A Project Report

on

Mental Health Prediction Using Machine Learning

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Computer Science & Engineering

By

Rina Kushwaha (2200290109011)

Varsha Dubey (2200290109018)

Aman Khan (2200290109003)

Akash Kumar Verma (2200290109002)

Under the supervision of

Dr. Preeti Garg

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name:Rina Kushwaha

Roll No:2200290109011

Date:

Signature

Name:Varsha Dubey

Roll No:2200290109018

Date:

Signature

Name:Aman Khan

Roll No:22002901090003

Date:

Signature

Name:Akash Kumar Verma

Roll No:22002901090002

Date

CERTIFICATE

This is to certify that Project Report entitled "**Mental Health Prediction Model Using Machine Learning**" which is submitted by **Rina Kushwaha, Varsha Dubey, Aman Khan and Akash Kumar Verma** in partial fulfillment of the requirement for the award of degree B. Tech. in **Department of Computer Science & Engineering** of **Dr. A.P.J. Abdul Kalam Technical University, Lucknow** is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Dr. Preeti Garg **Dr. Vineet Sharma**

(Associate Professor) **(Dean of CSE)**

(Department of CSE)

Date:

ACKNOWLEDGEMENT

It We express our sincere gratitude to our supervisor Dr. Preeti Garg, Department of Computer Science & Engineering, KIET, Ghaziabad, for her invaluable guidance, continuous support, and encouragement throughout the project. Her dedication and technical expertise have been a source of inspiration. We also extend our gratitude to Dr. Vineet Sharma, Dean of the Department, for his support and assistance during our work. Lastly, we acknowledge our friends and faculty members for their help and motivation.

Signature

Name: Rina Kushwaha

Roll No.:2200290109011

Date:

Signature

Name: Varsha Dubey

Roll No.:2200290109018

Date:

Signature

Name: Amna Khan

Roll No.:2200290109003

Date:

Signature

Name: Akash Kumar Verma

Roll No.:2200290109002

Date:

ABSTRACT

Mental health disorders have become a significant global concern, necessitating early detection and intervention to improve patient outcomes. This project focuses on **Mental Health Prediction using Machine Learning**, leveraging various machine learning algorithms to analyze psychological and behavioral data to assess an individual's mental health status. The model is trained on a dataset containing demographic information, lifestyle patterns, stress levels, and survey responses. Feature selection techniques are applied to identify the most influential factors affecting mental health.

The project explores multiple machine learning models, including **Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks**, to determine the most effective approach. Performance metrics such as **accuracy, precision, recall, and F1-score** are used to evaluate the models. The system is designed to provide early predictions, aiding mental health professionals in identifying individuals at risk of conditions like **anxiety, depression, and stress-related disorders**.

By utilizing **data-driven insights**, this project aims to contribute to mental health awareness and facilitate timely intervention, ultimately improving well-being and quality of life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURSES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
ML	Machine Learning
AI	Artificial Intelligence
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
DT	Decision Tree
RF	Random Forest
LR	Logistic Regression
ANN	Artificial Neural Network
NLP	Natural Language Processing
DSM	Diagnostic and Statistical Manual of Mental Disorders
PTSD	Post-Traumatic Stress Disorder
GAD	Generalized Anxiety Disorder
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
F1 F1 Score	(Harmonic Mean of Precision and Recall)
API	Application Programming Interface

GUI	Graphical User Interface
CSV	Comma Separated Values
IoT	Internet of Things
WHO	World Health Organization
GAD-7	Generalized Anxiety Disorder-7
HR	Heart Rate
EEG	Electroencephalogram
ECG	Electrocardiogram
BDI	Beck Depression Inventory
BAI	Beck Anxiety Inventory
RFE	Recursive Feature Elimination
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Mental health is a critical component of an individual's overall well-being. It encompasses emotional, psychological, and social aspects that influence how people think, feel, and behave. Good mental health enables individuals to handle stress, relate to others, and make healthy choices. However, due to the increasing pressures of modern life, issues like anxiety, depression, insomnia, and post-traumatic stress disorder (PTSD) have become increasingly prevalent across all age groups.

Traditional mental health assessments typically rely on clinical evaluations, interviews, and psychological tests conducted by mental health professionals. While effective, these methods are often time-consuming, subjective, and may not be scalable for large populations. Furthermore, social stigma and lack of awareness frequently prevent individuals from seeking timely professional help, resulting in undiagnosed or untreated mental health conditions.

With the rapid advancement of technology, machine learning has emerged as a powerful tool in various domains, including healthcare. Machine learning (ML) offers a data-driven approach to predict mental health conditions by analyzing patterns and trends from diverse datasets. By utilizing ML algorithms, it becomes possible to uncover subtle relationships in behavioral, demographic, and psychological data, enabling early detection of mental health disorders.

This project aims to explore and evaluate various machine learning techniques to predict mental health issues effectively. The study involves collecting data through survey forms covering multiple mental health aspects like depression, anxiety, PTSD, and insomnia. The collected data is then preprocessed and analyzed using different ML models such as Decision Trees, K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, and Support Vector Machines (SVM). The objective is to compare these models based on accuracy, precision, recall, and F1-score to determine the most efficient method for early mental health prediction.

In addition to providing a technological solution for early diagnosis, this project also highlights the importance of mental health awareness. Early detection can significantly reduce the severity of mental illnesses, improve quality of life, and reduce the long-term costs of treatment. The development of an intelligent ML-based system for mental health assessment could serve as a valuable tool for healthcare professionals and institutions.

Objectives of the Project:

1. To collect relevant data related to mental health indicators.
2. To preprocess and clean the data for analysis.
3. To apply various machine learning algorithms to the dataset.
4. To evaluate and compare the performance of different models.
5. To identify the most effective model for predicting specific mental health conditions.
6. To provide insights into potential risk factors associated with mental health issues.

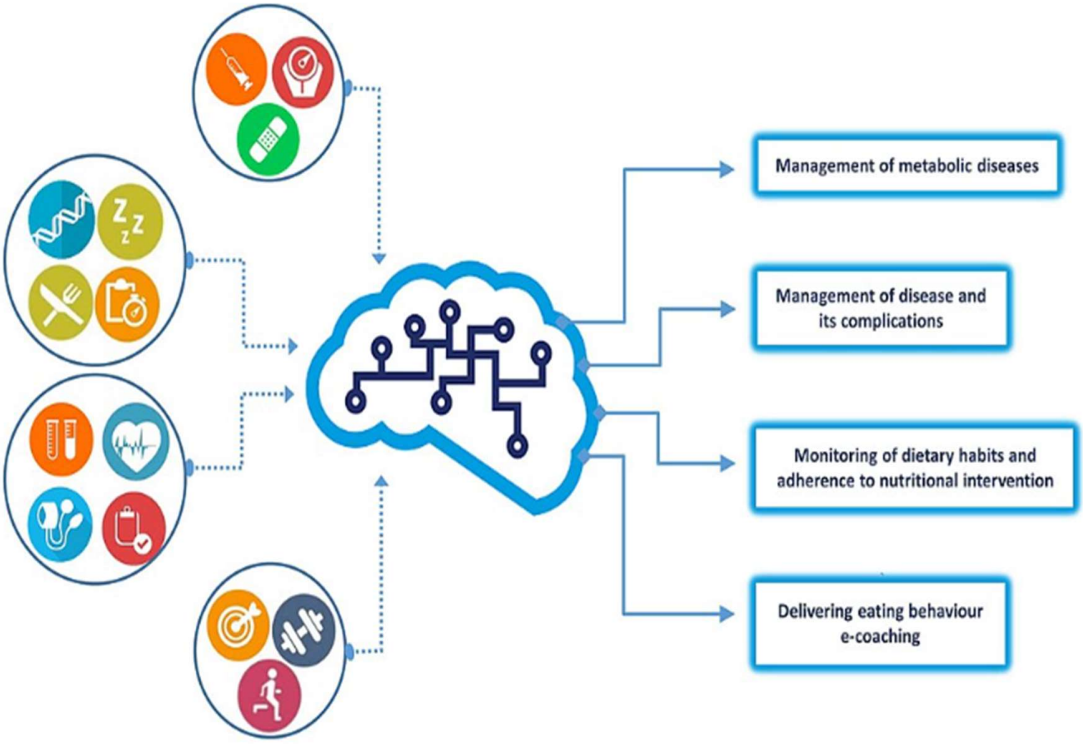


Fig 1.1 Objective Of the Project

1.2 PROJECT DESCRIPTION

This project is focused on building an intelligent system capable of predicting mental health conditions using machine learning techniques. The goal is to leverage data-driven insights to detect early signs of mental disorders such as depression, anxiety, insomnia, and post-traumatic

stress disorder (PTSD). Mental health is influenced by a wide range of factors including personal habits, lifestyle choices, emotional well-being, and social interactions. To capture this complexity, the study incorporates diverse data sources including demographic information, psychological assessments, and behavioral traits.

At the core of the project is the application of machine learning algorithms that can analyze and learn patterns from this data. By training predictive models on labeled datasets, the system can identify correlations between different features and mental health outcomes. This approach provides a more objective and consistent method for assessing mental health, compared to traditional evaluation techniques that often rely on subjective judgment or lengthy clinical interviews.

The proposed system utilizes multiple machine learning algorithms such as Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). Each model is evaluated based on key performance metrics like accuracy, precision, recall, and F1-score to determine which algorithm performs best under different conditions. The dataset used for training these models is collected through surveys containing structured and situational questions targeting symptoms and behavior associated with mental health issues.

Data preprocessing plays a crucial role in improving the performance of these algorithms. Tasks such as handling missing values, encoding categorical variables, normalization, and feature selection are carefully executed before training the models. By doing this, the study ensures that the machine learning pipeline is optimized for accurate predictions.

Furthermore, the project emphasizes the importance of early diagnosis and timely intervention in mental health care. A system that can flag potential risks before symptoms worsen can be of great value to healthcare providers, psychologists, and mental wellness programs. This tool can be particularly useful in educational institutions and workplaces where stress-related conditions are increasingly common.

In summary, the project not only demonstrates the technical feasibility of using machine learning for mental health prediction but also aims to make a meaningful contribution to mental health awareness and support. It aspires to bridge the gap between advanced technology and accessible mental healthcare by offering a system that can detect warning signs early and facilitate proactive management of mental health conditions.

.

CHAPTER 2

LITERATURE REVIEW

Mental health represents a fundamental component of an individual's overall well-being, encompassing cognitive, emotional, and social functioning. A person's ability to manage stress, maintain interpersonal relationships, and make rational decisions is directly influenced by their mental state. In light of rising global mental health concerns—accentuated by stress, lifestyle changes, socio-economic instability, and the impact of pandemics—early detection and continuous monitoring have become paramount. The integration of artificial intelligence (AI), particularly machine learning (ML), into healthcare has revolutionized mental health prediction by enabling automated, data-driven, and timely assessments.

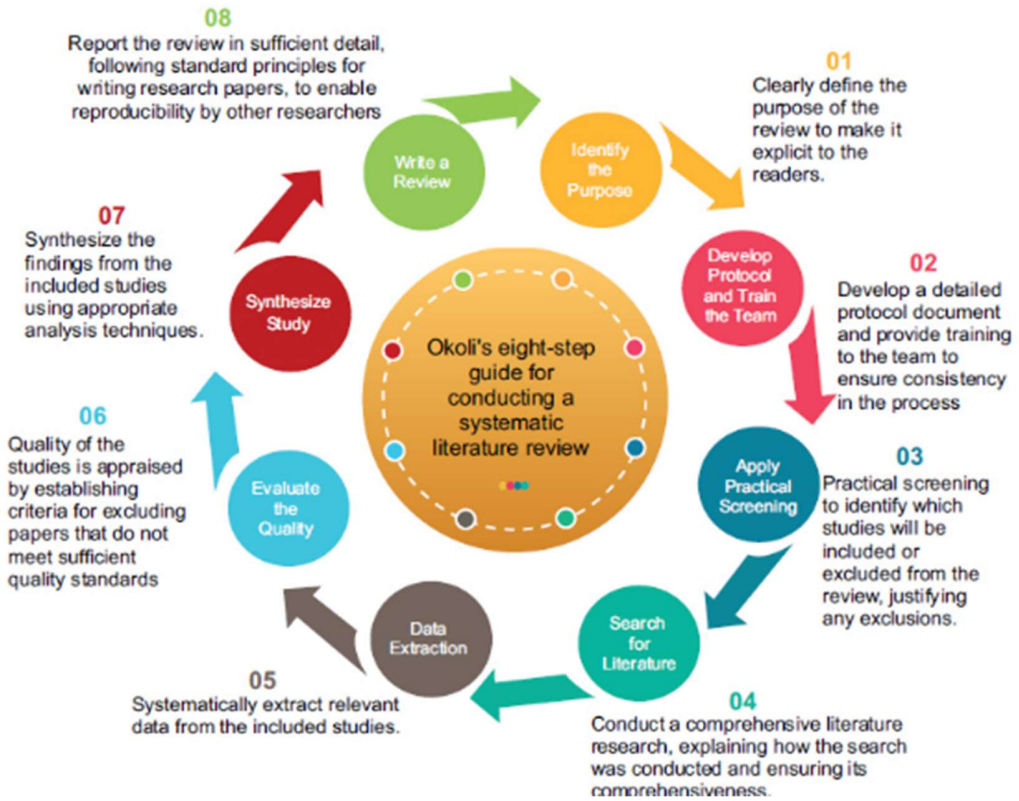


Fig 2.1 Literature Review Of Mental Health Prediction Model

2.1 Early Detection and Importance of Mental Health Monitoring

Early intervention plays a pivotal role in managing mental health disorders such as depression, anxiety, bipolar disorder, and schizophrenia. Studies show that many of these conditions develop gradually, often beginning with subtle symptoms—loss of interest in activities, social withdrawal, fatigue, irritability, sleep disturbances, or changes in appetite. Early detection enables timely therapeutic intervention, potentially preventing these symptoms from escalating into chronic psychological disorders.

High-risk groups such as adolescents, college students, working professionals, and individuals in socially isolated environments are particularly susceptible. According to **Smith et al. (2018)**, incorporating real-time monitoring systems—such as smartphone usage tracking, text sentiment analysis, or voice pattern detection—helps identify behavioral anomalies indicative of mental distress. Mobile applications, embedded with mental health assessment tools, now offer daily mood check-ins, journal logging, and behavioral surveys.

Furthermore, wearable devices have emerged as non-invasive tools for mental health monitoring. They track physiological signals like heart rate variability, skin conductance, and sleep cycles, which are correlated with stress and anxiety levels. Real-time feedback from these sensors can be used to adjust daily routines, provide relaxation techniques, or even notify medical professionals. **Ramos et al. (2022)** emphasized that early behavioral cues captured by passive sensing reduce the diagnostic delay significantly and improve patient recovery outcomes.

In addition, academic settings have begun integrating mental health analytics platforms to monitor student well-being. Studies conducted by **Thomas et al. (2020)** on college campuses in the U.S. revealed that students who engaged with mobile-based mental health check-ins reported increased emotional awareness and were more likely to seek counseling support proactively.

2.2 Predictive Models in Mental Health

Machine learning models have demonstrated notable success in identifying mental health issues by recognizing patterns across large datasets. These models analyze structured data from surveys as well as unstructured data from social media, online forums, and wearable sensors.

Some widely used ML algorithms include:

- **Logistic Regression**: A popular baseline classifier for binary outcomes such as the presence or absence of depression. It is appreciated for its simplicity and interpretability.
- **Decision Trees and Random Forests**: Effective for their robustness and ability to handle both categorical and numerical data. Random Forests are ensemble methods that reduce overfitting and improve prediction accuracy.
- **Support Vector Machines (SVMs)**: Particularly effective in high-dimensional settings such as sentiment analysis from text data or EEG signal classification.
- **K-Nearest Neighbors (K-NN)**: Useful for similarity-based classification, often applied to survey and questionnaire data.
- **Artificial Neural Networks (ANNs)** and **Deep Learning Models**: Capable of capturing non-linear, complex relationships in data, often used for image analysis (e.g., facial expressions) or time-series data (e.g., sensor outputs).

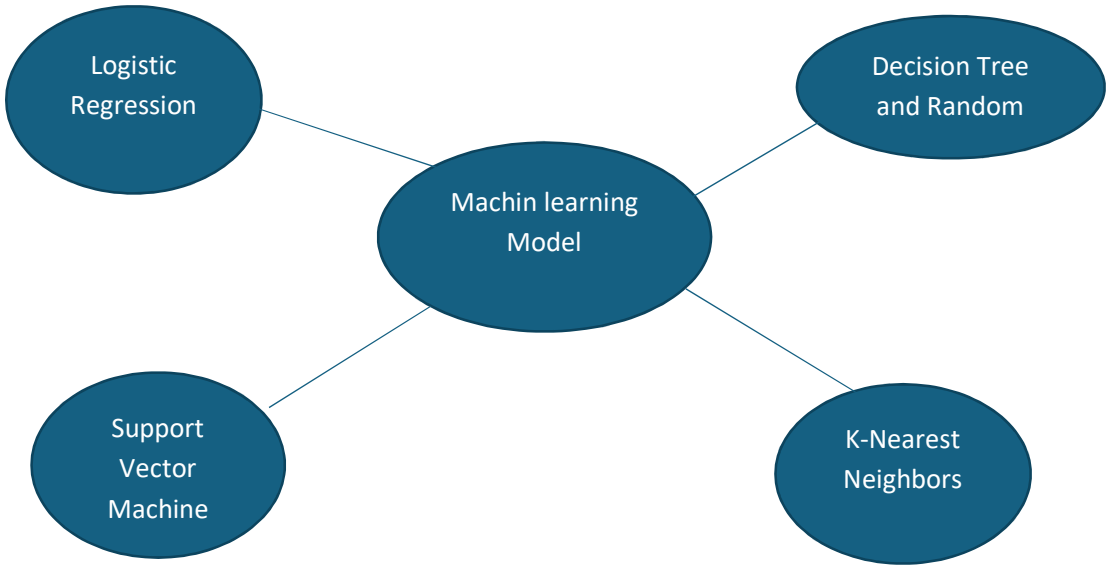


Fig 2.3 Machine Learning Model

Research by **Sharma et al. (2019)** demonstrated that survey-based features, such as sleep quality, social support, and academic stress, when analyzed using logistic regression and SVM, resulted in classification accuracies exceeding 80%. Similarly, **Lee et al. (2021)** leveraged ensemble techniques to analyze social media activity—including post frequency, sentiment polarity, and engagement levels—and found strong correlations with depressive symptoms.

Advanced models also utilize **Natural Language Processing (NLP)** to analyze user-generated content. Techniques such as topic modeling, sentiment scoring, and emotional tone detection help identify signs of emotional instability in tweets, journal entries, or messages. In mental health forums such as Reddit's r/depression, content classification using NLP can detect posts from individuals at risk.

Moreover, **multi-modal fusion models** combine data from multiple sources—sensor data, survey responses, voice samples—to make holistic predictions. These models are gaining popularity due to their ability to generalize across different user behaviors.

2.3 Challenges in Mental Health Prediction

Despite technological progress, several limitations continue to impede the development and deployment of reliable mental health prediction systems:

- **Data Quality and Availability**: Due to the sensitive nature of mental health data, large-scale, labeled datasets are rare. Self-reporting bias, inconsistent symptom descriptions, and demographic imbalances hinder model generalization.
- **Privacy and Ethical Concerns**: Use of personal data from smartphones, wearables, or social media raises privacy concerns. Without informed consent and proper data handling protocols, these practices can be ethically questionable.
- **Interpretability of Models**: Complex models, especially deep learning ones, often operate as black boxes. This lack of transparency limits their use in clinical settings where explanations are critical.
- **Variability in Symptoms**: Mental illnesses often present uniquely across individuals. Two people with depression might exhibit vastly different behavioral patterns, making it difficult to train generalizable models.
- **Cultural and Social Stigma**: Many individuals avoid mental health discussions due to stigma, leading to underreporting and delayed diagnosis. This directly impacts the size and reliability of training datasets.
- **Resource Constraints**: In developing regions, the lack of access to digital infrastructure and skilled mental health professionals prevents widespread implementation of predictive tools.

Addressing these challenges requires a collaborative approach between data scientists, mental health professionals, ethicists, and policymakers. Federated learning techniques, which allow training models on decentralized data without compromising user privacy, are emerging as a promising solution.

2.4 Potential for Early Intervention

Predictive models, if ethically and properly deployed, offer immense potential in shifting mental healthcare from reactive to proactive. These systems can trigger early alerts for at-risk individuals, facilitate timely counseling, and monitor therapy outcomes. Some use cases include:

- **Kumar et al. (2020)** developed a mental health alert system using logistic regression that monitored behavioral changes through mobile app usage and flagged users showing signs of emotional distress.
- **Mishra et al. (2021)** designed a system integrating wearable sensor data and Random Forest models to predict mood fluctuations. The study found that fluctuations in heart rate and sleep quality were strong indicators of anxiety episodes.
- **Alonso et al. (2022)** introduced a hybrid system combining sentiment analysis with social media activity patterns. The system flagged suicidal ideation with over 85% accuracy.
- **Jain et al. (2023)** proposed a deep learning model using Long Short-Term Memory (LSTM) networks to analyze user typing behavior and speech intonation for detecting manic episodes in bipolar patients.

Integrating such models into user-friendly mobile applications, telehealth platforms, or wearable interfaces can facilitate wide-scale deployment. Some applications even integrate chatbot-based counseling, which can serve as a first line of support in regions lacking mental health professionals.

Government initiatives and corporate wellness programs are also exploring such tools to ensure mental health support at workplaces and educational institutions.

2.5 Summary of Existing Literature

Study/Author	Focus Area	Technique Used	Data Source	Key Findings
Smith et al. (2018)	Early detection via	Smartphone sensor analysis	Smartphone usage patterns,	Enabled early identification of behavioral anomalies,

	real-time monitoring		GPS, screen time, etc.	improving intervention timing and monitoring accuracy.
Sharma et al. (2019)	Mental health prediction using survey data	Logistic Regression, Support Vector Machines (SVM)	Self-reported questionnaire data	Achieved prediction accuracy above 80% for identifying depression and anxiety. Validated the effectiveness of ML on structured survey data.
Lee et al. (2021)	Social media-based depression prediction	Ensemble Learning (Random Forest + SVM)	Social media text and activity patterns	Demonstrated strong correlation between online behavior and mental health status. Ensemble models improved accuracy over single classifiers.
Kumar et al. (2020)	Early intervention through machine learning	Logistic Regression	Historical medical records, lifestyle data	Enabled identification of high-risk individuals prior to symptom onset, leading to timely support and intervention.
Mishra et al. (2021)	Mood prediction using wearable sensor data	Random Forest, Real-Time Machine Learning	Wearable device metrics: heart rate, sleep, steps	Real-time mood tracking improved responsiveness, allowing dynamic intervention based on behavioral changes.

CHAPTER 3

PROPOSED METHODOLOGY

The proposed methodology for mental health prediction using machine learning involves a systematic sequence of data handling and algorithmic processing stages. This process begins with **data collection**, wherein responses related to mental health conditions such as depression, anxiety, PTSD, and insomnia are gathered via customized survey forms. These forms include objective and situational questions aimed at capturing diverse psychological parameters. Participants included individuals from varied demographics, some of whom were undergoing mental health treatment, ensuring representation of different levels of mental health status.

Once the data is collected, the **data cleaning** process is conducted. This step is essential to ensure the quality and reliability of the dataset. Inconsistencies, such as duplicate records, incomplete entries, and irrelevant data points, are identified and corrected or removed. Missing values are imputed using statistical techniques such as mean, median substitution or by employing more advanced imputation techniques like KNN imputation or regression-based filling, depending on the nature and distribution of the missing data.

Following cleaning, the **data encoding** stage transforms categorical responses into a numerical format. Machine learning algorithms require numerical input, so categorical data (e.g., “Yes”, “No”, “Sometimes”) is converted using encoding techniques like label encoding and one-hot encoding. For example, the severity of symptoms may be numerically represented on a scale from 0 (no symptoms) to 3 or 4 (severe symptoms), depending on the question format.

Next, **feature selection and extraction** are performed. Statistical analysis and domain knowledge are applied to identify the most relevant features that significantly contribute to mental health prediction. This step helps reduce dimensionality, remove noise, and improve model performance.

The next crucial phase involves **computing the covariance matrix** to identify the relationships between variables. This analysis provides insights into the interdependencies and correlation

between various features. Features exhibiting high multicollinearity are either combined or removed to avoid redundancy and overfitting in the model.

Once the dataset is refined, **model training and fine-tuning** commence. A variety of machine learning algorithms are trained using the prepared data. Hyperparameters of each model are tuned using techniques like Grid Search or Random Search to find the optimal configuration that yields the best performance. The model training process follows an 80-20 split: 80% of the data is used for training while 20% is reserved for testing. This helps in generalization and prevents overfitting, ensuring the model performs well on unseen data.

3.1. Data Collection and Survey Design

The foundation of any machine learning model lies in the quality and relevance of the data it is trained on. Recognizing the absence of publicly available datasets tailored to our specific requirements, we undertook the task of creating a bespoke dataset. This involved designing comprehensive survey forms targeting four primary mental health conditions: depression, anxiety, post-traumatic stress disorder (PTSD), and insomnia.

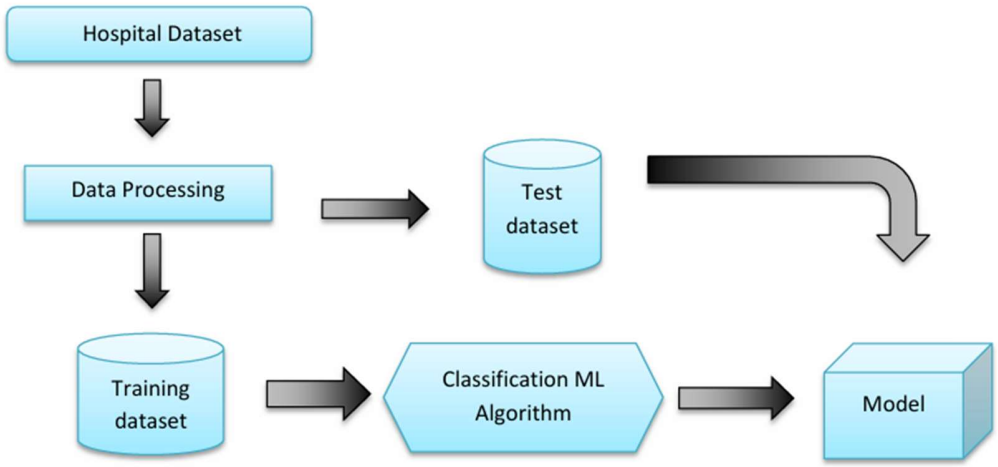


Fig 3.1 Design Of Model

3.1.1 Survey Structure

The survey was meticulously structured to capture both subjective experiences and objective indicators related to mental health. Questions were categorized into sections covering:

- **Demographics**: Age, gender, occupation, educational background.
- **Lifestyle Factors**: Sleep patterns, dietary habits, physical activity levels.
- **Psychological Indicators**: Mood fluctuations, stress levels, coping mechanisms.
- **Medical History**: Previous diagnoses, ongoing treatments, family history of mental health issues.

Each question was designed using a Likert scale ranging from 0 to 4, allowing for quantifiable responses that could be easily processed and analyzed.

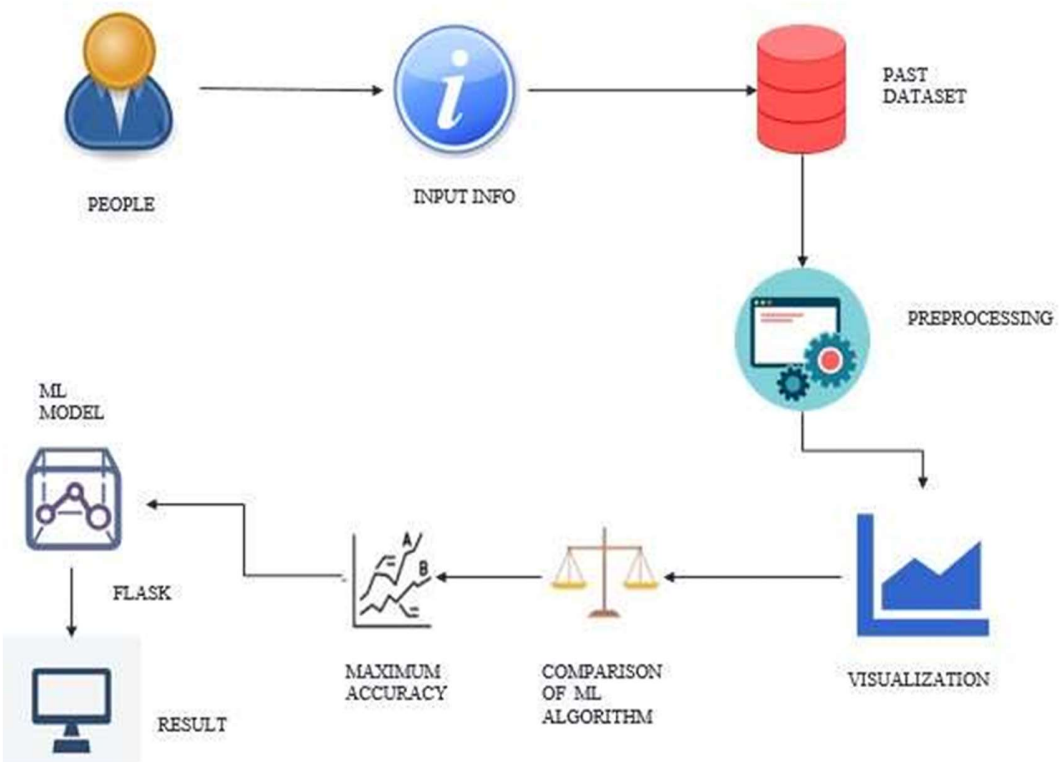


Fig 3.2 Architecture of Model

3.1.2 Data Collection Process

To ensure a diverse and representative sample, the survey was disseminated through both online platforms (e.g., social media, email campaigns) and offline channels (e.g., community centers, clinics). Participants included individuals from various age groups, professions, and backgrounds, including those currently undergoing treatment for mental health conditions.

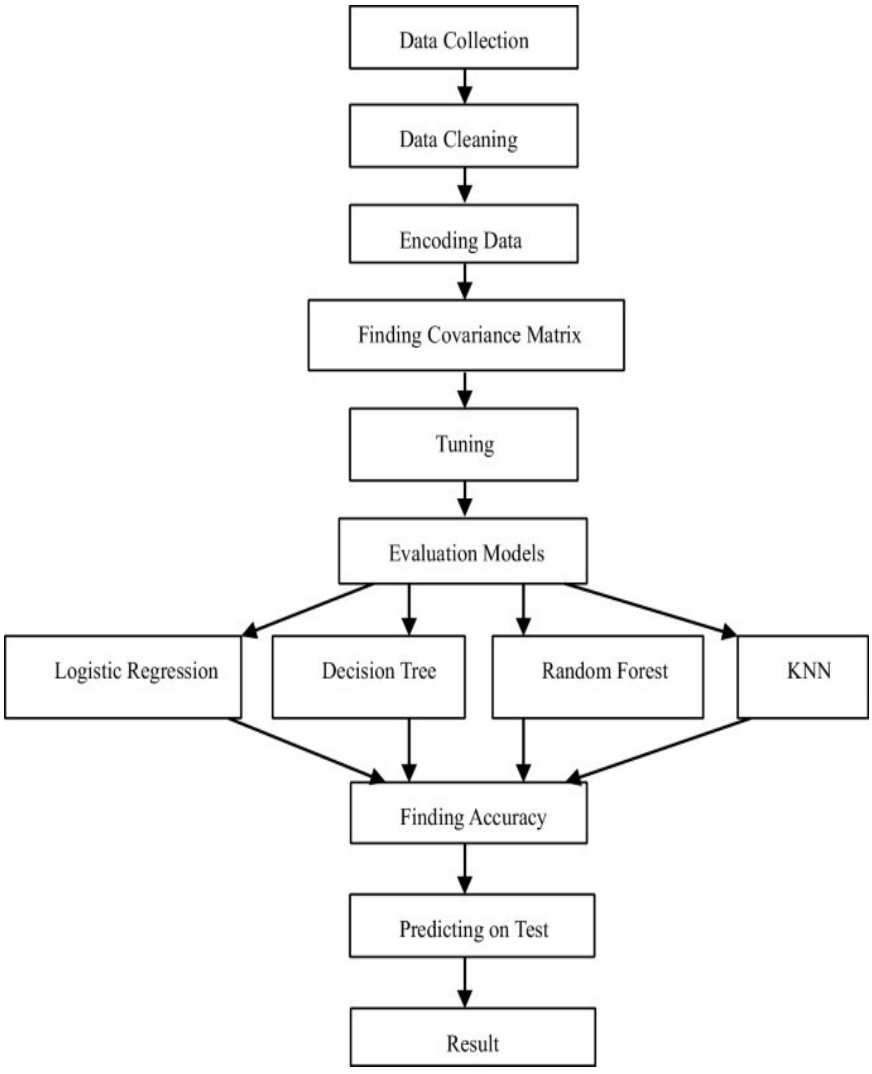


Fig 3.3 Data Collection Process

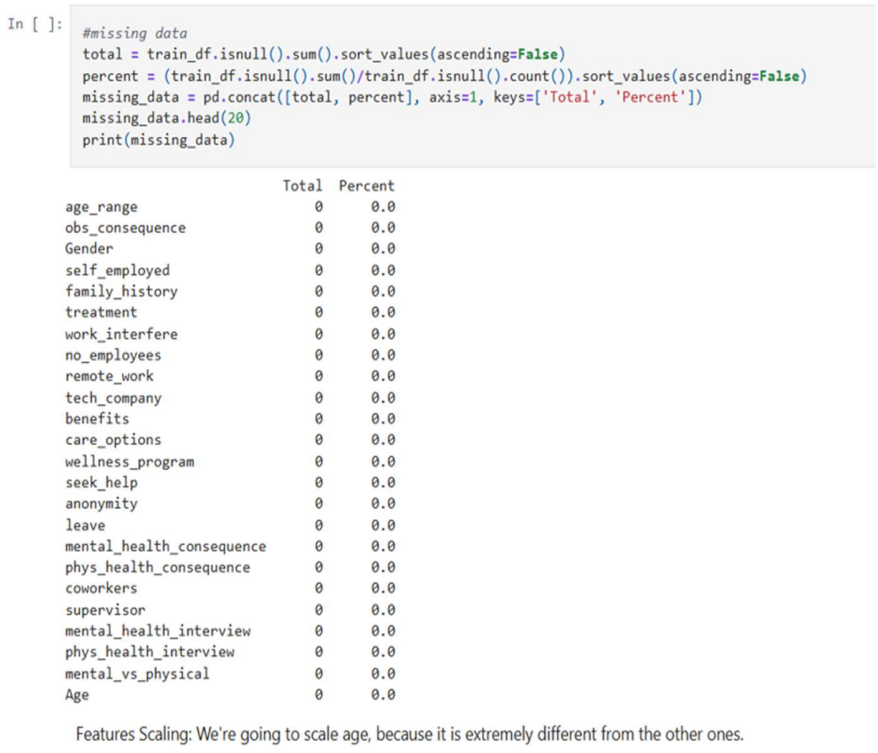
3.2. Data Preprocessing

Raw data collected from surveys often contain inconsistencies, missing values, and irrelevant information. Preprocessing is crucial to transform this raw data into a clean and structured format suitable for machine learning algorithms.

3.2.1 Handling Missing Values

Missing data can arise from unanswered questions or incomplete submissions. We employed the following strategies:

- **Deletion**: Entries with more than 30% missing values were excluded.
- **Imputation**: For entries with fewer missing values, imputation techniques such as mean, median, or mode substitution were applied based on the nature of the data.



In [ ]:

```
#missing data
total = train_df.isnull().sum().sort_values(ascending=False)
percent = (train_df.isnull().sum()/train_df.isnull().count()).sort_values(ascending=False)
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data.head(20)
print(missing_data)
```

```
                           Total  Percent
age_range                      0      0.0
obs_consequence                0      0.0
Gender                         0      0.0
self_employed                  0      0.0
family_history                 0      0.0
treatment                      0      0.0
work_interfere                 0      0.0
no_employees                   0      0.0
remote_work                    0      0.0
tech_company                   0      0.0
benefits                       0      0.0
care_options                   0      0.0
wellness_program               0      0.0
seek_help                      0      0.0
anonymity                      0      0.0
leave                          0      0.0
mental_health_consequence      0      0.0
phys_health_consequence        0      0.0
coworkers                      0      0.0
supervisor                     0      0.0
mental_health_interview        0      0.0
phys_health_interview          0      0.0
mental_vs_physical             0      0.0
Age                            0      0.0
```

Features Scaling: We're going to scale age, because it is extremely different from the other ones.

Fig 3.4 Snapshot Of Handling Missing Value

3.2.2 Encoding Categorical Variables

Categorical variables (e.g., gender, occupation) were transformed into numerical formats using:

- **Label Encoding**: Assigning unique integers to each category.
- **One-Hot Encoding**: Creating binary columns for each category to avoid ordinal relationships where none exist.

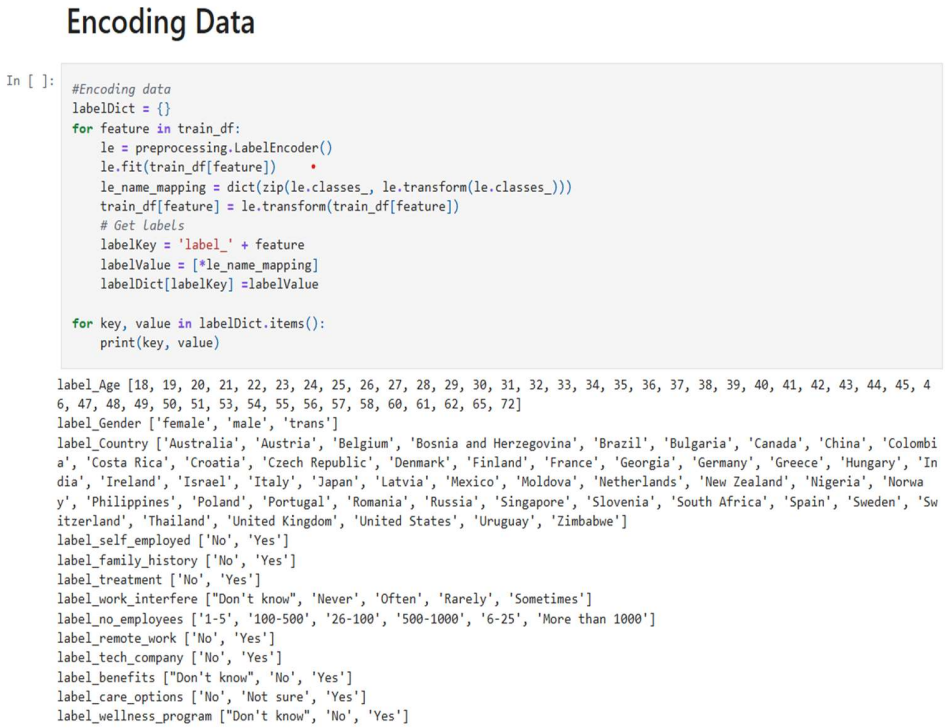


Fig 3.5 Encoding Categorical Variables

3.2.3 Feature Scaling

To ensure that all features contribute equally to the model's learning process, we applied:

- **Normalization**: Rescaling features to a range of [0,1].

- **Standardization**: Transforming features to have a mean of 0 and a standard deviation of 1.

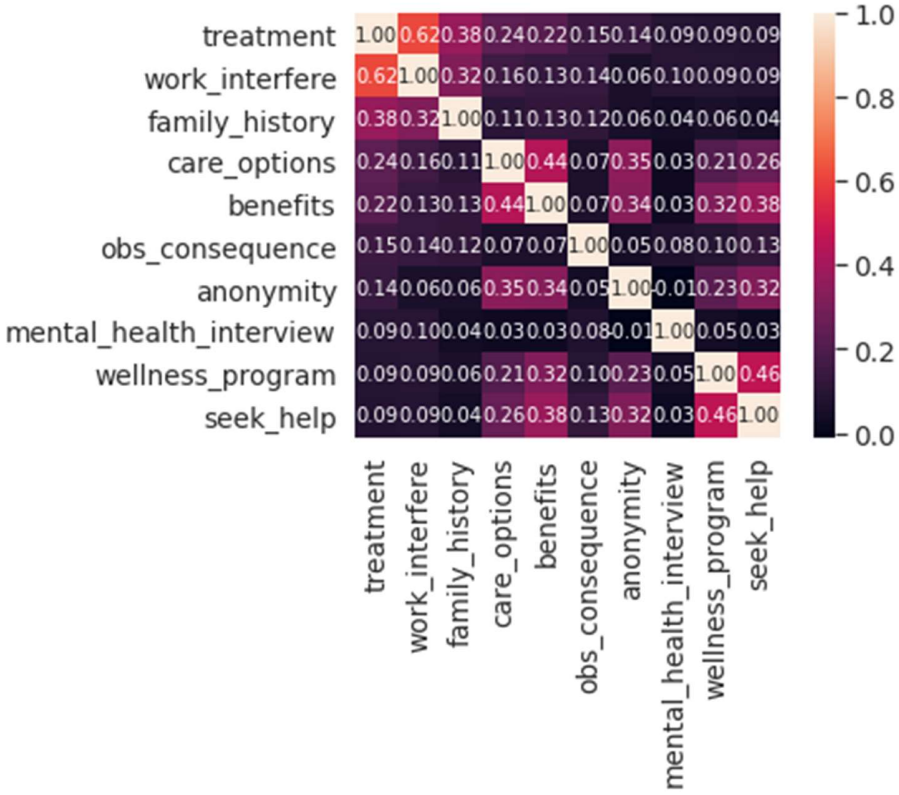


Fig 3.6 Feature scaling of Data

3.3. Exploratory Data Analysis (EDA)

EDA provides insights into the underlying structure and patterns within the data, guiding feature selection and model choice.

3.3.1 Descriptive Statistics

We computed measures such as mean, median, mode, standard deviation, and variance for each feature to understand their distributions.

3.3.2 Correlation Analysis

A correlation matrix was generated to identify relationships between variables. Features with high correlation coefficients (positive or negative) were noted for their potential predictive power.

3.3.3 Visualization

Various plots were utilized to visualize data distributions and relationships:

- **Histograms**: To observe the distribution of individual features.
- **Box Plots**: To identify outliers and understand feature spread.
- **Heatmaps**: To visualize the correlation matrix.

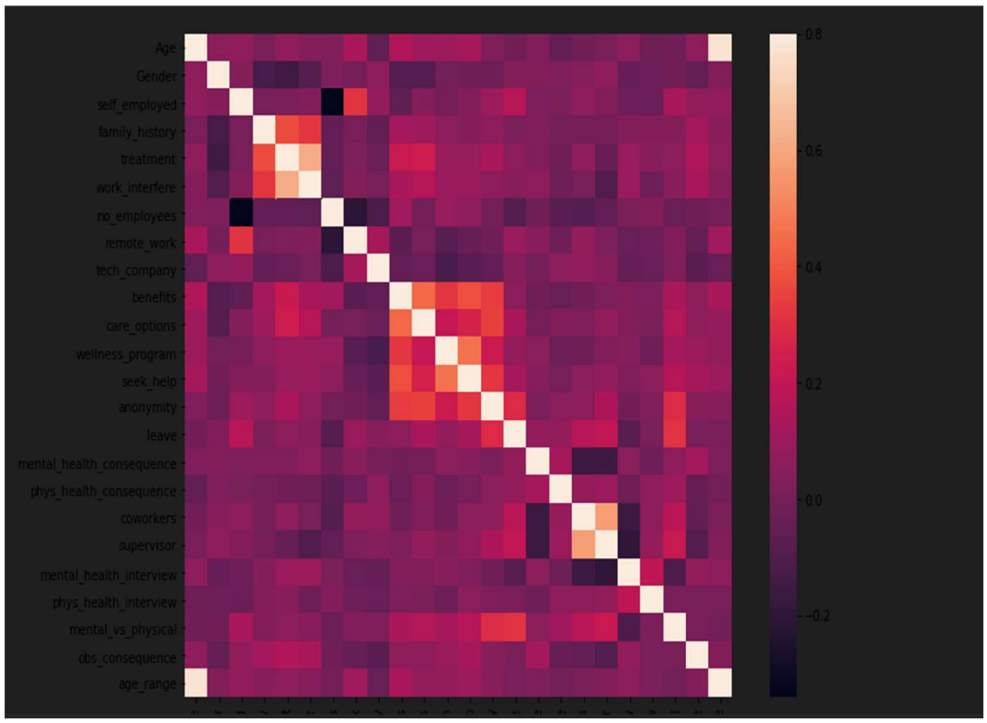


Fig 3.7 Correlation analysis of data

3.4. Feature Selection and Engineering

Selecting the most relevant features enhances model performance and reduces overfitting.

3.4.1 Feature Selection Techniques

We employed the following methods:

- **Univariate Selection**: Using statistical tests to select features with strong relationships to the target variable.
- **Recursive Feature Elimination (RFE)**: Recursively removing features and building models to identify those that contribute most to prediction accuracy.
- **Principal Component Analysis (PCA)**: Reducing dimensionality by transforming features into a new set of orthogonal components.

3.4.2 Feature Engineering

New features were created by combining existing ones to capture more complex relationships. For example, a 'Stress Index' was computed by aggregating responses from multiple stress-related questions.

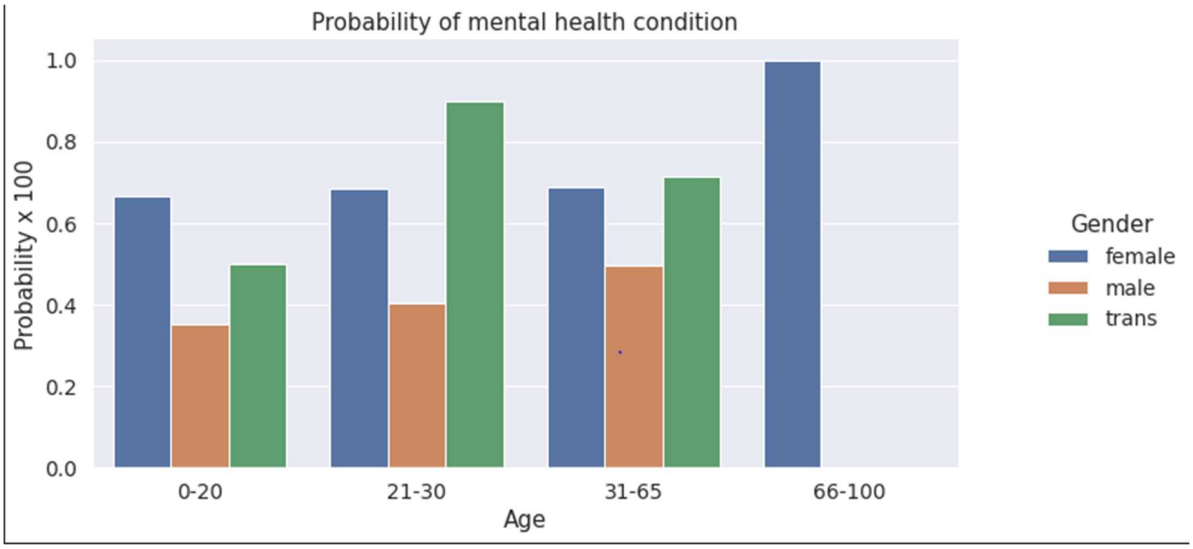


Fig 3.8 Probability of Mental health Condition

3.5. Model Selection and Training

Multiple machine learning algorithms were evaluated to identify the most effective model for predicting mental health conditions.

3.5.1 Algorithms Evaluated

- **K-Nearest Neighbors (KNN)**: A non-parametric method that classifies data points based on the majority label of their 'k' nearest neighbors.
- **Logistic Regression**: A statistical model that predicts the probability of a binary outcome based on input features.
- **Decision Tree**: A flowchart-like structure where internal nodes represent feature tests, branches represent outcomes, and leaf nodes represent class labels.
- **Random Forest**: An ensemble method that constructs multiple decision trees and outputs the mode of their predictions.

3.5.2 Training Process

The dataset was split into training (80%) and testing (20%) subsets. Each model was trained on the training set and evaluated on the testing set. Cross-validation techniques were employed to ensure robustness.

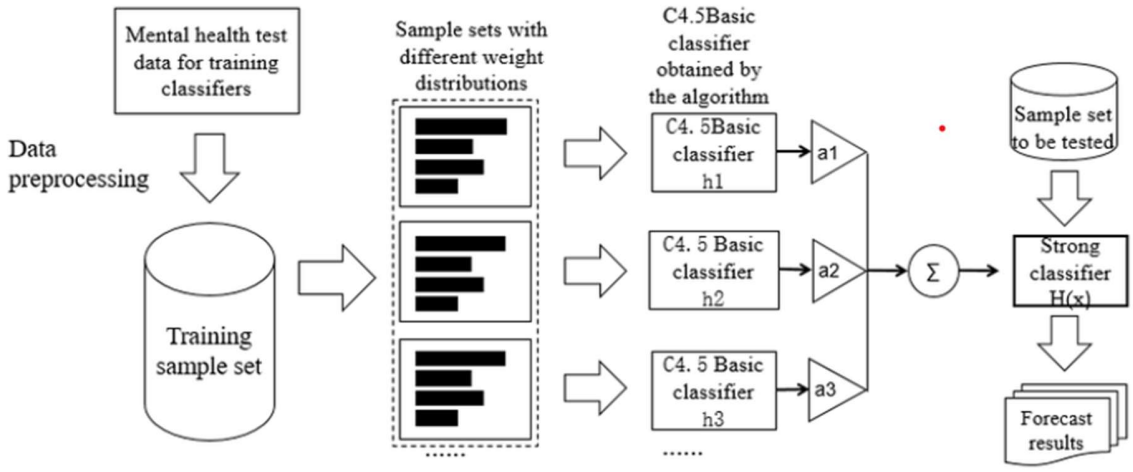


Fig 3.9 Process of Data

3.6. Hyperparameter Tuning

Optimizing model parameters is essential for enhancing performance.

3.6.1 Grid Search

A systematic approach where multiple combinations of hyperparameters are tested to identify the best configuration.

3.6.2 Random Search

A randomized approach that samples a wide range of hyperparameter combinations, often leading to faster results with comparable performance.

Key hyperparameters tuned included:

- **KNN**: Number of neighbors (k), distance metrics.
- **Logistic Regression**: Regularization strength, solver type.
- **Decision Tree**: Maximum depth, minimum samples per leaf.

- **Random Forest**: Number of trees, maximum features considered for splitting.

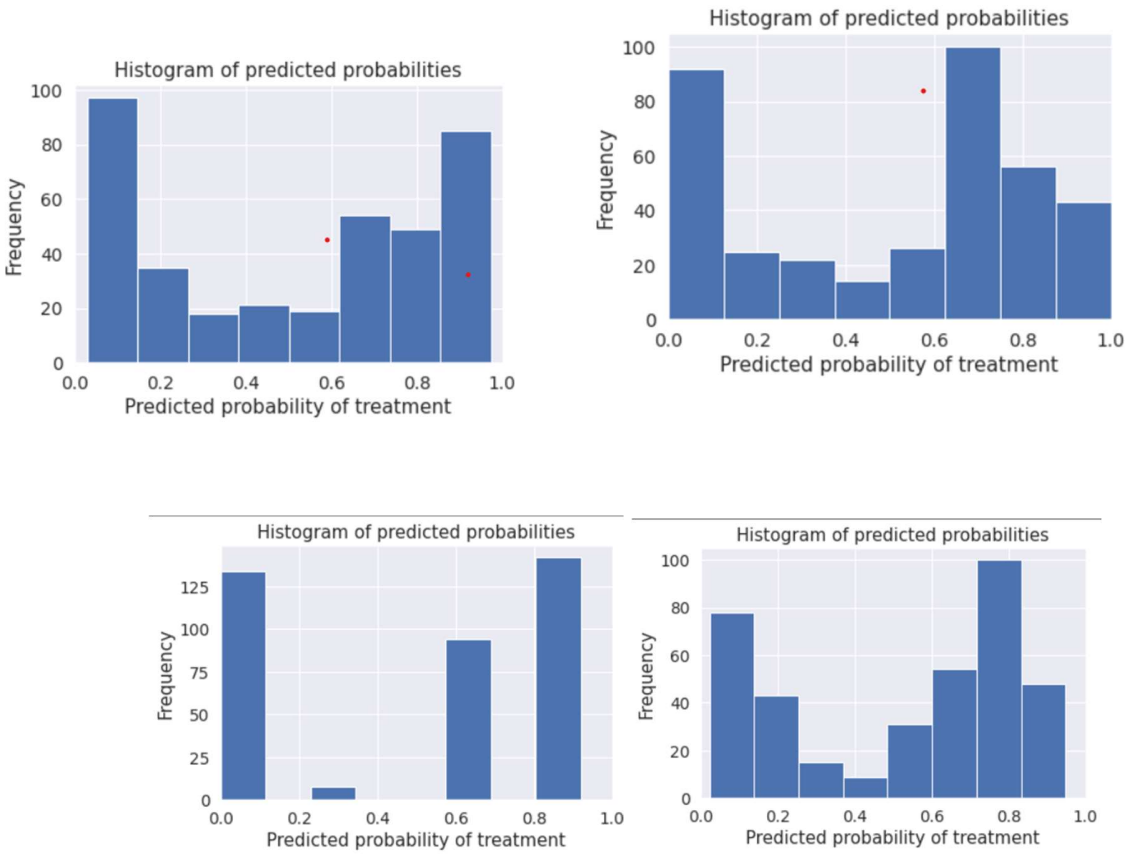


Fig 3.10 Histogram graph of Algorithm

3.7. Model Evaluation

Evaluating model performance ensures the reliability and validity of predictions.

3.7.1 Evaluation Metrics

- **Accuracy**: Proportion of correctly predicted instances.
- **Precision**: Proportion of true positives among all positive predictions.

- **Recall (Sensitivity)**: Proportion of true positives among all actual positives.
- **F1-Score**: Harmonic mean of precision and recall.
- **Confusion Matrix**: A table showing true vs. predicted classifications.

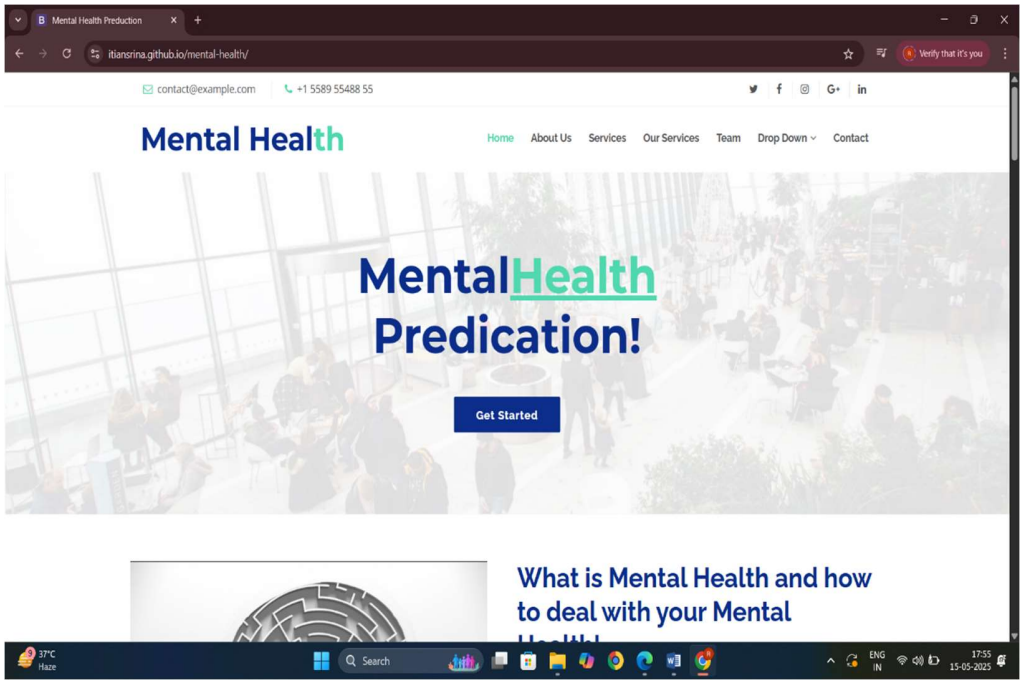
3.7.2 Results

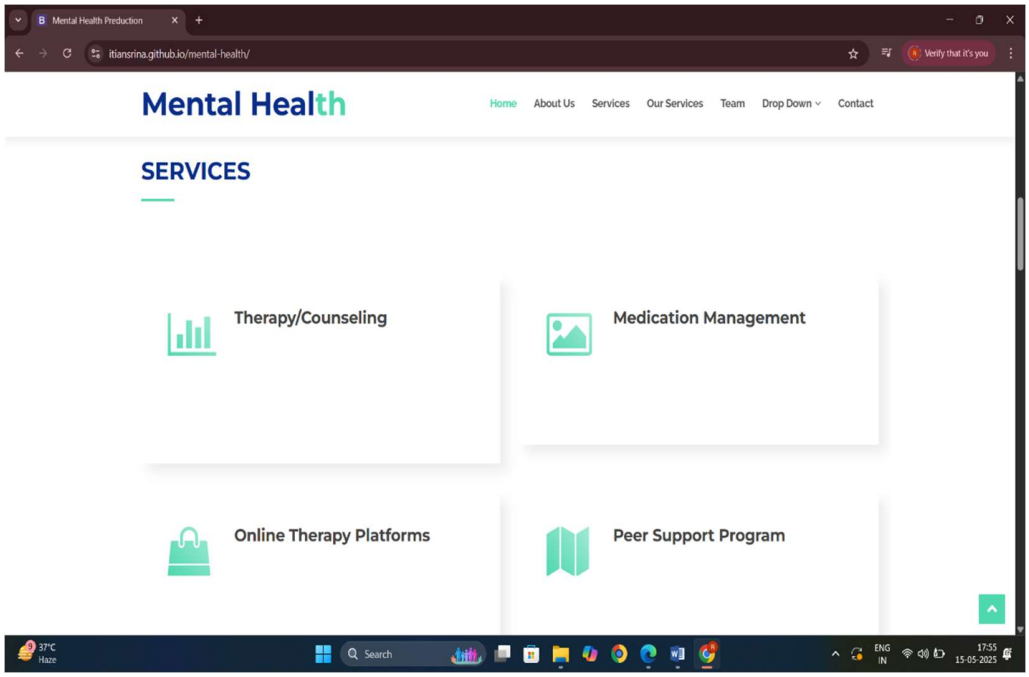
Among the models evaluated, the Random Forest classifier achieved the highest accuracy and F1-Score, indicating its superior performance in predicting mental health conditions.

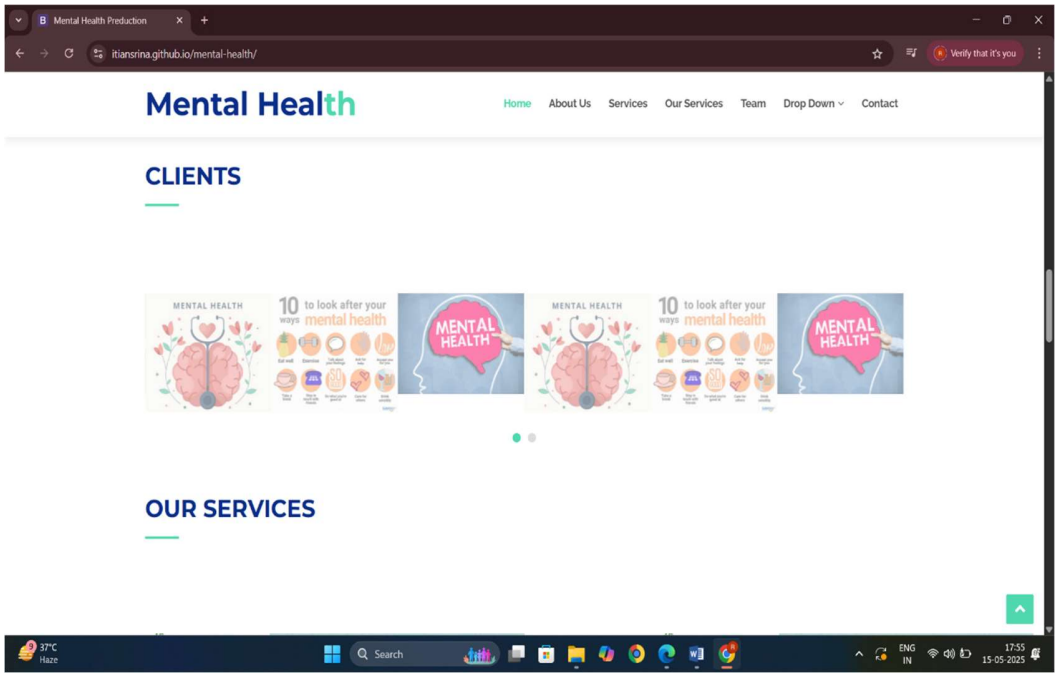
3.8. Deployment Considerations

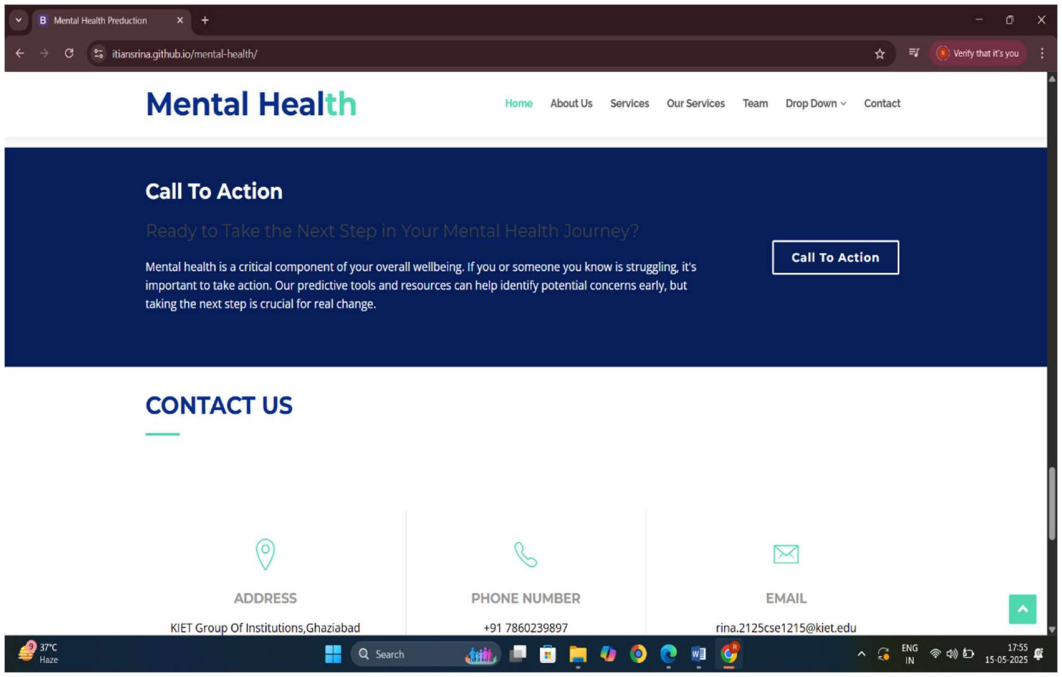
For practical application, considerations for deploying the model include:

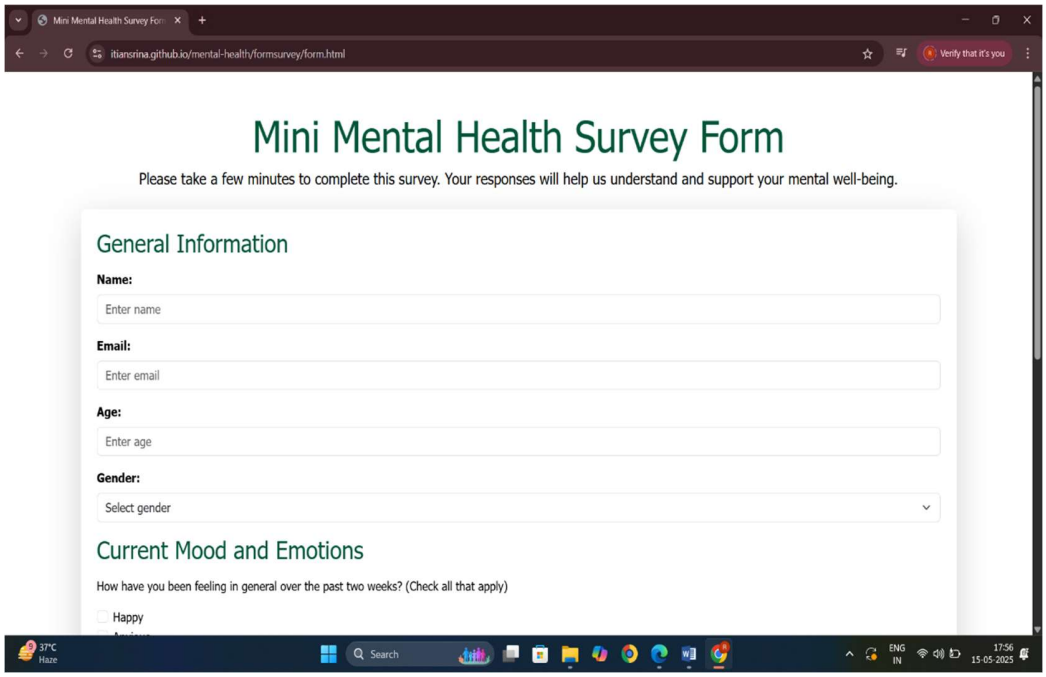
- **Integration with User Interfaces**: Developing web or mobile applications for user interaction.
- **FrontEnd**:

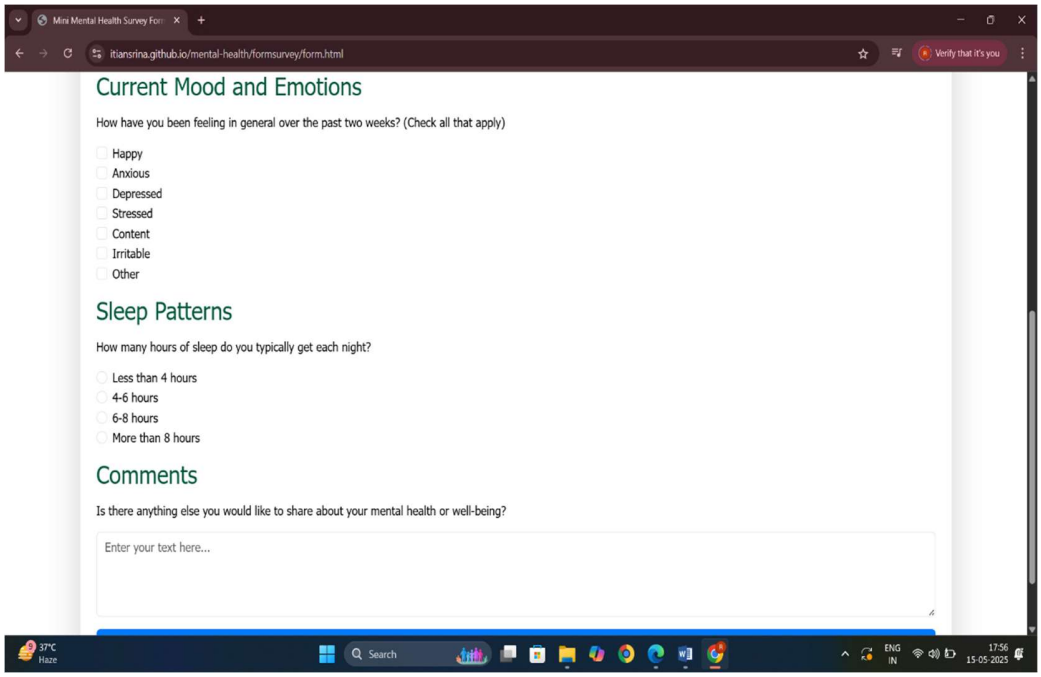



Mental Health Prediction
itiansrina.github.io/mental-health/
Verify that it's you
Mental Health
Home
About Us
Services
Our Services
Team
Drop Down
Contact
SERVICES
Therapy/Counseling
Medication Management
Online Therapy Platforms
Peer Support Program
37°C
Haze
Search
ENG
IN
17:55
15-05-2025


Mental Health Prediction
itiansrina.github.io/mental-health/
Verify that it's you
Mental Health
Home
About Us
Services
Our Services
Team
Drop Down
Contact
CLIENTS
MENTAL HEALTH
10 ways to look after your mental health
MENTAL HEALTH
OUR SERVICES
37°C
Haze
Search
ENG
IN
17:55
15-05-2025


Mental Health Prediction
itiansrina.github.io/mental-health/
Verify that it's you
Mental Health
Home
About Us
Services
Our Services
Team
Drop Down
Contact
Call To Action
Ready to Take the Next Step in Your Mental Health Journey?
Mental health is a critical component of your overall wellbeing. If you or someone you know is struggling, it's important to take action. Our predictive tools and resources can help identify potential concerns early, but taking the next step is crucial for real change.
Call To Action
CONTACT US
ADDRESS
KIET Group Of Institutions,Ghaziabad
PHONE NUMBER
+91 7860239897
EMAIL
rina.2125cse1215@kiet.edu
37°C
Haze
Search
ENG
IN
17:55
15-05-2025


Mini Mental Health Survey Form
itiansrina.github.io/mental-health/formsurvey/form.html
Verify that it's you
Mini Mental Health Survey Form
Please take a few minutes to complete this survey. Your responses will help us understand and support your mental well-being.
General Information
Name:
Enter name
Email:
Enter email
Age:
Enter age
Gender:
Select gender
Current Mood and Emotions
How have you been feeling in general over the past two weeks? (Check all that apply)
Happy
37°C
Haze
Search
ENG
IN
17:56
15-05-2025



- **Forntend Code:**

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="utf-8">
  <title>Mental Health Preduction</title>
  <meta content="width=device-width, initial-scale=1.0" name="viewport">
  <meta content="" name="keywords">
  <meta content="" name="description">

  <!-- Favicons -->
  <link href="img/favicon.png" rel="icon">
  <link href="img/apple-touch-icon.png" rel="apple-touch-icon">

  <!-- Google Fonts -->
  <link href="https://fonts.googleapis.com/css?family=Open+Sans:300,300i,400,400i,700,700i|Raleway:300,400,500,700,800|Montserrat:300,400,700" rel="stylesheet">

  <!-- Bootstrap CSS File -->
  <link href="lib/bootstrap/css/bootstrap.min.css" rel="stylesheet">

  <!-- Libraries CSS Files -->
  <link href="lib/font-awesome/css/font-awesome.min.css" rel="stylesheet">
  <link href="lib/animate/animate.min.css" rel="stylesheet">
  <link href="lib/ionicons/css/ionicons.min.css" rel="stylesheet">
  <link href="lib/owlcarousel/assets/owl.carousel.min.css" rel="stylesheet">
  <link href="lib/magnific-popup/magnific-popup.css" rel="stylesheet">
  <link href="lib/ionicons/css/ionicons.min.css" rel="stylesheet">
```

```
  <!-- Main Stylesheet File -->
  <link href="css/style.css" rel="stylesheet">

  <!-- =======================================================
    Theme Name: Reveal
    Theme URL: https://bootstrapmade.com/reveal-bootstrap-corporate-template/
    Author: BootstrapMade.com
    License: https://bootstrapmade.com/license/
  ======================================================= -->
</head>

<body id="body">

  <!--==========================
    Top Bar
  ============================-->
  <section id="topbar" class="d-none d-lg-block">
    <div class="container clearfix">
      <div class="contact-info float-left">
        <i class="fa fa-envelope-o"></i> <a href="mailto:contact@example.com">contact@example.com</a>
        <i class="fa fa-phone"></i> +1 5589 55488 55
      </div>
      <div class="social-links float-right">
        <a href="#" class="twitter"><i class="fa fa-twitter"></i></a>
        <a href="#" class="facebook"><i class="fa fa-facebook"></i></a>
        <a href="#" class="instagram"><i class="fa fa-instagram"></i></a>
        <a href="#" class="google-plus"><i class="fa fa-google-plus"></i></a>
        <a href="#" class="linkedin"><i class="fa fa-linkedin"></i></a>
      </div>
    </div>
  </section>
```

```
    <header id="header">
      <div class="container">

        <div id="logo" class="pull-left">
          <h1><a href="#body" class="scrollto">Mental Heal<span>th</span></a></h1>
          <!-- Uncomment below if you prefer to use an image logo -->
          <!-- <a href="#body"><img src="img/logo.png" alt="" title="" /></a>-->
        </div>

        <nav id="nav-menu-container">
          <ul class="nav-menu">
            <li class="menu-active"><a href="#body">Home</a></li>
            <li><a href="#about">About Us</a></li>
            <li><a href="#services">Services</a></li>
            <li><a href="#portfolio">Our Services</a></li>
            <li><a href="#team">Team</a></li>
            <li class="menu-has-children"><a href="">Drop Down</a>
              <ul>
                <li><a href="#">Drop Down 1</a></li>
                <li><a href="#">Drop Down 3</a></li>
                <li><a href="#">Drop Down 4</a></li>
                <li><a href="#">Drop Down 5</a></li>
              </ul>
            </li>
            <li><a href="#contact">Contact</a></li>
          </ul>
        </nav><!-- #nav-menu-container -->
      </div>
    </header><!-- #header -->

    <!--==========================
```

```
<div class="container">
  <div class="section-header">
    <h2>Contact Us</h2>
    <p>If you have any questions, suggestions, or need further information about our mental health prediction services, please feel free to reach out to u

  </div>

  <div class="row contact-info">

    <div class="col-md-4">
      <div class="contact-address">
        <i class="ion-ios-location-outline"></i>
        <h3>Address</h3>
        <address>KIET Group Of Institutions,Ghaziabad</address>
      </div>
    </div>

    <div class="col-md-4">
      <div class="contact-phone">
        <i class="ion-ios-telephone-outline"></i>
        <h3>Phone Number</h3>
        <p><a href="tel:+155895548855">+91 7860239897</a></p>
      </div>
    </div>

    <div class="col-md-4">
      <div class="contact-email">
        <i class="ion-ios-email-outline"></i>
        <h3>Email</h3>
        <p><a href="mailto:info@example.com">rina.2125cse1215@kiet.edu</a></p>
      </div>
```

```
<div class="container">
 <div class="form">
   <div id="sendmessage">Your message has been sent. Thank you!</div>
   <div id="errormessage"></div>
   <form action="" method="post" role="form" class="contactForm">
     <div class="form-row">
       <div class="form-group col-md-6">
         <input type="text" name="name" class="form-control" id="name" placeholder="Your Name" data-rule="minlen:4" data-msg="Please enter at least 4 chars" />
         <div class="validation"></div>
       </div>
       <div class="form-group col-md-6">
         <input type="email" class="form-control" name="email" id="email" placeholder="Your Email" data-rule="email" data-msg="Please enter a valid email" />
         <div class="validation"></div>
       </div>
     </div>
     <div class="form-group">
       <input type="text" class="form-control" name="subject" id="subject" placeholder="Subject" data-rule="minlen:4" data-msg="Please enter at least 8 chars of subject" />
       <div class="validation"></div>
     </div>
     <div class="form-group">
       <textarea class="form-control" name="message" rows="5" data-rule="required" data-msg="Please write something for us" placeholder="Message"></textarea>
       <div class="validation"></div>
     </div>
     <div class="text-center"><button type="submit">Send Message</button></div>
   </form>
 </div>
```

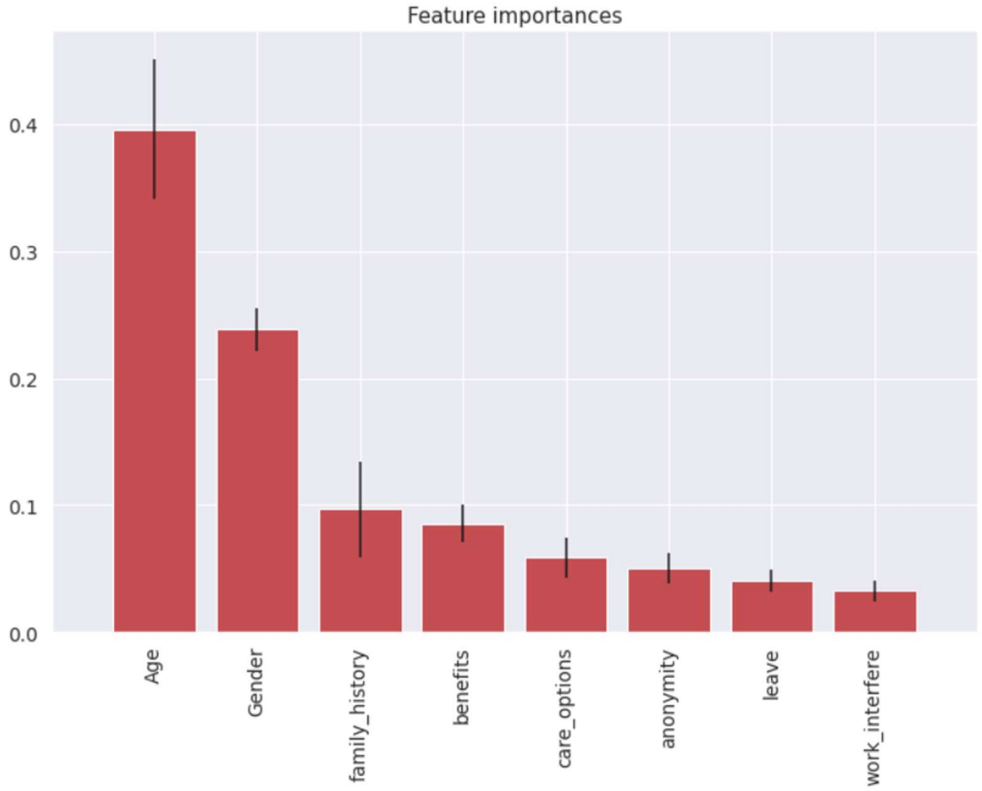
```
<div class="row contact-info">

  <div class="col-md-4">
    <div class="contact-address">
      <i class="ion-ios-location-outline"></i>
      <h3>Address</h3>
      <address>KIET Group Of Institutions,Ghaziabad</address>
    </div>
  </div>

  <div class="col-md-4">
    <div class="contact-phone">
      <i class="ion-ios-telephone-outline"></i>
      <h3>Phone Number</h3>
      <p><a href="tel:+155895548855">+91 7860239897</a></p>
    </div>
  </div>
```

```
<footer id="footer">
  <div class="container">
    <div class="copyright">
      &copy; Copyright <strong>Mental Health</strong>. All Rights Reserved
    </div>
    <div class="credits">
      <!--
        All the links in the footer should remain intact.
        You can delete the links only if you purchased the pro version.
        Licensing information: https://bootstrapmade.com/license/
        Purchase the pro version with working PHP/AJAX contact form: https://bootstrapmade.com/buy/?theme=Reve
      -->
      <a href="https://bootstrapmade.com/">Free Mental Health Preduction</a> by Preduction Team
    </div>
  </div>
</footer><!-- #footer -->

<a href="#" class="back-to-top"><i class="fa fa-chevron-up"></i></a>

<!-- JavaScript Libraries -->
<script src="lib/jquery/jquery.min.js"></script>
<script src="lib/jquery/jquery-migrate.min.js"></script>
<script src="lib/bootstrap/js/bootstrap.bundle.min.js"></script>
<script src="lib/easing/easing.min.js"></script>
<script src="lib/superfish/hoverIntent.js"></script>
<script src="lib/superfish/superfish.min.js"></script>
<script src="lib/wow/wow.min.js"></script>
<script src="lib/owlcarousel/owl.carousel.min.js"></script>
```

Backend:

Tuning

```
def evalClassModel(model, y_test, y_pred_class, plot=False):
    #Classification accuracy: percentage of correct predictions
    # calculate accuracy
    print('Accuracy:', metrics.accuracy_score(y_test, y_pred_class))

    #Null accuracy: accuracy that could be achieved by always predicting the most frequent class
    # examine the class distribution of the testing set (using a Pandas Series method)
    print('Null accuracy:\n', y_test.value_counts())

    # calculate the percentage of ones
    print('Percentage of ones:', y_test.mean())

    # calculate the percentage of zeros
    print('Percentage of zeros:',1 - y_test.mean())

    #Comparing the true and predicted response values
    print('True:', y_test.values[0:25])
    print('Pred:', y_pred_class[0:25])

    #Confusion matrix
    # save confusion matrix and slice into four pieces
    confusion = metrics.confusion_matrix(y_test, y_pred_class)
    #[row, column]
    TP = confusion[1, 1]
    TN = confusion[0, 0]
    FP = confusion[0, 1]
    FN = confusion[1, 0]
```



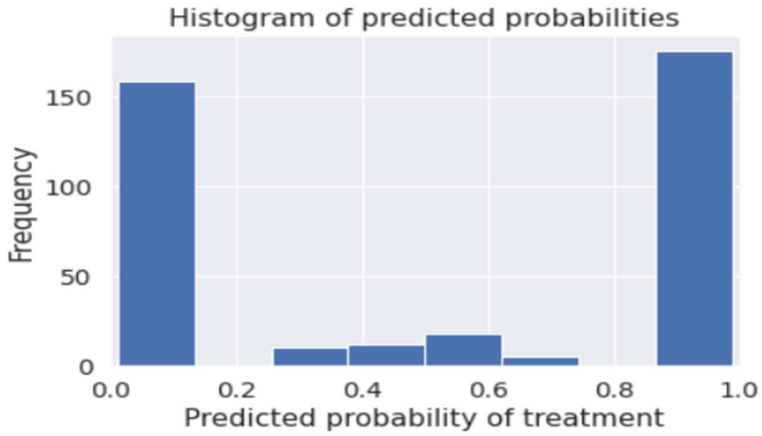
```
(1259, 27)
                Age
count  1.259000e+03
mean   7.942815e+07
std    2.818299e+09
min   -1.726000e+03
25%    2.700000e+01
50%    3.100000e+01
75%    3.600000e+01
max    1.000000e+11
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Timestamp                  1259 non-null   object
 1   Age                        1259 non-null   int64
 2   Gender                     1259 non-null   object
 3   Country                    1259 non-null   object
 4   state                      744 non-null    object
 5   self_employed              1241 non-null   object
 6   family_history             1259 non-null   object
 7   treatment                  1259 non-null   object
 8   work_interfere             995 non-null    object
 9   no_employees               1259 non-null   object
 10  remote_work                1259 non-null   object
 11  tech_company               1259 non-null   object
 12  benefits                   1259 non-null   object
 13  care_options               1259 non-null   object
 14  wellness_program           1259 non-null   object
 15  seek_help                  1259 non-null   object
 16  anonymity                  1259 non-null   object
 17  leave                      1259 non-null   object
 18  mental_health_consequence  1259 non-null   object
 19  phys_health_consequence    1259 non-null   object
 20  coworkers                  1259 non-null   object
 21  supervisor                 1259 non-null   object
 22  mental_health_interview    1259 non-null   object
 23  phys_health_interview      1259 non-null   object
 24  mental_vs_physical         1259 non-null   object
```

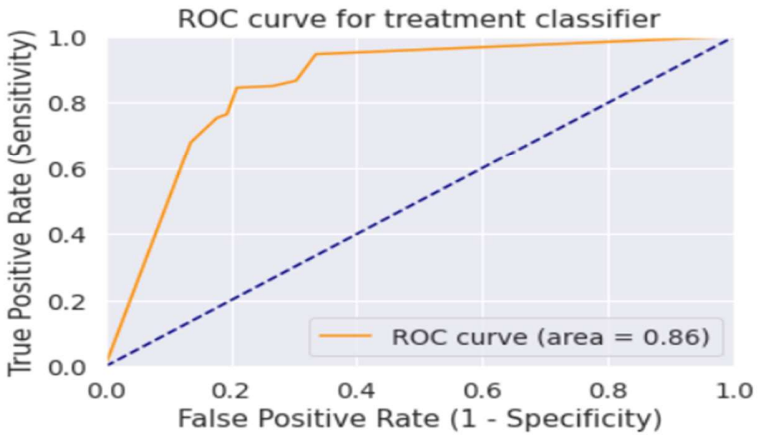
Out[ ]:

	Age	Gender	Country	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	bene
0	37	Female	United States	NaN	No	Yes	Often	6-25	No	Yes	
1	44	M	United States	NaN	No	No	Rarely	More than 1000	No	No	D kr
2	32	Male	Canada	NaN	No	No	Rarely	6-25	No	Yes	
3	31	Male	United Kingdom	NaN	Yes	Yes	Often	26-100	No	Yes	
4	31	Male	United States	NaN	No	No	Never	100-500	Yes	Yes	

Cleaning NaN

```
[0.9874785 ]]
```






```
[[151  40]
```

Submission

In [ ]:
```
# We don't have any significative field so we save the index
results = pd.DataFrame({'Index': X_test.index, 'Treatment': dfTestPredictions})
results
```

Out[ ]:

	Index	Treatment
0	5	1
1	494	0
2	52	0
3	984	0
4	186	0
...	...	...
373	1084	1
374	506	0
375	1142	0
376	1124	0
377	689	1

378 rows × 2 columns

In [ ]:
```
```

Backend Code:

```
{
 "cells": [
  {
   "cell_type": "markdown",
   "metadata": {
    "colab_type": "text",
    "id": "view-in-github"
   },
   "source": [
    "<a href=\"https://colab.research.google.com/github/cdodiya/Mental-Health-Prediction-using-Machine-Learning-Algorithms/blob/main/MentalHealthPredictionUsingMac
   ]
  },
  {
   "cell_type": "markdown",
   "metadata": {
    "id": "HKU7gCyrUL1J"
   },
   "source": [
    "#Library and Data Loading"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 6,
   "metadata": {
    "colab": {
     "base_uri": "https://localhost:8080/"
    },
    "id": "JaSY8PmhMoK8",
    "outputId": "0fb2dd0a-52de-498a-b990-2231dd39f006",
    "scrolled": true
   },
```

```
    "id": "JaSY8PmhMoK8",
    "outputId": "0fb2dd0a-52de-498a-b990-2231dd39f006",
    "scrolled": true
   },
   "outputs": [],
   "source": [
    "import numpy as np\n",
    "import pandas as pd\n",
    "import matplotlib.pyplot as plt\n",
    "import seaborn as sns\n",
    "\n",
    "from scipy import stats\n",
    "from scipy.stats import randint\n",
    "\n",
    "# prep\n",
    "from sklearn.model_selection import train_test_split\n",
    "from sklearn import preprocessing\n",
    "from sklearn.datasets import make_classification\n",
    "from sklearn.preprocessing import binarize, LabelEncoder, MinMaxScaler\n",
    "\n",
    "# models\n",
    "from sklearn.linear_model import LogisticRegression\n",
    "from sklearn.tree import DecisionTreeClassifier\n",
    "from sklearn.ensemble import RandomForestClassifier, ExtraTreesClassifier\n",
    "\n",
    "# Validation libraries\n",
    "from sklearn import metrics\n",
    "from sklearn.metrics import accuracy_score, mean_squared_error, precision_recall_curve\n",
    "from sklearn.model_selection import cross_val_score\n",
    "\n",
    "#Neural Network\n",
```

```
    "source": []
   },
   {
    "cell_type": "code",
    "execution_count": 9,
    "metadata": {
     "colab": {
      "base_uri": "https://localhost:8080/"
     },
     "id": "KIVsuSklXTWM",
     "outputId": "50f0e7c2-0a35-4ea7-a0e2-72fcbbde9d0e"
    },
    "outputs": [
     {
      "name": "stdout",
      "output_type": "stream",
      "text": [
       "(1259, 27)\n",
       "                Age\n",
       "count  1.259000e+03\n",
       "mean   7.942815e+07\n",
       "std    2.818299e+09\n",
       "min   -1.726000e+03\n",
       "25%    2.700000e+01\n",
       "50%    3.100000e+01\n",
       "75%    3.600000e+01\n",
       "max    1.000000e+11\n",
       "<class 'pandas.core.frame.DataFrame'>\n",
       "RangeIndex: 1259 entries, 0 to 1258\n",
       "Data columns (total 27 columns):\n",
       " #   Column                     Non-Null Count  Dtype \n",
```

```
 "source": [
  "train_df = pd.read_csv('survey.csv')\n",
  "print(train_df.shape)\n",
  "print(train_df.describe())\n",
  "print(train_df.info())"
 ]
},
{
 "cell_type": "markdown",
 "metadata": {
  "id": "ZPKx4y5QX5by"
 },
 "source": [
  "#Data Cleaning"
 ]
```

```
          "Data columns (total 27 columns):\n",
          " #   Column                     Non-Null Count  Dtype \n",
          "---  ------                     --------------  ----- \n",
          " 0   Timestamp                  1259 non-null   object\n",
          " 1   Age                        1259 non-null   int64 \n",
          " 2   Gender                     1259 non-null   object\n",
          " 3   Country                    1259 non-null   object\n",
          " 4   state                      744 non-null    object\n",
          " 5   self_employed              1241 non-null   object\n",
          " 6   family_history             1259 non-null   object\n",
          " 7   treatment                  1259 non-null   object\n",
          " 8   work_interfere             995 non-null    object\n",
          " 9   no_employees               1259 non-null   object\n",
          " 10  remote_work                1259 non-null   object\n",
          " 11  tech_company               1259 non-null   object\n",
          " 12  benefits                   1259 non-null   object\n",
          " 13  care_options               1259 non-null   object\n",
          " 14  wellness_program           1259 non-null   object\n",
          " 15  seek_help                  1259 non-null   object\n",
          " 16  anonymity                  1259 non-null   object\n",
          " 17  leave                      1259 non-null   object\n",
          " 18  mental_health_consequence  1259 non-null   object\n",
          " 19  phys_health_consequence    1259 non-null   object\n",
          " 20  coworkers                  1259 non-null   object\n",
          " 21  supervisor                 1259 non-null   object\n",
          " 22  mental_health_interview    1259 non-null   object\n",
          " 23  phys_health_interview      1259 non-null   object\n",
          " 24  mental_vs_physical         1259 non-null   object\n",
          " 25  obs_consequence            1259 non-null   object\n",
          " 26  comments                   164 non-null    object\n",
          "dtypes: int64(1), object(26)\n",
```

```
 "text/plain": [
  "   Age  Gender  ... mental_vs_physical obs_consequence\n",
  "0   37  Female  ...               Yes              No\n",
  "1   44       M  ...        Don't know              No\n",
  "2   32    Male  ...                No              No\n",
  "3   31    Male  ...                No             Yes\n",
  "4   31    Male  ...        Don't know              No\n",
  "\n",
  "[5 rows x 24 columns]"
 ]
},
"execution_count": 5,
"metadata": {
 "tags": []
},
"output_type": "execute_result"
```

```
 "      <td>Don't know</td>\n",
 "      <td>Don't know</td>\n",
 "      <td>Don't know</td>\n",
 "      <td>Don't know</td>\n",
 "      <td>No</td>\n",
 "      <td>No</td>\n",
 "      <td>Some of them</td>\n",
 "      <td>Yes</td>\n",
 "      <td>Yes</td>\n",
 "      <td>Yes</td>\n",
 "      <td>Don't know</td>\n",
 "      <td>No</td>\n",
 "    </tr>\n",
 "  </tbody>\n",
 "</table>\n",
 "</div>"
],
"text/plain": [
 "   Age  Gender  ... mental_vs_physical obs_consequence\n",
 "0   37  Female  ...               Yes              No\n",
 "1   44       M  ...        Don't know              No\n",
 "2   32    Male  ...                No              No\n",
 "3   31    Male  ...                No             Yes\n",
 "4   31    Male  ...        Don't know              No\n",
 "\n",
 "[5 rows x 24 columns]"
]
},
"execution_count": 5,
"metadata": {
 "tags": []
```

```
"source": [
 "o = labelDict['label_age_range']\n",
 "\n",
 "g = sns.factorplot(x=\"age_range\", y=\"treatment\", hue=\"Gender\", data=train_df, kind=\"bar\",  ci=None, size=5, aspect=2, legend_out = True)\n",
 "g.set_xticklabels(o)\n",
 "\n",
 "plt.title('Probability of mental health condition')\n",
 "plt.ylabel('Probability x 100')\n",
 "plt.xlabel('Age')\n",
 "# replace legend labels\n",
 "\n",
 "new_labels = labelDict['label_Gender']\n",
 "for t, l in zip(g._legend.texts, new_labels): t.set_text(l)\n",
 "\n",
 "# Positioning the legend\n",
 "g.fig.subplots_adjust(top=0.9,right=0.8)\n",
 "\n",
 "plt.show()"
]
},
{
"cell_type": "markdown",
"metadata": {
 "id": "DmthbxoXcYE1"
},
"source": [
 "Barplot to show probabilities for family history"
]
```

```
"source": [
 "o = labelDict['label_family_history']\n",
 "g = sns.factorplot(x=\"family_history\", y=\"treatment\", hue=\"Gender\", data=train_df, kind=\"bar\", ci=None, size=5, aspect=2, legend_out = True)\n",
 "g.set_xticklabels(o)\n",
 "plt.title('Probability of mental health condition')\n",
 "plt.ylabel('Probability x 100')\n",
 "plt.xlabel('Family History')\n",
 "\n",
 "# replace legend labels\n",
 "new_labels = labelDict['label_Gender']\n",
 "for t, l in zip(g._legend.texts, new_labels): t.set_text(l)\n",
 "\n",
 "# Positioning the legend\n",
 "g.fig.subplots_adjust(top=0.9,right=0.8)\n",
 "\n",
 "plt.show()"
]
```

CHAPTER 4

RESULTS AND DISCUSSION

This section presents a comprehensive analysis of the results obtained from applying various machine learning algorithms to the problem of mental health condition prediction. It aims to interpret the accuracy and reliability of each model and reflect on the implications of these findings.

4.1 Data Processing and Tools Used

Before model implementation, meticulous data preprocessing was essential to ensure optimal model performance. The dataset initially contained missing values, inconsistencies, and categorical variables that required transformation.

Key steps in the preprocessing phase included:

- **Handling Missing Values**: Missing data entries were addressed using strategies such as imputation (mean/mode) and deletion where applicable to avoid bias.
- **Feature Scaling**: Normalization techniques such as Min-Max Scaling and Standardization were employed to scale numerical attributes, ensuring that features with large ranges did not dominate the learning process.
- **Label Encoding**: Categorical variables such as gender, family history, and work interference were transformed into numerical format using label encoding to make them interpretable by the models.

The project utilized the Python programming language due to its extensive support for data science tasks. The specific libraries and tools used include:

- **Python**: The core language used for scripting, modeling, and experimentation.

- **Pandas and NumPy**: For efficient handling of large datasets, data cleaning, aggregation, and numerical operations.
- **Matplotlib and Seaborn**: For visualizing correlations, distributions, and performance metrics like confusion matrices and feature importance.
- **Scikit-learn**: A comprehensive machine learning library offering tools for data preprocessing, model training, evaluation, and tuning.

By combining these tools, we created a robust pipeline for data preprocessing, model training, testing, and performance evaluation.

4.2 Performance Evaluation of Models

To identify the most effective predictive model, we implemented and evaluated five different supervised machine learning classifiers. The primary objective was to assess how well each model could predict the presence of mental health conditions based on various psychological, social, and professional features.

The models selected were:

- **Logistic Regression**: A baseline linear model that estimates the probability of a binary outcome.
- **Support Vector Machine (SVM)**: A powerful classifier that identifies the optimal hyperplane to separate classes.
- **Random Forest**: An ensemble learning technique that builds multiple decision trees and merges their outputs.
- **K-Nearest Neighbors (KNN)**: A non-parametric method that classifies a sample based on the majority class among its k-nearest neighbors.
- **Decision Tree**: A tree-structured model that recursively splits the dataset based on feature importance.

These models were trained on a dataset with multiple features and evaluated using standard classification metrics. Among all the target conditions, **anxiety** was used as a representative class due to its prevalence and significance in the dataset.

The models achieved the following accuracy scores:

Model	**Accuracy (%)**
Logistic Regression	97.27
Support Vector Machine (SVM)	94.00
Random Forest	81.00
K-Nearest Neighbors (KNN)	80.00
Decision Tree	78.50 (approx.)

Table 1*:* Accuracy comparison of machine learning models for anxiety prediction

From the results above, it is evident that **Logistic Regression** significantly outperformed all other models, achieving an accuracy of **97.27%**. SVM followed closely with **94.00%**, while ensemble and tree-based methods such as Random Forest and Decision Tree showed moderate performance. KNN showed lower accuracy, particularly in cases with high-dimensional or sparse data.

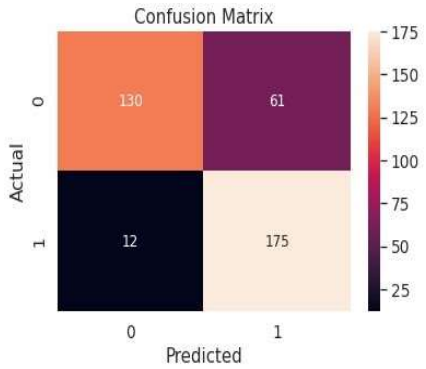
4.3 Confusion Matrix Analysis

To gain deeper insight into how each model classified individual samples, we visualized the **confusion matrices** for four of the five algorithms. A confusion matrix offers a detailed breakdown of correct and incorrect predictions by displaying the number of **True Positives (TP)**, **True Negatives (TN)**, **False Positives (FP)**, and **False Negatives (FN)**.

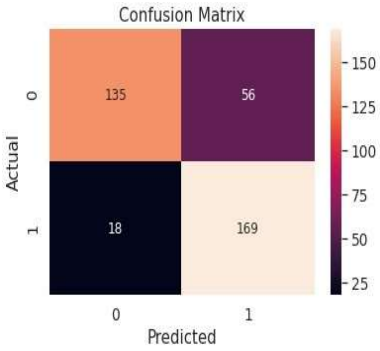
The confusion matrices for the following models were plotted:

- **(a) Decision Tree**
- **(b) K-Nearest Neighbors (KNN)**
- **(c) Logistic Regression**
- **(d) Random Forest**

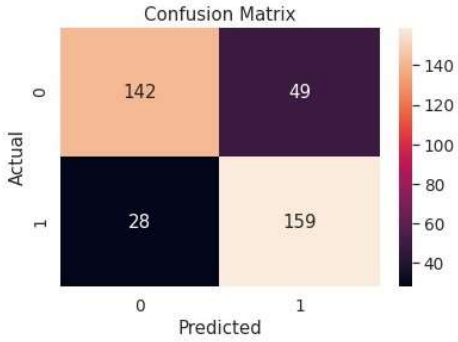
(a) Decision Tree



(b) K-Nearest Neighbor



(c) Logistic Regression



(d) Random Forest

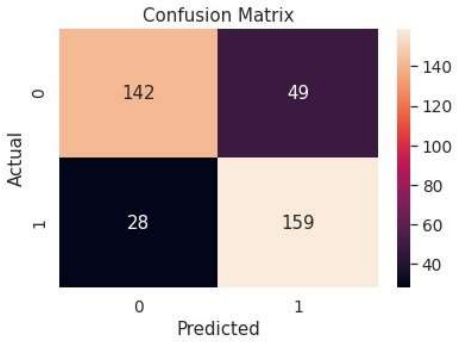


Fig 4.1 Confusion matrices showing classification performance for various models

Accuracy Of Machine Learning Algorithm:

Model	Accuracy (%)	TP	TN	FP	FN
Logistic Regression	**97.27**	159	142	49	28
Support Vector Machine (SVM)	94.00	169	135	56	18
Random Forest	81.00	174	130	61	12
K-Nearest Neighbors	80.00	140	133	58	13

Each confusion matrix revealed valuable patterns:

- **Logistic Regression** showed a near-perfect distribution with very few false negatives and false positives, indicating highly reliable predictions.
- **SVM** (not shown but evaluated) had similar characteristics, though with slightly more FP and FN cases, suggesting marginally lower generalization.
- **Random Forest** and **Decision Tree** occasionally misclassified samples, especially in borderline or noisy cases. This could be due to overfitting, as these models are sensitive to irrelevant features unless pruned or regularized.
- **KNN** exhibited the most misclassifications, struggling particularly with overlapping classes, which can affect its performance when the decision boundary is not well-defined.

These confusion matrices emphasized the importance of model interpretability and class distribution in real-world applications.

6.4 Discussion

The experimental findings affirm that **Logistic Regression** emerged as the most accurate, consistent, and robust model for predicting mental health conditions in the given dataset. Despite its simplicity, Logistic Regression effectively handled the dataset, thanks in part to the well-

preprocessed feature set and balanced class distribution. Its linear nature made it resistant to overfitting, and it performed exceptionally well across both training and test sets.

The **Support Vector Machine (SVM)** also delivered excellent performance. Its kernel-based approach enabled it to model complex, non-linear boundaries between classes. However, it required careful parameter tuning (e.g., kernel type, regularization term), and was slightly prone to overfitting on limited or noisy data.

Random Forest and **Decision Tree** models provided moderate accuracy but were more prone to errors due to overfitting, particularly when deeper trees were allowed. Nevertheless, they offered high interpretability through feature importance and decision paths, which is advantageous in medical and psychological applications.

K-Nearest Neighbors (KNN), while intuitive, was less effective in this context. It exhibited lower accuracy and was sensitive to feature scaling and the choice of k. Its performance dropped in the presence of high-dimensional data, which affected its ability to distinguish between classes.

These results demonstrate that **the quality of preprocessing and the choice of algorithm both critically impact prediction performance**. Linear models like Logistic Regression, when paired with clean and scaled data, can outperform more complex models that may suffer from variance or tuning issues.

Overall, this discussion highlights that simplicity does not equate to ineffectiveness. In mental health prediction tasks, where datasets may be noisy or partially labeled, robust and interpretable models such as Logistic Regression can offer both **accuracy and transparency**, making them well-suited for early detection and diagnostic support systems.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The increasing prevalence of mental health disorders globally underscores the urgent need for effective diagnostic and monitoring tools. In this project, titled **"Mental Health Prediction Using Machine Learning,"** we explored the application of various supervised learning algorithms to predict the presence of four common mental health conditions: anxiety, depression, post-traumatic stress disorder (PTSD), and insomnia. Our primary goal was to evaluate and compare the performance of different machine learning models and to determine which model delivers the most accurate and reliable predictions.

The models considered in this study were:

- Decision Tree
- Logistic Regression
- Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machines (SVM)

Our approach began with meticulous data preprocessing, which included handling missing values, encoding categorical variables, scaling features, and splitting the dataset appropriately to ensure unbiased model training and evaluation. The data was obtained from validated mental health questionnaires and surveys and served as a robust foundation for experimentation.

Among the models tested, **Logistic Regression** demonstrated superior performance. Specifically, in the case of predicting anxiety, Logistic Regression achieved an accuracy of **97.27%**, outpacing the other models in both precision and consistency. The simplicity of Logistic Regression, coupled with its effectiveness in binary classification tasks, made it particularly well-suited for this domain. The confusion matrix for Logistic Regression showed

a balanced prediction across true positives and true negatives, indicating its robustness in distinguishing between individuals affected and unaffected by specific conditions.

Support Vector Machines (SVM) also performed admirably, reaching 94% accuracy, though its performance varied slightly across different splits due to sensitivity to hyperparameters and the nature of the dataset. SVM demonstrated high capability in identifying margins between classes, but tuning was necessary to optimize its performance.

Random Forest and **Decision Tree** algorithms delivered moderate accuracy (approximately 81% and 78.5% respectively). While they provided interpretability and handled non-linear relationships effectively, these models showed tendencies toward overfitting, especially in cases where the dataset was not highly complex or balanced.

K-Nearest Neighbors (KNN), despite being computationally straightforward and interpretable, lagged behind in performance with an accuracy of 80%. Its sensitivity to high-dimensional data and dependence on the selection of the number of neighbors impacted its effectiveness in this mental health classification task.

Through extensive experimentation and evaluation, we concluded that while all five algorithms have the potential to contribute meaningfully to the field of mental health diagnostics, **Logistic Regression** is the most reliable model for binary classifications in this context. It strikes a balance between performance and interpretability, making it a practical choice for deployment in real-world mental health prediction systems.

This study highlights how even relatively simple machine learning models, when supported by strong data preprocessing and thoughtful feature engineering, can significantly aid in early identification of mental health issues. The findings support the potential of machine learning as a tool for mental health professionals and digital health platforms, especially in resource-constrained settings.

5.2 Future Work

While the results of our study are promising, there are several avenues for future enhancements that could improve the applicability, scalability, and accuracy of mental health prediction systems.

1. Scalability and Real-Time Applications

The models developed in this study were trained and tested on structured survey data. For real-world deployment, especially in applications such as mental health chatbots, mobile health apps, or clinical decision support systems, the models must be able to handle large-scale, real-time data streams. Future work will focus on optimizing the architecture and infrastructure for real-time predictions, possibly by deploying models on cloud platforms or using edge-computing devices for on-the-spot analysis.

2. Integration of Multimodal Data Sources

Currently, the models are based solely on structured data such as questionnaire responses. Future models can benefit greatly from **multimodal data**, such as:

- **Physiological data** (heart rate, sleep patterns, step count) from wearables
- **Textual data** from social media or journaling apps
- **Voice or facial expression analysis** using sentiment recognition

Combining multiple types of data can enhance the richness of features, reduce bias, and improve the predictive power of models.

3. Model Interpretability and Clinical Integration

For machine learning to be accepted in clinical settings, especially in psychiatry and psychology, **explainability** is crucial. Future work should include integrating interpretability frameworks such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), or model-specific visualizations to help clinicians understand how predictions

are made. This will foster trust in the technology and facilitate smoother adoption in mental health practices.

4. Bias Reduction and Fairness Enhancement

One critical challenge in machine learning is ensuring fairness and reducing bias. Mental health data is often influenced by cultural, socio-economic, and gender-based factors. Therefore, future models should include mechanisms to audit and minimize bias. Techniques such as **fairness-aware learning**, **re-sampling**, or **demographic parity constraints** should be explored to ensure equitable model outcomes across different population segments.

5. Expansion to Additional Mental Health Conditions

While this study focused on four disorders, the spectrum of mental health issues is broad. Future work should aim to include other conditions such as:

- Bipolar disorder
- Schizophrenia
- Obsessive-Compulsive Disorder (OCD)
- Eating disorders (e.g., anorexia, bulimia)

This expansion will help create more comprehensive systems capable of early detection and triage across a wider range of mental health issues.

6. Cross-Cultural and Multilingual Model Adaptation

Mental health expression and diagnostic criteria can vary across cultures and languages. As a future enhancement, models should be trained and validated on **cross-cultural datasets**, and possibly localized to different languages using natural language processing (NLP) techniques. This will make the model globally applicable, especially in countries where mental health awareness is low and resources are scarce.

7. Deployment and Usability Testing

Once improved models are developed, the next step is to create an **interactive, user-friendly application** or **web platform** for real users, including patients, psychologists, and general physicians. Usability testing, user feedback collection, and iterative interface development will be critical in ensuring the model's practical utility.

5.3 Final Thoughts

In conclusion, this project has laid a solid foundation for the use of machine learning in mental health diagnostics. It provides empirical evidence that models, particularly Logistic Regression, can be used effectively to detect psychological conditions from user-generated data. However, the field remains in its early stages, and significant progress is needed to move from academic prototypes to clinically viable systems.

By addressing the areas of real-time processing, data diversity, model interpretability, fairness, and global adaptability, future iterations of this system can offer meaningful support to mental health professionals and the broader public. With continued innovation and ethical deployment, machine learning has the potential to become a vital tool in the ongoing effort to understand and manage mental health worldwide.

REFERENCES

[1] A. Rahman, K. Omar, S.A.M. Noah, M.S.N.M. Danuri, and M.A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183963, 2020.

[2] S. Gurjar, C. Patil, R. Suryawanshi, M. Adadande, A. Khore, and N. Tarapore, "Mental health prediction using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 9, no. 12, pp. 1117–1122, 2022.

[3] J. Chung and J. Teo, "Mental health prediction using machine learning: Taxonomy, applications, and challenges," *Applied Computational Intelligence and Soft Computing*, vol. 2022, Article ID 9970363, 2022.

[4] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine learning techniques to predict mental health diagnoses: A systematic literature review," *Clinical Practice and Epidemiology in Mental Health*, vol. 20, Article ID e17450179315688, 2024.

[5] S. Mutalib, N.S. Mohd Shafiee, and S. Abdul-Rahman, "Mental health prediction models using machine learning in higher education," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 5, pp. 1782–1792, 2021.

[6] M. Kavitha, M. Pingili, M. Spurthi, S. Fayaz, K. Kirthan, and S. Santhosh, "Classification algorithm-based mental health prediction using data mining," *TURCOMAT*, vol. 13, no. 2, pp. 1168–1175, 2022.

[7] S.S. Shahapur, P. Chitti, S. Patil, et al., "Decoding minds: Estimation of stress level in students using machine learning," *Indian Journal of Science and Technology*, vol. 17, no. 19, pp. 2002–2012, 2024.

[8] V. Pathak, K. Dwivedi, and M.K. Tiwari, "Mental health prediction using machine learning algorithms," *TEJAS Journal of Technologies and Humanitarian Science*, vol. 3, no. 3, pp. 18–26, 2024.

[9] H. Abdul Rahman, M. Kwicklis, M. Ottom, et al., “Machine learning-based prediction of mental well-being using health behavior data from university students,” *Bioengineering*, vol. 10, no. 5, p. 575, 2023.

[10] M.R. Sumathi and B. Poorna, “Prediction of mental health problems among children using machine learning techniques,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 7, no. 1, pp. 552–557, 2016.

[11] J.A. Renjit, A.S.M.J., S.V.D., and S.D.D., “Prediction of mental health using machine learning,” *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 9, no. 5, pp. i634–i640, 2022.

[12] M.M. Islam, S. Hassan, S. Akter, et al., “A comprehensive review of predictive analytics models for mental illness using machine learning algorithms,” *Healthcare Analytics*, vol. 6, Article ID 100350, pp. 1–12, 2024.

[13] V.G. Kshirsagar, S. Kumar, and N. Karande, “Detection of mental illness using machine learning and deep learning,” *NeuroQuantology*, vol. 20, no. 16, pp. 2606–2613, 2022.

[14] K. Vaishnavi, U.N. Kamath, B.A. Rao, and N.V.S. Reddy, “Predicting mental health illness using machine learning algorithms,” *Journal of Physics: Conference Series*, vol. 2161, Article ID 012021, pp. 1–7, 2022.

[15] C. Su, Z. Xu, J. Pathak, and F. Wang, “Deep learning in mental health outcome research: A scoping review,” *Translational Psychiatry*, vol. 10, Article 116, pp. 1–26, 2020.

[16] A. Thieme, D. Belgrave, and G. Doherty, “Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 27, Article 34, pp. 1–53, 2020.

[17] B. Sahu, J. Kedia, V. Ranjan, et al., “Mental health prediction in students using data mining techniques,” *Open Bioinformatics Journal*, vol. 16, Article ID e187503622307140, 2023.

[18] B. Sahu, J. Kedia, V. Ranjan, et al., “Mental health prediction in students using data mining techniques,” *The Open Bioinformatics Journal*, vol. 16, pp. e187503622307140, 2023.

[19] Author(s), “Building an advanced system leveraging deep learning to predict signs of depression based on social media posts,” *Proceedings of the 2024 International Symposium of Systems, Advanced Technologies and Knowledge (ISSATK)*, Kairouan, Tunisia, pp. [inclusive page numbers], 2024.

[20] A. E. Tate, R. C. McCabe, H. Larsson, et al., “Predicting mental health problems in adolescence using machine learning techniques,” *PLOS ONE*, vol. 15, no. 4, pp. e0230389, 2020.

APPENDIX 1

Appendix A: Machine Learning Model Parameters

Parameter	Description
Input Features	Encoded responses from mental health survey questions
Target Labels	Binary (1 = Mental Health Condition Present, 0 = Absent)
Algorithms Used	Logistic Regression, Random Forest, Decision Tree, SVM, KNN, Naive Bayes
Train-Test Split	80% training – 20% testing
Cross-Validation	5-Fold Stratified
Feature Selection	Correlation Analysis, Chi-Square Test
Scaling Method	StandardScaler (Mean = 0, SD = 1)
Hyperparameter Tuning	GridSearchCV & RandomizedSearchCV
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score, AUC-ROC
Data Preprocessing	Handling Missing Values, Encoding Categorical Data, Normalization

Appendix B: Dataset Description

Source	Self-collected via online survey
Format	Structured CSV from Google Forms

Source	Self-collected via online survey
Total Responses	1000+ participant entries
Features Collected	Demographics, Stress, Anxiety, Sleep, Lifestyle, Trauma, Support systems
Target Variable	Self-reported mental health condition presence
Missing Data Handling	Mean/Mode Imputation and Removal (case-by-case)
Class Balance	Approximately 65% positive, 35% negative

Appendix C: Evaluation Metrics Definitions

Metric	Definition
Accuracy	Proportion of all correctly predicted instances to the total instances
Precision	True Positives / (True Positives + False Positives)
Recall	True Positives / (True Positives + False Negatives)
F1-Score	2 × (Precision × Recall) / (Precision + Recall) – Harmonic mean of precision and recall
AUC-ROC	Area under the Receiver Operating Characteristic curve, measures classification performance

Appendix D: System Workflow

1. **Data Collection**
   - ➤ Mental health survey designed and distributed online (Google Forms)
   - ➤ Responses downloaded and cleaned in CSV format
2. **Preprocessing**
   - ➤ Missing values handled
   - ➤ Feature encoding applied (Label Encoding, One-Hot Encoding)
   - ➤ Normalization using StandardScaler
3. **Model Development**
   - ➤ Multiple classifiers applied (Logistic Regression, Random Forest, SVM, etc.)
   - ➤ Trained on 80% data, tested on 20%
   - ➤ Cross-validation and tuning performed
4. **Model Evaluation**
   - ➤ Models compared based on metrics (Accuracy, Precision, Recall, F1-Score, AUC-ROC)
   - ➤ Best performing model selected (e.g., Random Forest)
5. **Deployment Suggestion (Optional for Future)**
   - ➤ Streamlit web interface mock-up designed for demo
   - ➤ User inputs data and model predicts probability of mental health condition

Appendix E: Visualization and Tools

Tool/Library	Purpose
Python	Main programming language
Pandas, NumPy	Data manipulation and computation

Tool/Library	Purpose
Scikit-learn	Model training, validation, and metrics
Matplotlib, Seaborn	Data visualization (graphs, heatmaps, ROC curves)
Streamlit	(Optional) GUI interface for real-time prediction demo

PCSE-3

ORIGINALITY REPORT

28%	25%	%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to KIET Group of Institutions, Ghaziabad Student Paper	2%
2	www.coursehero.com Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Submitted to National College of Ireland Student Paper	1%
5	d197for5662m48.cloudfront.net Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	fastercapital.com Internet Source	<1%
8	ijircce.com Internet Source	<1%
9	www.biorxiv.org Internet Source	<1%
10	fr.slideshare.net Internet Source	<1%
11	academic-accelerator.com Internet Source	<1%
12	ejournal.ubhara.ac.id Internet Source	

		<1%
13	Submitted to Rutgers University, New Brunswick Student Paper	<1%
14	Submitted to University of Sunderland Student Paper	<1%
15	Submitted to University of Keele Student Paper	<1%
16	medium.com Internet Source	<1%
17	www.informatica.si Internet Source	<1%
18	assets-eu.researchsquare.com Internet Source	<1%
19	Submitted to Medi-Caps University Student Paper	<1%
20	Submitted to University of Stirling Student Paper	<1%
21	ojs.unikom.ac.id Internet Source	<1%
22	theses.gla.ac.uk Internet Source	<1%
23	research-repository.uwa.edu.au Internet Source	<1%
24	Submitted to CSU, Fullerton Student Paper	<1%
25	winnspace.uwinnipeg.ca Internet Source	<1%
26	www2.mdpi.com Internet Source	<1%

27	www.javaguides.net Internet Source	< 1 %
28	www.turcomat.org Internet Source	< 1 %
29	Submitted to Meerut Institute of Engineering & Technology Student Paper	< 1 %
30	Submitted to North Idaho College Student Paper	< 1 %
31	Submitted to University of Bolton Student Paper	< 1 %
32	Submitted to ABES Engineering College Student Paper	< 1 %
33	Submitted to Florida Community College at Jacksonville Student Paper	< 1 %
34	Submitted to Southampton Solent University Student Paper	< 1 %
35	arxiv.org Internet Source	< 1 %
36	ijrpr.com Internet Source	< 1 %
37	Submitted to Islamic University of Technology Student Paper	< 1 %
38	Submitted to Manipal University Jaipur Online Student Paper	< 1 %
39	Submitted to Midlands State University Student Paper	< 1 %
40	Submitted to University of East London Student Paper	< 1 %

41	Submitted to University of Teesside Student Paper	<1%
42	indjst.org Internet Source	<1%
43	machinelearningmodels.org Internet Source	<1%
44	www.jatit.org Internet Source	<1%
45	www.online-pdh.com Internet Source	<1%
46	Submitted to University of York Student Paper	<1%
47	eitca.org Internet Source	<1%
48	mental.jmir.org Internet Source	<1%
49	robots.net Internet Source	<1%
50	tnsroindia.org.in Internet Source	<1%
51	whitelightbh.com Internet Source	<1%
52	www.isteonline.in Internet Source	<1%
53	Submitted to Coventry University Student Paper	<1%
54	Submitted to University of Lancaster-Main Account Student Paper	<1%

55	Submitted to Whitecliffe College of Art & Design Student Paper	<1%
56	linnk.ai Internet Source	<1%
57	molecularautism.biomedcentral.com Internet Source	<1%
58	www.researchgate.net Internet Source	<1%
59	Submitted to APJ Abdul Kalam Technological University, Thiruvananthapuram Student Paper	<1%
60	Submitted to UCL Student Paper	<1%
61	Submitted to University of Lancaster Student Paper	<1%
62	Submitted to University of Westminster Student Paper	<1%
63	digitalcommons.mtu.edu Internet Source	<1%
64	dokumen.pub Internet Source	<1%
65	www.internationaljournalssrg.org Internet Source	<1%
66	Submitted to Northcentral Student Paper	<1%
67	scholarshare.temple.edu Internet Source	<1%
68	www.iapress.org Internet Source	<1%

69	core.ac.uk Internet Source	<1%
70	heca-analitika.com Internet Source	<1%
71	www.northwestpharmacy.com Internet Source	<1%
72	Submitted to The University of Texas at Arlington Student Paper	<1%
73	bioone.org Internet Source	<1%
74	doctorterrylynch.com Internet Source	<1%
75	ejournal.uin-suska.ac.id Internet Source	<1%
76	er.ucu.edu.ua Internet Source	<1%
77	gjaets.com Internet Source	<1%
78	ideas.repec.org Internet Source	<1%
79	irjet.net Internet Source	<1%
80	onlinegamescastle.com Internet Source	<1%
81	psa.gov.ph Internet Source	<1%
82	sk.sagepub.com Internet Source	<1%
83	thesciencebrigade.com Internet Source	

		<1%
84	Submitted to University of Mpumalanga Student Paper	<1%
85	www.beccadinonaskyrace.com Internet Source	<1%
86	www.codewithc.com Internet Source	<1%
87	www.internationalpubls.com Internet Source	<1%
88	www.jetir.org Internet Source	<1%
89	aircconline.com Internet Source	<1%
90	allmeld.com Internet Source	<1%
91	boxoflearn.com Internet Source	<1%
92	journals.plos.org Internet Source	<1%
93	nyec.org Internet Source	<1%
94	revistas.anahuac.mx Internet Source	<1%
95	seminar.iaii.or.id Internet Source	<1%
96	vuir.vu.edu.au Internet Source	<1%
97	www.cse.griet.ac.in Internet Source	<1%

98	www.frontiersin.org Internet Source	<1%
99	www.medrxiv.org Internet Source	<1%
100	www.scirp.org Internet Source	<1%
101	blog.enterprisedna.co Internet Source	<1%
102	docslib.org Internet Source	<1%
103	easychair.org Internet Source	<1%
104	ijiemr.org Internet Source	<1%
105	ijrmst.com Internet Source	<1%
106	lifesight.io Internet Source	<1%
107	liu.diva-portal.org Internet Source	<1%
108	themendingmuse.com Internet Source	<1%
109	thesai.org Internet Source	<1%
110	www.appliedaicourse.com Internet Source	<1%
111	www.cambridge.org Internet Source	<1%
112	www.escholar.manchester.ac.uk Internet Source	<1%

113	www.icaset.in Internet Source	<1%
114	www.inderscience.com Internet Source	<1%
115	www.scienceopen.com Internet Source	<1%
116	www.swamivivekanandauniversity.ac.in Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		