diagram of patient	21
6.5	Activity diagram of therapist	22
6.6	Sequence diagram of admin	22
6.7	Sequence diagram of patient	23
6.8	Sequence diagram of therapist	23

LIST OF ABBREVIATIONS

Terminology	Expansion
ASP	Active Server Page
AUC	Area under the ROC Curve
CSS	Cascading Style Sheet
DBMS	Database Management System
DFD	Data Flow Diagram
DT	Decision Tree
EES	Expert Embedded System
ERD	Entity Relationship Diagram
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hyper Text Transfer Protocol
IDE	Integrated Development Environment
IT	Information Technology
KNN	K Nearest Neighbour
LMT	Logistic Model Trees
NB	Naïve Bayes
NBM	Naïve Bayes Multinational
RF	Random Forest
SDLC	System Development Life Cycle
SQL	Structure Query Language
SVM	Support Vector Machine
XML	Extensible Markup Language

CHAPTER 1

1. INTRODUCTION

Around the world 970 million individuals are affected with mental illness. The increase of mental problems and the need for effective medical health care have led to an investigation of machine learning that can be applied in mental health problems. This paper presents a recent systematic review of machine learning approaches in predicting mental health problems. We will discuss the challenges, limitations and future directions for the application of machine learning in the mental health field. As the mental health is as important as someone's physical health it's each one's responsibility to take care of it. Our project focus on this issue and give prediction about user's mental health.

Machine learning is a technique that aims to construct systems that can improve through experience by using advanced statistical and probabilistic techniques. It is believed to be significantly useful tool to help in predicting mental health. It is allowing many researches to acquire important information from the data, provide personalized experiences and develop automated intelligent systems. The widely used algorithms in the field of machine such as support vector machine, random forest, and artificial neural networks have been utilized to forecast and categorize the future events. Supervised learning in machine in machine learning in the most widely applied approach in many types of research, studies, and experiments, especially in predicting illness in the medical field. In supervised learning, the terms, attributes, and values should be reflected in all data instances.

In this paper, the main objective is to provide a systematic literature review, critical review, and summary of the machine learning techniques that are being used to predict, diagnose, and identify mental health problems. Moreover, this paper will propose future avenues for research on this topic. It would also give attention to the challenges and limitations of applying the machine learning techniques in this area.

CHAPTER 2

2. LITERATURE SURVEY

According to Yadav, S., Kaim, T., Gupta, S., Bharati, U., and Priyadarshi, P., the model could be integrated into existing government database or several public health entities to examine the risk factors and prevalence of depression. Early forecast and identification of depression from the data provided from various surveys. This model only focus on predicting depression and it doesn't focus on predicting the specified type by analysing behavioural anomalies given by data set. This model boarded only to depression prediction as it can also predict several other mental illnesses as it use wide range of data set.

According to U. S. Reddy, A. V. Thota and A. Dharun, "machine learning techniques for stress prediction in working employees," analyse stress patterns in working adults and to narrow down the factors that strongly determine the stress levels. By using decision trees, prominent features that influence stress were identified as gender, family history and availability of health benefits in the workplace. This technique lack of accuracy in cross validation, which is the major aspect of forming a prediction model is to ensure its stability.

According to A.Gruenerbl, G.Bahle, J.Weppner,P .Olesky and P.Luknowicz,'Towards smartphone based monitoring of bipolar disorder', in proceedings of the second ACM workshop on mobile systems, application and services for healthcare. In this paper, the system, which based on smartphone sensing is able to recognize depressive and manic States and detect state changes of patients suffering from bipolar disorder.

According to U.Haetal, A Wearable EEG-HEG-HRV Multimodal system with simultaneous monitoring of tEs for mental health management'. The system can measure the brain activity from the various signal domain for improved accuracy on it enables the simultaneous monitoring of stimulation. This proposed wearable system can be used for an accurate, safe and effective closed loop stress management.

CHAPTER 3

3. PROPOSED SYSTEM

3.1 INTRODUCTION

Around the world 970 million individuals are affected with mental illness. The increase of mental problems and the need for effective medical health care have led to an investigation of machine learning that can be applied in mental health problems. This paper presents a recent systematic review of machine learning approaches in predicting mental health problems. We will discuss the challenges, limitations and future directions for the application of machine learning in the mental health field. As the mental health is as important as someone's physical health it's each one's responsibility to take care of it. Our project focus on this issue and give prediction about user's mental health.

Machine learning is a technique that aims to construct systems that can improve through experience by using advanced statistical and probabilistic techniques. It is believed to be a significantly useful tool to help in predicting mental health. It is allowing many researchers to acquire important information from the data, provide personalized experiences and develop automated intelligent systems. The widely used algorithms in the field of machine such as support vector machine, random forest, and artificial neural networks have been utilized to forecast and categorize the future events. Supervised learning in machine learning is the most widely applied approach in many types of research, studies, and experiments, especially in predicting illness in the medical field. In supervised learning, the terms, attributes, and values should be reflected in all data instances.

In this paper, the main objective is to provide a systematic literature review, critical review, and summary of the machine learning techniques that are being used to predict, diagnose, and identify mental health problems. Moreover, this paper will propose future avenues for research on this topic. It would also give attention to the challenges and limitations of applying the machine learning techniques in this area.

3.2 ARCHITECTURAL DESIGN

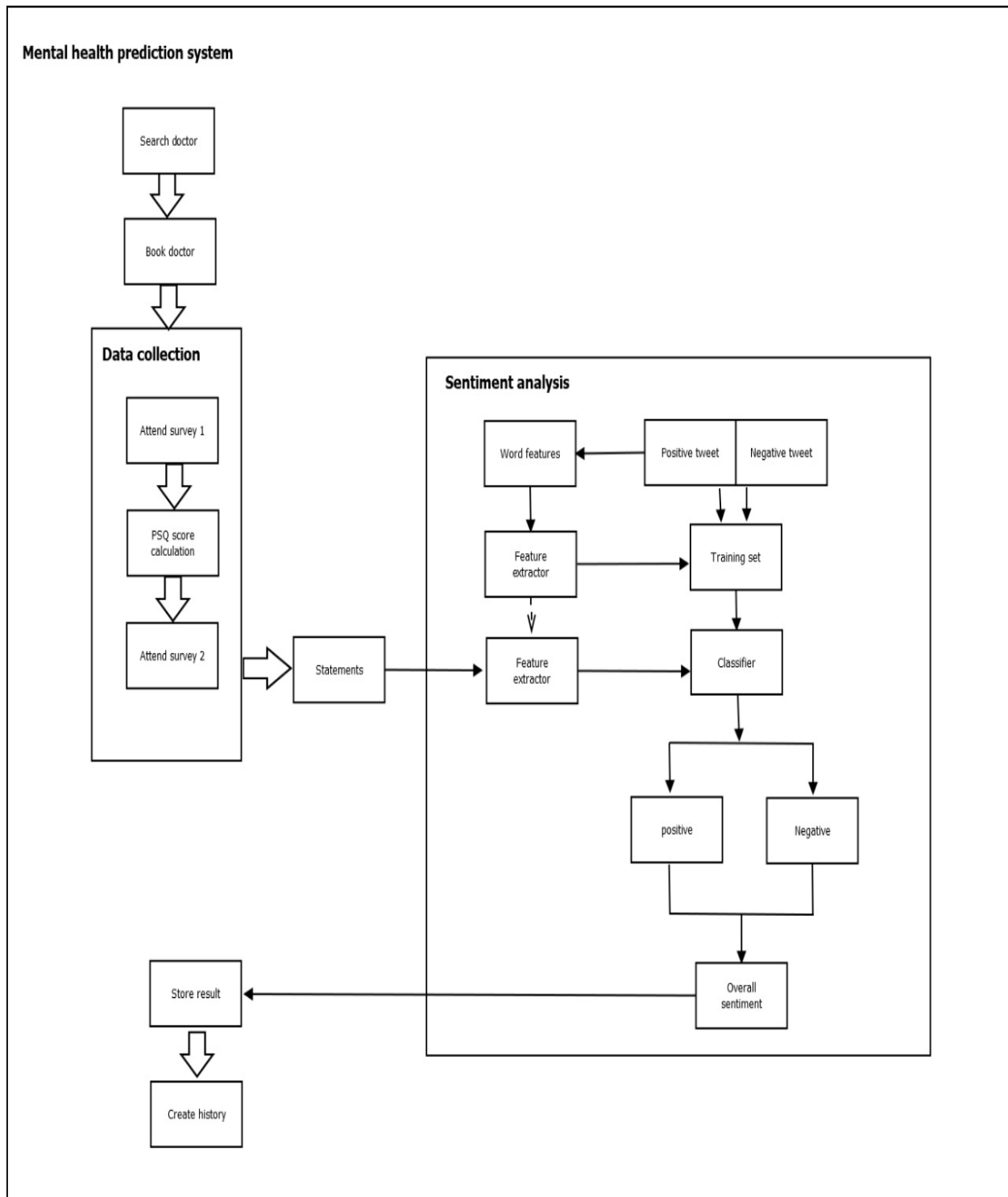


Figure 3.2 Architectural design

- **ADMIN**

Admin may include implementing security protocols, modifying programs, creating backups, resolving software problems, updating content and more. Admin manages the details of therapist and add the details of therapist like name, address, email, contact, qualification, experience, and location. Admin can view all the details of the patients registered. Admin can also delete their account if found to be fake or caught for some malpractices.

- **USER**

A Person or individual who wants to use the app and who wants to get a relieve from his current situation, or someone who wants to talk out his feelings. Patient need to register to the system to access the features. Only registered users can login to the site. Registered patients can attend a survey containing objective type and descriptive type questions. Both the tests are used for analysis the mental health of patient. The answers of the descriptive type questions are used for sentimental analysis and this result is also used for analysis. Patients can view their booking details and booking history.

- **THERAPIST**

Admin manages the details of therapist. Therapists can update their personal details after login the system. Therapist can view which all patients are booked on each date.

- **DATA PROCESSING**

Variable with more than 50% missingness were removed from analysis (202 variables excluded). Redundant variables were also removed (134 excluded). Additionally, variables with no variance were removed (32 excluded) and those with variance near zero were combined into one variable, if possible, e.g., dust, mold, and pollen allergy collapsed into allergy. Ultimately, 85 variables were determined to be suitable for analysis. As most machine learning techniques require complete datasets, missing values were imputed with tree-based imputation with the r package mice.

- **FEATURE EXTRACTION**

There is an immense need to analysis and monitor a person's mental health as justified in our previous work. We utilize deep learning feature extraction algorithm like sentiment analysis to analysis mental health of person from a set of questions. Using polarity, the statement to the questions is tested as positive or negative

- **MENTAL HEALTH RECOGNATION**

Registered patients can attend a survey which is objective and descriptive type. Both the tests are used to analysis the mental health of patient. The answers of the descriptive type questions are used for sentimental analysis and this result is also used for analysis.

3.3 ADVANTAGES

The use of proposed system will avoid the problem of the existing and we also get a new system. The proposed system offers:

- Easy to store data in the database.
- Administrator can add/remove new therapists and patients.
- Increase processing speed.
- Improved security
- Patients and therapists can update their personal details.
- A simple to do list is provided for the user.
- Deep learning and hybrid classifiers used for improving the accuracy.

CHAPTER 4

4. SYSTEM DESIGN

4.1 REGISTRATION

Patients need to register to the system to access the features. Only registered users can login to the site. During registration details such as name, address, email, contact, location etc are stored in the database.

4.2 LOGIN

All three users-admin, therapist and patient need to login to the system to access their home page. Email id and password are the user credentials used. This is created during the registration of each request

4.3 MANAGE THERAPIST

Admin manages the details of therapist. This is done to prevent the submission of fake details to the system. Admin add the details of therapist like name, address, email, contact and location. This location is used during the searching process. These details are manually collected by the admin and added to the system. Admin can also delete a therapist if needed

4.4 MANAGE PATIENT

Admin can view all the details of the patients registered in the site. Admin can also delete their account if found to be fake or caught for some malpractices.

4.5 UPDATE PROFILE

Patients and therapists can update their personal details after login to the system.

4.6 ATTEND SURVEY

Registered patients can attend a survey containing objective type and descriptive type questions. Both the tests are used for analysing the mental health of patient. The answers of the descriptive type questions are used for sentimental analysis and this result is also used for analysis.

4.7 BOOK THERAPIST

After the test, patients can find out the nearest therapist based on their location. They can also book for the therapist.

4.8 VIEW BOOKING

Patients can view their booking details and booking history. Therapist can view which all patients are booked on each date.

4.9 VIEW HISTORY

Patients, therapist and admin can view the history of the questions and their answers given during each test of patients. This is used by the therapist for further treatments.

CHAPTER 5

5. SYSTEM REQUIREMENTS

5.1 HARDWARE REQUIREMENTS

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware, a hardware requirements list is often accompanied by a hardware compatibility list (HCL), especially in case of operating systems. An HCL lists tested, compatible and sometimes incompatible hardware devices for a particular operating system or application. The following sub-sections discuss the various aspects of hardware requirements.

- Processor: 1.4GHz or above processors are considered good. They can handle heavier tasks.
- Memory (RAM): Minimum 2 GB; Recommended 4 GB or above.
- Storage: 1GB free disk space for workstations (2-5 GB Recommended)-1GB data storage capacity (2GB recommended).
- Minimum screen resolution is required is 738x844.

5.2 SOFTWARE REQUIREMENTS

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The requirements can be obvious or hidden, known or unknown, expected or unexpected from client's point of view.

- Android version: 5 or above
- Languages used are HTML, CSS, JAVASCRIPT, Bootstrap in the frontend.
- Web-application development using python as backend and Django as python framework.
- MVT Architecture is used.

- SQL lite is used for database.

5.2.1 FRONT END

Simply put, the front end of the web-app is what the user experiences. You may be familiar with the terms “HTML, CSS, JS, BOOTSTRAP and PYTHON”; the front end is where the user experiences and interacts with these things. A front-end developer will be focused on what happens with the app, rather than what goes on behind the scenes, which is typically described as the backend. The web-app development process involves cooperation between front end and backend developers in order to finish as a complete web-application. There are typically two types of front-end developments: Native and hybrid. Native front-end developments are typically designed either for the iOS or Android platforms, whereas hybrids use both and can be compatible with all types of operating systems.

5.2.2 BACK END

The back end of a web-app is meant by the framework that we using and which is responsible for the working and storing of the data of the web-app. We use Django in back end. Django is an open-source framework for backend web applications based on Python one of the top web development languages. Its main goals are simplicity, flexibility, reliability, and scalability. Django has its own naming system for all functions and components. Django is an MVT web framework that is used to build web applications. The huge Django web-framework comes with so many “batteries included” that developers often get amazed as to how everything manages to work together. The principle behind adding so many batteries is to have common web functionalities in the framework itself instead of adding latter as a separate library.

CHAPTER 6

6. SOLUTION METHODOLOGY

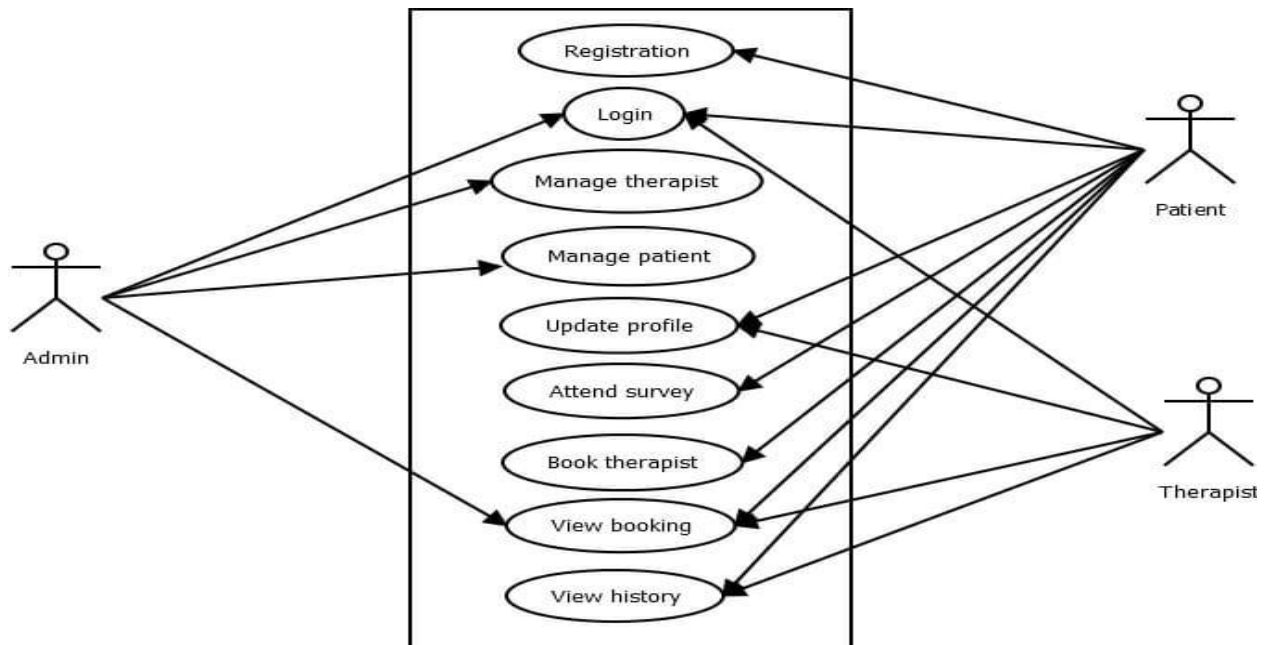


Figure 6.1 Use case diagram

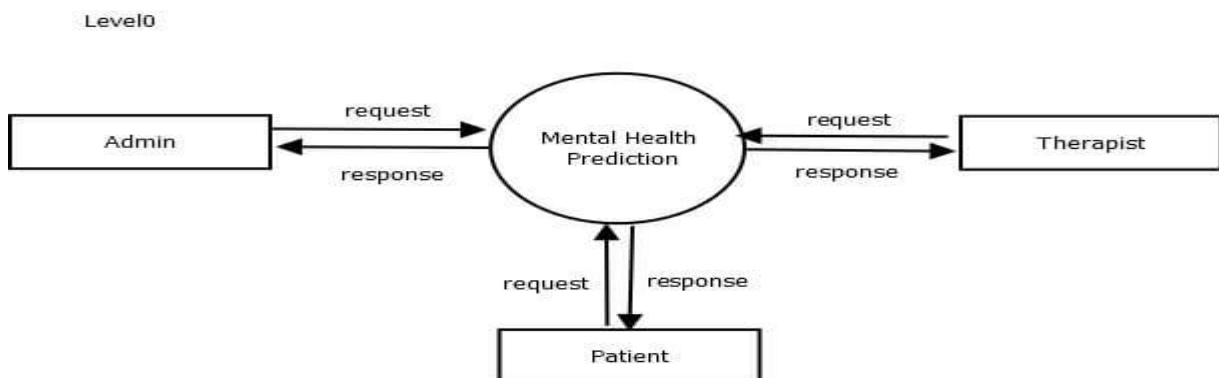


Figure 6.2 Flow chart

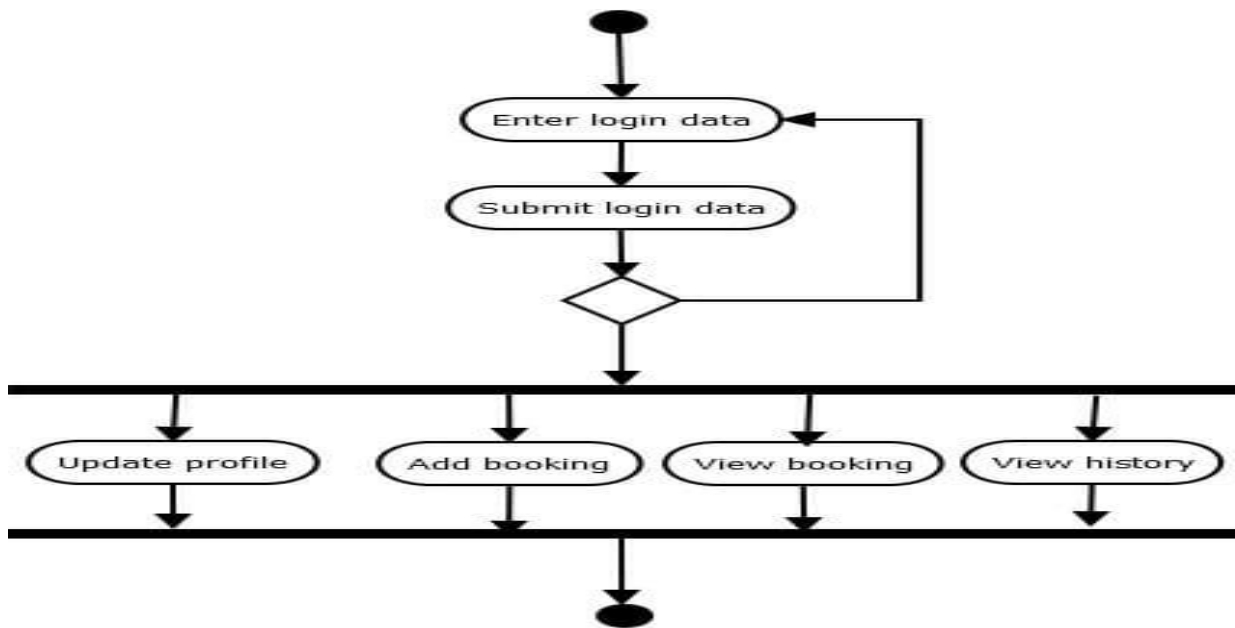


Figure 6.3 Activity Diagram of Patient

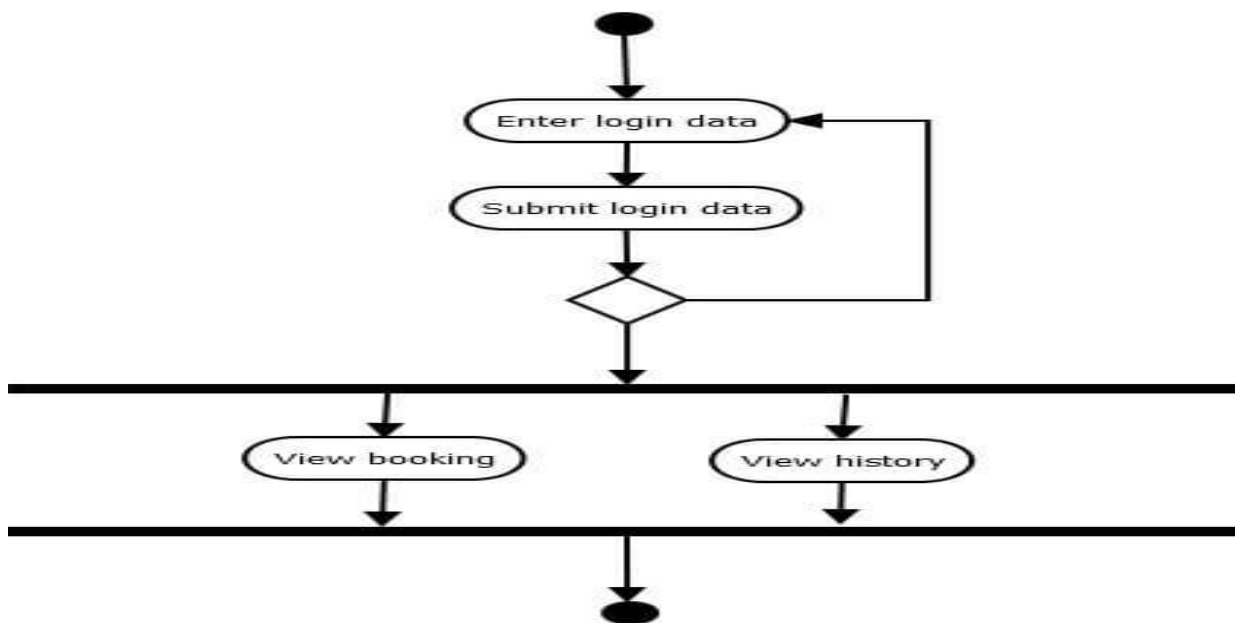


Figure 6.4 Activity Diagram of Therapist

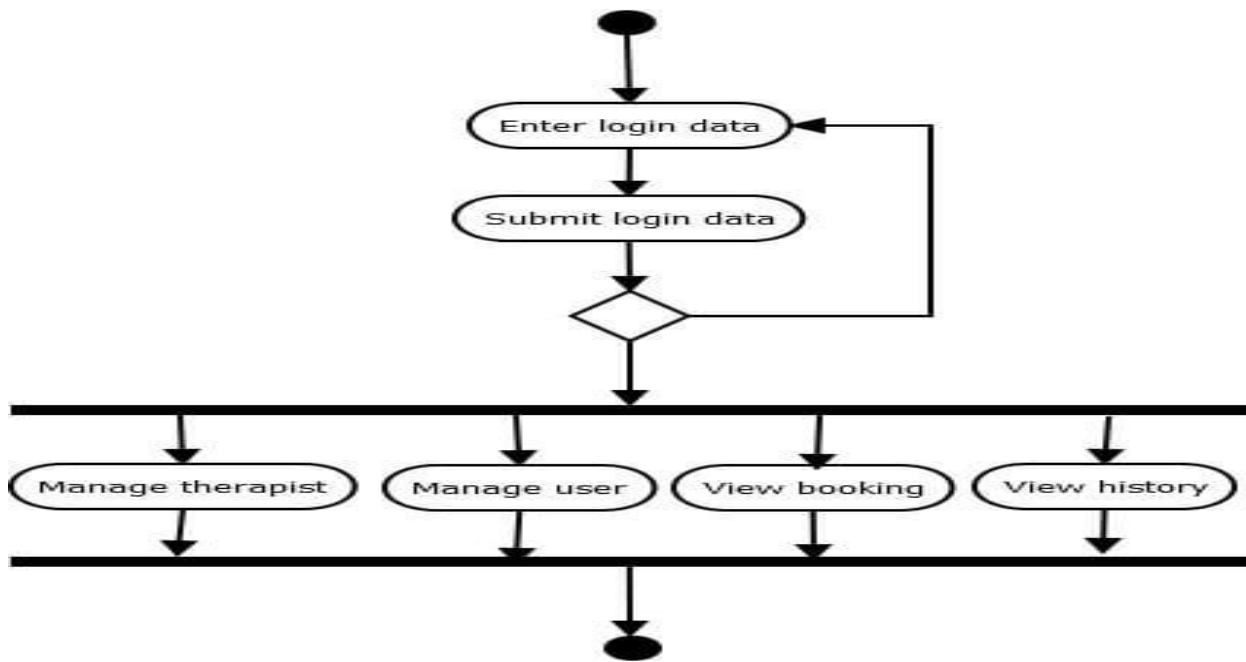


Figure 6.5 Activity Diagram of Admin

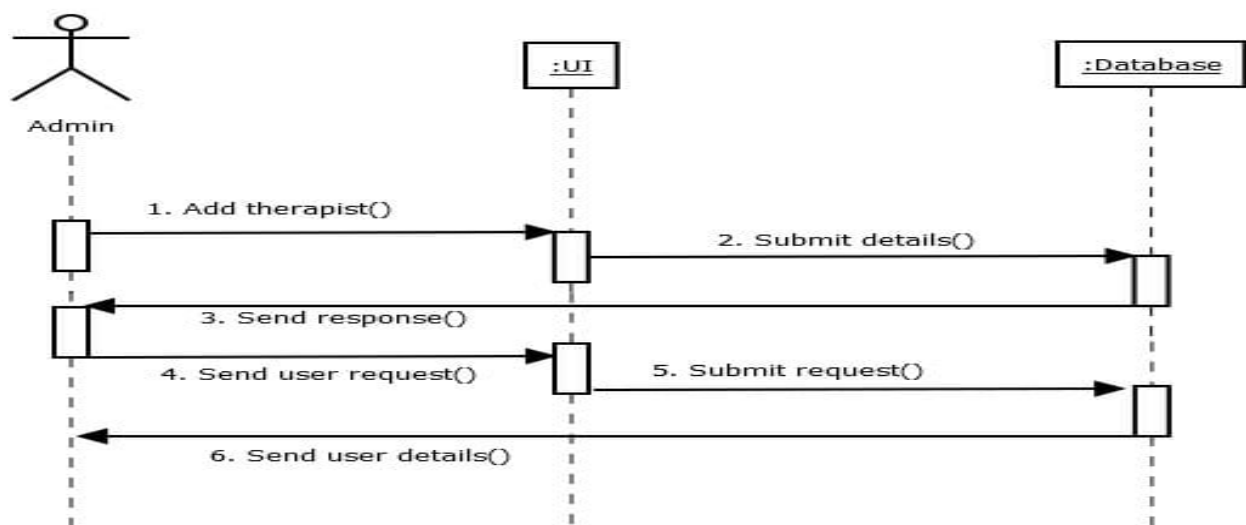


Figure 6.6 Sequence Diagram of Admin

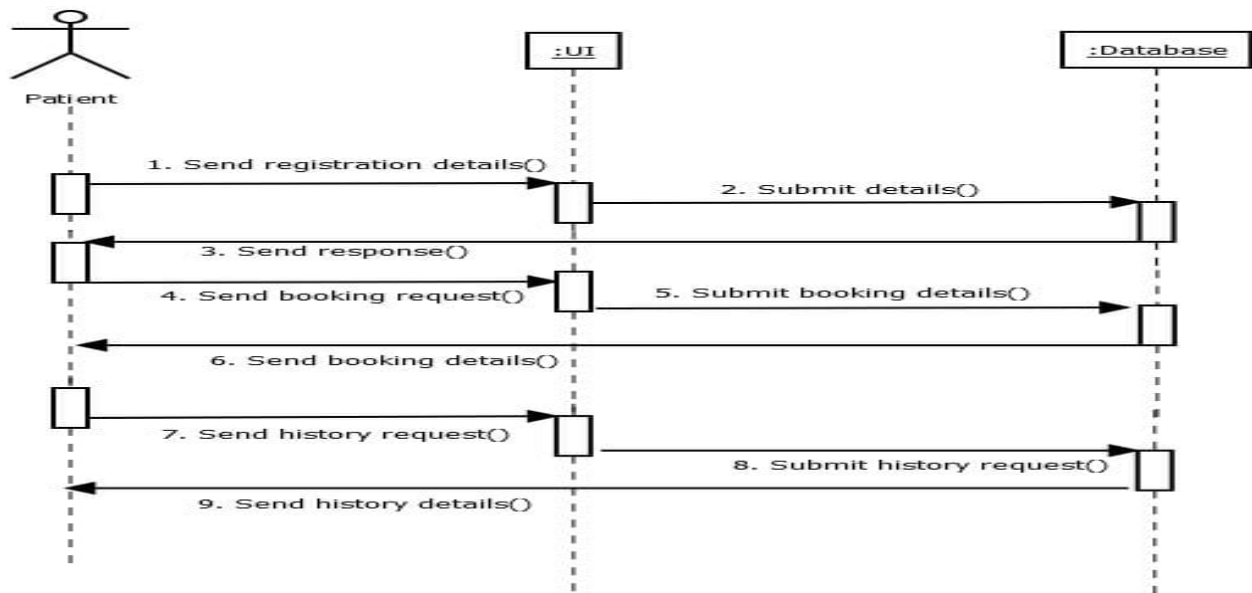


Figure 6.7 Sequence Diagram of Patient

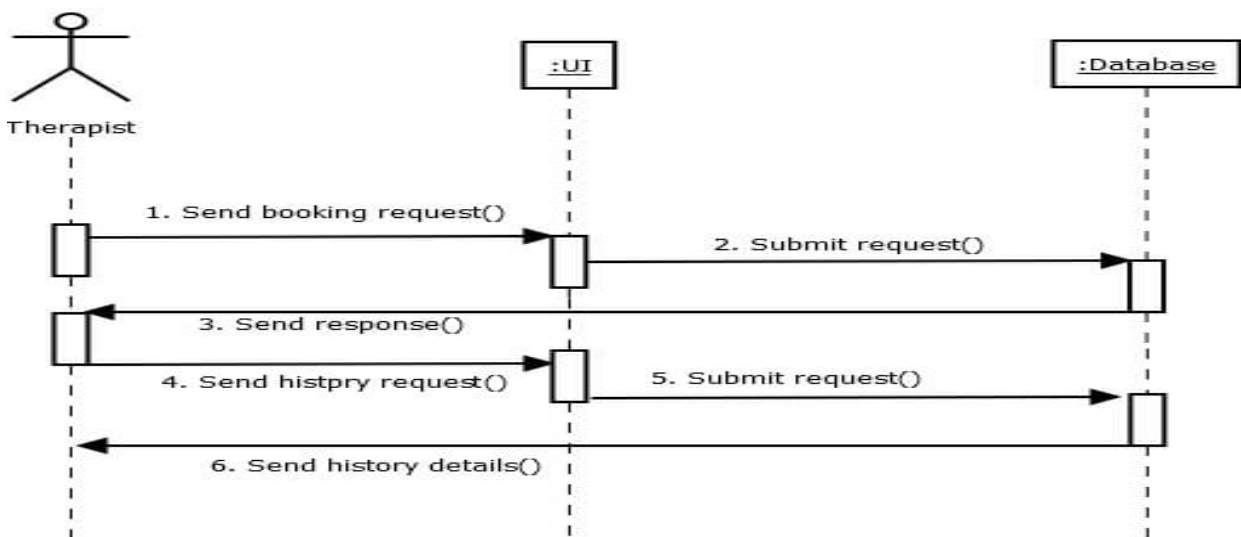


Figure 6.8 Sequence Diagram of Therapist

CHAPTER 7

7. SYSTEM IMPLEMENTATION

Common mental disorders such as depression, anorexia, dementia, posttraumatic stress disorder or schizophrenia affect millions of people around the world. Most people believe that mental disorders are not usual or happen to other people that have specific personal damage. When in fact, mental disorders are prevalent and familiar. Many families think they are not prepared to face the fact that some loved one has a mental problem. A mental disorder causes different interferences in the thinking and behaviour of the affected person. These interferences could vary from mild to severe, and could result in an inability to live routines in daily life and ordinary demands. Common mental disorders such as depression and anorexia affect millions of people around the world. They may be related to a single incident causing excessive stress on the person or by a series of different stressful events. It is also well known that mental disorders tend to increase in countries experiencing generalized violence or recurrent natural disasters. Mental health has always been an important and challenging issue. The modernized (hectic) lifestyle and workload take a toll over people over time making them more prone to mental disorders like mood disorder and anxiety disorder. Thus, the risk mental health problems increase in working professionals. Previous work they have addressed to utilize the data from mental health survey 2019 that contains the data of working professionals for both tech and non-tech company employees. We process data to find the features influencing the mental health of employees or features that can help to predict the mental health of the employee the feature can be either personal or professional. We apply multiple machine learning algorithms to find the model with the best accuracy. We take precision and recall as the measure to check the performance of different ML models. This information presents an opportunity for us to extend the use of ML algorithms to predict the mental health status of user and also assign a therapist if needed.

7.1 DATA PRE-PROCESSING

Variables with more than 50% missingness were removed from analysis (202 variables excluded). Redundant variables were also removed (134 excluded). Additionally, variables with no variance were removed (32 excluded) and those with variance near zero were combined into one variable, if possible, e.g., dust, mold, and pollen allergy collapsed into allergy.

Ultimately, 85 variables were determined to be suitable for analysis. As most machine learning techniques require complete datasets, missing values were imputed with tree-based imputation with the R package mice.

7.2 DATASET EXTRACTION

The bug reports dataset is extracted from the repository. The bug reports were classified into five levels: lowest, low, medium, high and highest. Although software engineers usually follow a specific guideline on how to assign severity of reported bugs, however, the categorization process seems to be evaluated imprecisely. In this research work, we treat lowest and low severity as non-severe, while high and highest severity as severe bugs. The medium class, as it is investigated, is the default option for reporting a bug and it seems that a large number of reporters report any confusing bug as a medium. The datasets mainly consist of three features including bug ID, a short description (i.e., one-line summary) and the severity level of the bug. The description of each bug is represented as a short text (snippet). From the bug reports, the severity and the short description of the bug report were mainly used for the prediction process. Too many words are repetitive in the dataset where the number of distinct words is significantly fewer than the total number of all words exists within the dataset (765 out of 9016).

7.3 STATISTICAL ANALYSIS

All analyses were performed in R. First, a learning curve was plotted with the entire dataset in order to check if our study was sufficiently powered. Then, we split our data into a training-set (60% of the sample), a tune-set (10%), and a test-set (30%). Splitting data allows for more accurate determination in how the model will perform in a new dataset and helps alleviate overfitting, i.e., to fit the training data too closely to accurately predict other datasets. Stratified random sampling was used to ensure that the twin pairs would not get separated between the datasets, thus avoiding potential overfitting. Additionally, we preserved an equal distribution of the outcome between each set. Descriptive statistics were created for each set to determine the quality of the partition.

7.4 PERFORMANCE OF PREDICTIONS

The performance of predictions from considered models were determined by the area under the receiver operating characteristic curve (AUC). We created prediction models using several machine learning techniques: Support vector(sum), K- nearest neighbour (KNN), Decision Tree, Random Forest, Naïve Bayes. To determine which produced the best fitting model for a test set. Using cross validation, each technique trained multiple models using the training set and tested their performance on a subset of the training set. The model with the lowest error was then tested using the tune set. Once the performance in the tune set was deemed satisfactory, the final models were then fitted to the test set. Parameter tuning was guided in part by standard practice when available, however a majority of the tuning took place through the random search function in R package mlr. Random search was completed using cross-validation with 3 iterations Decision Tree, 50 times. Variable importance was calculated for tree-based model: random forest. Confidence intervals at 95% were created for each AUC by bootstrapping predictions 10,000 times. Positive Predictive and Negative predictive values were obtained for the best performing model.

7.5 MACHINE LEARNING ALGORITHMS

In this section, different machine learning algorithms for predicting the severity level of the bugs are described and utilized in this study. These algorithms are mainly concerned to deal with unstructured data such as text. In this study, machine learning algorithms namely: Support vector (SVM), K- nearest neighbour (KNN), Decision Tree, Random Forest, Naïve Bayes algorithms are used to classify severity levels of bug reports. These algorithms are described in the following subsections:

- **NAÏVE BAYES:**

It is a probability-based classifier which applies Bayes theorem. For Naïve Bayes features should have strong independence. The Naïve Bayes classifier estimates the class conditional probability of class label with the assumption that all attributes are independent [34]. The NB has been widely used in the domain of text classification due to its simplicity and effectiveness. On the other hand, NBM is similar to the naïve Bayes classifier except that it considers a weight for each feature during probability calculations. The weight of each feature can be determined by specific distributions or as a parametric model. The parameters can be estimated based on

the training data. For the case of NBM, the distribution of data is assumed as a multinomial model.

- **SUPPORT VECTOR MACHINE (SVM)**

SVM takes data points as input and gives the output as a hyperplane. It divides the classes using a plane (hyperplane) also known as the decision boundary. Where the decision boundary must maximize the distance of the nearest element of each class. Decision boundary separates the points into different classes. Let n be number of data points

$$(x_1 * y_1) \dots (x_n * y_n) \quad (1)$$

Here x is the real vector and y represents the class (0 or 1). The maximum margin hyperplane can be defined as

$$w * x - b \quad (2)$$

Where w is the normal vector and b is the offset of hyperplane along.

- **DECISION TREE:**

The decision tree (DT) is a machine learning predictive approach used for classification and regression problems. It creates a prediction model by learning decision rules from data on a tree-like model. A DT model has a tree structure in which each node represents a test on a given variable and each branch represents the outcome of that test. Leaf nodes of the tree represent a class label (decision). The path from the root to leaf nodes represents a classification rule. Different DT based algorithms have been implemented including J48, random tree, random forest, Logistic model trees, and many others. Generally, decision tree classifiers have good accuracy. It is a typical inductive approach to learn knowledge from data. In this study, three DT algorithms were applied. These algorithms include J48, Random Forest (RF), and Logistic Model Trees (LMT)

- **K-NEAREST NEIGHBOURS (KNN)**

As it is a supervised learning algorithm, we will need labelled data for training the model. When classifying a data point, we look at K nearest neighbours of the data the majority will decide which class the data point will belong. Where nearest distance is in form of Euclidean distance, whose formula is

After this, the input data point will be assigned to the class which have the highest probability. The probability can be represented as

In KNN we have to be prudent when selecting the value of K. A smaller value of k usually leads to an irregular decision boundary and high value of k leads a smoother decision boundary.

- **RANDOM FOREST**

It is an ensemble model which means it uses many machine learning algorithms to increase its performance as compared to other machine learning algorithms. Random forest randomly picks up a subset from the training data set and generate various decision trees. It will predict the class of test class objects using the decision trees.

7.6 EXPERIMENTAL SETTINGS

The validation approach used in our study to evaluate the performance of different classifiers is the k-fold cross validation approach. This approach was used to generate a test set and avoid the over-fitting problem. The cross-validation approach performs independent tests without requiring separate test data samples and without reducing the data samples used to build prediction models. In the k-fold cross-validation process, the original data is classified into k groups, so that each group is used once as a validation set and the remaining data as a training set. In this study, the 10-fold cross-validation was used to evaluate the performance of different classifiers. Stratified based sampling was used to generate training and testing sets. In each case, a prediction model was trained using 9 folds of the data and tested on the remaining fold and the results were averaged.

7.7 PERFORMANCE EVALUTION

Once a classifier is trained successfully and a prediction model is generated, performance evaluation measures must be applied on a separate dataset called test set to evaluate the performance of the generated model. In our experiments, the performance of the generated prediction models is evaluated using performance measures. Each measure provides a different perspective of performance evaluation and a broader set of performance results to compare.

The accuracy measures how correctly a classifier predicts class labels. It is calculated as the percentage of true positive and true negative rates to the number of all instance.

CHAPTER 8

8. RESULTS AND DISCUSSION

This is a web-app for predicting user's mental health status using Machine Learning and if needed user can have a booking of a therapist. Input is the answers which are given by the user both objective and descriptive type questionnaire. Using Machine Learning Algorithms, our program predicts user's mental health status. He/she/they can successfully sign up and create their own account. Then they can book their slots. After booking the slots they can attend a set of objective and descriptive questions.

OBJECTIVE QUESTIONS

1. You feel rested
2. You feel that too many demands are being made on you
3. You are irritable or grouchy
4. You have too many things to do
5. You feel lonely or isolated
6. You find yourself in a situation of conflict
7. You feel you are doing things you really like
8. You feel tired
9. You feel you may not manage to attain your goals
10. You feel calm
11. You have too many decisions to make
12. You feel frustrated
13. You are full of energy
14. You feel tense
15. Your problems seem to be piling up

16. You feel you are in a hurry
17. You feel safe and protected
18. You have many worries
19. You are under pressure from other people
20. You feel discouraged
21. You enjoy yourself
22. You are afraid of the future
23. You feel you are doing things because you have to not because you want to
24. You feel criticized and judged
25. You are light-hearted
26. You feel mentally exhausted
27. You have trouble relaxing
28. You feel loaded down with responsibility
29. You have enough time for yourself
30. You feel under pressure from deadlines

DESCRIPTIVE QUESTIONS

How is your sleep?

How is your energy?

What do you prefer - to stay at home rather than going out and doing new things? Why?

What makes you irritated? Why?

How do you feel about yourself?

What makes you relaxed?

When both the objective and descriptive questions have been answered then the predicted results are available with respect to the responses provided by user. If in need of therapist the booking remains. In the test result, there will be

Sentiment	Negative/positive
Negativity	0-5
Positivity	0-5
Neutral	0-5

Following the test, He/she can connect to nearby therapist or counsellor of their taste. It also provides a daily routine plan for user. Each user's personality data is stored as history.

- **Feature extraction**

There is an immense need to analyse and monitor a person's mental health as justified in our previous work. We utilize deep learning feature extraction algorithm like sentiment analysis to analyse mental health of person from a set of questions. Using polarity, the statement to the question is tested as positive or negative.

- **Mental health recognition**

Registered patients can attend a survey which is objective and descriptive type. Both the tests are used for analysing the mental health of patient. The answers of the descriptive type questions are used for sentimental analysis and this result is also used for analysis.

- **Mental health classification**

The mental health condition is analysed based on the sentiment analysis. Each objective type answer is tested using training set. Positivity is tested based on polarity. If the polarity is above +0.5 the statement is positive, below -0.5 statement is negative and between is neutral statement. The descriptive type statement gives the level of mental condition as high, low and moderate.

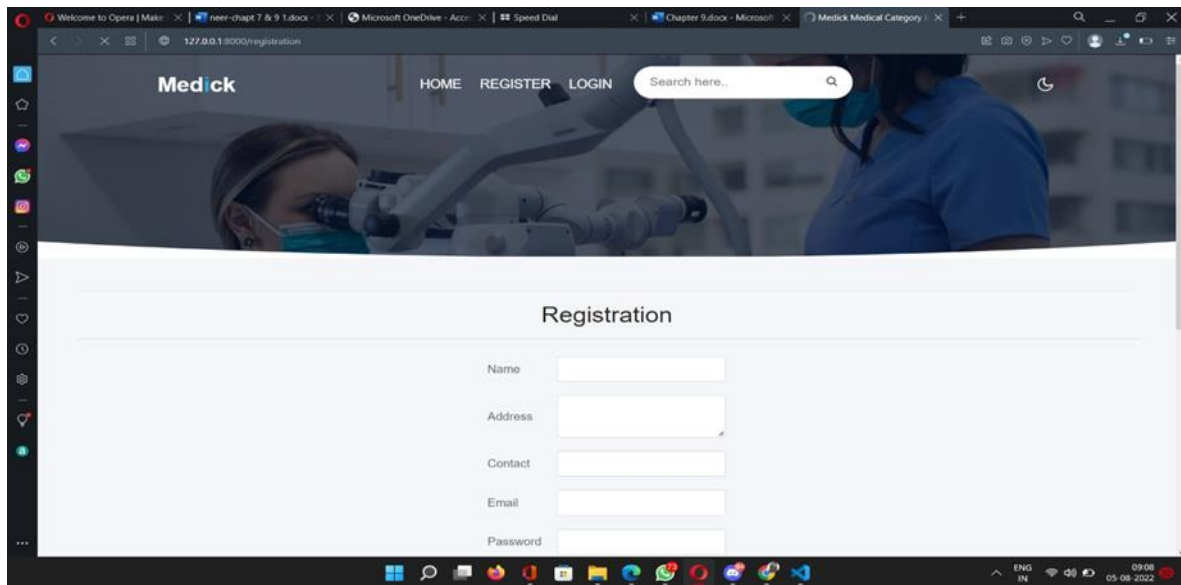
- **Recommendation**

There is a set of activities provided irrespective of the mental condition of the user. Based on the test result output we have the sentiment of user and the negativity, positivity and the

neutral measures also the mental health status (high/low). On the base of this result user also get a to-do-list for 30 days.

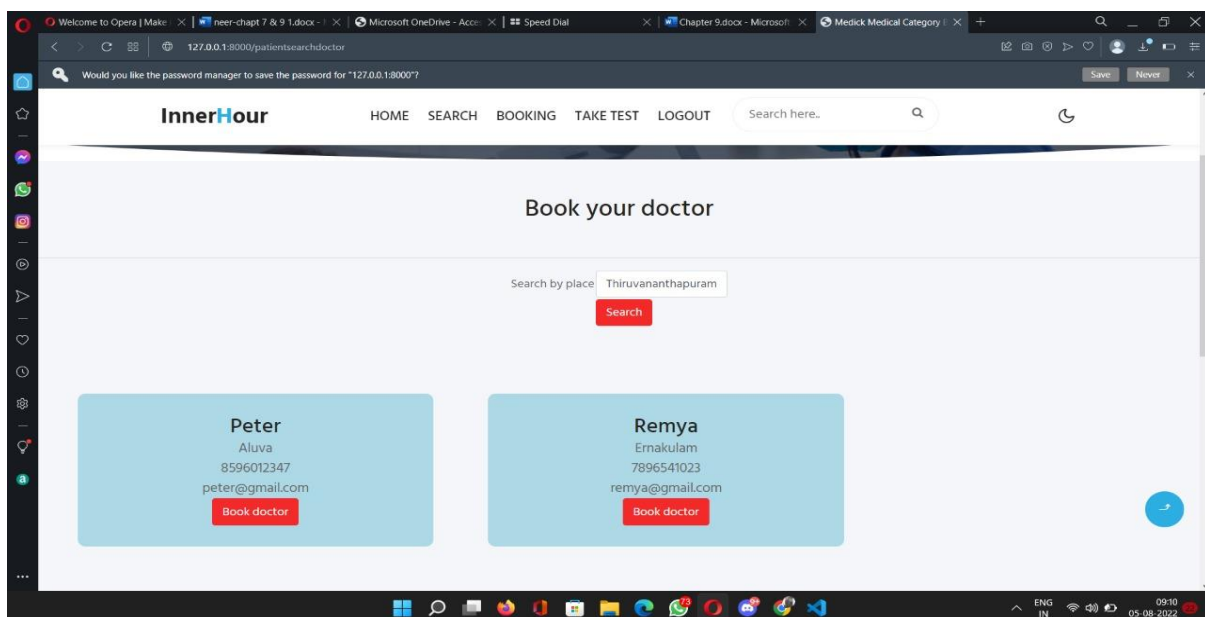
- **Easy handling**

Well-designed application that runs in the background and provide the recommendation based on their social media activities.



The screenshot shows a web browser window displaying the 'Medick' registration page. The page has a header with the 'Medick' logo, navigation links for 'HOME', 'REGISTER', and 'LOGIN', and a search bar. Below the header is a large image of a person using a microscope. The main content area is titled 'Registration' and contains a form with the following fields: Name, Address, Contact, Email, and Password. The browser's address bar shows the URL '127.0.0.1:8000/registration'.

Figure 8.1 Registration



The screenshot shows a web browser window displaying the 'InnerHour' 'Book your doctor' page. The page has a header with the 'InnerHour' logo, navigation links for 'HOME', 'SEARCH', 'BOOKING', 'TAKE TEST', and 'LOGOUT', and a search bar. Below the header is a large image of a person. The main content area is titled 'Book your doctor' and contains a search bar with the text 'Search by place Thiruvananthapuram' and a red 'Search' button. Below the search bar are two doctor profiles, each with a 'Book doctor' button. The first profile is for 'Peter Aluva' with contact number '8596012347' and email 'peter@gmail.com'. The second profile is for 'Remya Ernakulam' with contact number '7896541023' and email 'remya@gmail.com'. The browser's address bar shows the URL '127.0.0.1:8000/patientsearchdoctor'.

Figure 8.2 Booking Therapist

InnerHour HOME SEARCH BOOKING TAKE TEST LOGOUT Search here.

Questionnaire

	Never	Sometimes	Often	Usually
1 You feel rested	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2 You feel that too many demands are being made on you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3 You are irritable or grouchy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4 You have too many things to do	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5 You feel lonely or isolated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6 You find yourself in a situation of conflict	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7 You feel you are doing things you really like	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8 You feel tired	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9 You feel you may not manage to attain your goals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10 You feel calm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11 You have too many decisions to make	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12 You feel frustrated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 8.3 Questionnaire

InnerHour HOME SEARCH BOOKING TAKE TEST LOGOUT Search here.

How is your sleep?
not good

How is your energy?
low

What do you prefer - to stay at home rather than going out and doing new things? Why?
to stay and avoid people

What makes you irritated? Why?
peoples

How do you feel about yourself?
awful

Figure 8.4 Questionnaire

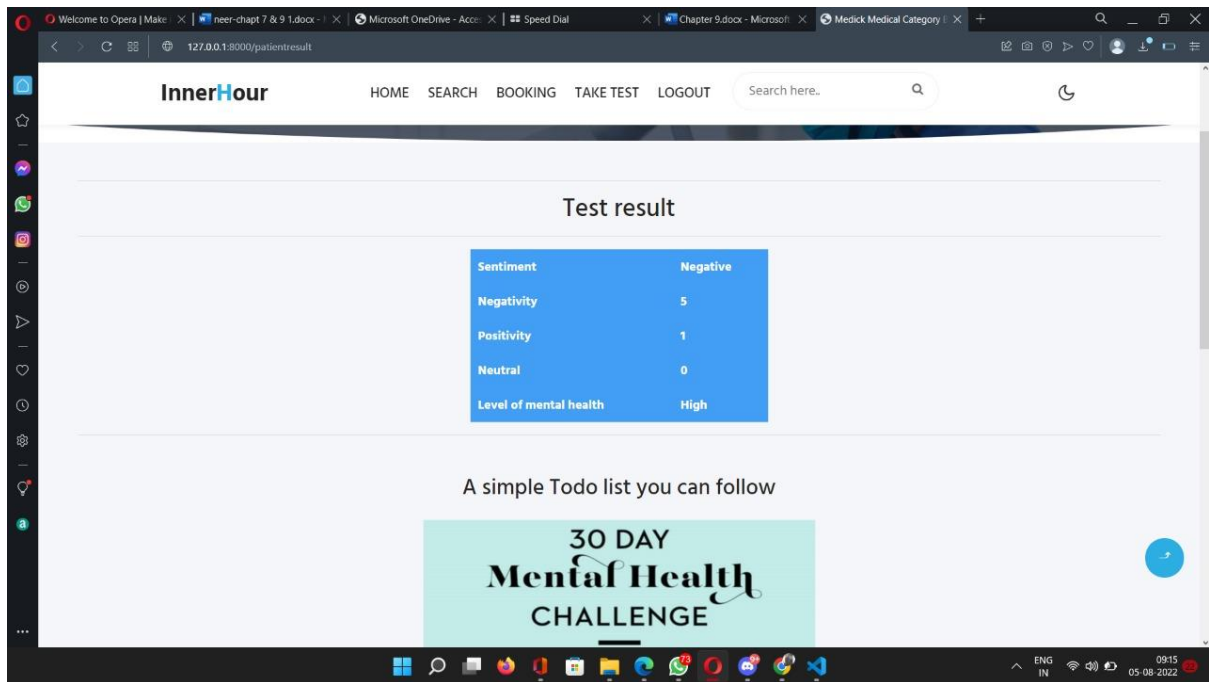


Figure 8.5 Result

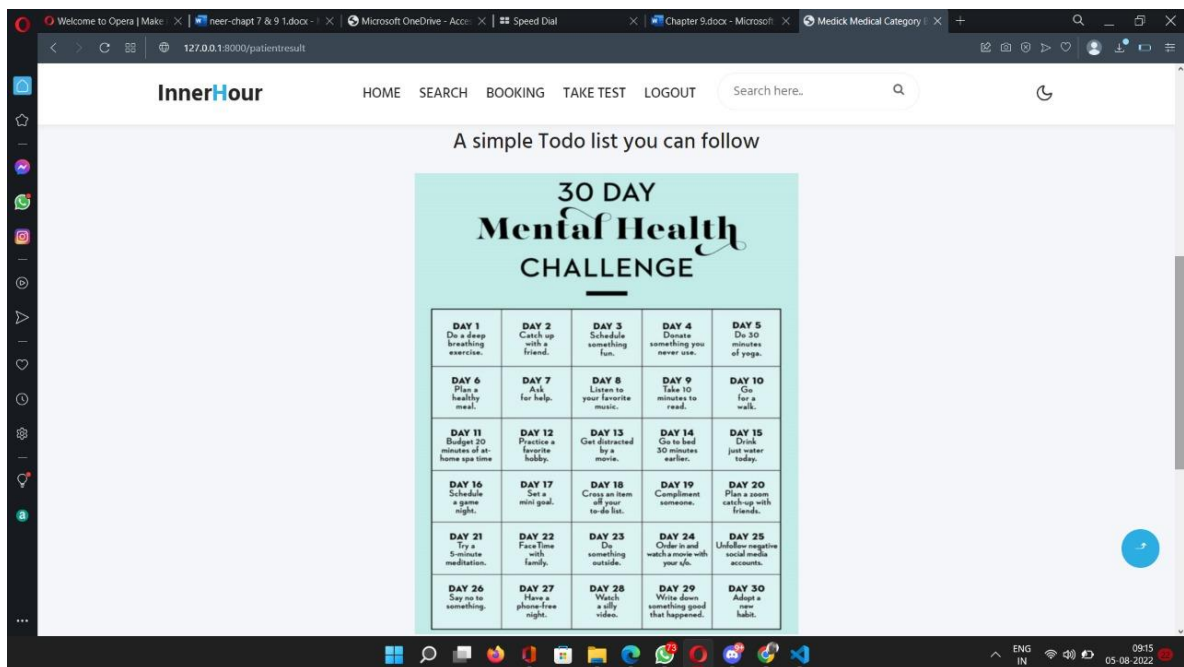


Figure 8.6 To Do list

CHAPTER 9

9. CONCLUSION AND FUTURE HOPE

9.1 CONCLUSION

Mental health care is always treated as a taboo in our society, where people fear to express the concerns of their mental health. As mental health is as important as someone's physical health, we proposed a new system which enables user to evaluate their mental health status and facilitate them to take an appointment with a therapist nearby easily without getting worried about privacy due to the high confidentiality of the system. The system is developed with high accuracy, high flexibility, lowcost with advanced technology and machine learning algorithms that enables good performance. The system is easy to develop and implement. We focused on designing a user friendly web application for which enables user to know about their mental health status. On the other hand, the inclusion of a dynamic analysis over emotions, improved the detection of users that presents signs of anorexia and depression, showing the usefulness of considering the changes of emotions over time. It is worth mentioning the simplicity and interpretability of both representations, then creating a more straightforward analysis of the results. Finally, the capability to model the emotional behaviour of users using their data provided by user give an opportunity for future wellness facilitating technologies. This kind of technology can serve as warning systems that provide wide-area analysis and information related to a mental disorder respecting user privacy. We believe that it's important to respect user data and individual privacy so the usage of this data are strictly for research and analysis only, and the misuse or mishandling of the information is prohibited

9.2 FUTURE SCOPE

Many different techniques and algorithms had been introduced and proposed to test and solve the mental health problems of user. There are still many solutions that can be refined .In addition, there are still many problems to be discover and tested using a wide variety of settings in machine learning for the mental health domain. As classifying the mental health data is generally a very challenging problem, the features used in the machine learning algorithms will significantly affect the performance of the classification. The existing studies and research show that machine learning can be a useful tool in helping understand psychiatric disorders. Besides that, it may also help distinguish and classify the mental health problems among patients for further treatment. Newer approaches that use data that arise from the integration of various

sensor modalities present in advanced devices have proven to be a convenient resource to recognize the mood state and responses from patients among others. It is noticeable that most of the research and studies are still struggling to validate the results because of insufficiency of acceptable validated evidence, especially from the external sources. Besides that, most of the machine learning techniques might not have the same performance across all the problems. The performance of the machine learning models will vary depending on the data samples obtained and the features of the data. Moreover, machine learning models can also be affected by preprocessing activities such as data cleaning and parameter tuning in order to achieve optimal results. Hence, it is very important for researchers to investigate and analyze the data with various machine learning algorithms to choose the highest accuracy among the machine learning algorithms. Not only that, challenges and limitations faced by the researchers need to be managed with proper care to achieve satisfactory results that could improve the clinical practice and decision making.

REFERENCES

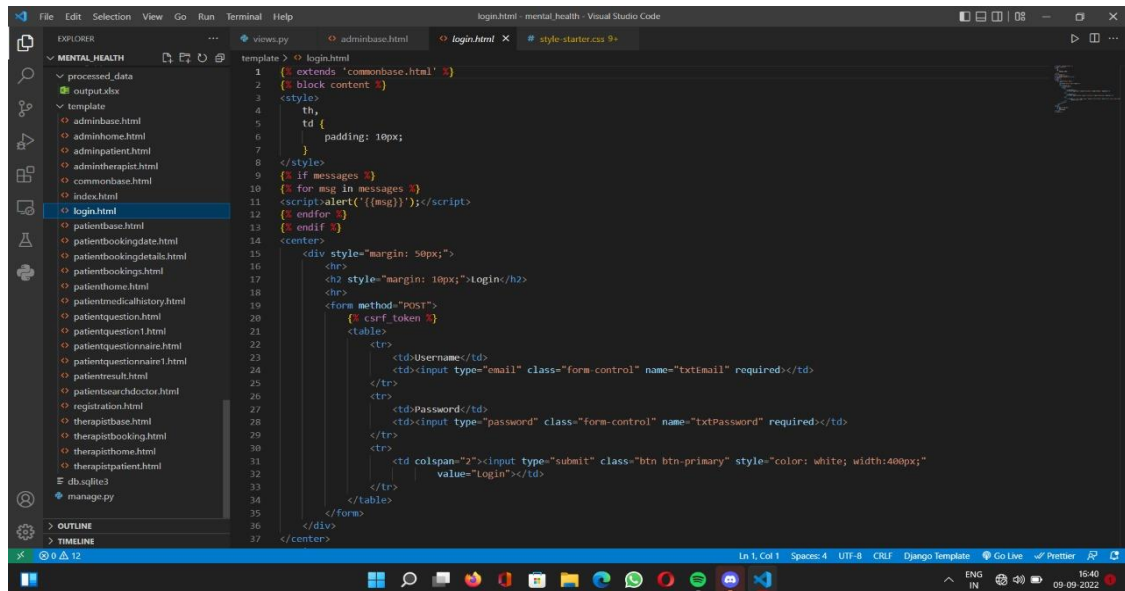
- [1] Rahul Katarya .,Saurav Maan .,"Predicting Mental health disorders using Machine Learning for employees in technical and non-technical companies", 2020 IEEE INTERNATIONAL CONFERENCE ON ADVANCES AND DEVELOPMENTS IN ELECTRICAL AND ELECTRONICS ENGINEERING (ICADEE 2020)
- [2] A. Grünerbl et al., "Smartphone-based recognition of states and state changes in bipolar disorder patients," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 1, pp. 140–148, 2015, Doi: 10.1109/JBHI.2014.2343154.
- [3] U. Ha et al., "A Wearable EEG-HEG-HRV Multimodal System with Simultaneous Monitoring of Tes for Mental Health Management," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 6, pp. 758– 766, 2015, Doi: 10.1109/TBCAS.2015.2504959.
- [4] Yadav, S., Kaim, T., Gupta, S., Bharti, U., & Priyadarshi, P. (2020). "Predicting Depression From Routine Survey Data using Machine Learning". *IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*.
- [5] U. S. Reddy, A. V. Thota, and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE Int. Conf. Compute. Intell. Compute. Res. ICCIC 2018, Doi: 10.1109/ICCIC.2018.8782395
- [6] Interpreting Depression From Question-Wise Long-Term Video Recording of SDS Evaluation. Published in: *IEEE Journal of Biomedical and Health Informatics* (Volume: 26, Issue: 2, February 2022)
- [7] AI to Detect Social Media users Depression Polarity Score Published in: 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)
- [8] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, "Review of machine learning algorithms for diagnosing mental illness," *Psychiatry Investigation*, vol. 16, no. 4, pp. 262–269, 2019.
- [9] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PLoS One*, vol. 15, no. 4, Article ID e0230389, 2020.

[10] P. Morillo, H. Ortega, D. Chauca, J. Proaño, D. Vallejo-Huanga, and M. Cazares, “Psycho web: a machine learning platform for the diagnosis and classification of mental disorders,” in *Advances in Neuroergonomics and Cognitive Engineering*, pp. 399–410, Springer International Publishing, Berlin, Germany, 2019.

APPENDIX 1

SAMPLE CODE

1. LOGIN PAGE

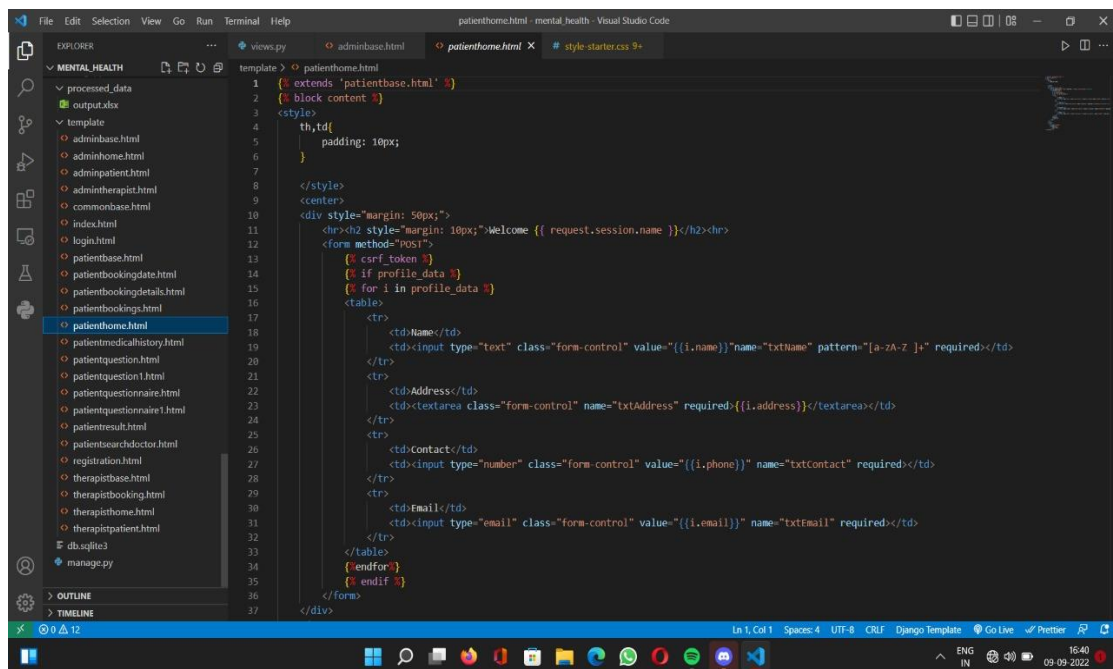


```

1 {% extends 'commonbase.html' %}
2 {% block content %}
3 <style>
4     th,
5     td {
6         padding: 10px;
7     }
8 </style>
9 {% if messages %}
10 {% for msg in messages %}
11 <script>alert('{{msg}}');

```

2. PATIENT HOME



```

1 {% extends 'patientbase.html' %}
2 {% block content %}
3 <style>
4     th,td{
5         padding: 10px;
6     }
7 </style>
8 <center>
9 <div style="margin: 50px;">
10 <hr><h2 style="margin: 10px;">Welcome {{ request.session.name }}</h2><hr>
11 <form method="POST">
12     {% csrf_token %}
13     {% if profile_data %}
14     {% for i in profile_data %}
15     <table>
16     <tr>
17         <td>Name</td>
18         <td><input type="text" class="form-control" value="{{i.name}}" name="txtName" pattern="[a-zA-Z ]+" required></td>
19     </tr>
20     <tr>
21         <td>Address</td>
22         <td><textarea class="form-control" name="txtAddress" required>{{i.address}}</textarea></td>
23     </tr>
24     <tr>
25         <td>Contact</td>
26         <td><input type="number" class="form-control" value="{{i.phone}}" name="txtContact" required></td>
27     </tr>
28     <tr>
29         <td>Email</td>
30         <td><input type="email" class="form-control" value="{{i.email}}" name="txtEmail" required></td>
31     </tr>
32     </table>
33     </form>
34     {% endfor %}
35     </div>
36 </div>
37 </center>

```


3. PATIENT MEDICAL HISTORY

```

1 # Generated by Django 4.0.5 on 2022-08-01 04:53
2
3 from django.db import migrations, models
4 import django.db.models.deletion
5
6
7 class Migration(migrations.Migration):
8
9     dependencies = [
10         ('mental_app', '0001_initial'),
11     ]
12
13     operations = [
14         migrations.CreateModel(
15             name='Medicalhistory',
16             fields=[
17                 ('id', models.BigAutoField(auto_created=True, primary_key=True, serialize=False, verbose_name='ID')),
18                 ('history', models.CharField(max_length=500)),
19                 ('hereditary', models.CharField(max_length=500)),
20                 ('bookingid', models.ForeignKey(on_delete=django.db.models.deletion.CASCADE, to='mental_app.booking')),
21             ],
22         ),
23     ]

```

```

1 # Generated by Django 4.0.5 on 2022-08-01 04:53
2
3 from django.db import migrations, models
4 import django.db.models.deletion
5
6
7 class Migration(migrations.Migration):
8
9     dependencies = [
10         ('mental_app', '0001_initial'),
11     ]
12
13     operations = [
14         migrations.CreateModel(
15             name='Medicalhistory',
16             fields=[
17                 ('id', models.BigAutoField(auto_created=True, primary_key=True, serialize=False, verbose_name='ID')),
18                 ('history', models.CharField(max_length=500)),
19                 ('hereditary', models.CharField(max_length=500)),
20                 ('bookingid', models.ForeignKey(on_delete=django.db.models.deletion.CASCADE, to='mental_app.booking')),
21             ],
22         ),
23     ]

```

4. THERAPIST DISTRICT

The screenshot shows the Visual Studio Code interface with a file explorer on the left and a code editor on the right. The file explorer shows a project structure with folders like 'data', 'mental_app', and 'mental_health'. The code editor displays a Django migration file named '0003_therapist_district.py'. The code defines a migration class that adds a 'district' field to the 'therapist' model.

```

1 # Generated by Django 4.0.5 on 2022-08-01 05:03
2
3 from django.db import migrations, models
4
5
6 class Migration(migrations.Migration):
7
8     dependencies = [
9         ('mental_app', '0002_medicalhistory'),
10    ]
11
12     operations = [
13         migrations.AddField(
14             model_name='therapist',
15             name='district',
16             field=models.CharField(default='', max_length=100),
17         ),
18    ]
19

```

5. DATA COMPARISON

The screenshot shows the Visual Studio Code interface with a file explorer on the left and a code editor on the right. The file explorer shows a project structure with folders like 'data', 'mental_app', and 'mental_health'. The code editor displays a JSON array of tweet data from a file named 'tweetdata.txt'. The data includes fields like 'created_at', 'id', 'id_str', and 'text'.

```

359 [{"created_at": "Fri Jun 02 00:05:18 +0000 2017", "id": "870431089055092737", "id_str": "870431089055092737", "text": "Woman with extreme #anxiet
360
361 [{"created_at": "Fri Jun 02 00:05:20 +0000 2017", "id": "870431096973938690", "id_str": "870431096973938690", "text": "@anaeun_ sorry it's just a
362
363 [{"created_at": "Fri Jun 02 00:05:20 +0000 2017", "id": "870431097015992320", "id_str": "870431097015992320", "text": "RT @TheGoogleFactz: Art ag
364
365 [{"created_at": "Fri Jun 02 00:05:20 +0000 2017", "id": "870431097229901824", "id_str": "870431097229901824", "text": "RT @Kon_K: Men, seeking h
366
367 [{"created_at": "Fri Jun 02 00:05:20 +0000 2017", "id": "870431098387484672", "id_str": "870431098387484672", "text": "damn straight boi https://
368
369 [{"created_at": "Fri Jun 02 00:05:21 +0000 2017", "id": "870431101893853185", "id_str": "870431101893853185", "text": "RT @humansofny: \u201cI do
370
371 [{"created_at": "Fri Jun 02 00:05:21 +0000 2017", "id": "870431102107832321", "id_str": "870431102107832321", "text": "lob) I have a lot of posit
372
373 [{"created_at": "Fri Jun 02 00:05:21 +0000 2017", "id": "870431104741679104", "id_str": "870431104741679104", "text": "RT @lipglock: From: depres
374
375 [{"created_at": "Fri Jun 02 00:05:22 +0000 2017", "id": "870431106541260801", "id_str": "870431106541260801", "text": "well I can relate. ive bee
376
377 [{"created_at": "Fri Jun 02 00:05:22 +0000 2017", "id": "870431107052965888", "id_str": "870431107052965888", "text": "RT @mememtribute: me trying
378
379 [{"created_at": "Fri Jun 02 00:05:22 +0000 2017", "id": "870431107258482689", "id_str": "870431107258482689", "text": "RT @ImpactADHD: These unex
380
381 [{"created_at": "Fri Jun 02 00:05:22 +0000 2017", "id": "870431106709032960", "id_str": "870431106709032960", "text": "RT @oraltwjnk: Depression
382
383
384
385
386
387
388
389
390
391
392
393
394
395

```

6. MENTAL HEALTH EVALUATION

The screenshot shows a Visual Studio Code window with the following details:

- File Explorer (Left):** Displays a project named 'MENTAL_HEALTH'. The file structure includes:
 - data
 - dictionary.tsv (selected)
 - tweetdata.txt
 - mental_app
 - __pycache__
 - migrations
 - static
 - _init_.py
 - admin.py
 - apps.py
 - models.py
 - tests.py
 - views.py
 - mental_health
 - __pycache__
 - asgi.py
 - settings.py
 - urls.py
 - wsgi.py
 - processed_data
 - output.xlsx
 - template
 - db.sqlite3
 - manage.py
- Main Editor:** Displays the content of 'dictionary.tsv'. The file contains a list of words and their associated sentiment scores, such as:

Word	Score	Part of Speech	Sentiment
strongsubj	1	activist	neutral
strongsubj	1	activist	neutral
strongsubj	1	actual	neutral
strongsubj	1	actuality	neutral
strongsubj	1	actually	neutral
strongsubj	1	acumen	positive
strongsubj	1	adamant	negative
strongsubj	1	adamantly	negative
weaksubj	1	adaptable	positive
weaksubj	1	adaptability	positive
weaksubj	1	adaptive	positive
weaksubj	1	addict	negative
weaksubj	1	addiction	negative
weaksubj	1	adept	positive
weaksubj	1	adeptly	positive
weaksubj	1	adequate	positive
weaksubj	1	adherence	positive
weaksubj	1	adherent	positive
weaksubj	1	adhesion	positive
strongsubj	1	admirable	positive
strongsubj	1	admirer	positive
strongsubj	1	admirable	positive
strongsubj	1	admirably	positive
strongsubj	1	admiration	positive
strongsubj	1	admire	positive
strongsubj	1	admiring	positive
strongsubj	1	admiringly	positive
weaksubj	1	admission	positive
weaksubj	1	admission	positive
strongsubj	1	admit	positive
strongsubj	1	admittedly	positive
strongsubj	1	admonish	negative
strongsubj	1	admonisher	negative
strongsubj	1	admonishingly	negative
strongsubj	1	admonishment	negative
strongsubj	1	admonition	negative
strongsubj	1	adolescents	neutral
- Status Bar (Bottom):** Shows 'Ln 1, Col 1', 'Spaces: 4', 'UTF-8', 'CRLF', 'Plain Text', 'Go Live', 'Prettier', and the date '05-08-2022'.