CHAPTER 1

INTRODUCTION

Mental health is a crucial aspect of overall well-being, yet it often goes unnoticed until serious dilemmas present themselves. In today world everyone lives in, many individuals experience different forms of mental health disorders, for example, anxiety, depression, and stress-related disorders. Early detection with timely intervention can prove to lead to better outcomes. However, the traditional assessment methods consume a lot of time and do not reach everyone in need.

The "Mental Health Prediction Using Machine Learning" project addresses this gap by using machine learning to deliver a user-friendly platform to classify mental health conditions. Utilizing the presented web application, a structured questionnaire will be applied to collect responses in regard to emotional or psychological states of users. Several trained machine learning models, including Random Forest, Gradient Boosting, Logistic Regression, and K-Nearest Neighbors (KNN), will be used in an attempt to analyse the data in a way to predict possible mental health issues.

The nature of questions it allows them to provide would be such that the user obtains all the information he requires but simultaneously, the user is in an enabling environment. Upon filling the questionnaire, users have the ability of choosing different predictive models which can enable them to get insight into their mental health status. The application empowers individuals with knowledge about their mental health and promotes awareness thus creating the impulse to see a professional.

this project aims to provide a valuable tool for early detection and proactive management of mental health issues, ultimately contributing to improved mental health outcomes in the community.

CHAPTER 2

Literature Review

This chapter gives an insight about the description of the survey of literature, existing and proposed system, Hardware and Software Requirements along with Tools and Technologies used in implementing the project.

2.1 Understanding Mental Health

Mental health is a vital part of our overall well-being, influencing how we think, feel, and interact with the world around us. It's not just about the absence of mental illness; it's about having the emotional resilience to cope with life's challenges and the ability to connect with others. Sadly, mental health issues—like anxiety, depression, and stress—are more common than many of us realize. In fact, research shows that about 1 in 5 adults will experience a mental health condition at some point in their lives.

The impact of mental health extends beyond the individual, affecting families, friends, and entire communities. For instance, studies have found that people with mental health disorders often face additional health challenges, which can complicate their lives even further. This reality highlights the importance of addressing mental health proactively and comprehensively.

2.2 Machine Learning in Mental Health

In recent years, machine learning has emerged as a powerful ally in the quest to understand and predict mental health outcomes. Researchers are increasingly turning to technology to analyse complex data sets and uncover patterns that might indicate mental health issues. For example, a fascinating study by Reece and Danforth (2017) explored how algorithms can analyse social media activity to predict feelings of depression and anxiety. By examining the language and behaviour of users, these algorithms can provide insights into their mental states, showcasing how technology can serve as an early warning system.

Another noteworthy study by Ribeiro et al. (2016) demonstrated how machine learning can classify individuals based on their mental health status using survey data. The results were promising algorithms like support vector machines and random forests showed remarkable accuracy in predicting mental health conditions. This ability to analyse large amounts of data and identify hidden patterns makes machine learning an invaluable resource in the mental health field.

Furthermore, integrating machine learning with electronic health records (EHRs) is opening new doors for mental health predictions. For example, a study by Dilsizian and Siegel (2014) found that machine learning models could effectively predict hospital readmissions for patients with mental health disorders. This kind of predictive capability can help healthcare providers allocate resources more effectively and improve patient care.

2.3 Problem statement:

Issues of depression, anxiety and stress are becoming common among people rather than individuals suffering from them only; however, they impact on their quality living standards within a community setting (i.e. Within families where parents are always busy working). Therefore, conventional approaches which rely on self-reported data obtained through clinical interviews often lack timeliness or accuracy hence making it necessary to have a machine learning model to predict and suggest the diagnosis.

2.4 Questionnaires in Mental Health Assessment

Questionnaires have long been a staple in mental health assessment, offering a structured way to gather information about individuals' emotional and psychological well-being. Tools like the Beck Depression Inventory (BDI) and the Generalized Anxiety Disorder 7-item scale (GAD-7) help clinicians quantify symptoms and track changes over time.

One of the great things about questionnaires is their accessibility. They can be administered in various settings, including online, making it easier for people to participate. Research by Hohls et al. (2017) highlighted how online questionnaires can reach a broader audience, breaking down barriers to mental health assessment. Additionally, mobile apps that incorporate questionnaires are making it even easier for individuals to monitor their mental health in real-time (Fleming et al., 2019).

However, crafting an effective questionnaire requires careful thought. It's essential to design questions that are clear and unbiased, as poorly worded questions can lead to inaccurate responses and misinterpretation of findings. McDowell (2006) emphasized the importance of straightforward question design to ensure that the data collected is reliable.

2.5 Technical Features of the Proposed System

- **User-Friendly Interface:** The proposed system will feature an intuitive and accessible user interface designed to enhance the user experience. Participants will first navigate through a series of questions related to their mental health. After submitting their responses, they will have the opportunity to review their answers. If they wish, they can easily go back to edit any of their responses before finalizing their submission.
- **Personalized Predictions and Recommendations:** Once participants have completed their questionnaire, they will receive a personalized prediction regarding their mental health condition. This prediction will include a description of any identified disorders, providing valuable insights into their mental well-being. Additionally, the system will offer medication recommendations tailored to their situation, helping them understand potential next steps in their mental health journey.
- **Resource Links for Further Information:** To empower participants with knowledge, the system will also provide links to reputable resources where they can learn more about their predicted condition. By clicking on these links, users will gain access to comprehensive information that can help them make informed decisions about their mental health.

This user-centric approach aims to create a supportive environment where individuals feel comfortable engaging with their mental health assessment and are empowered to seek additional information and support as needed.

2.6 Tools and Technologies Used

- **Visual Studio Code:** For developing the application, I utilized Microsoft's Visual Studio Code, a powerful and versatile source-code editor. This tool provided essential features such as debugging, syntax highlighting, and intelligent code completion, which streamlined my workflow while working with Flask for the backend in Python. Additionally, I created HTML documents for the home page and results page within the same environment, benefiting from its customizable themes and extensions that enhanced my coding experience.
- **Jupyter Notebook:** To train the machine learning model, I employed Jupyter Notebook, which allowed me to write and execute Python code in an interactive manner. This platform is particularly well-suited for data analysis and model training, enabling me to visualize results and iterate on my machine learning algorithms effectively.
- **Google Forms:** For data collection, I used Google Forms to design and distribute the mental health questionnaire. This user-friendly tool made it easy to gather responses from participants, ensuring an efficient and organized collection process.
- **Google Sheets:** After collecting the data, I turned to Google Sheets for data preprocessing. This tool provided a convenient way to clean, organize, and analyse the responses before feeding them into the machine learning model.
- **Excel:** Finally, I used Excel for storing the processed data. This familiar and robust spreadsheet application allowed me to manage and maintain the dataset effectively, ensuring that the information was readily accessible for further analysis and reporting.

By leveraging these tools and technologies, I was able to create a cohesive and efficient workflow throughout the various stages of the project, from data collection to model training and application development.

2.7 Hardware and Software Requirements

When deploying this application, it's important to consider the minimum and recommended hardware and software requirements to ensure optimal performance. The application is developed using Anaconda, and the following specifications will guide users in setting up their systems effectively.

2.7.1 Hardware Requirements

Table 2.1 Recommended Hardware Requirements

Component	Requirement
Processor	Intel Core i5 (2.5 GHz) or equivalent
RAM	4GB
Disk Space	200GB

Table 2.1 presents the recommended hardware requirements for deploying the application effectively. With this configuration, users can expect smooth performance and an overall better experience while using the application.

2.7.2 Software Requirements

To run this application, users will need compatible software environments. The following operating systems are supported:

- Windows
- macOS
- Linux

The application is developed using the following tools and technologies:

- Programming Language: Python
- Framework: Flask
- Frontend Technologies: HTML, CSS
- Development Tools: Anaconda, Visual Studio Code, Jupyter Notebook

CHAPTER 3

Methodology

3.1 Questionnaire Development

The journey of our project began with creating a structured questionnaire aimed at assessing mental health symptoms. We focused on identifying key symptoms associated with various mental health disorders, which helped shape our questions. Here's how we went about it:

3.1.1 Design Process

To design the questionnaire, we researched common symptoms related to mental health disorders. By reviewing existing literature, we compiled a list of symptoms that people typically experience, such as anxiety, sadness, and changes in sleep patterns. With these insights, we crafted questions that would accurately capture these symptoms in a clear and relatable way.

3.1.2 Hypothesis

- **Hypothesis 1:** Younger individuals are more likely to report feeling nervous, anxious, or experience panic attacks compared to older individuals.

- **Hypothesis 2:** Individuals who have trouble concentrating are also likely to report trouble sleeping.

- **Hypothesis 3:** Individuals who are having trouble with work are more likely to feel hopeless and tired.

- **Hypothesis 4:** People who feel hopeless are more prone to feelings of anger and overreacting in situations.

- **Hypothesis 5:** Introverts are more likely to avoid social activities and report addicted to social media.

3.1.2 Questionnaires

1. How old are you?
2. Do you feel often nervous?
3. Do you feel often panic?
4. Do you get sudden bursts of rapid breathing?
5. Do you get sudden bursts of excessive Sweating?
6. Do you have trouble concentrating
7. Do you have trouble in sleeping?
8. Are you having trouble with work?
9. Do you feel hopeless?
10. Do you feel angry?
11. Do you overreact?
12. Have you changed your eating pattern?
13. Do you ever get suicidal thoughts?
14. Are you always tired?
15. Do you have a close friend(s)?
16. Are you addicted to social media?
17. Have you gained weight of recently?
18. Are you an introvert?
19. Do you ever get pop ups of a stressful memory?
20. Do you frequently experience nightmares?
21. Do you often avoid people or activities?
22. Do you often find yourself feeling negative?
23. Do you have trouble focusing on a task?
24. Do you often blame yourself for an unfortunate event?
25. Do you ever get hallucinations?
26. Do you exhibit repetitive behaviour?
27. Do you get seasonal depression?
28. Have you recently experienced an increase in energy?

3.1.3 Importance of Questionnaires in Mental Health Prediction

- **Direct Data Collection:** These questionnaires allow us to directly collect information from individuals regarding their mental health status. By asking specific questions related to emotional and psychological well-being, we gain insights into their symptoms, experiences, and overall mental health.
- **Personal Experiences:** Through these questionnaires, individuals can express their opinions and personal experiences related to mental health. This includes reporting feelings of anxiety, depression, stress, and other psychological states.
- **Identifying Critical Issues:** The questionnaires help identify which mental health issues are most critical to individuals. This understanding can guide targeted interventions and support strategies.
- **Learning Patterns with Machine Learning:** Machine learning algorithms can learn patterns from the questionnaire-based data to predict mental health outcomes. This capability enables early detection and proactive management of mental health issues

3.1.4 Challenges and Considerations

- **Bias:** It is crucial to ensure that these questionnaires are unbiased and do not favour any specific demographic group or mental health condition. This helps maintain the integrity of the data collected.
- **Sample Size:** Collecting a representative sample size is essential to generalize findings and ensure that the results reflect the broader population.
- **Privacy:** Respecting respondent’s privacy and ensuring anonymity is critical. This fosters trust and encourages individuals to provide honest and accurate responses.

3.2 Data Collection

With our questionnaire ready, I have decided to use Google Forms for data collection. This platform made it easy and accessible for participants to share their experiences while ensuring a smooth process. Here’s how we managed the data collection:

3.2.1 Population and Sampling Methods

Our target population included all the age groups people from children to older. To ensure a diverse representation, we used a stratified sampling method, dividing the population based on age groups. However, we only collected age as demographic information to respect participants' privacy. This approach allowed us to maintain anonymity while still gathering valuable insights.

3.2.2 Administering the Questionnaire

We distributed the questionnaire through Google Forms, which provided a user-friendly interface for participants. We shared the link via social media and community groups, encouraging individuals to participate by highlighting the importance of mental health awareness. The online format allowed participants to respond at their convenience, increasing participation rates.

3.2.3 Ethical Considerations

Throughout the data collection process, we prioritized ethical considerations. Participants were fully informed about the study’s purpose, procedures, and any potential risks before giving their consent. We emphasized that their responses would be kept confidential and anonymized. And we did not collect any personal information from the respondents.

3.2.4 Overview of Dataset

- **Age Distribution**

The dataset comprises a diverse range of participants categorized by age. Below is a summary of the age distribution:

- Ages 4 to 10 ⇨ 31 participants
- Ages 11 to 20 ⇨ 87 participants
- Ages 21 to 30 ⇨ 371 participants
- Ages 31 to 40 ⇨ 100 participants
- Ages 41 to 50 ⇨ 10 participants
- Ages 51 to 60 ⇨ 3 participants
- Ages 61 to 70 ⇨ 2 participants
- Ages 71 to 80 ⇨ 2 participants
- Ages 81 to 90 ⇨ 1 participant
- Ages 91 to 100 ⇨ 1 participant
- Ages 101 to 120 ⇨1 participant

This age distribution illustrates a wide range of participants, with the majority falling within the age group of 21 to 30, indicating a focus on young adults.

- **Disorder Prevalence**

The dataset also categorizes participants based on various mental health disorders. Below are the counts of the most common disorders identified:

- Post-Traumatic Stress Disorder (PTSD): 177 participants
- General Health Concern: 123 participants
- Obsessive-Compulsive Disorder (OCD): 122 participants
- Panic Disorder, Post-Traumatic Stress Disorder (PTSD): 119 participants
- Attention-Deficit/Hyperactivity Disorder (ADHD): 118 participants
- Panic Disorder: 81 participants
- Generalized Anxiety Disorder (GAD): 70 participants
- Generalized Anxiety Disorder (GAD), Major Depressive Disorder (Depression): 49 participants
- Social Anxiety Disorder: 43 participants
- Schizophrenia: 43 participants

- Borderline Personality Disorder (BPD): 39 participants
- Major Depressive Disorder (Depression): 38 participants
- Bipolar Disorder: 36 participants
- Seasonal Affective Disorder (SAD): 25 participants
- General Health Check: 11 participants

✈**Grand Total of Participants: 1,094**

This overview of disorders indicates that PTSD is the most frequently reported condition, emphasizing the need for resources and support for individuals experiencing trauma-related symptoms. The dataset reflects a broad spectrum of mental health challenges, underscoring the importance of comprehensive assessment and intervention strategies.

3.3 Data Processing

Once we collected the responses, it was time to process the data and prepare it for analysis. Here’s how we approached this phase:

3.3.1 Data Cleaning

Data cleaning was essential to ensure we had accurate and reliable data. We carefully reviewed the responses to identify any missing values, duplicate entries, or outliers. Any incomplete responses were excluded from the dataset to maintain its integrity.

3.3.2 Data Preprocessing Techniques

After cleaning the data, we used Excel to process and scale the responses. We employed the SUBSTITUTE formula to replace specific values in the dataset, making it easier to standardize responses. This allowed us to clearly scale the data, ensuring consistency across participants’ answers. For instance, we could convert qualitative responses into numerical values that could be more easily analysed.

Formula:- SUBSTITUTE(text, old_text, new_text, [instance_num])

Example:- SUBSTITUTE(A1, “Yes”, "2")

3.4 Data Labelling for Model Training

After collecting the data, the next crucial step in our methodology is to prepare the dataset for training the machine learning models. This preparation involves labelling the data based on the symptoms associated with each mental health disorder.

- **Identifying Symptoms:** Each disorder is characterized by a unique set of symptoms. We conducted a thorough review of existing literature and clinical guidelines to identify and define these symptoms accurately. This ensures that our model is trained on relevant and clinically recognized indicators.
- **Labelling the Data:** Using the identified symptoms, we categorized and labelled the collected data. Each participant's responses were analysed to determine which disorder(s) they align with based on their reported symptoms.

For example, if a participant reported symptoms consistent with Post-Traumatic Stress Disorder (PTSD), their data would be labelled accordingly. This process was repeated for each disorder represented in our dataset.

- **Creating Labelled Datasets:** The labelled data was then organized into separate datasets for each disorder. This organization allows the model to learn specific patterns and characteristics associated with each condition, improving the accuracy of predictions. Additionally, we ensured that the dataset is balanced, where possible, to avoid bias in the model's learning process.
- **Training the Model:** Once the dataset is labelled, it is ready for model training. The labelled data serves as the foundation for teaching the machine learning algorithms to recognize and differentiate between the various mental health disorders based on the symptoms presented by users. I have used multiple machine learning models, including Logistic Regression and Random Forest, to evaluate their effectiveness in predicting mental health conditions accurately.

3.4.1 Overview of Chosen Algorithms

The final phase of our methodology involved using machine learning algorithms to analyse the processed data and predict mental health outcomes. Here's how we approached this part:

I have selected several machine learning algorithms to train our models:

- **Logistic Regression:**

  Our Logistic Regression model achieved an accuracy of 80%. This model effectively identifies relationships between the features in the dataset and the likelihood of a participant experiencing specific mental health disorders.

  Logistic Regression is a statistical method used for binary classification problems, although it can be extended to multi-class classification. It estimates the probability that a given input point belongs to a particular category by applying a logistic function to the linear combination of input features.

  Here is how Logistic Regression Works

  The process of implementing Logistic Regression involves the following steps:

  - **Model Initialization:** The model initializes with coefficients (weights) assigned to each feature, which are adjusted during the training process to minimize the error.
  - **Calculating the Probability:** For each input instance, the model computes a weighted sum of the input features. This sum is then passed through the logistic function (also known as the sigmoid function), transforming the output into a probability value between 0 and 1.
  - **Thresholding:** A threshold (commonly set at 0.5) is applied to the probability to determine the predicted class. If the predicted probability is greater than or equal to the threshold, the instance is classified into one category; otherwise, it is classified into another.
  - **Training the Model:** The model is trained using a labelled dataset where the actual outcomes are known. During training, the model uses optimization techniques like gradient descent to adjust the coefficients

iteratively, minimizing a loss function (often the log loss) that quantifies the difference between predicted probabilities and actual outcomes.

- **Evaluating Performance:** Once trained, the model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in predicting mental health disorders.

```
Accuracy: 0.8053 (+/- 0.0293)
Precision: 0.8118 (+/- 0.0258)
Recall: 0.8053 (+/- 0.0293)
F1: 0.8011 (+/- 0.0299)
```

➢ Decision Tree

Our Decision Tree model exhibited accuracy of 72%. The model effectively captured patterns within the data, showcasing its robust predictive capabilities.

A decision tree is a flowchart-like structure used to make decisions or predictions. It consists of nodes representing decisions or tests on attributes, branches representing the outcome of these decisions, and leaf nodes representing final outcomes or predictions. Each internal node corresponds to a test on an attribute, each branch corresponds to the result of the test, and each leaf node corresponds to a class label or a continuous value.

Here is how Decision Trees Work

The process of creating a decision tree involves:

- Selecting the Best Attribute: Using a metric like Gini impurity, entropy, or information gain, the best attribute to split the data is selected.
- Splitting the Dataset. The dataset is split into subsets based on the selected attribute.
- Repeating the Process. The process is repeated recursively for each subset, creating a new internal node or leaf node until a stopping criterion is met (e.g., all instances in a node belong to the same class or a predefined depth is reached).

```
Accuracy: 0.7202 (+/- 0.0428)
Precision: 0.7214 (+/- 0.0400)
Recall: 0.7202 (+/- 0.0428)
F1: 0.7142 (+/- 0.0429)
```

➢ Random Forest

Our Random Forest model achieved an accuracy of 85%. This ensemble learning method effectively improves predictive performance by combining multiple decision trees, thereby enhancing stability and accuracy.

Random Forest is a collection of decision trees, where each tree is built using a random subset of the data and features. The final prediction is made by aggregating the predictions from all the individual trees, either through majority voting (for classification) or averaging (for regression).

The process of implementing a Random Forest involves the following steps:

- **Bootstrapping:** Random subsets of the training data are created using bootstrapping (sampling with replacement). Each subset is used to train a separate decision tree.
- **Feature Randomness:** At each node of the decision tree, a random subset of features is selected to determine the best split. This randomness helps to ensure that the trees are diverse and reduces the risk of overfitting.
- **Building Multiple Trees:** Each tree is grown to its full depth without pruning, allowing it to capture complex patterns in the data.
- **Aggregating Predictions:** Once all trees are trained, the model aggregates their predictions. For classification tasks, the predicted class is the majority vote among the trees. For regression tasks, the final prediction is the average of the predictions from all trees.
- **Evaluating Performance:** The model's performance is evaluated using various metrics, such as accuracy, precision, and recall, to ensure its effectiveness in predicting mental health disorders

```
Accuracy: 0.7998 (+/- 0.0145)
Precision: 0.8248 (+/- 0.0111)
Recall: 0.7998 (+/- 0.0145)
F1: 0.7893 (+/- 0.0119)
```

➢ Gradient Boosting

Our Gradient Boosting model achieved an accuracy of 85%. This ensemble technique builds a strong predictive model by combining the outputs of several weak learners, specifically decision trees, in a sequential manner.

Gradient Boosting focuses on minimizing the loss function of the model by adjusting the predictions incrementally. Each new tree is trained to correct the errors made by the previous trees, leading to improved accuracy.

The process of implementing Gradient Boosting involves the following steps:

- **Initialization:** The model starts with an initial prediction, often the mean of the target variable for regression tasks or the log-odds for classification tasks.
- **Calculating Residuals:** The residuals (errors) from the initial prediction are calculated, representing the difference between the actual outcomes and the predicted values.
- **Training Weak Learners:** A new decision tree (weak learner) is trained on the residuals. This tree attempts to capture the patterns in the errors made by the previous predictions.
- **Updating Predictions:** The predictions are updated by adding the new tree's predictions, scaled by a learning rate (a parameter that controls how much influence the new tree has).
- **Repeating the Process:** Steps 2 to 4 are repeated for a specified number of iterations or until the residuals are minimized to a satisfactory level.
- **Evaluating Performance:** The model's performance is assessed using metrics such as accuracy, precision, and recall, ensuring its effectiveness in predicting mental health disorders.

```
Accuracy: 0.8619 (+/- 0.0284)
Precision: 0.8718 (+/- 0.0291)
Recall: 0.8619 (+/- 0.0284)
F1: 0.8586 (+/- 0.0289)
```

3.4.2 Model Training and Validation Processes

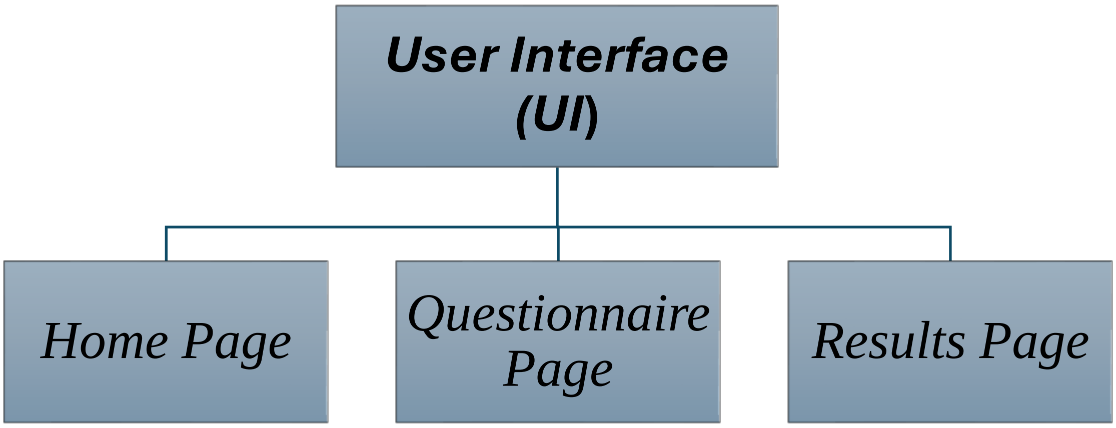
To train our models, we divided the dataset into training and testing sets using a stratified random sampling approach. The training set allowed us to teach the models, while the testing set helped us validate their performance. We employed cross-validation techniques, such as k-fold cross-validation, to ensure our models were robust and less prone to overfitting.

```
Logistic Regression Cross-Validation Scores: [0.81142857 0.85142857 0.84571429 0.74857143 0.76      ]
Mean Cross-Validation Score for Logistic Regression: 0.8034
Decision Tree Cross-Validation Scores: [0.74285714 0.72       0.68571429 0.67428571 0.72      ]
Mean Cross-Validation Score for Decision Tree: 0.7086
Random Forest Cross-Validation Scores: [0.79428571 0.79428571 0.81142857 0.78857143 0.77142857]
Mean Cross-Validation Score for Random Forest: 0.7920
Gradient Boosting Cross-Validation Scores: [0.84571429 0.88       0.82857143 0.79428571 0.82857143]
Mean Cross-Validation Score for Gradient Boosting: 0.8354
```

CHAPTER 4

System Design

- **User Interface (UI)**



1. Home Page: Upon entering the system, users can see Home Page that provides an overview of the application. A prominent button labelled "Predict Your Mental Health" invites users to begin their assessment journey. Clicking this button directs them to the Questionnaire Page.

2. Questionnaire Page: The Questionnaire Page presents users with a series of carefully curated questions related to mental health. Each question comes with multiple-choice options for users to select their answers easily. As users navigate through the questions, they can take their time to consider their responses. Once all questions are answered, users are given the opportunity to review their answers before final submission. This feature ensures that they can make adjustments if needed, promoting accurate responses.

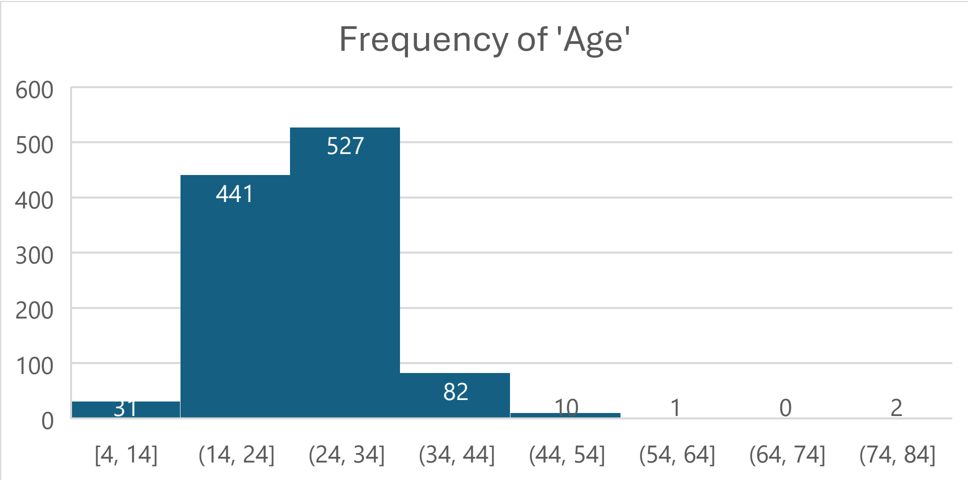
3. Results Page: After completing the questionnaire, users are redirected to the Results Page, where they can view predictions generated by various machine learning models, such as: Logistic Regression, Random Forest, Gradient Boost, the results provide insights into the user's mental health status based on their responses, helping them understand their condition better.

CHAPTER 5

Data Analysis

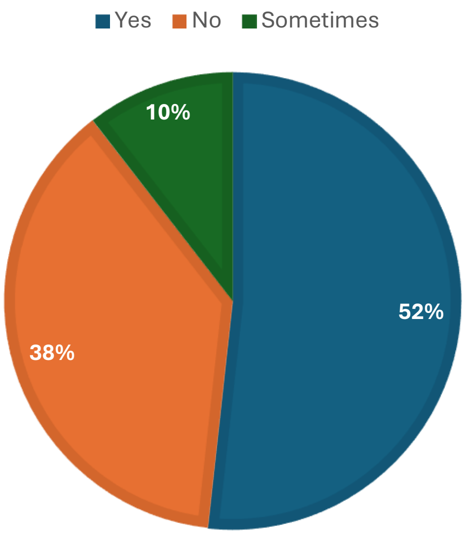
Here we are going to discuss the data set distribution for the attributes one by one starting from Age to the responses of people for all questions.

- **Age**



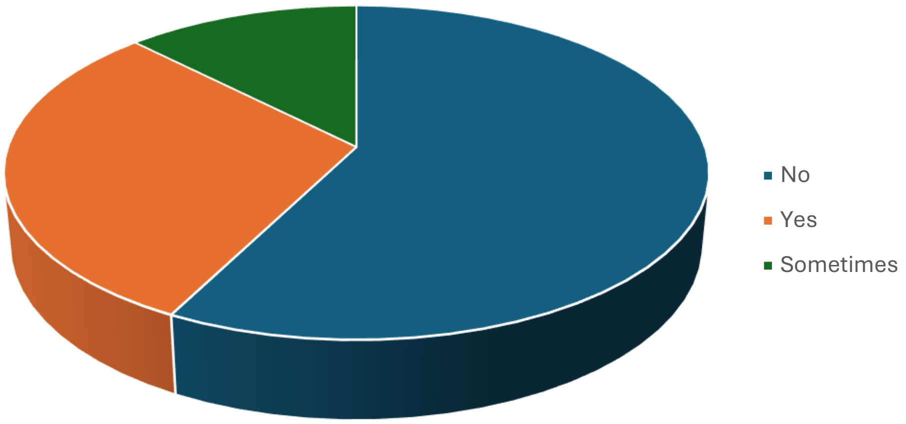
- **Do you feel nervous?**

Nervous	Count of Nervous
Yes	566
No	413
Sometimes	115
Grand Total	**1094**



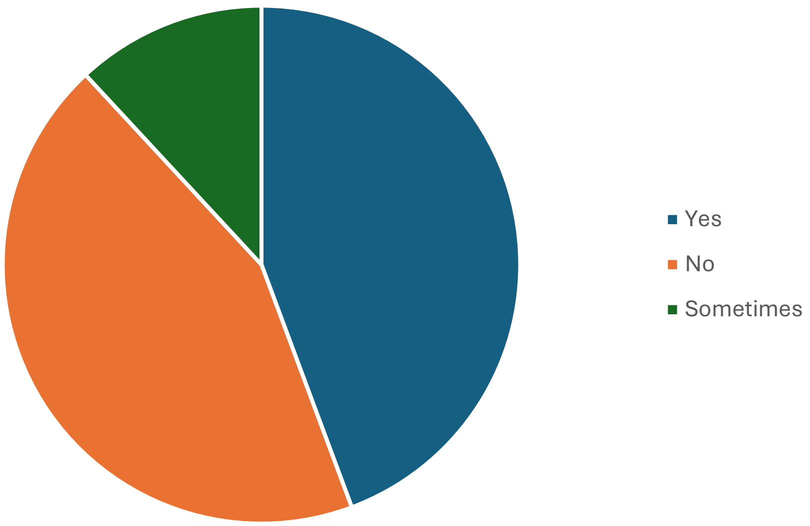
➢ Do you feel panic?

Panic	Count of Panic
No	630
Yes	328
Sometimes	136
Grand Total	**1094**



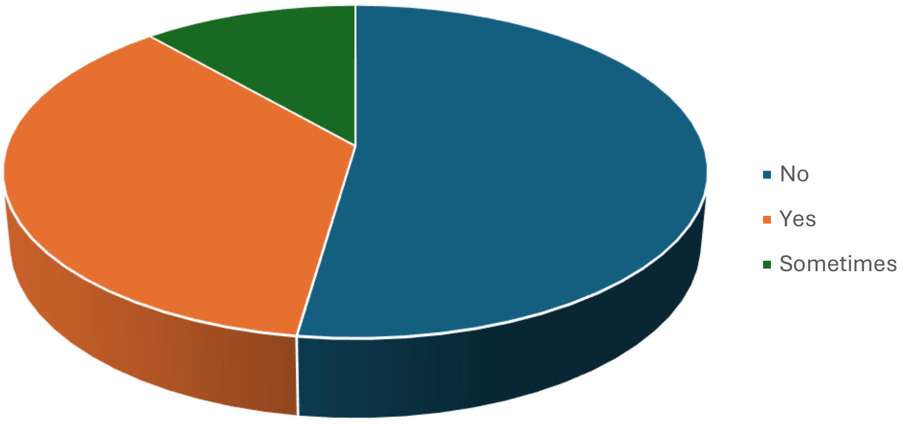
➢ Do you get sudden bursts of rapid breathing?

Rapid breathing	Count of Rapid breathing
Yes	484
No	478
Sometimes	130
Grand Total	**1092**



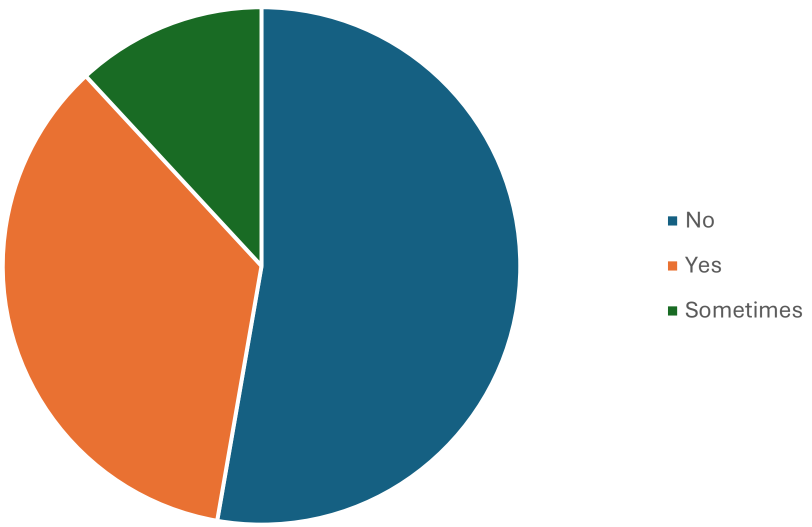
➢ **Do you get sudden bursts of excessive Sweating?**

Excessive Sweat	Count of Excessive Sweat
No	572
Yes	397
Sometimes	125
Grand Total	**1094**



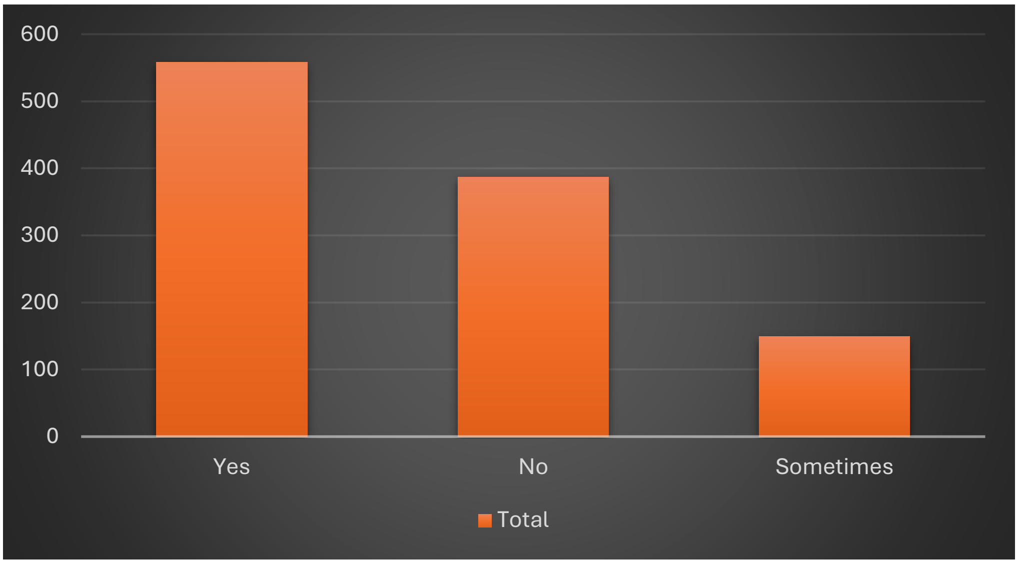
➢ **Do you have trouble focusing on a task?**

Trouble focus on task	Count of Trouble focus on task
No	577
Yes	387
Sometimes	130
Grand Total	**1094**



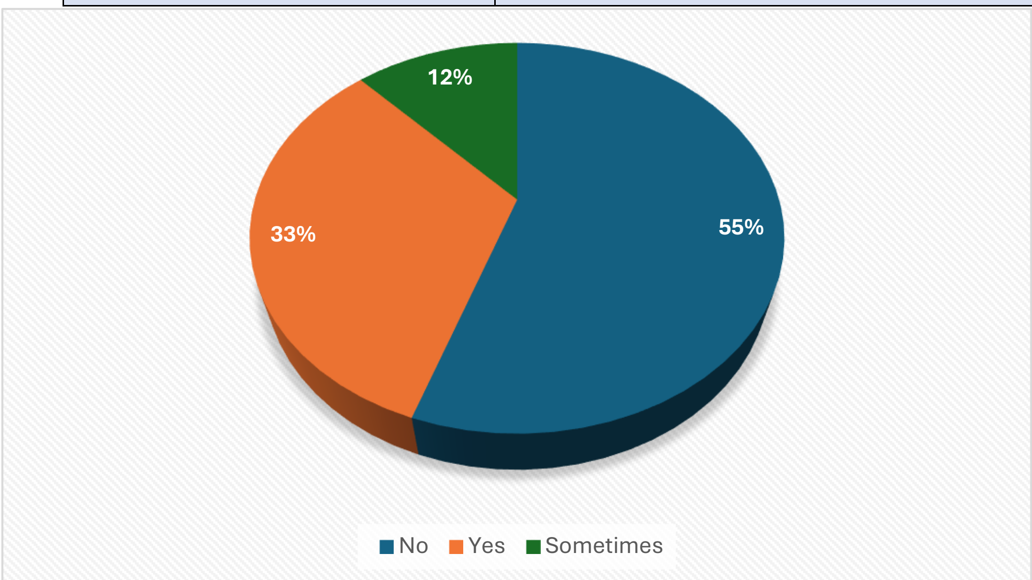
➢ Do you have trouble in sleeping?

Trouble sleeping	Count of Trouble sleeping
Yes	558
No	387
Sometimes	149
Grand Total	**1094**



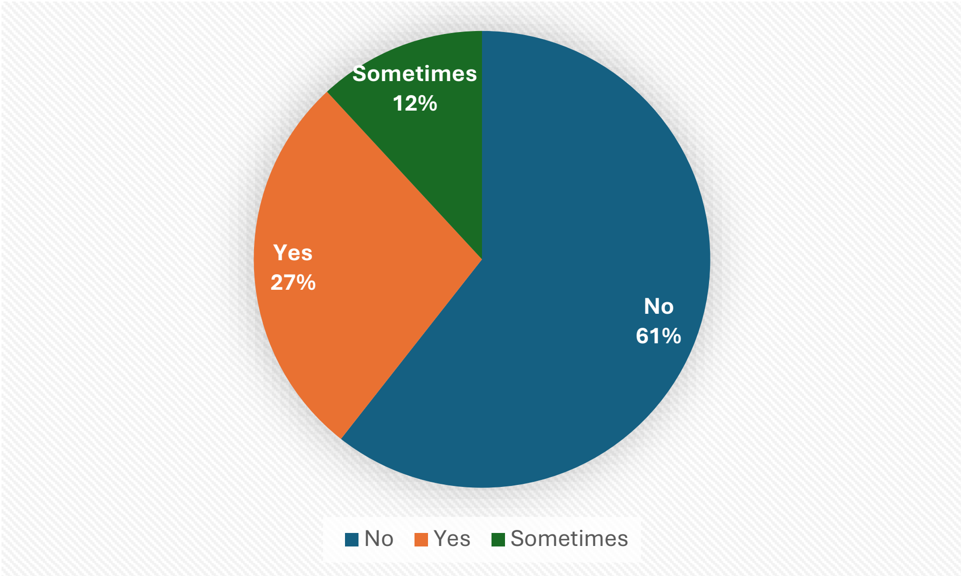
➢ Are you having trouble with work?

Trouble work	Count of Trouble work
No	605
Yes	358
Sometimes	131
Grand Total	**1094**



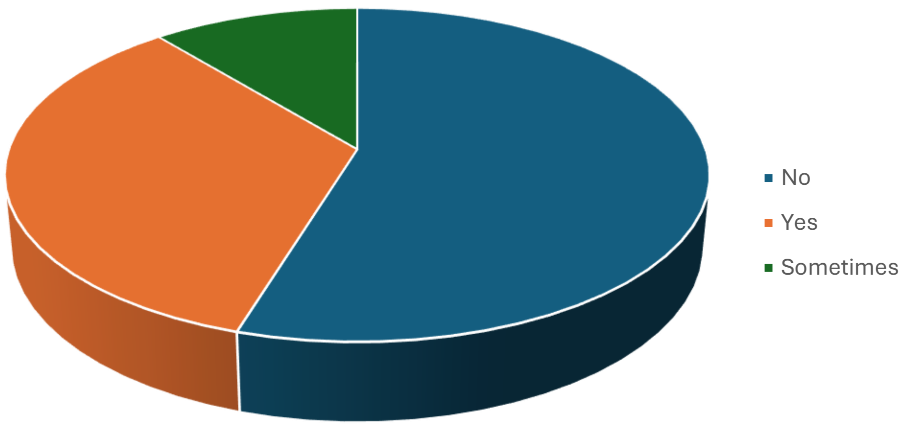
- **Do you feel hopeless?**

Hopeless	Count of Hopeless
No	663
Yes	301
Sometimes	130
Grand Total	**1094**



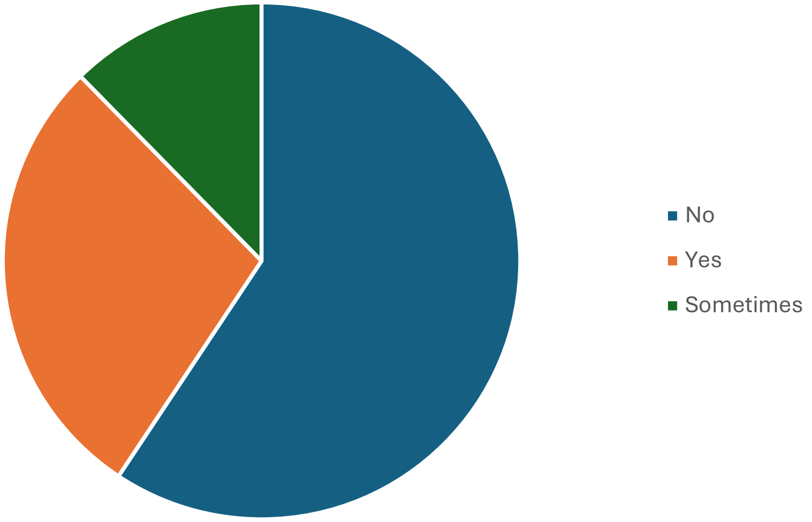
- **Do you feel angry?**

Anger	Count of Anger
No	598
Yes	374
Sometimes	120
Grand Total	**1092**



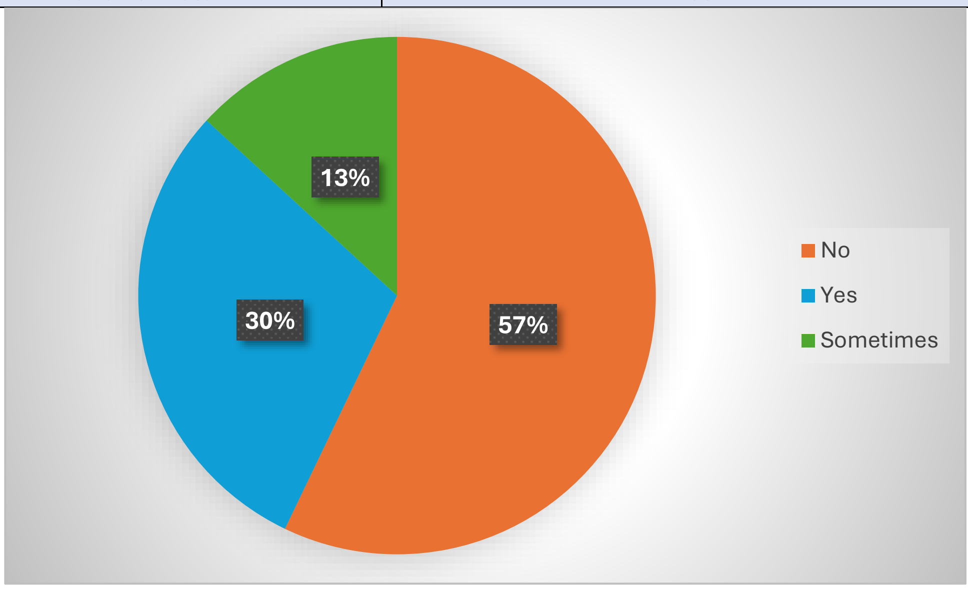
➢ Do you overreact?

Overreact	Count of Overreact
No	649
Yes	310
Sometimes	135
Grand Total	**1094**



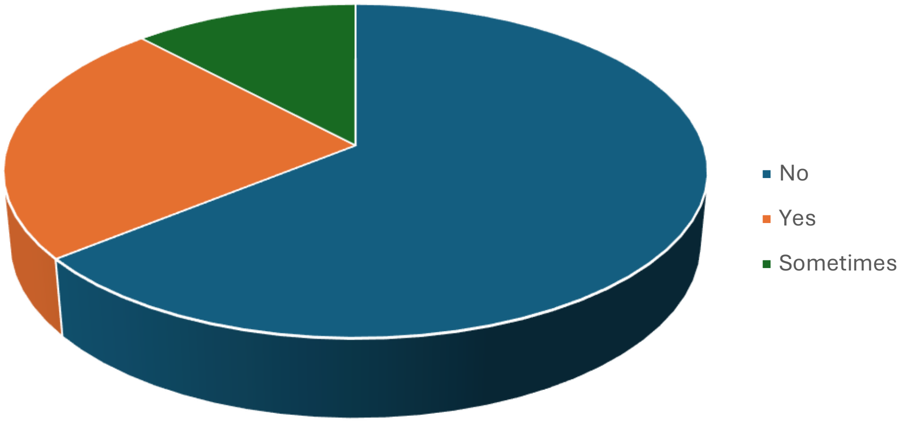
➢ Have you changed your eating pattern?

Eating pattern	Count of Eating pattern
No	625
Yes	325
Sometimes	144
Grand Total	**1094**



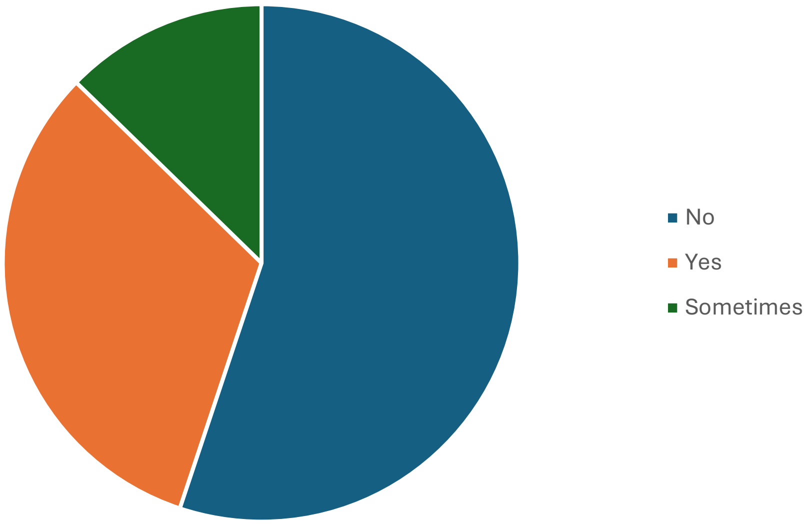
➢ Do you ever get suicidal thoughts?

Suicidal	Count of Suicidal
No	702
Yes	261
Sometimes	131
Grand Total	**1094**



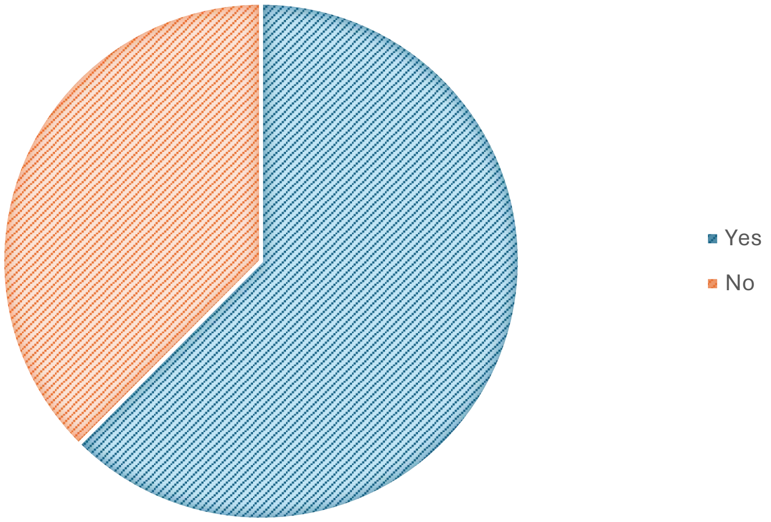
➢ Are you always tired?

Tired	Count of Tired
No	603
Yes	352
Sometimes	139
Grand Total	**1094**



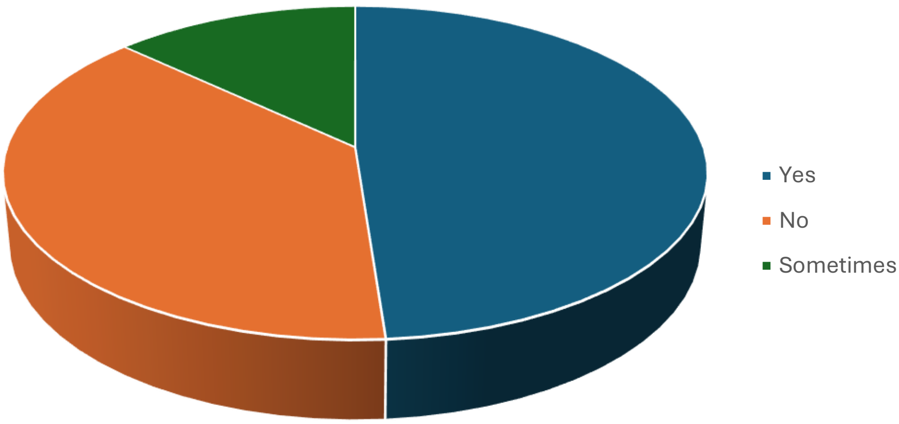
➢ Do you have a close friend(s)?

Close friend	Count of Close friend
Yes	682
No	410
Grand Total	**1092**



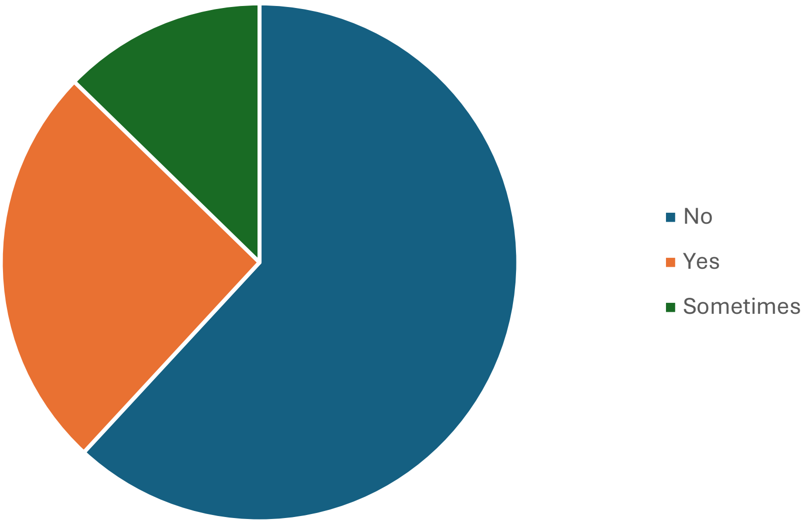
➢ Are you addicted to social media?

Social media	Count of Social media
Yes	534
No	417
Sometimes	143
Grand Total	**1094**



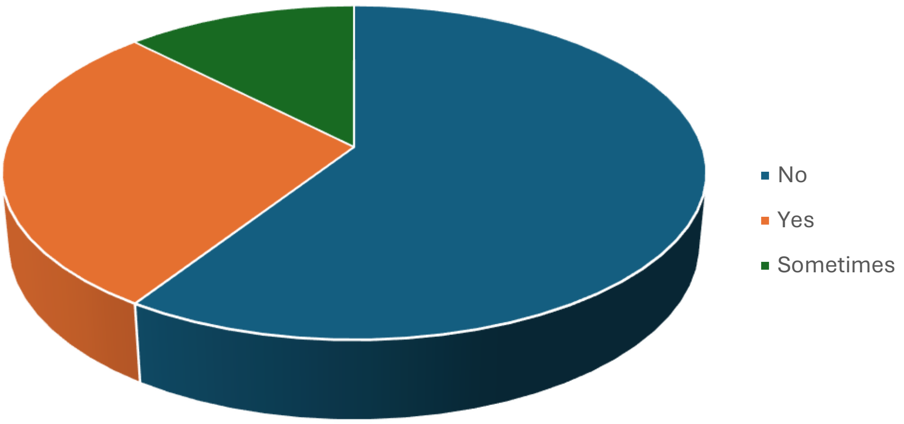
➢ Have you gained weight of recently?

Weight gain	Count of Weight gain
No	677
Yes	278
Sometimes	139
Grand Total	**1094**



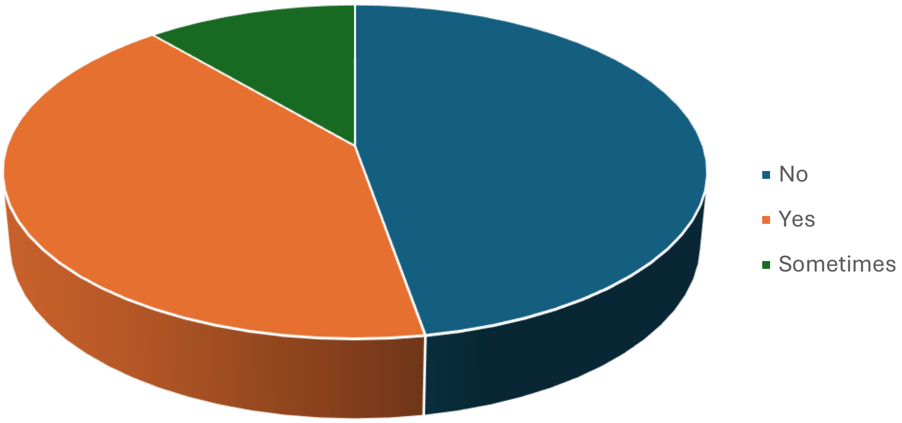
➢ Are you an introvert?

Introvert	Count of Introvert
No	648
Yes	311
Sometimes	135
Grand Total	**1094**



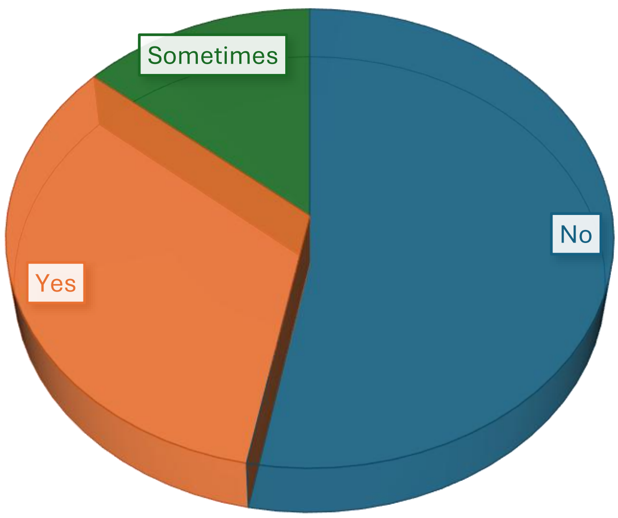
➢ Do you ever get pop ups of a stressful memory?

Stressful memory	Count of Stressful memory
No	517
Yes	454
Sometimes	123
Grand Total	**1094**



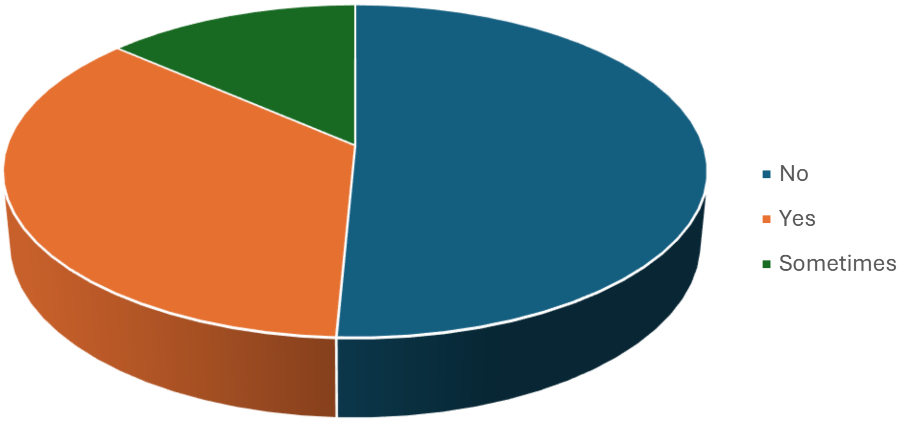
➢ Do you frequently experience nightmares?

Nightmare	Count of Nightmare
No	579
Yes	363
Sometimes	150
Grand Total	**1092**



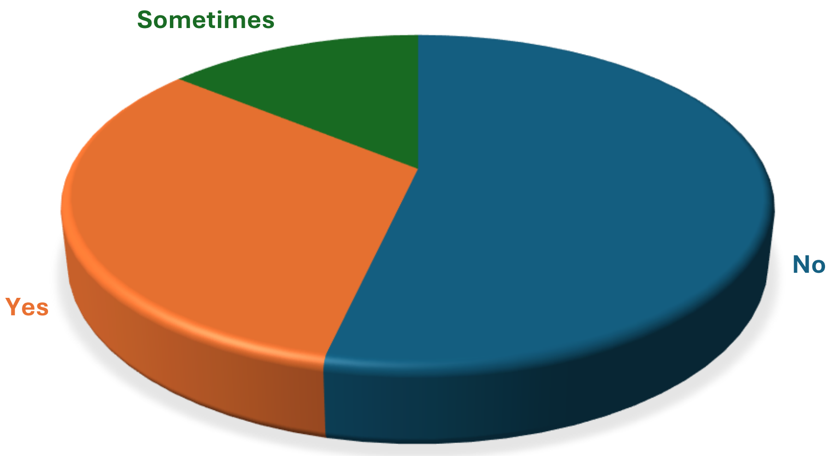
- **Do you often avoid people or activities?**

Avoid	Count of Avoid
No	554
Yes	390
Sometimes	148
Grand Total	**1092**



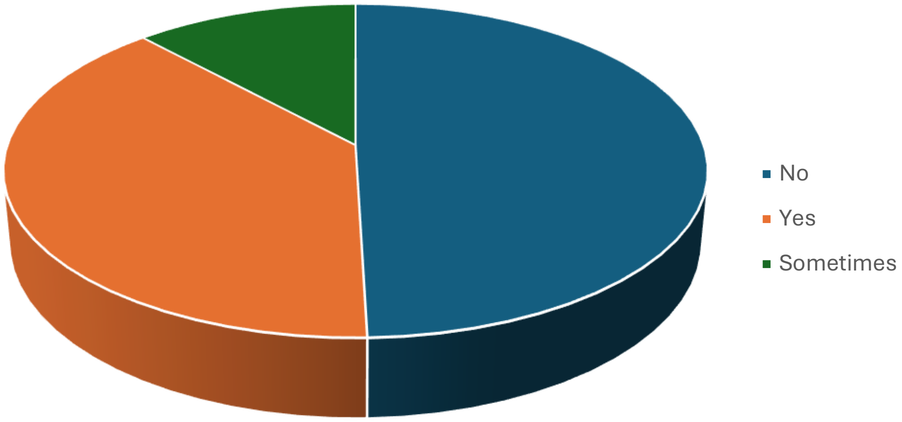
- **Do you often find yourself feeling negative?**

Feeling negative	Count of Feeling negative
No	586
Yes	354
Sometimes	154
Grand Total	**1094**



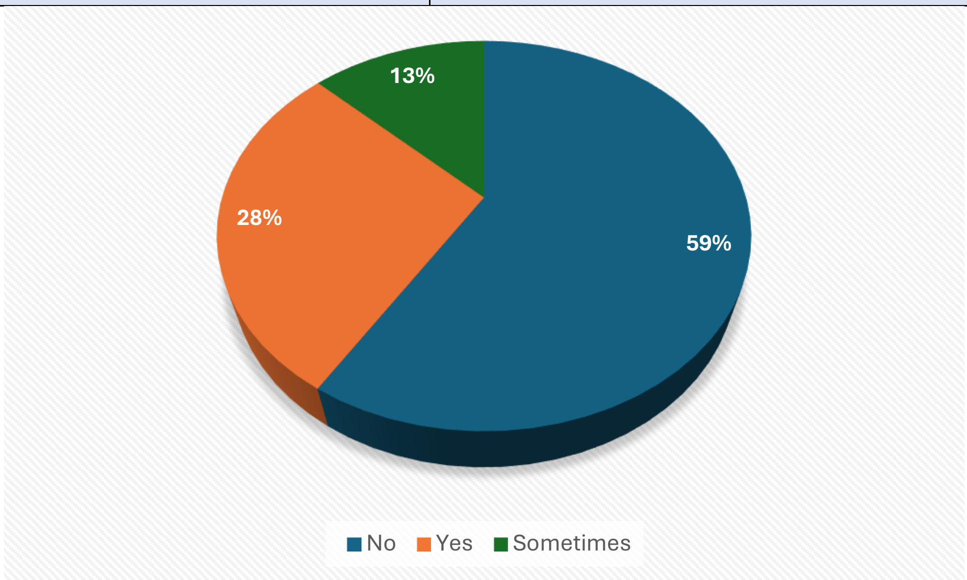
➢ Do you have trouble concentrating

Trouble concentrating	Count of Trouble concentrating
No	542
Yes	422
Sometimes	130
Grand Total	**1094**



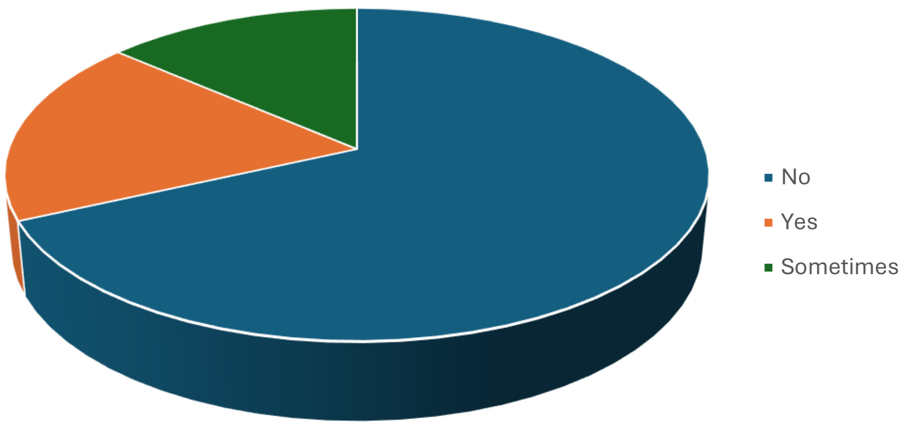
➢ Do you often blame yourself for an unfortunate event?

Blame yourself	Count of Blame yourself
No	644
Yes	310
Sometimes	140
Grand Total	**1094**



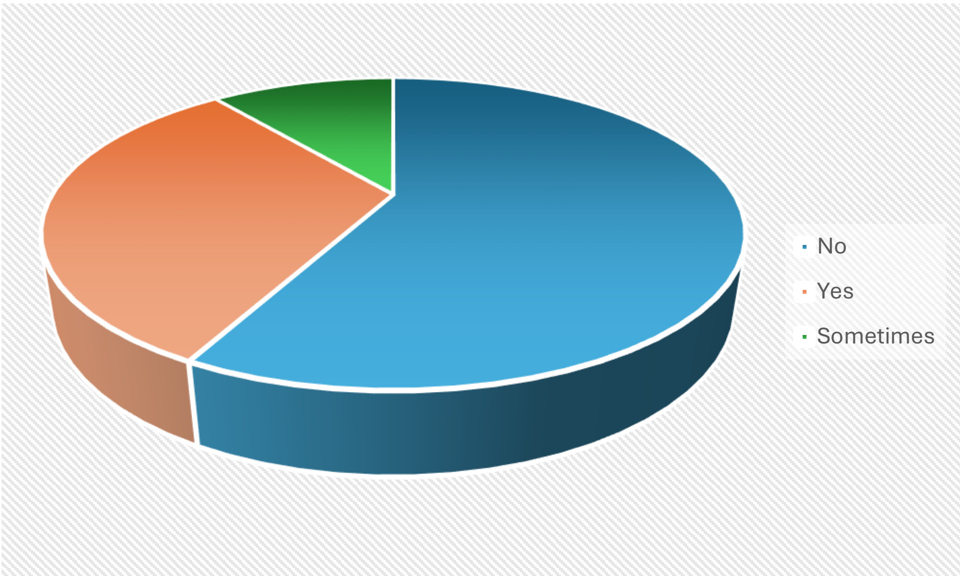
➢ Do you ever get hallucinations?

Hallucinations	Count of Hallucinations
No	746
Yes	199
Sometimes	149
Grand Total	**1094**



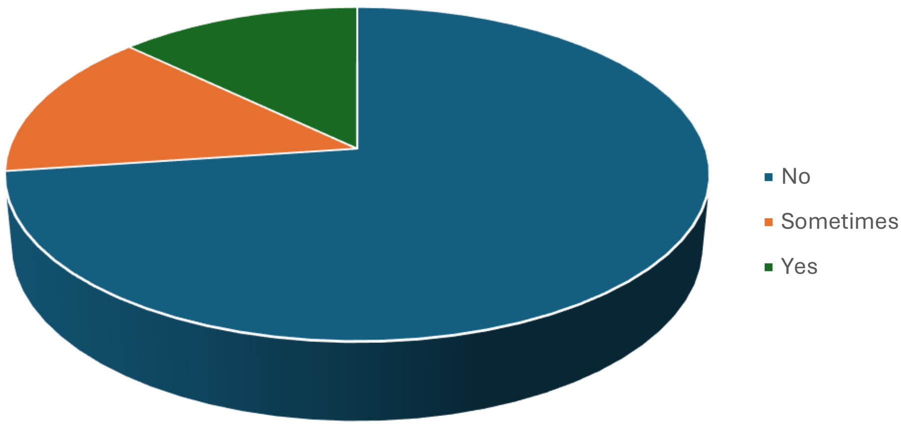
➢ Do you exhibit repetitive behaviour?

Repetitive behaviour	Count of Repetitive behaviour
No	630
Yes	345
Sometimes	117
Grand Total	**1092**



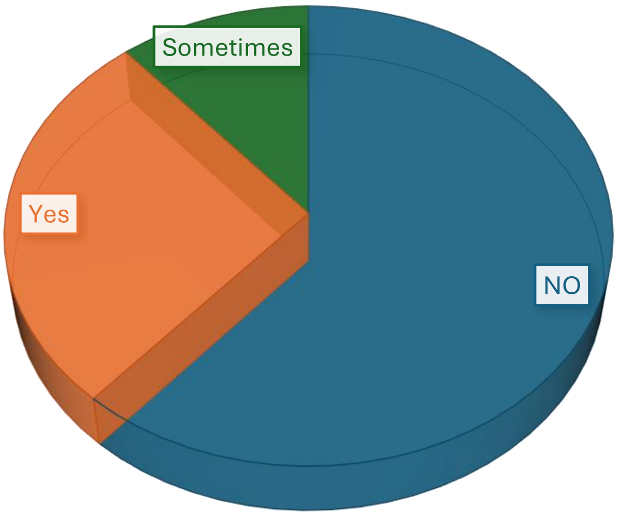
➢ Do you get seasonal depression?

Seasonal depression	Count of Seasonal depression
No	797
Sometimes	156
Yes	141
Grand Total	**1094**



➢ Have you recently experienced an increase in energy?

Increase energy	Count of Increase energy
NO	672
Yes	299
Sometimes	123
Grand Total	**1094**



CHAPTER 6

Evaluation

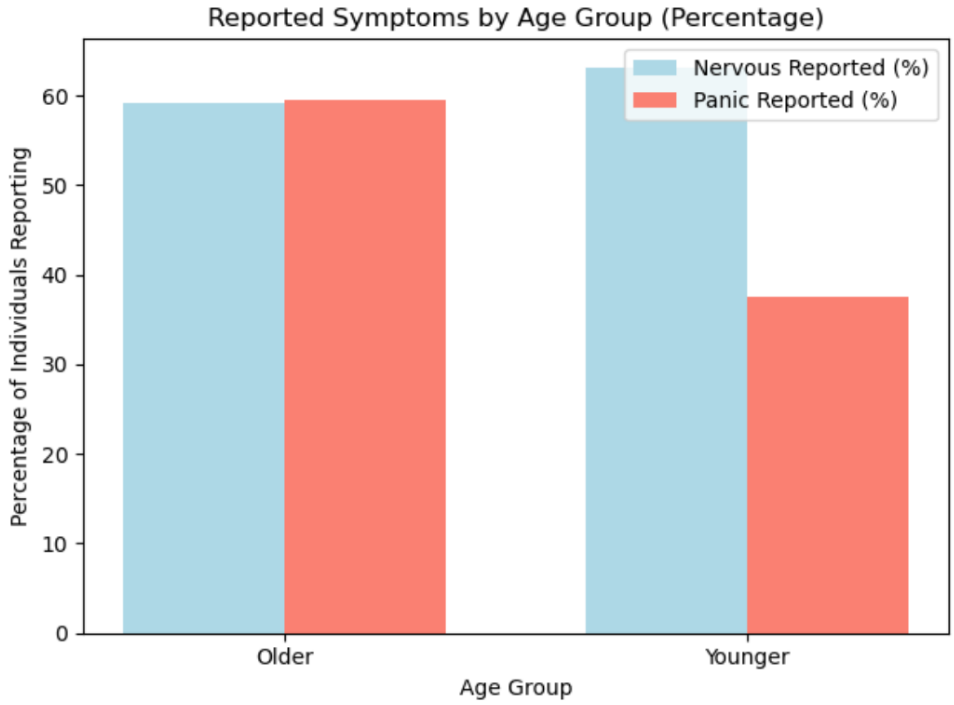
6.1 Performance of the models

Below is the output provided when tested with any random tuple from the data frame. Further we have also provided an option to test the models, with newly provided data.

```
Evaluating Logistic Regression...
Accuracy: 0.8053 (+/- 0.0293)
Precision: 0.8118 (+/- 0.0258)
Recall: 0.8053 (+/- 0.0293)
F1: 0.8011 (+/- 0.0299)


Evaluating Decision Tree...
Accuracy: 0.7202 (+/- 0.0428)
Precision: 0.7214 (+/- 0.0400)
Recall: 0.7202 (+/- 0.0428)
F1: 0.7142 (+/- 0.0429)


Evaluating Random Forest...
Accuracy: 0.7998 (+/- 0.0145)
Precision: 0.8248 (+/- 0.0111)
Recall: 0.7998 (+/- 0.0145)
F1: 0.7893 (+/- 0.0119)


Evaluating Gradient Boosting...
Accuracy: 0.8619 (+/- 0.0284)
Precision: 0.8718 (+/- 0.0291)
Recall: 0.8619 (+/- 0.0284)
F1: 0.8586 (+/- 0.0289)
```

6.2 Hypothesis Statement and analysis results

- **Hypothesis 1:** Younger individuals are more likely to report feeling nervous or experience panic attacks compared to older individuals.



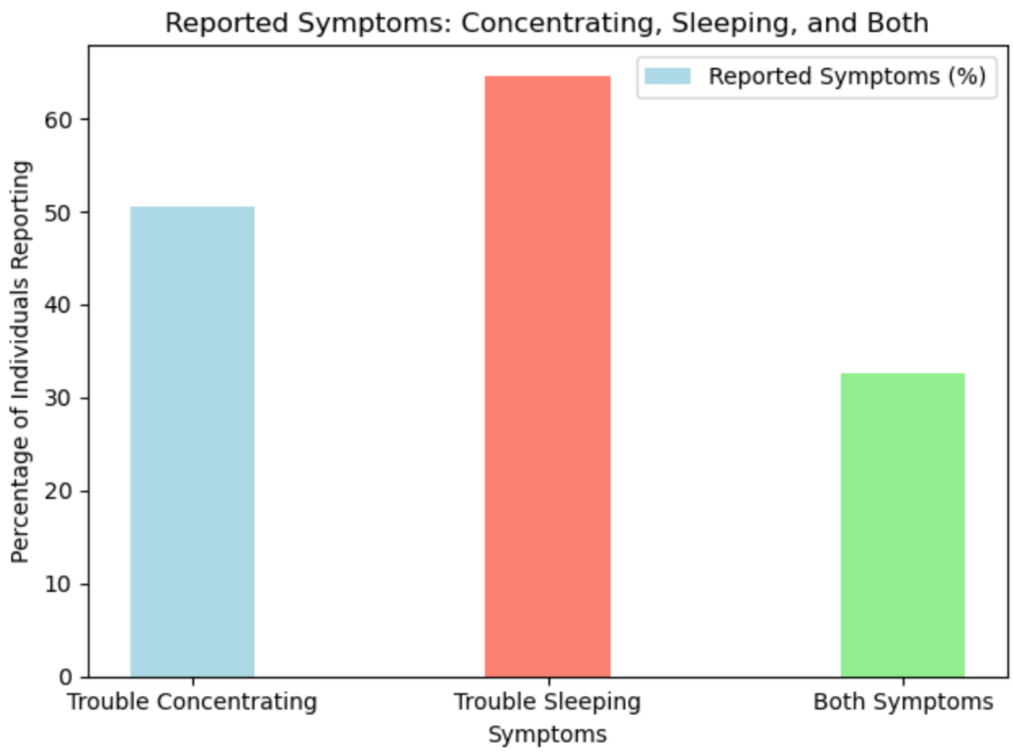
The results indicate that a significant proportion of both age groups reported symptoms of anxiety, but there are notable differences:

Nervous Feelings: A higher percentage of younger individuals (63.15%) reported feeling nervous compared to older individuals (59.09%). This suggests that younger individuals are more likely to experience nervous feelings than their older counterparts.

Panic Attacks: Conversely, the reporting of panic attacks was slightly higher among older individuals (59.50%) than younger individuals (37.56%). This finding indicates that older individuals may experience panic attacks more frequently than younger individuals.

The analysis reveals that younger individuals are more likely to report feeling nervous, while older individuals have a higher tendency to report experiencing panic attacks. These findings underscore the importance of targeted mental health interventions tailored to the specific needs of different age groups.

- **Hypothesis 2:** Individuals who have trouble concentrating are also likely to report trouble sleeping.



Summary of Symptoms Reporting:

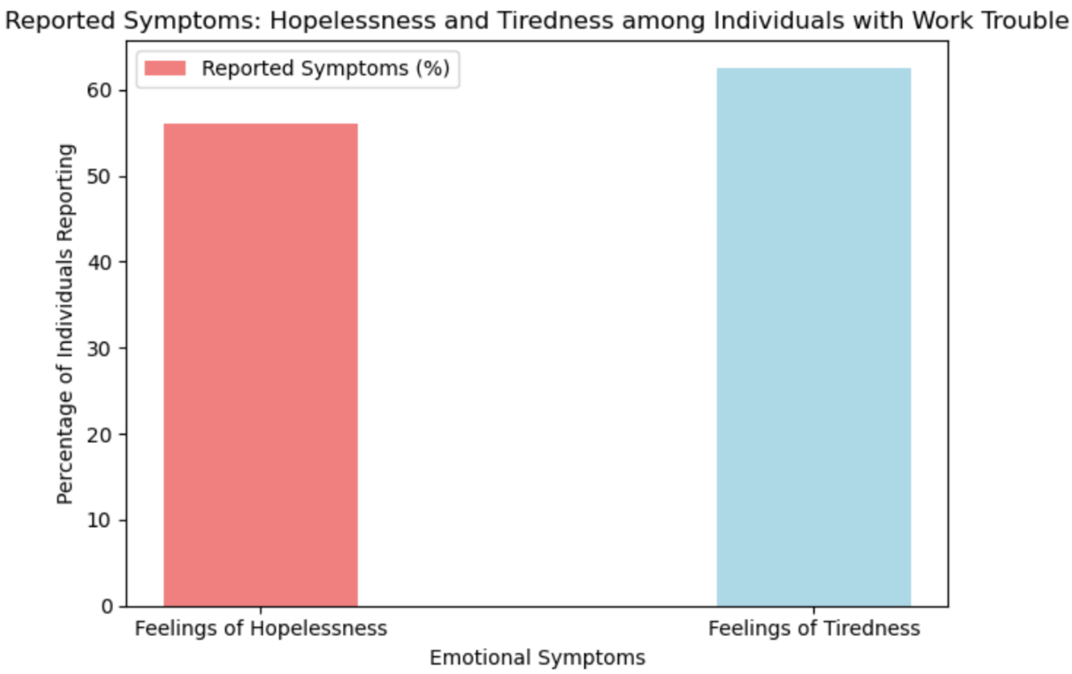
Total Individuals: 1094

Individuals reporting both trouble concentrating and sleeping: 356 (32.54%)

The results indicate that a significant proportion of individuals reported experiencing both trouble concentrating and trouble sleeping, highlighting a noteworthy relationship between these symptoms.

Trouble Concentrating and Sleeping: Among the total participants, 32.54% reported experiencing both trouble concentrating and trouble sleeping. This suggests that a considerable number of individuals facing concentration challenges are also likely to struggle with sleep issues.

➢ **Hypothesis 3:** Individuals who are having trouble with work are more likely to feel hopeless and tired.



Summary of Symptoms Reporting:

Total Individuals: 1094

Individuals reporting trouble with work: 489

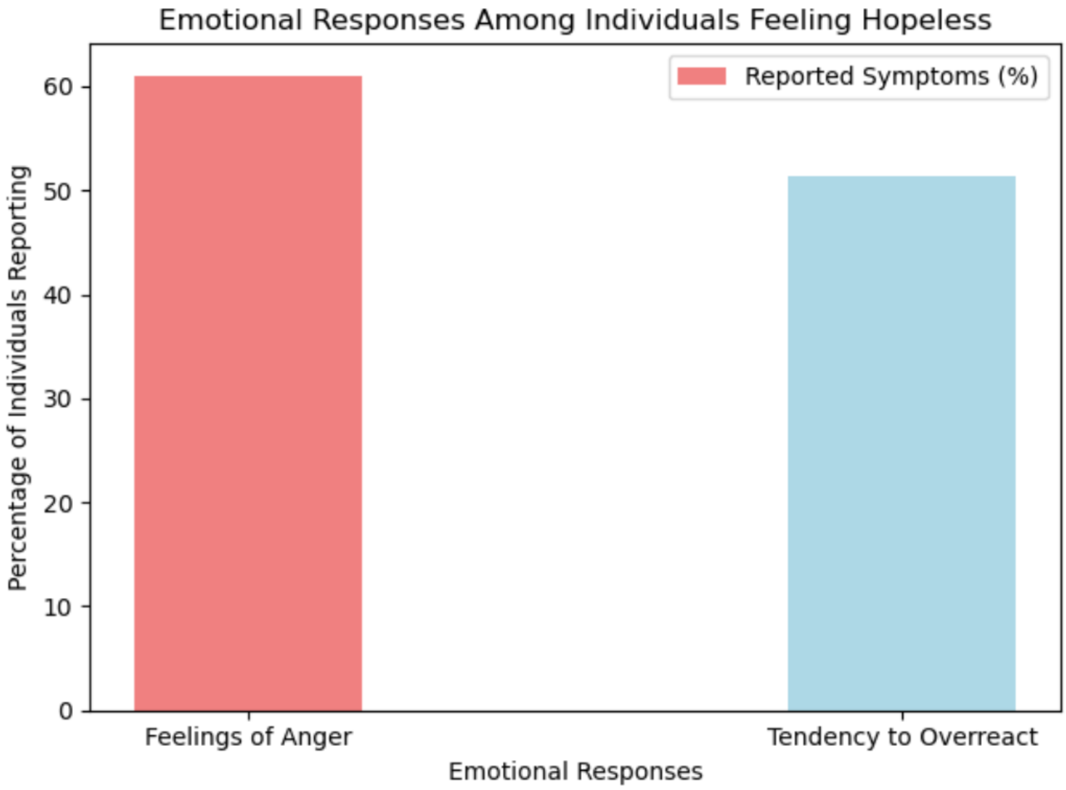
Among those, individuals feeling hopeless: 274 (56.03%)

Among those, individuals feeling tired: 306 (62.58%)

The findings indicate that a significant proportion of individuals who reported difficulties with work also experience feelings of hopelessness and tiredness. Specifically, over half (56.03%) of those with work troubles reported feeling hopeless, while an even higher percentage (62.58%) reported feeling tired.

Given these results, we can accept the hypothesis. The data strongly supports the assertion that individuals having trouble with work are indeed more likely to feel hopeless and tired.

- **Hypothesis 4:** People who feel hopeless are more prone to feelings of anger and overreacting in situations.



Summary of Symptoms Reporting:

Total Individuals: 1094

Individuals reporting feelings of hopelessness: 431

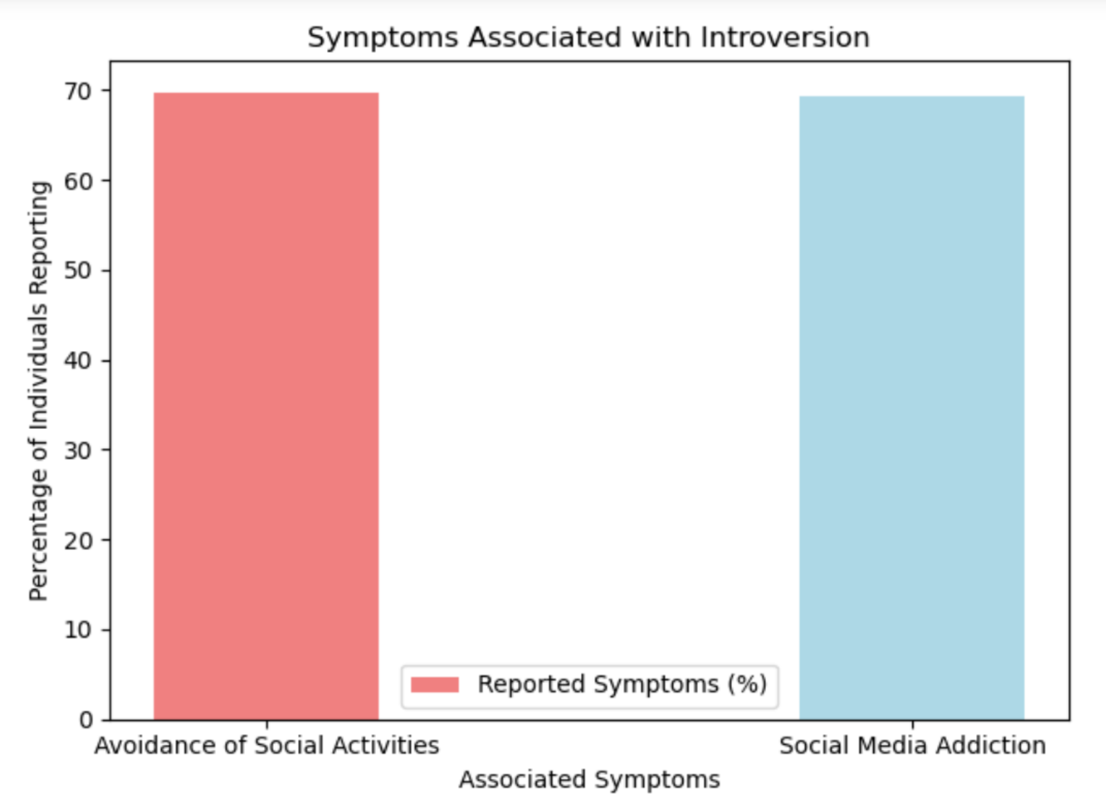
Among those, individuals feeling angry: 263 (61.02%)

Among those, individuals who tend to overreact: 221 (51.28%)

The findings strongly support the hypothesis that individuals experiencing feelings of hopelessness are indeed more likely to report feelings of anger and a tendency to overreact. Specifically: A significant majority (61.02%) of those feeling hopeless also reported feelings of anger. and more than half (51.28%) of individuals feeling hopeless indicated that they tend to overreact in situations.

Given these results, we can confidently accept the hypothesis. The data clearly illustrates the connection between hopelessness and elevated emotional responses such as anger and overreaction.

➢ **Hypothesis 5:** Introverts are more likely to avoid social activities and report addicted to social media.



Summary of Symptoms Reporting:

Total Individuals: 1094

Individuals identified as introverts: 446

Among those, individuals avoiding social activities: 311 (69.73%)

Among those, individuals reporting social media addiction: 309 (69.28%)

The findings strongly support the hypothesis that introverts are indeed more likely to avoid social activities and report addiction to social media. Specifically: A significant majority (69.73%) of introverts reported avoiding social activities. almost as many (69.28%) reported being addicted to social media.

Given these results, we can confidently accept the hypothesis. The data illustrates a clear association between introversion and both the avoidance of social activities and a tendency to engage in social media addiction.

CHAPTER 7

IMPLEMENTATION

7.1 Code Snippets

```
from flask import Flask, request, render_template, redirect, url_for, session
import joblib

app = Flask(__name__)
app.secret_key = 'your_secret_key'

model1 = joblib.load('model/gradient_boosting_model.pkl')
model2 = joblib.load('model/random_forest_model.pkl')
model3 = joblib.load('model/logistic_regression_model.pkl')

questions = [
    {"question": "How old are you?", "type": "number"},
    {"question": "Do you feel often nervous?", "type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you feel often panic?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you get sudden bursts of rapid breathing?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you get sudden bursts of excessive Sweating?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you have trouble focusing on a task?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you have trouble in sleeping?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Are you having trouble with work?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you feel hopeless?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you feel angry?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Do you overeact?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Have you changed your eating pattern?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]},
    {"question": "Have you recently experience an increase in energy?","type": "multiple_choice","options": ["Yes", "Sometimes", "No"]}
]
```

```python
@app.route('/')
def home():
    session.clear()
    return render_template('home.html')

@app.route('/questionnaire', methods=['GET', 'POST'])
def questionnaire():
    if 'question_index' not in session:
        session['question_index'] = 0
    if 'answers' not in session:
        session['answers'] = []

    question_index = session['question_index']

    if request.method == 'POST':
        # Check if the user clicked a "Back" button
        if 'back' in request.form:
            # Clear the current answer if going back
            if session['answers']:
                session['answers'].pop()
            session['question_index'] = max(0, session['question_index'] - 1)
        else:
            # Save the answer and move to the next question
            answer = request.form.get('answer')
            session['answers'].append(answer)
            session['question_index'] += 1
        return redirect(url_for('questionnaire'))

    if question_index < len(questions):
        question = questions[question_index]
        progress_percentage = (question_index / len(questions)) * 100
        return render_template('question.html', question=question,
progress_percentage=progress_percentage)
    else:
        return redirect(url_for('submit'))


@app.route('/submit', methods=['GET', 'POST'])
def submit():
    if request.method == 'POST':
        if 'back' in request.form:
            session['answers'].pop()
            session['question_index'] -= 1
            return redirect(url_for('questionnaire'))
        # Directly use Model 1 for prediction
        return redirect(url_for('predict'))
```

```
    answers = session.get('answers', [])
    return render_template('submit.html',questions=questions, answers=answers)

@app.route('/predict', methods=['GET', 'POST'])
def predict():
    models = ['Gradient Boost', 'Random Forest', 'Logistic Regression']  # Ensure
all models are listed

    # Determine the selected model
    selected_model = request.form.get('model') if request.method == 'POST' else
request.args.get('model')

    answers = session.get('answers', [])
    if not answers:
        return "No answers found in session."

    try:
        age = int(answers[0])
        processed_answers = [age] + [2 if ans == 'Yes' else 1 if ans == 'Sometimes'
else 0 for ans in answers[1:]]
    except (ValueError, IndexError) as e:
        return f"Error processing answers: {e}"

    # Select the correct prediction model
    if selected_model == 'Gradient Boost':
        prediction = model1.predict([processed_answers])[0]
    elif selected_model == 'Random Forest':
        prediction = model2.predict([processed_answers])[0]
    else:
        prediction = model3.predict([processed_answers])[0]

    return render_template('result.html', prediction=prediction,
answers=processed_answers, models=models, selected_model=selected_model)



if __name__ == '__main__':
    app.run(debug=True)
```

Home page:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Mental Health Prediction</title>
</head>
 <style>
       body {
          display: flex;
          justify-content: center;
          align-items: center;
          height: 100vh;
          margin: 0;
          font-family: Arial, sans-serif;
          background-image: url('static/photo.png');
          background-size: cover;
          background-position: center;
       }
</style>
<body>
   <div class="container">
       <div class="title">Mental Health Prediction</div>
       <a href="{{ url_for('questionnaire') }}" class="btn">Predict your mental
health</a>
   </div>
</body>
</html>
```

figure 7.1: Home Page

Questionnaire page:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <title>Question</title>
</head>
<body>
   <div class="progress-bar-container">
      <div class="progress-bar">
         <div class="progress"></div>
      </div>
   </div>
   <form method="POST" action="">
      {% if question['type'] == 'multiple_choice' %}
         <div class="options-container">
            {% for option in question['options'] %}
               <button type="submit" name="answer" value="{{ option }}"
class="btn">{{ option }}</button>
            {% endfor %}
         </div>
      {% endif %}
      {% if session['question_index'] == questions|length - 1 %}
         <button type="submit" class="predict-button">Predict</button>
      {% endif %}
   </form>
</body>
</html>
```

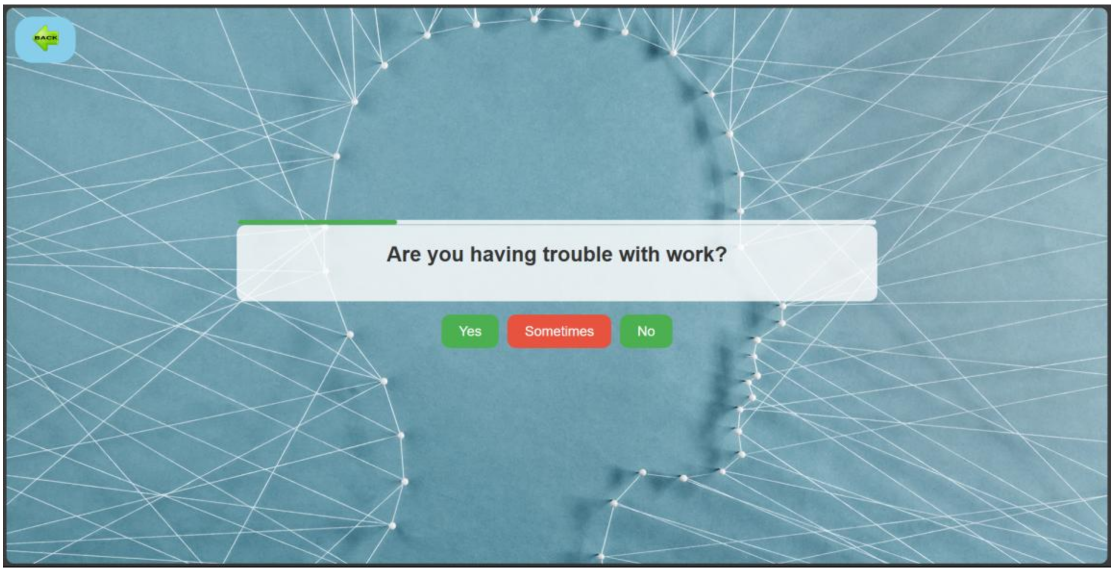


figure 7.2: Questionnaire page

Result Page:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>Prediction Result</title>
</head>
<body>
   <div class="container">
      <form method="post" action="/predict">
      </form>
      <div class="result-box">
         <h1>Prediction Result</h1>
         <p><strong>{{ prediction }}</strong> </p>
      </div>
      <form action="{{ url_for('home') }}" method="get">
         <button type="submit" class="button">Go Back Home</button>
      </form>
   </div>
</body>
</html>
```

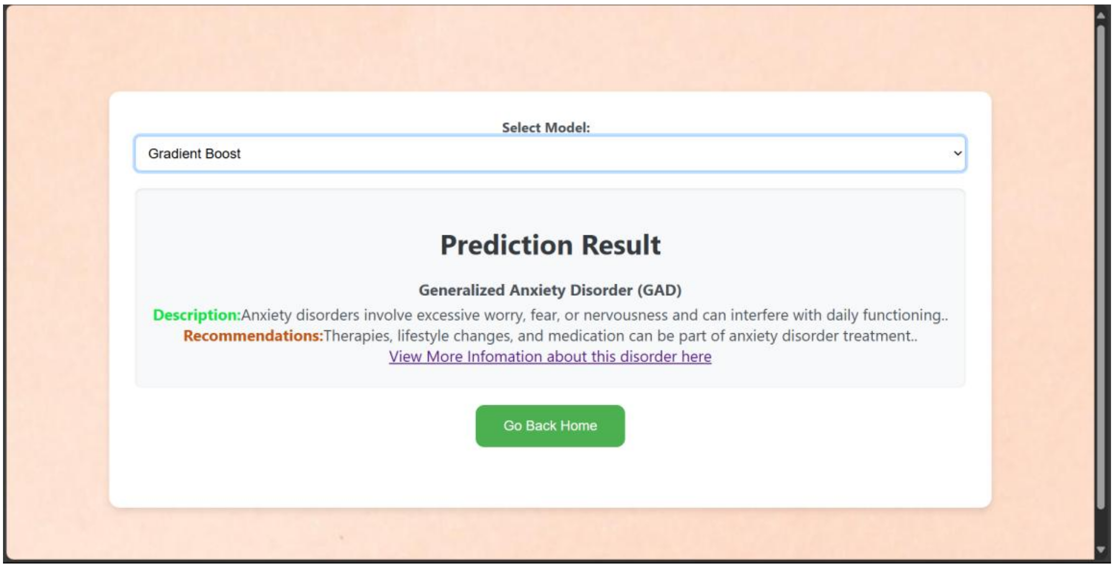


figure 7.3: **Result Page**

CHAPTER 8

CONCLUSION

In our study of mental health, we discovered that various factors influence how individuals experience and manage their well-being. It's clear that everyone, regardless of their personality type, faces challenges that can impact their mental health in different ways.

In wrapping up this project, we've uncovered some important connections between different feelings and behaviours in people. Our study found that younger folks often feel nervous, while older people might have panic attacks more often. There's also a strong link between having trouble concentrating and not being able to sleep well. Stress at work seems to make people feel hopeless and tired, and those who feel hopeless are more likely to get angry and overreact. Introverts, on the other hand, tend to avoid social gatherings and might spend a lot of time on social media.

To make use of these findings, we built a simple web page using the Flask framework. It asks users questions about their feelings and behaviours and then gives them feedback on their mental health. This tool aims to help people understand their mental health better and encourage them to seek help if needed.

Ultimately, this project highlights how complicated mental health can be. By turning our research into a practical tool, we hope to make it easier for people to recognize and address their mental health needs. There's still room for improvement, and with more research, this tool could become even more helpful for users and mental health professionals.

CHAPTER 9

Future Enhancements

Looking ahead, there are several exciting opportunities to enhance the web application and improve its effectiveness in predicting and analysing mental health conditions:

1. Integration of More Predictive Models: By incorporating additional machine learning models, we can analyse a wider range of data and improve the accuracy of predictions. This could include models that consider different factors such as lifestyle, social interactions, and historical mental health data, allowing for a more comprehensive assessment.

2. User Interface Improvements: Enhancing the user interface (UI) will make the web application more user-friendly and engaging. This could involve simplifying the layout, using more intuitive navigation, and incorporating visually appealing elements such as charts and graphs to present results. A responsive design would also ensure the application functions well on various devices, including smartphones and tablets.

3. Personalized Feedback and Resources: Providing users with personalized feedback based on their responses can help them understand their mental health better. Including links to relevant resources, such as articles, support groups, and professional services, can empower users to take proactive steps towards their mental well-being.

By implementing these enhancements, aim to create a more robust and supportive tool that meets the evolving needs of users seeking to understand and improve their mental health.

CHAPTER 10

BIBLIOGRAPHY

- Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges by Jetli Chung and Jason Teo from ResearchGate https://www.researchgate.net/publication/357624760_Mental_Health_Prediction_Using_Machine_Learning_Taxonomy_Applications_and_Challenges

- National Institutes of Health https://medlineplus.gov/mentalhealth.html

- World health organization https://www.who.int/health-topics/mental-health#tab=tab_1

- Student Mental Health Survey from Kaggle https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey