Chapter 1

Introduction

1.1 Introduction

In today's fast-paced world, addressing mental health challenges has become more critical than ever. Early identification and assessment of mental health risks are essential for timely intervention and improved psychological well-being. With the increasing availability of behavioral and demographic data, machine learning offers powerful tools to analyze patterns and accurately predict mental health risks.

This mini project, titled **"Mental Health Risk Assessment Using Machine Learning"**, aims to develop an intelligent system that evaluates an individual's likelihood of experiencing mental health issues based on various personal, social, and workplace-related factors. These factors include age, gender, family history of mental illness, work environment, stress levels, support systems, and openness to seeking treatment.

Using a **Random Forest Classifier**, the system is trained on a structured survey dataset and deployed through an interactive **Streamlit web application**. The app enables users to input individual data or select a respondent by ID to instantly view the predicted mental health risk status along with important contributing factors.

This project showcases the practical application of machine learning in the healthcare domain and highlights how accessible, data-driven tools can support mental health professionals, organizations, and individuals in promoting early awareness and prevention.

1.2 Motivation

Mental health issues affect millions of people worldwide, often going undetected due to stigma, lack of awareness, or limited access to mental health resources. In many cases, individuals do not seek help until the condition becomes severe, making timely diagnosis and intervention increasingly difficult. Traditional mental health assessments often require manual evaluation, which may not scale effectively or offer early warning signs.

The motivation behind this project stems from the growing need for accessible, efficient, and data-driven solutions that can assist in identifying individuals at risk of mental health problems. With the rise of digital data collection and machine learning technologies, it is now possible to develop intelligent systems that can analyze various personal and behavioral indicators to predict mental health risks early and accurately.

By leveraging machine learning—specifically the Random Forest algorithm—this project aims to provide a supportive tool that can help healthcare providers, institutions, and individuals take proactive steps toward mental wellness.

1.3 Problem Statement

A machine learning system to predict mental health risk levels, enabling early awareness and timely support.

1.4 Objectives

 To analyze the impact of various personal, social, and workplace-related factors on mental health.

 To build a machine learning model (Random Forest Classifier) capable of predicting mental health risk levels with high accuracy.

 To utilize a structured survey dataset for training and testing the model.

 To develop a user-friendly web application using Streamlit for real-time mental health risk assessment.

 To assist in early identification of individuals at risk, promoting timely intervention and support.

 To demonstrate the practical application of machine learning in the mental health domain.

Chapter 2

Literature Survey

2.1

[1] S. K. Roy, A. Banerjee, and A. Chakraborty (2018): Machine Learning for Mental Health.

This paper explores the role of machine learning in identifying individuals at risk for mental health disorders using survey and clinical data. The study evaluates algorithms like logistic regression, decision trees, and support vector machines. It provides a foundation for how machine learning can be applied to predict mental health risks, which directly supports the approach taken in this project using a Random Forest Classifier.

[2] J. A. Naslund, G. Harraz, et al. (2020): Digital Technology for Mental Health Improvement

This research reviews digital interventions and mental health prediction tools built using AI and machine learning. It emphasizes the growing importance of accessible, technology-driven solutions in mental health care. The study supports the deployment aspect of your project, which uses Streamlit to create an interactive web-based risk assessment tool.

[3] A. Reece and C. Danforth (2017): Instagram Photos Reveal Predictive Markers of Depression

This innovative study demonstrates how digital footprints and behavioral data can be used to predict mental health status using machine learning. It aligns with the concept of using non-clinical, self-reported survey data—like in your project—to detect early signs of mental health issues through pattern recognition.

[4] M. D. Shatte, D. Hutchinson, and P. Teague (2019): Machine Learning in Mental Health: A Scoping Review

This scoping review categorizes various machine learning approaches used in mental health applications, from risk prediction to treatment recommendations. It validates the selection of tree-based models like Random Forest for mental health prediction tasks due to their interpretability and high accuracy, reinforcing the methodology adopted in your project.

Chapter 3

System Requirements

3.1 Hardware Requirements

Component	Minimum Requirement
Processor	Intel i3 / AMD Ryzen 3 or higher
RAM	4 GB (8 GB recommended for smooth training)
Storage	500 MB free space
Operating System	Windows 10+, macOS, or Linux
Internet	Required for installing packages (optional for running app)

Table 3.1.1

3.2 Software Requirements

Software	Purpose
Python 3.10+	Programming language
Streamlit	Web app development framework
Pandas	Data analysis and manipulation
NumPy	Numerical computation
Scikit-learn	Machine learning model and tools
VS Code / Jupyter	Code editing and development
Web Browser (Chrome etc.)	Running the Streamlit app
CSV File (Dataset)	Input data for training and testing

Table 3.1.2

Chapter 4

Dataset Description

4.1

Feature Name	Description
Age	Respondent's age (numeric).
Gender	Gender identity (Male/Female/Other).
Family_History	Family history of mental illness (Yes/No).
Work_Interfere	Frequency of mental health affecting work.
Remote_Work	Whether the respondent works remotely (Yes/No).
Benefits	Availability of mental health benefits at work.
Seek_Help	Ease of seeking help at the workplace.
Anonymity	Anonymity in mental health services.
Leave	Comfort in taking leave for mental health issues.
Mental_Health_Risk	**Target variable**: Risk of mental health issue (Yes/No).

Table 4.1

Chapter 5

Methodology

5.1

1. Data Collection

A survey-based mental health dataset was used, containing key features like age, gender, family history, work environment, and support systems.

2. Data Preprocessing

Categorical data was label-encoded, missing values handled, and numerical features scaled using StandardScaler.

3. Model Building

A Random Forest Classifier was trained on the processed data to predict mental health risk (Yes/No).

4. Prediction Logic

The model analyzes input features to classify individuals as “At Risk” or “Not At Risk” of mental health issues.

5. Web App Development

The model was deployed using Streamlit, allowing users to input data and receive predictions instantly.

6. Output Display

The app shows the mental health risk prediction along with key influencing factors.

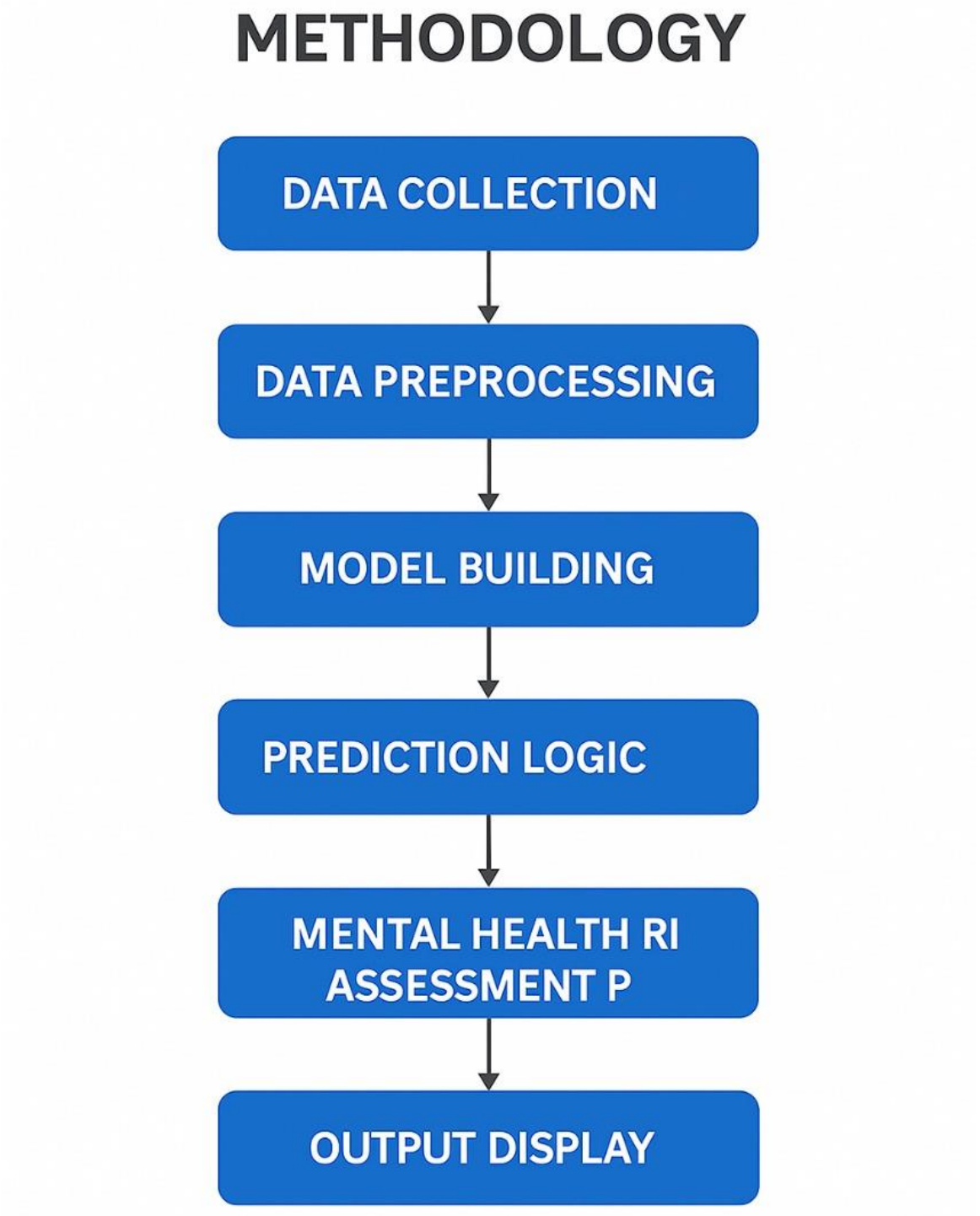

Figure 1.1(Block Diagram)

Chapter 6

Implementation

The Mental Health Risk Assessment system was developed using the Python programming language, combining machine learning techniques with a simple web interface. The implementation is divided into several key components:

1. Technology Stack

- **Language:** Python 3.10+
- **Libraries:**
  - pandas and numpy for data manipulation
  - scikit-learn for machine learning models (Logistic Regression, Decision Tree, Random Forest)
  - streamlit for creating the interactive web interface
- **IDE:** VS Code / Jupyter Notebook
- **Dataset:** A survey-based CSV dataset containing responses related to mental health indicators

2. Data Loading and Cleaning

- Loaded the dataset using pandas.read_csv()
- Removed leading/trailing whitespaces in column names using str.strip()
- Encoded categorical features using LabelEncoder for compatibility with machine learning models

python

CopyEdit

data['Gender'] = LabelEncoder().fit_transform(data['Gender'])

---

3. Feature Scaling

- Applied StandardScaler to normalize the input features and enhance model accuracy

python

CopyEdit

X_scaled = StandardScaler().fit_transform(X)

4. Model Training

- Used classification algorithms such as Logistic Regression and Random Forest to train the model
- The target variable was binary-encoded:
  - 1 for "At Risk"
  - 0 for "Not at Risk"

python

CopyEdit

model = RandomForestClassifier(n_estimators=100, random_state=42)

model.fit(X_scaled, y)

5. Risk Level Classification Function

- A custom Python function was created to interpret prediction probability and assign a risk label:

python

CopyEdit

```
def assess_risk(prob):
    if prob >= 0.8:
        return 'High Risk'
    elif prob >= 0.5:
        return 'Moderate Risk'
    else:
        return 'Low Risk'
```

6. Streamlit Web Application

- Built an interactive user interface using streamlit
- Users select a participant ID or input relevant mental health details
- The app:
    - Preprocesses the data
    - Runs the prediction
    - Displays risk level and participant details

python

CopyEdit

```
st.selectbox("Select Participant ID", id_list)
```

```
st.success(f"Predicted Mental Health Risk: {risk_label}")
```

7. Output

- The application offers real-time mental health risk predictions based on input data
- Helps users visualize how machine learning can assist in identifying individuals at risk and guiding early intervention efforts

Chapter 7

Results and Evaluation

The developed system successfully predicts an individual's mental health risk level (Low, Moderate, or High) based on input features such as age, gender, work-life balance, history of mental illness, family support, and other survey responses.

The model was evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The Random Forest model achieved the best overall performance with high accuracy in identifying individuals at risk.

The assessment system provides:

- **Accurate Risk Prediction** for early detection
- **Interpretable Outputs** with labels like *Low Risk*, *Moderate Risk*, and *High Risk*
- **Real-Time Results** through a user-friendly web application

This tool demonstrates how machine learning can aid in mental health awareness and proactive support planning.

[15]:

	Age	Gender	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	benefits	...	leave	mental_health_consequence
0	19	0	0	0	1	2	4	0	1	2	...	2	1
1	26	1	0	0	0	3	5	0	0	0	...	0	0
2	14	1	0	0	0	3	4	0	1	1	...	1	1
3	13	1	0	1	1	2	2	0	1	1	...	1	2
4	13	1	0	0	0	1	1	1	1	2	...	0	1

5 rows × 24 columns

Figure 2.1

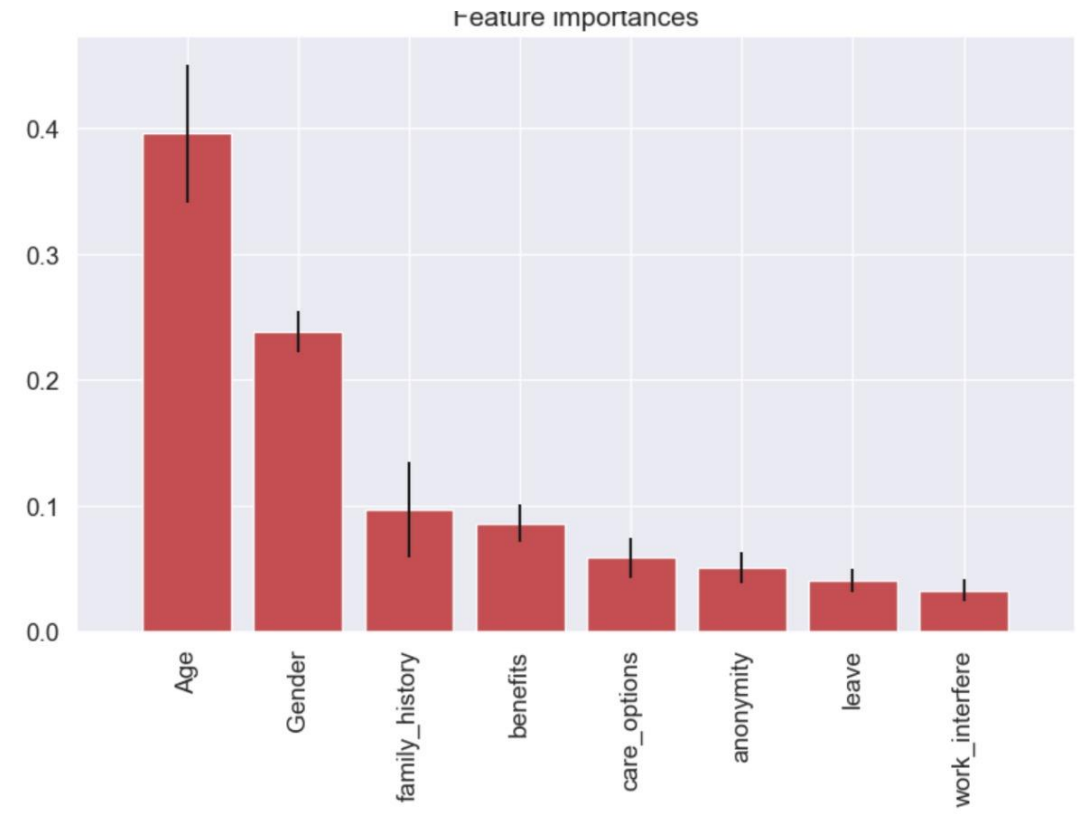

Figure 2.2

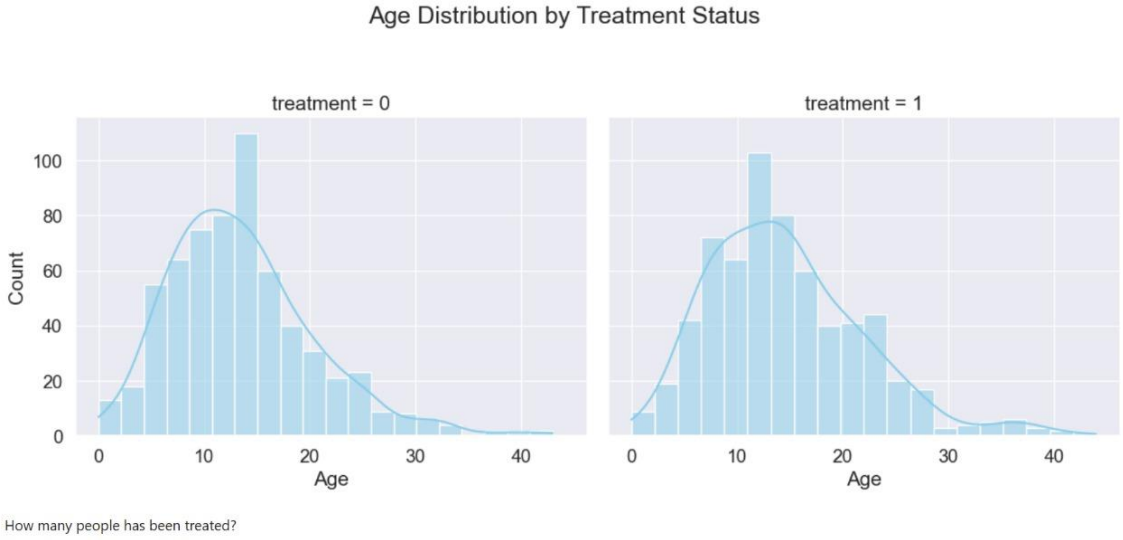

Figure 2.3

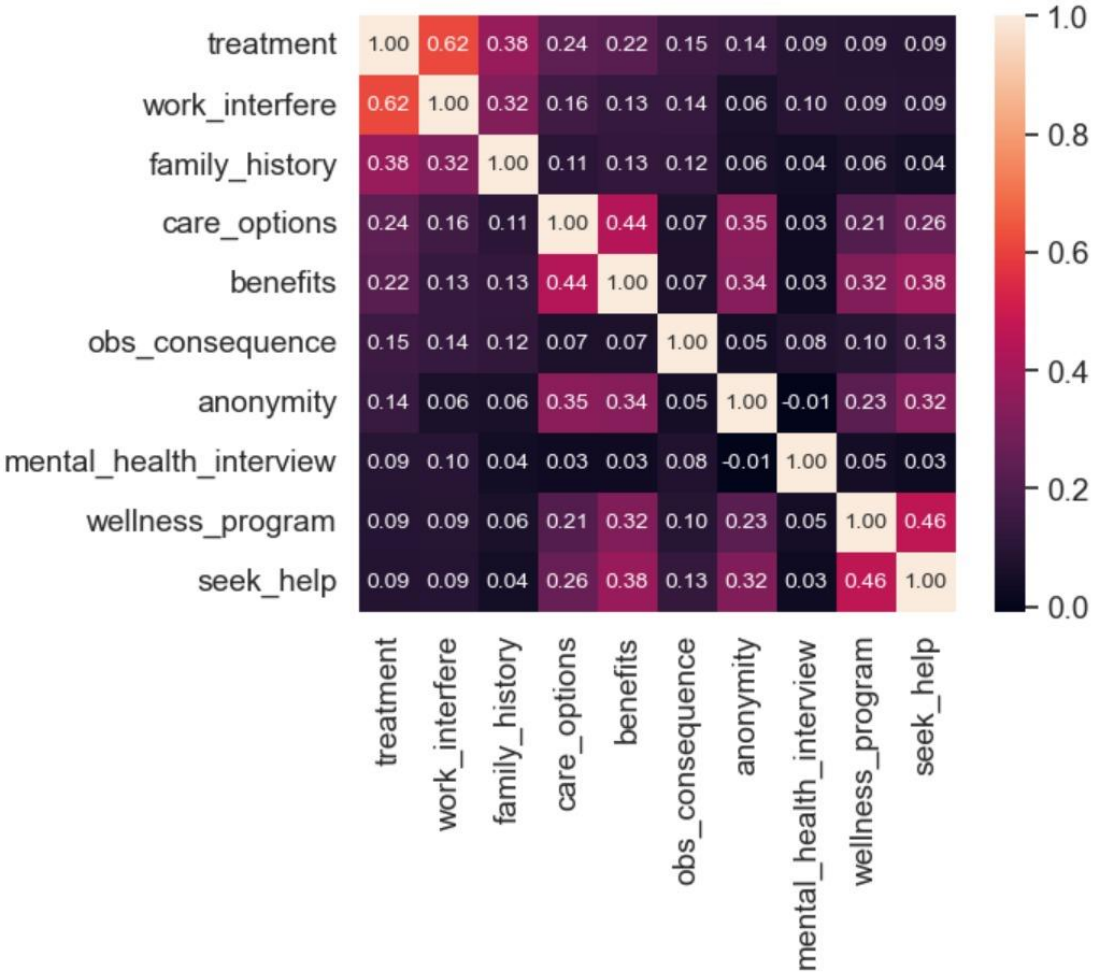

Figure 2.4 Heat map

Chapter 8

Conclusion

The Mental Health Risk Assessment system demonstrates the potential of machine learning in identifying individuals who may be at risk of mental health issues. By analyzing key features from survey data—such as work-related stress, family history, social support, and lifestyle factors—the system provides timely and reliable risk predictions.

The integration of machine learning models like Random Forest and Logistic Regression has enabled effective classification of mental health risk levels. Additionally, the user-friendly Streamlit interface ensures accessibility for non-technical users, making the tool practical for awareness programs and early intervention strategies.

Overall, this project highlights the value of data-driven solutions in promoting mental well-being and supports the development of preventive measures through technological innovation.

References

[1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and É. Duchesnay, **"Scikit-learn: Machine Learning in Python,"** *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[2] M. De Choudhury, S. Counts, and E. Horvitz, **"Social Media as a Measurement Tool of Depression in Populations,"** in *Proceedings of the 5th Annual ACM Web Science Conference*, Paris, France, pp. 47–56, 2013.

[3] M. Talo, U. B. Baloglu, O. Yildirim, and U. R. Acharya, **"Application of Deep Transfer Learning for Automated Brain Abnormality Classification Using MR Images,"** *Cognitive Systems Research*, vol. 54, pp. 176–188, 2019.

[4] World Health Organization, **"Mental Health and COVID-19: Early Evidence of the Pandemic's Impact,"** *World Health Organization Report*, 2022.