Mental Illness Prediction Using ML

**1. INTRODUCTION**

1.1 Project Overview

Mental Health First Aid teaches participants how to notice and support an individual who may be experiencing a mental health or substance use concern or crisis and connect them with the appropriate employee resources.

Employers can offer robust benefit packages to support employees who go through mental health issues. That includes Employee Assistance Programs, Wellness programs that focus on mental and physical health, Health and Disability Insurance or flexible working schedules or time off policies.Organisations that incorporate mental health awareness help to create a healthy and productive work environment that reduces the stigma associated with mental illness, increases the organization’s mental health literacy and teaches the skills to safely and responsibly respond to a co-workers mental health concern.

1.2 Purpose

The main purpose of the Mental Health Prediction system is to predict whether a person needs to seek Mental health treatment or not based on inputs provided by them.

**2. LITERATURE SURVEY**

2.1 Existing problem

The current landscape of workplace mental health poses significant challenges, with many professionals experiencing stress, burnout, and mental health issues. Identifying and addressing these issues is crucial for both the well-being of employees and the overall productivity of organizations. Despite increased awareness, there is a need for a systematic and proactive approach to predict and prevent mental health issues among working professionals.

2.2 References

To develop a comprehensive understanding of mental health challenges in the workplace, this project draws on a variety of reputable sources and studies. Key references include academic research articles, reports from mental health organizations, and relevant publications on the intersection of technology and mental health. By leveraging insights from these authoritative sources, the project aims to build a robust foundation for its predictive model and intervention strategies.

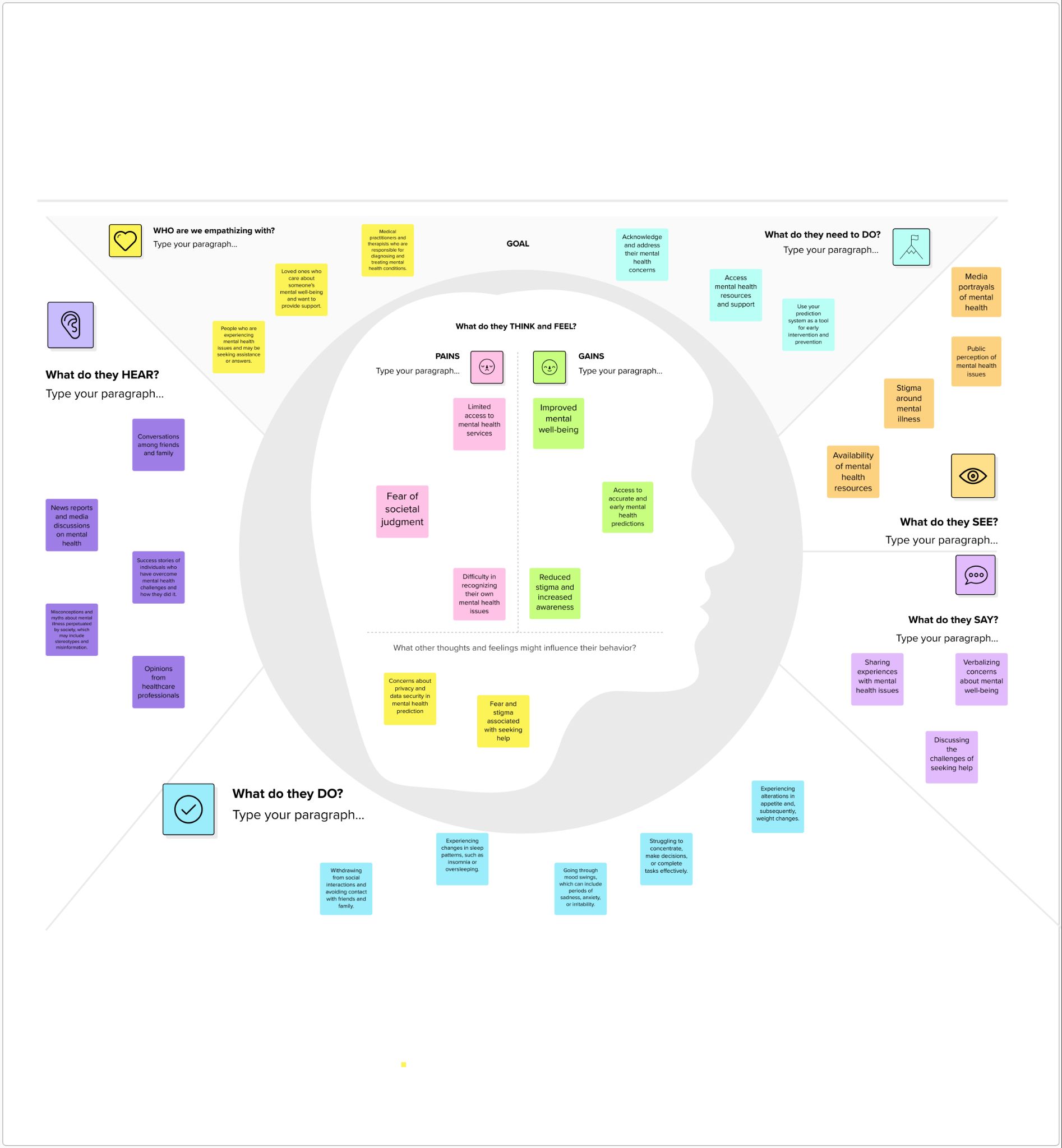
2.3 Problem Statement Definition

The problem at hand is to develop an effective and proactive solution for predicting mental health issues in working professionals. The goal is to create a predictive model that utilizes relevant features and indicators to assess the mental well-being of employees. By doing so, the project aims to identify potential issues early, allowing for timely intervention and support. The defined problem statement serves as a guide for the project's scope, outlining the specific challenges and objectives that the predictive model seeks to address.

**3. IDEATION & PROPOSED SOLUTION**

3.1 Empathy Map Canvas

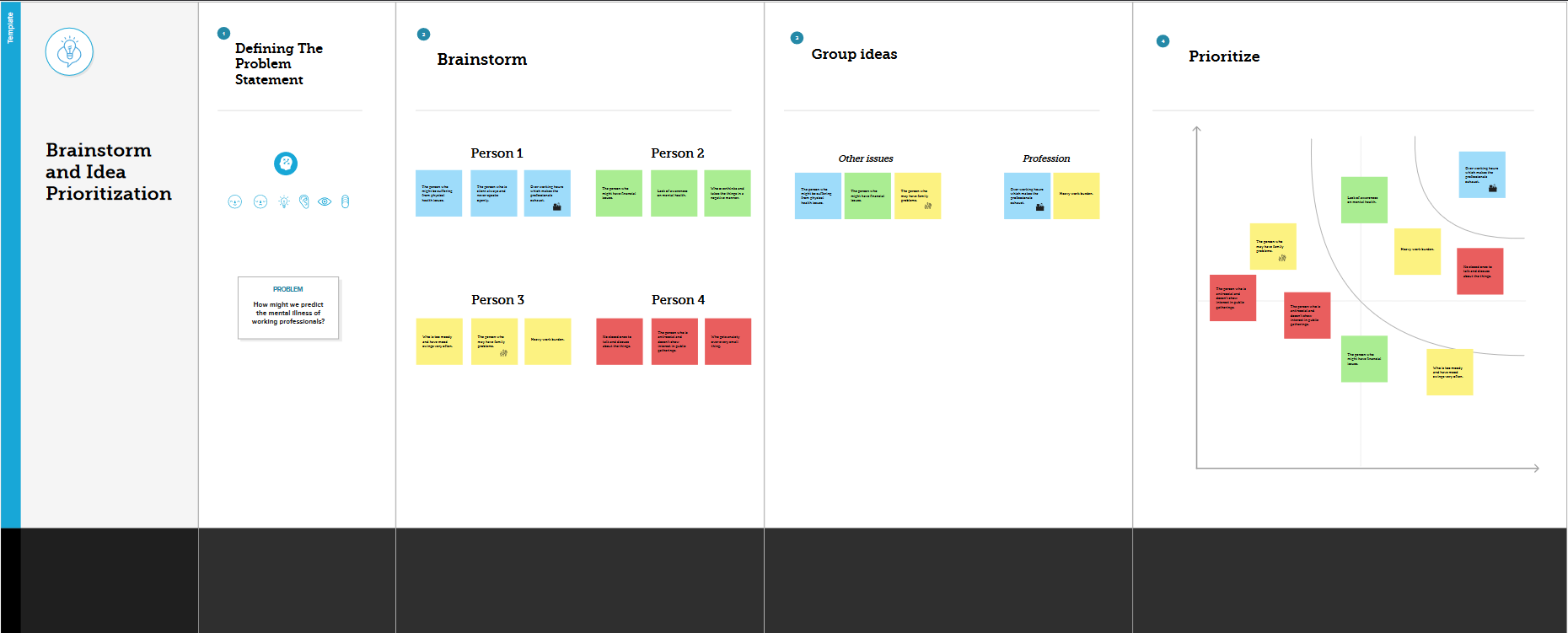
In the context of our project focused on predicting mental illness in working professionals, the Empathy Map Canvas emerges as a valuable tool to delve into the emotional and experiential aspects of our target audience. The Empathy Map serves as a visual framework that guides us in understanding the thoughts, feelings, behaviors, and pain points of the individuals we aim to support. By empathizing with the experiences of working professionals, we can gain deeper insights into their daily challenges, stressors, and triggers. Through the Empathy Map Canvas, we seek to cultivate empathy and align our project with the genuine concerns and experiences of the individuals we aim to assist.



In conclusion, the Empathy Map Canvas serves as a compass in our project, guiding us toward a deeper understanding of the emotional landscape of working professionals. By capturing the nuances of their experiences, fears, and aspirations, the canvas empowers us to develop a more empathetic and user-focused solution for predicting mental illness. As we navigate the complexities of mental health in the workplace, the insights gleaned from the Empathy Map Canvas become a cornerstone for creating a model that genuinely addresses the needs of our users, fostering a culture of well-being and support within professional settings.

3.2 Ideation & Brainstorming

In the ideation and brainstorming phase of our project, creativity takes the spotlight as we explore innovative solutions for predicting mental illness in working professionals. This phase marks a collaborative journey where diverse perspectives converge to generate a spectrum of ideas. By fostering an open and dynamic environment, we aim to unearth novel approaches, drawing inspiration from various domains to shape a robust and effective solution. The ideation and brainstorming process become the catalyst for ideating beyond conventional boundaries, paving the way for a project that is not only technically sound but also genuinely impactful in the realm of workplace mental health.



In concluding our ideation and brainstorming phase, we celebrate the richness of diverse ideas that have emerged. The collaborative energy and creativity harnessed during this phase lay the foundation for a project that embraces innovation. As we transition to the next stages, we carry with us the spirit of exploration and a reservoir of potential solutions.

**4. REQUIREMENT ANALYSIS**

4.1 Functional requirement

1. Mental Health Assessment Form:

- The website should provide a user-friendly form with relevant fields to gather information about the user's well-being, including work-related stressors and mood indicators.

2. Data Input and Submission:

- Users must be able to input their responses into the form and submit it for analysis.

3. Algorithm Integration:

- Implement an algorithm that analyzes the user's responses to predict potential mental health concerns.

4. Display Predicted Results:

- After form submission, display the predicted results based on the algorithm's analysis. This could include an overview of potential mental health issues or stressors.

4.2 Non-Functional requirements

1. User Privacy:

- The system should prioritize user privacy by ensuring that no personally identifiable information is stored in a database or cloud storage. User data should be processed securely and anonymously.

2. Website Performance:

- The website should have minimal load times to provide a smooth and responsive user experience.

3. Usability:

- Design the user interface to be intuitive and accessible, accommodating users with different levels of technological proficiency.

4. Algorithm Accuracy:

- Measure and maintain the accuracy of the prediction algorithm. Regularly assess its performance and make adjustments as needed.

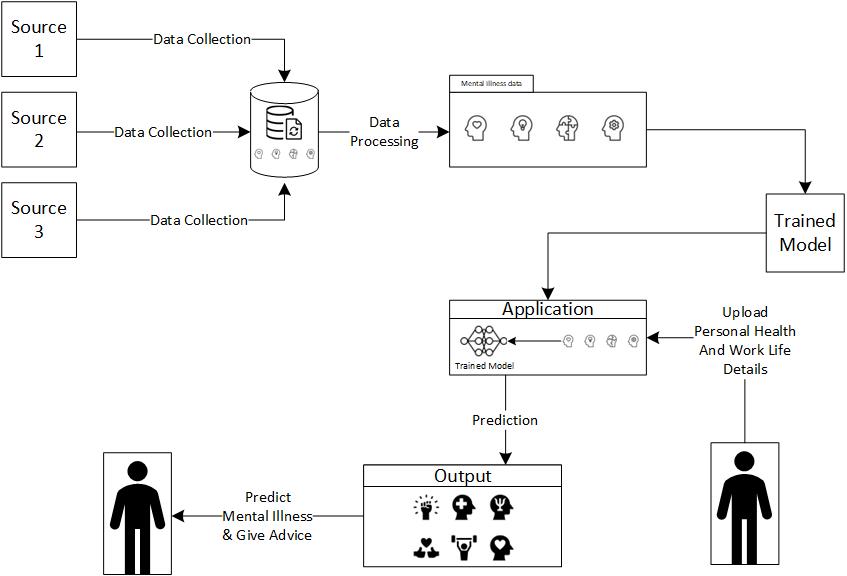
5. Responsiveness:

- Ensure that the website is accessible and functional on various devices, including desktops, tablets, and mobile phones.

**5. PROJECT DESIGN**

5.1 Data Flow Diagrams & User Stories

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



**User Stories:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional Requirement** | **User Story Number** | **User story/Talk** | **Acceptance criteria** | **Priority** | **Release** |
| Working Professionals  (Mobile user) | Project setup &  Infrastructure | USN-1 | Create a version-controlled repository on Git and set up a basic development environment | Properly created a repository in GitHub | High | Sprint-1 |
|  | development  environment | USN-2 | Set up the development environment with the required tools and frameworks to start the prediction of mental illness in working professionals project. | successfully configured  with all necessary tools  and frameworks | High | Sprint-1 |
|  | Data collection | USN-3 | Gather a diverse dataset containing different types of mental illnesses for training the deep  learning model. | Gathered a diverse  dataset  depicting various types  of mental illnesses. | High | Sprint-2 |
|  | Data Preprocessing | USN-4 | Preprocess the collected dataset by data cleaning, handling imbalances and splitting it into training and validation sets. | Preprocessed the  dataset | High | Sprint-2 |
|  | Model Development | USN-5 | Explore and evaluate different deep learning architectures (e.g., CNNs)  to select the most suitable model for mental illness prediction. | We could explore  various DL models | High | Sprint-3 |
|  | Training | USN-6 | train the selected deep learning model using the preprocessed dataset  and monitor its performance on the validation set. | We could do validation | Medium | Sprint-3 |
|  | Model Deployment &  Integration | USN-7 | deploy the trained deep learning model as a web service | We could check the  scalability | Medium | Sprint-4 |
|  | Testing & quality  assurance | USN-8 | fine-tune the model  hyperparameters and optimize its performance | We could create web application | Medium | Sprint-5 |

5.2 Solution Architecture

It assists in predicting the mental illness of working professionals by leveraging Classification algorithms for real time data prediction. It enhances how a working professional feels about their mental health, it improves the efficiency of working professionals by inculcating the importance of being healthy mentally in their working profession as it also affects their personal life as well.

Not only predicting the mental illness of a working professional, it suggests some general tips that might help the working professionals tackling their mental health in a basic way.

● Data Gathering

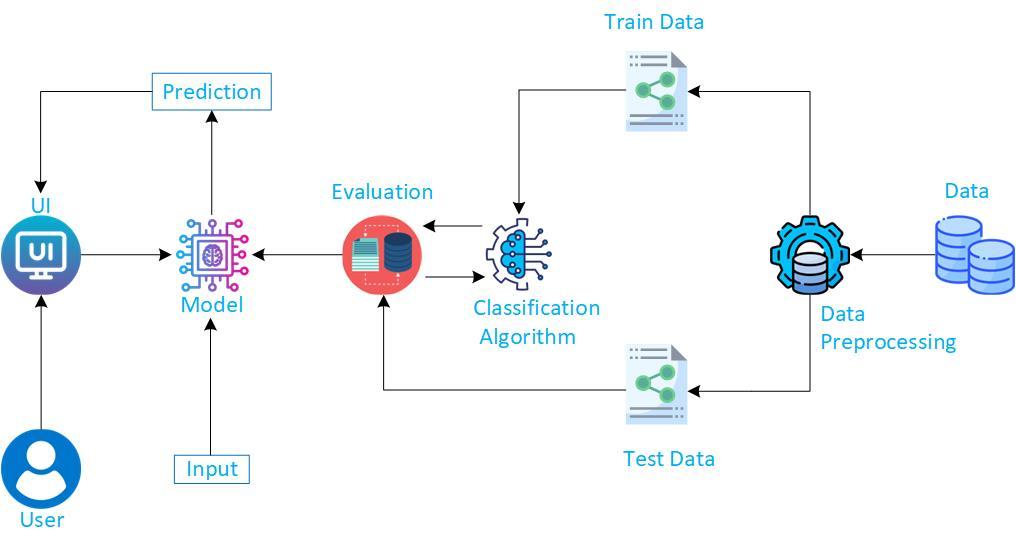
● Data Preprocessing

● Model Building

● Mental Illness Prediction

● Real Time Analysis

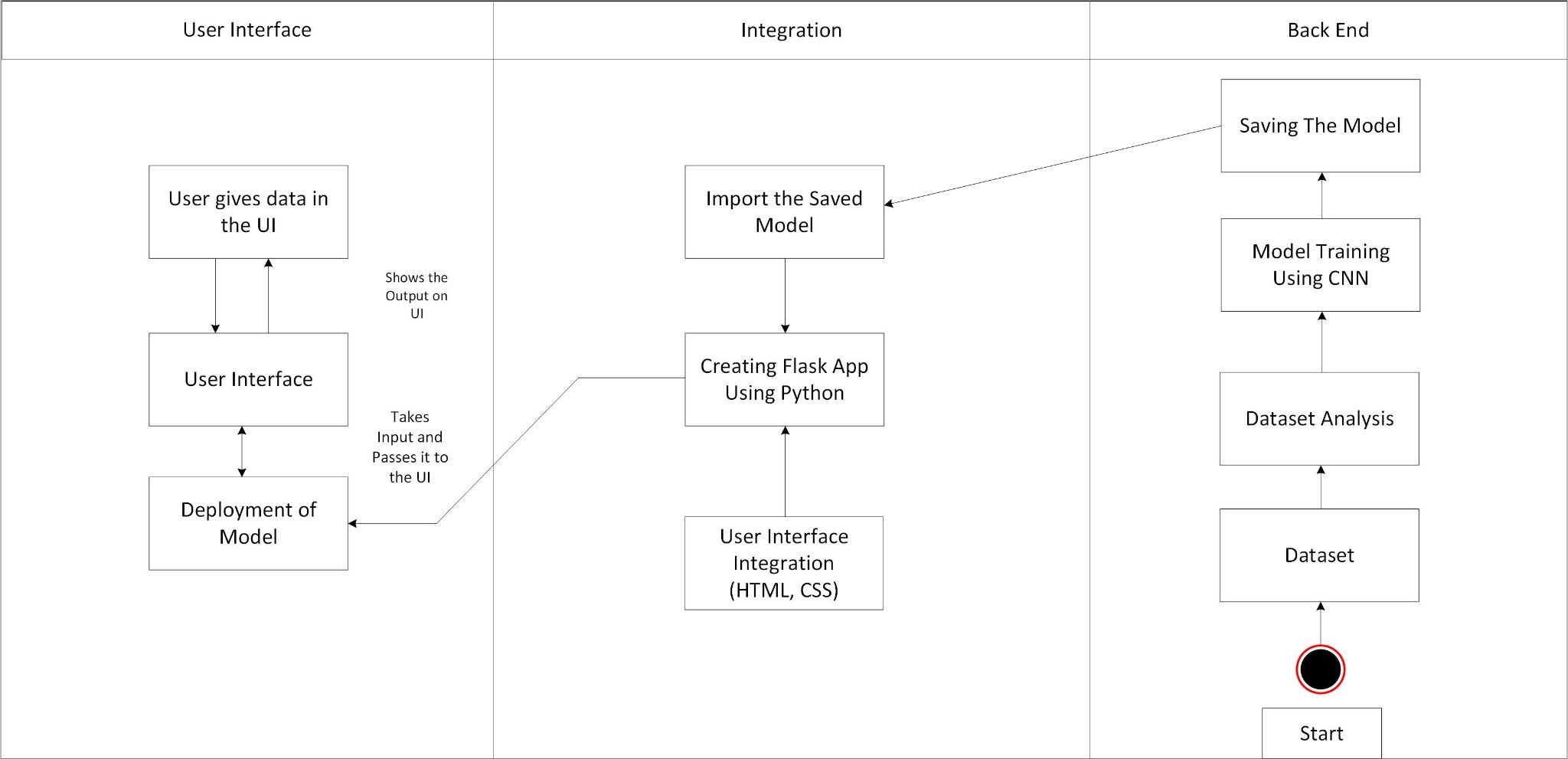
**Solution Architecture Diagram:**

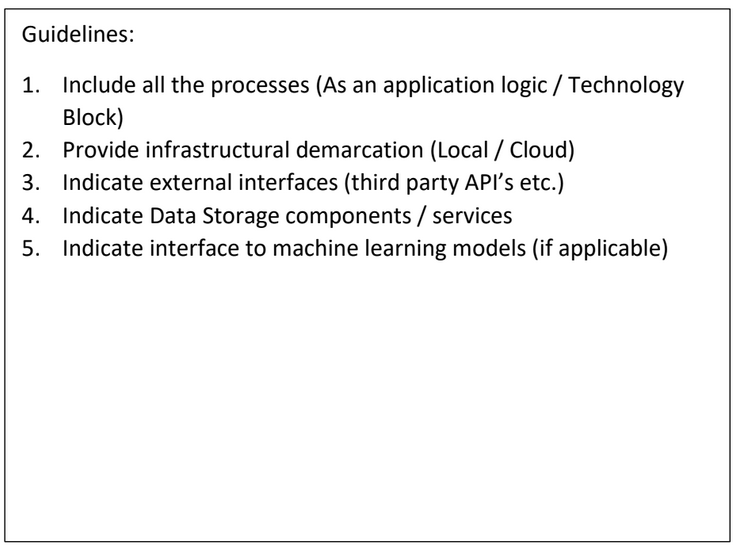


**6. PROJECT PLANNING & SCHEDULING**

6.1 Technical Architecture

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2





**Table-1 : Components & Technologies:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Component** | **Description** | **Technology** |
| 1. | User Interface | Web UI | HTML, CSS, JavaScript / Angular Js / React Js etc. |
| 2. | Application Logic-1 | Machine Learning Algorithm to Analyze Responses | Python |
| 3. | Database | Collect the Dataset Based on the Problem Statement | File Manager |
| 4. | File Storage/ Data | File storage requirements for Storing the dataset | Local System, Google Drive Etc |
| 5. | Frame Work | Used to Create a web Application, Integrating Frontend and Back End | Python Flask, Django etc |
| 6. | Deep Learning Model | To Predict Mental Illness In Employees | CNN, Transfer Learning etc. |

**Table-2: Application Characteristics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | List the open-source frameworks used | Flask (Python) |
| 2. | Performance | Design consideration for the performance of the  application | Caching |

References:

https://c4model.com/

https://www.leanix.net/en/wiki/ea/technical-architecture

https://aws.amazon.com/architecture

https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90d

6.2 Sprint Planning & Estimation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional Requirement** | **User Story Number** | **User story/Talk** | **Story Point** | **Priority** | **Team Member** |
| Sprint-1 | Project setup &  Infrastructure | USN-1 | Create a version-controlled repository on Git and set up a basic development environment | 1 | High | Pavani |
| Sprint-1 | Development  Environment | USN-2 | Set up the development environment with the required tools and frameworks to start the prediction of mental illness in working professional project. | 2 | High | Sindhu |
| Sprint-2 | Data Collection | USN-3 | Gather a diverse dataset containing different types of mental illnesses for training the deep learning model. | 2 | High | Pavani |
| Sprint-2 | Data Preprocessing | USN-4 | Preprocess the collected dataset by data cleaning, handling imbalances and splitting it into training and validation sets. | 4 | High | Sindhu |
| Sprint-3 | Model Development | USN-5 | Explore and evaluate different deep learning architectures (e.g CNNs) to select the most suitable model for mental illness prediction. | 4 | High | Vishnu |
| Sprint-3 | Training | USN-6 | train the selected deep learning model using the preprocessed dataset and monitor its performance on the validation set. | 4 | Medium | Pavan |
| Sprint-4 | Model Deployment &  Integration | USN-7 | Deploy the trained deep learning model as a web service | 1 | Medium | Vishnu |
| Sprint-5 | Testing & Quality  Assurance | USN-8 | Fine-tune the model hyperparameters and optimize its performance | 2 | Medium | Pavan |

6.3 Sprint Delivery Schedule

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Total Story Point** | **Duration** | **Sprint Start Date** | **Sprint End Date**  **(Planned)** | **Story Points**  **Completed**  **(as on Planned End Date)** | **Sprint Release Date**  **(Actual)** |
| Sprint - 1 | 3 | 3 days | 25 oct 2023 | 27 oct 2023 | 20 | 27 oct 2023 |
| Sprint - 2 | 6 | 8 days | 28 oct 2023 | 4 nov 2023 |  |  |
| Sprint - 3 | 8 | 4 days | 5 nov 2023 | 8 nov 2023 |  |  |
| Sprint - 4 | 1 | 7 days | 9 nov 2023 | 15 nov 2023 |  |  |
| Sprint - 5 | 2 | 6 days | 16 nov 2023 | 21 nov 2023 |  |  |

Let’s calculate the team’s average velocity (AV) per iteration unit (story points per day)

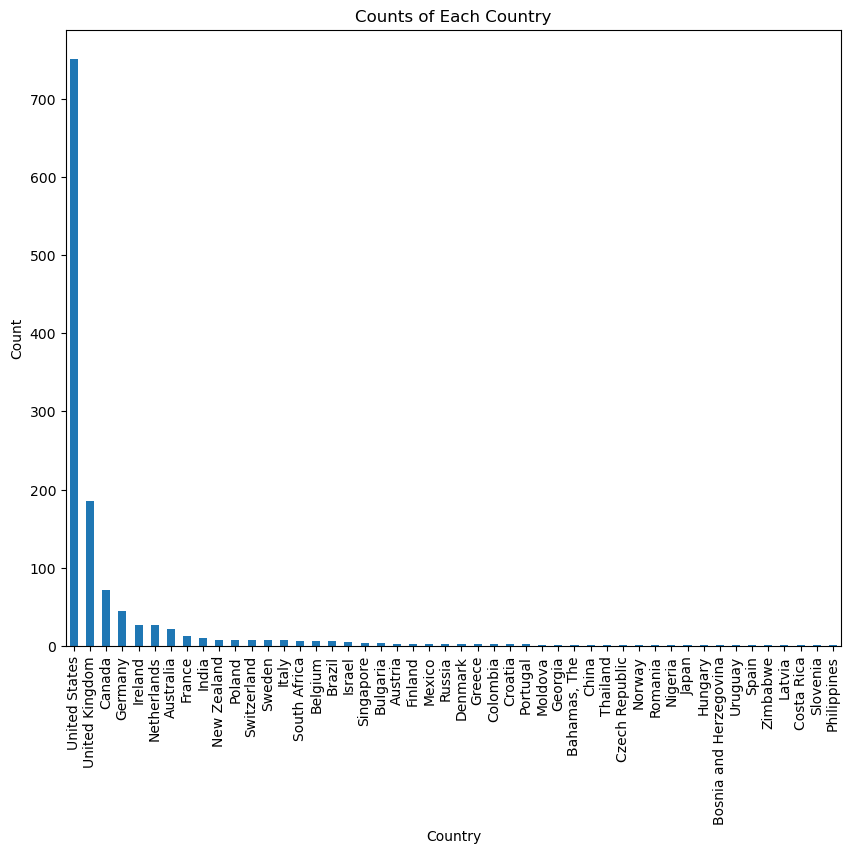
|  |
| --- |
|  |

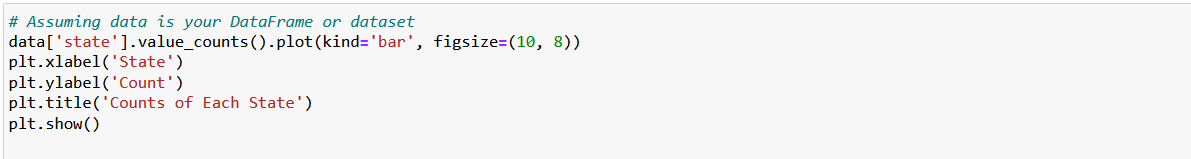
**AV = 20/28 = 0.714**

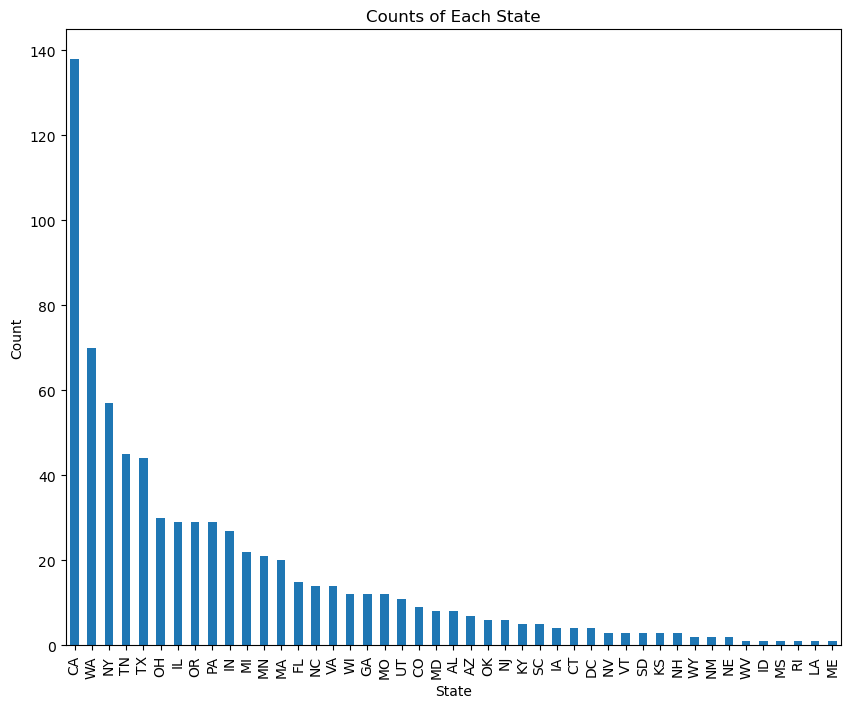
**7. CODING & SOLUTIONING (Explain the features added in the project along with code)**

7.1 Data Preprocessing:

The below picture shows the distribution of countries

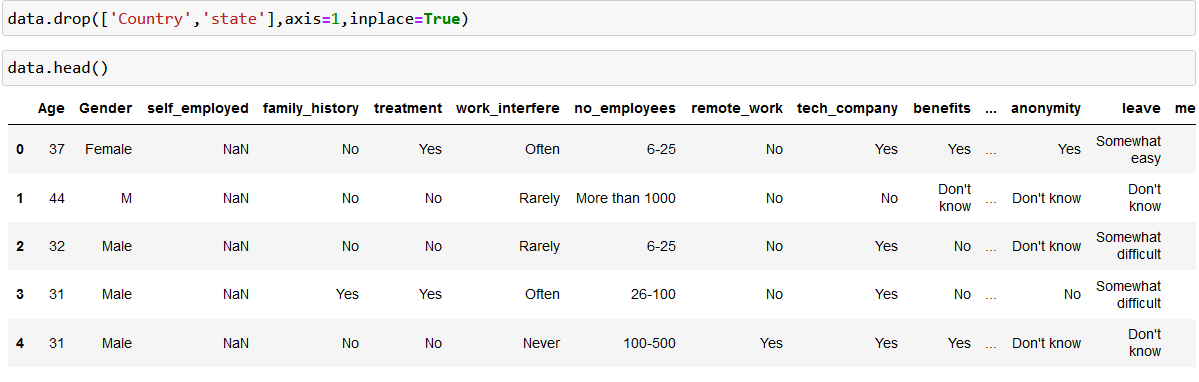






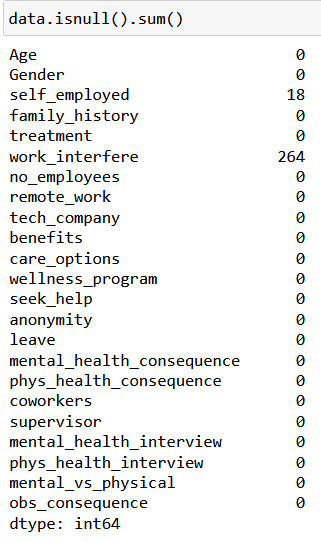
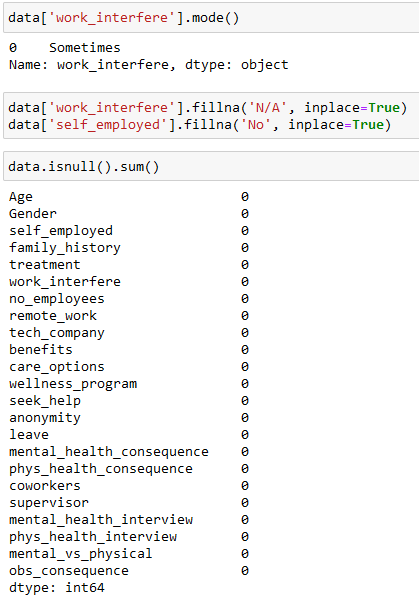
Since the countries are not evenly distributed, keeping this column will induce bias in

our model. so these columns are removed

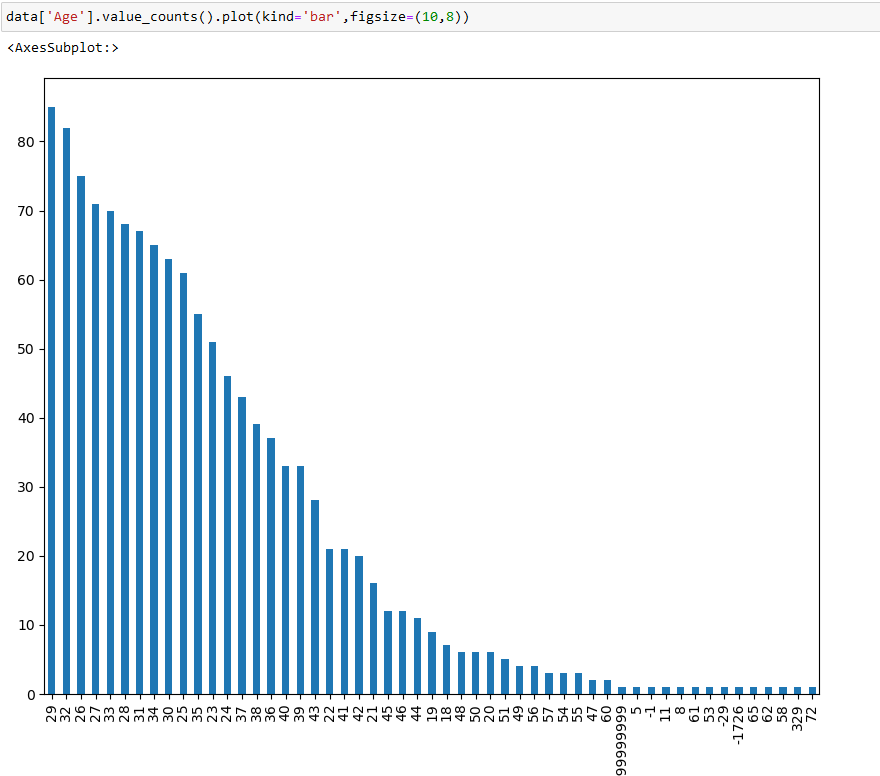


Checking and Clearing Null Values:

Checking - Clearing -

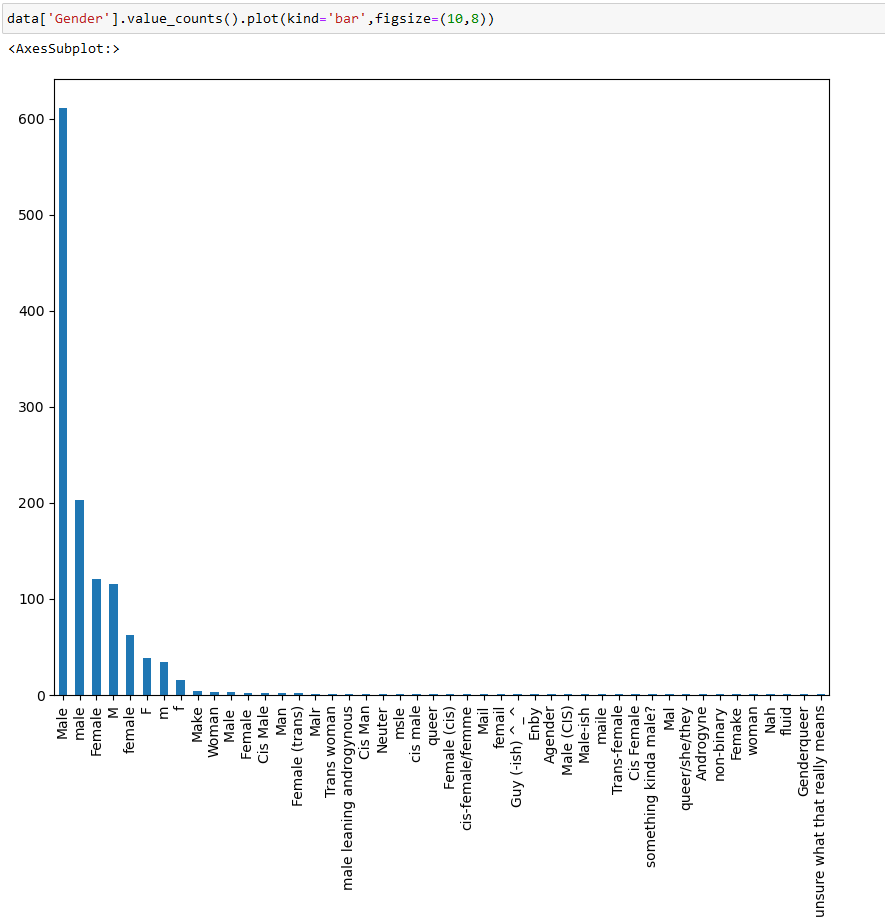
 

Outlier Removal:

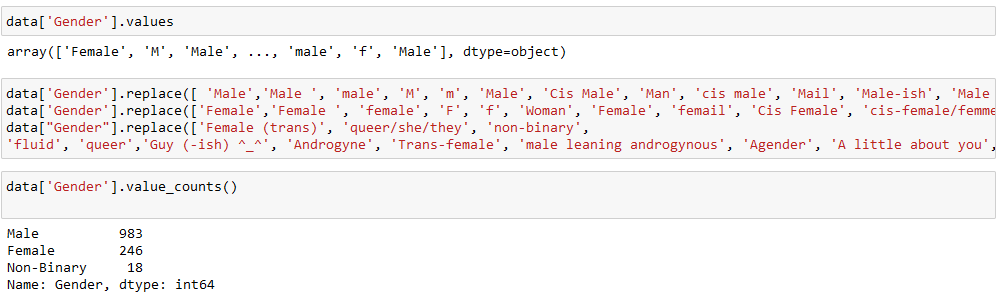


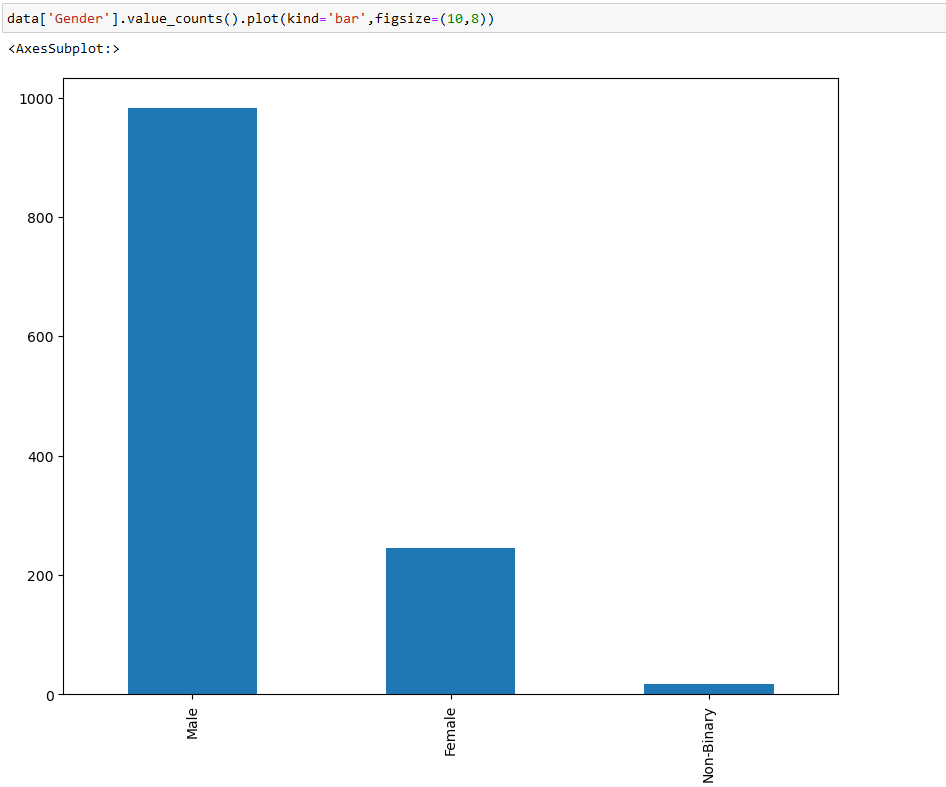


Checking Gender:



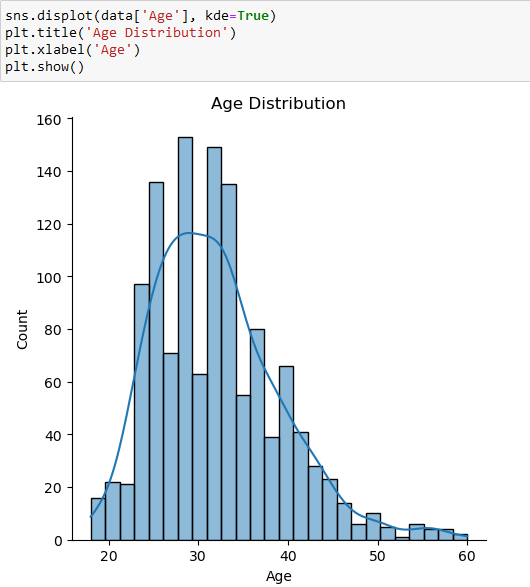
It is observed that different names are used for the same category of gender. Let us group them into 3 major categories- Male, Female and Non-Binary using the .replace() function.

****

****

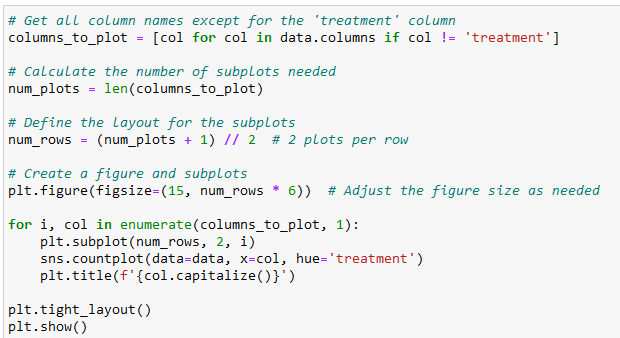
7.2 Data Analysis and Visualisation

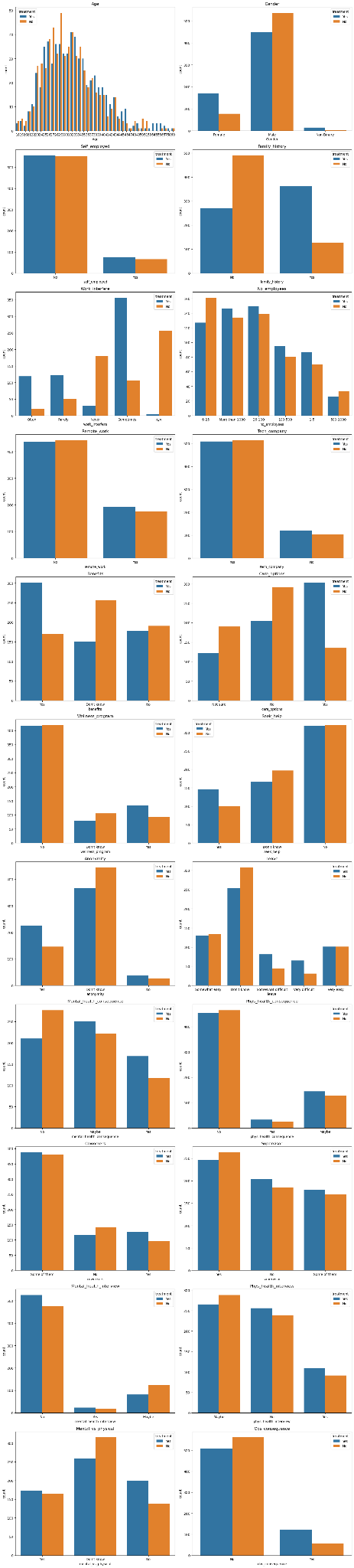
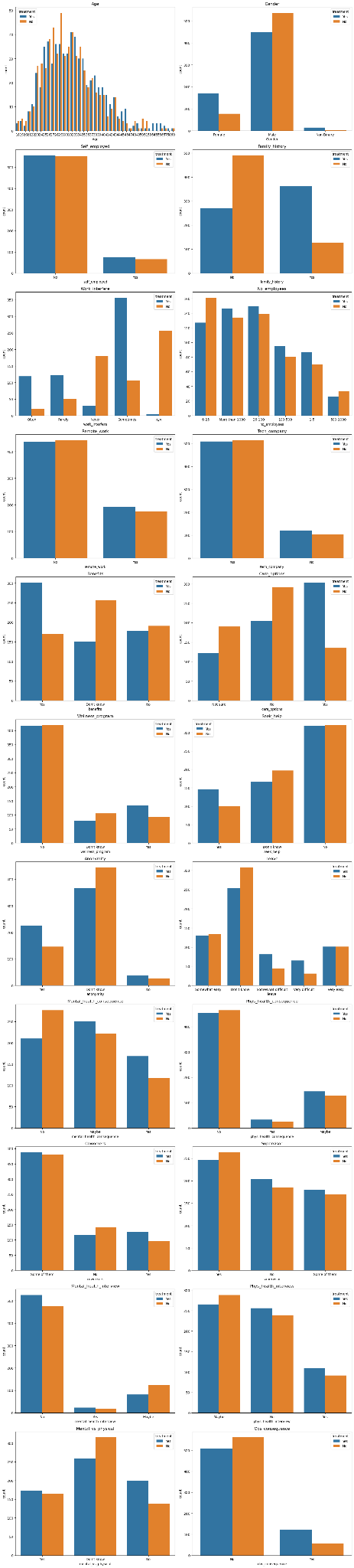
Univariate analysis:



Bivariate Analysis:

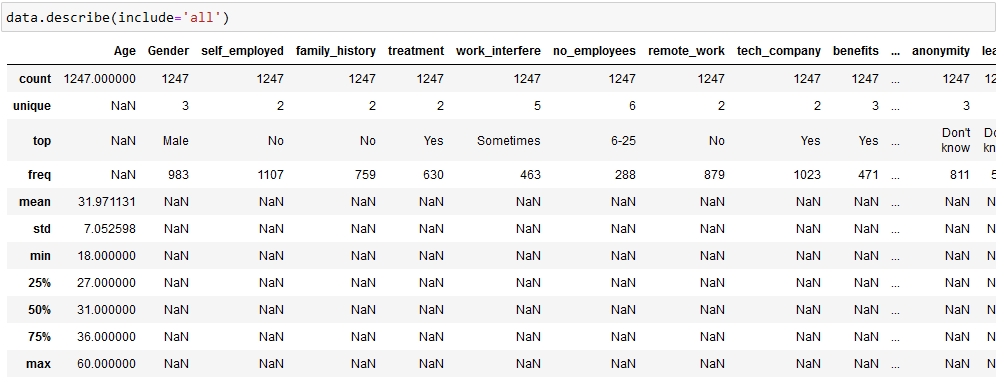
We use bivariate analysis to find the relation between two features. Here we are visualizing the relationship of various features with respect to treatment, which is our target variable.





Descriptive analysis:

With this describe function we can understand the unique, top and frequent values of categorical features. Also, we can find mean, std, min, max and percentile values of numerical features.



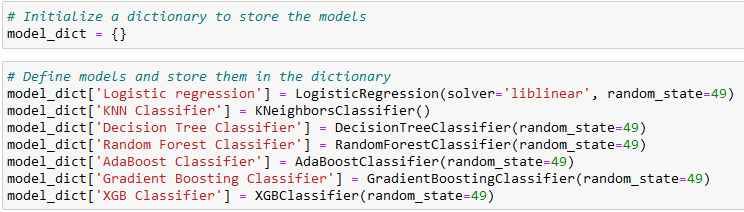
7.3 Handling Categorical Values:

Ordinal encoding for our features and label encoding for our target. Divide our data into features and targets.



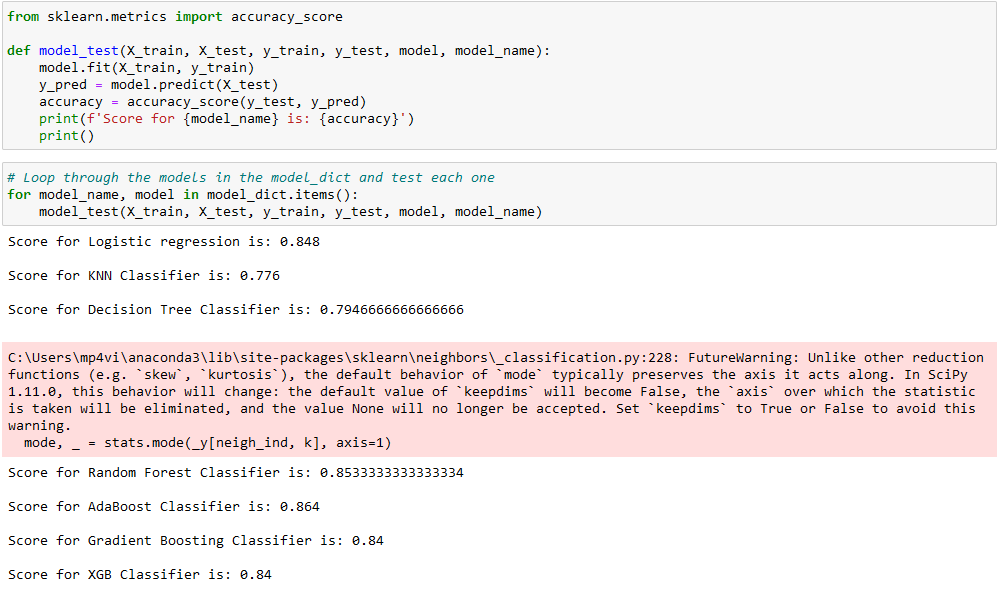
7.4 Comparing Accuracy of Various Models:

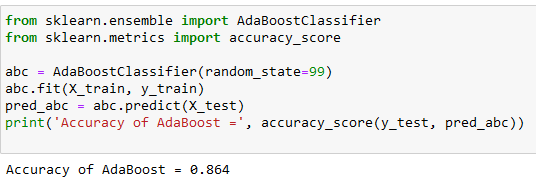
We will be considering multiple models to train our data and choose the one that performs the best.



We will obtain y\_pred by using .predict()

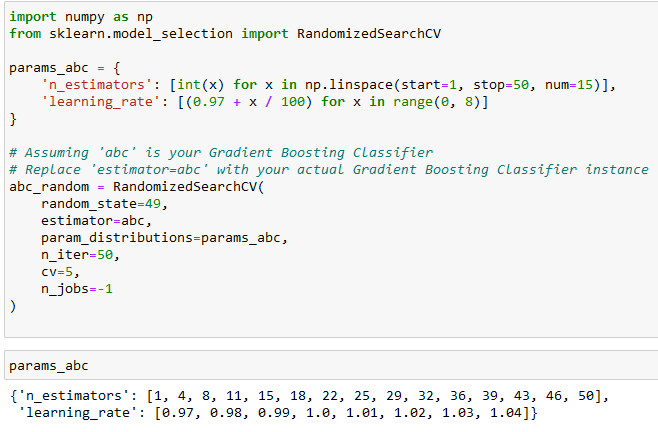
function and compute the accuracy score for every model by iterating through the dictionary.

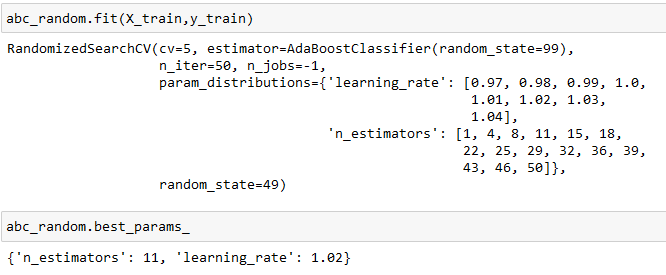




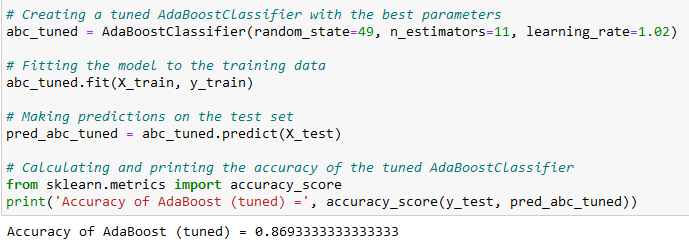
7.5 Hyperparameter Tuning of Selected Model

To further improve the model performance, we are going to carry out a process known as hyperparameter tuning. Every model will have multiple hyperparameters. Of these, we will be tuning n\_estimators and learning\_rate. For hyperparameter tuning, we can either use GridSearchCV or RandomizedSearchCV. RandomizedSearchCV is more fast, efficient and preferred so we will be using it for our project.

****

****

So, our model will perform the best if n\_estimators are equal to 11 and learning\_rate is equal to 1.02. Let us add these values to train our model, make predictions and check accuracy



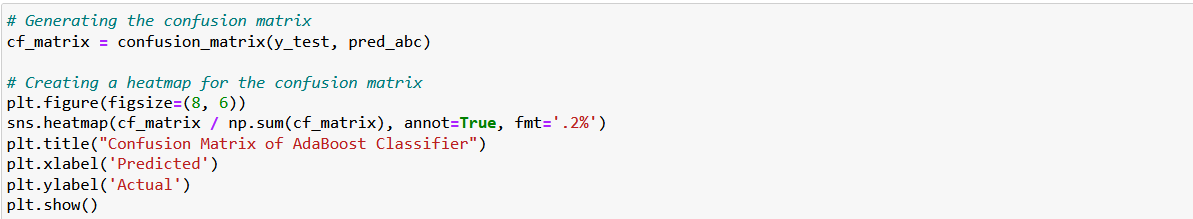
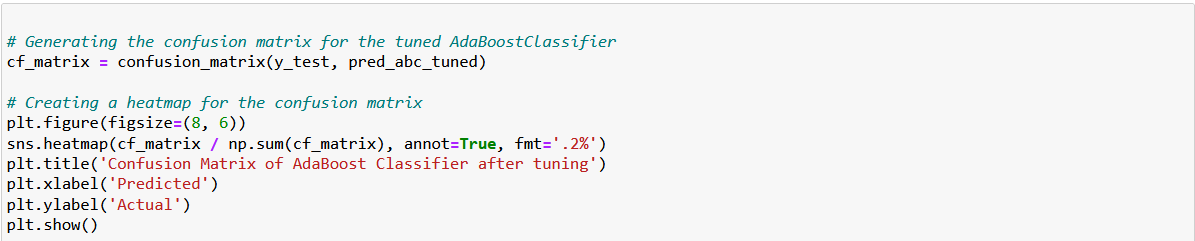
We observe that the accuracy has increased approximately by 0.5%. Though this is not a very great improvement, it is at least better than our previous model.

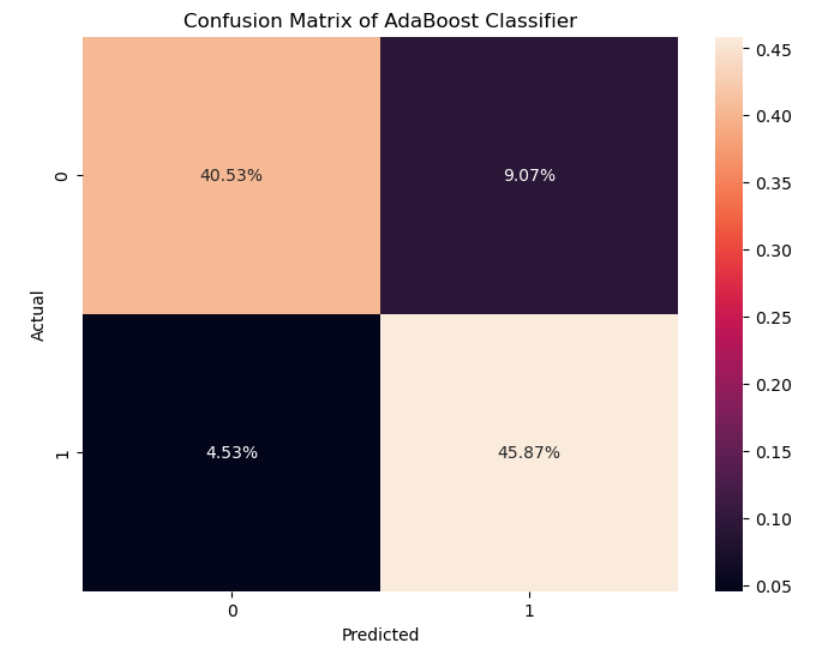
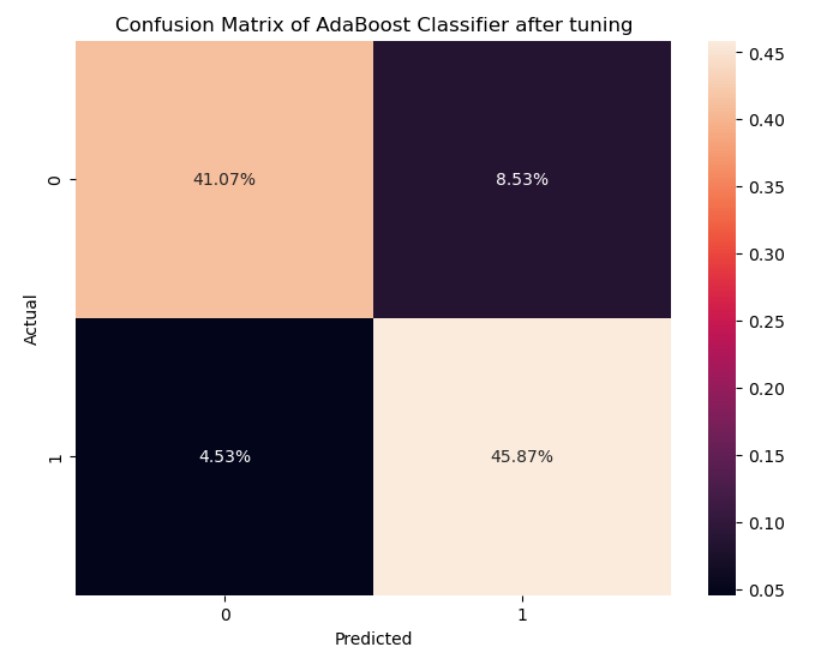
**8. PERFORMANCE TESTING**

8.1 Performance Metrics

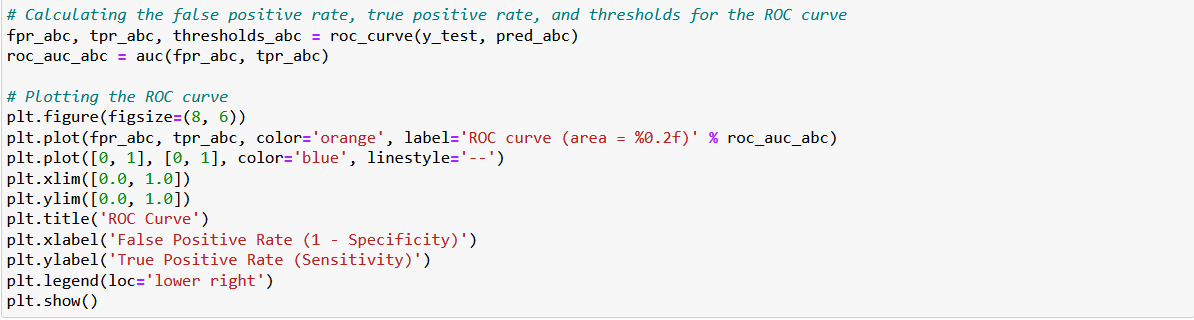
We will compare the confusion matrix, ROC curve and classification report for both models. In order to obtain these, we will be using the confusion\_matrix(), roc\_curve() and classification\_report() functions from sklearn.metrics.

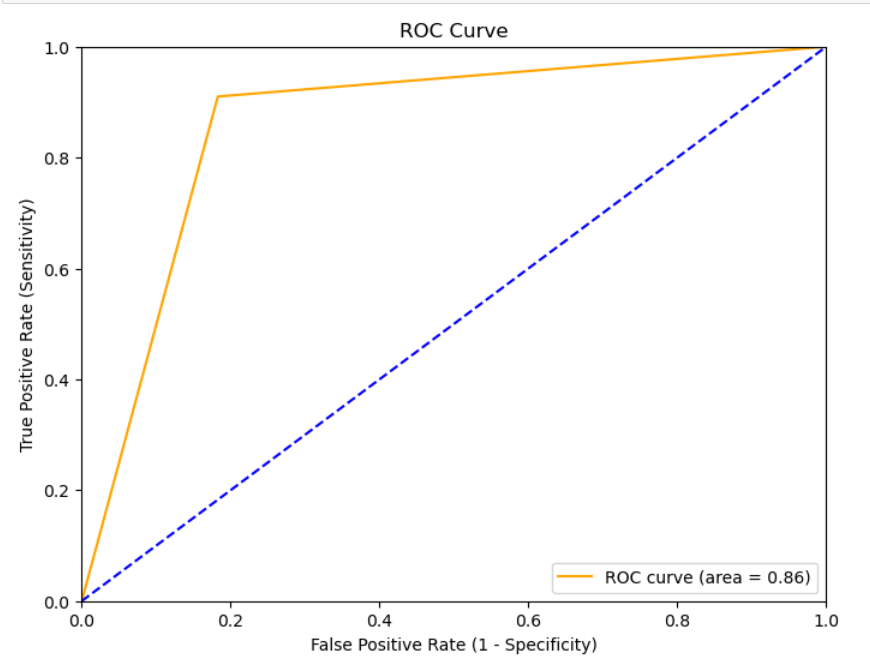
Confusion matrix:

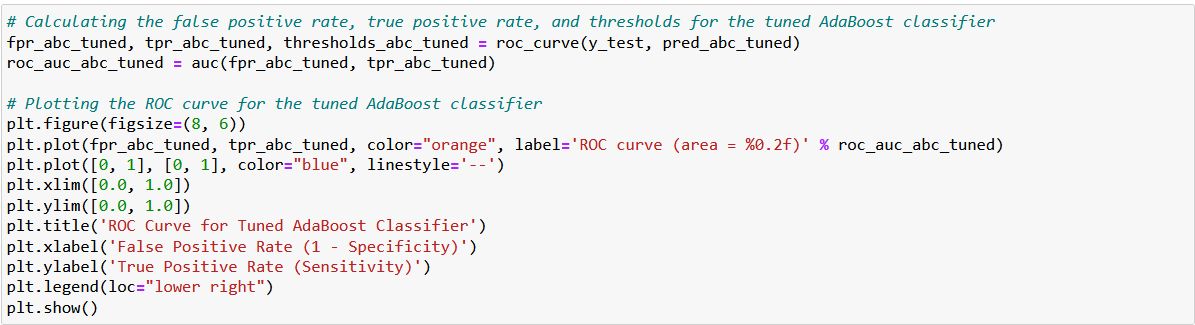
 

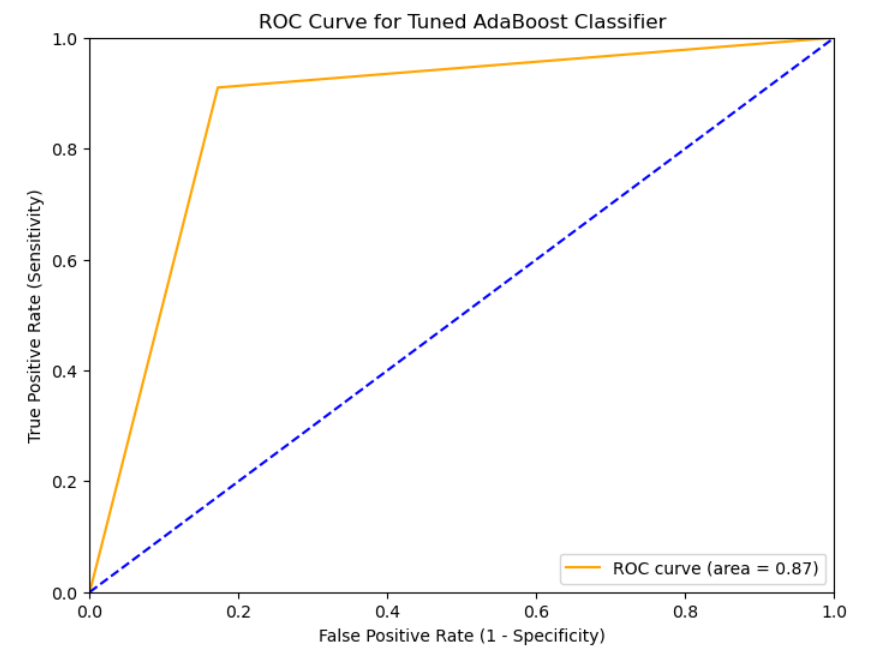
 

ROC Curve:

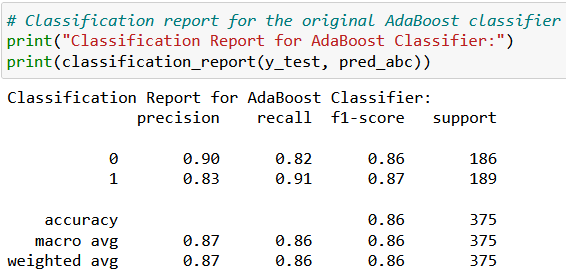
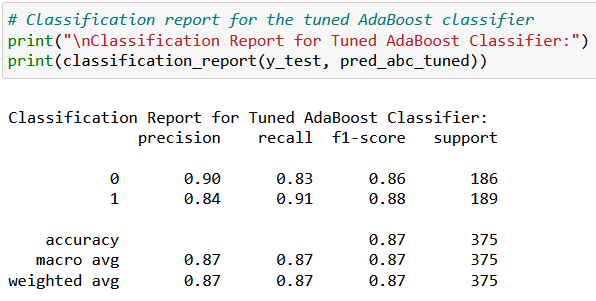








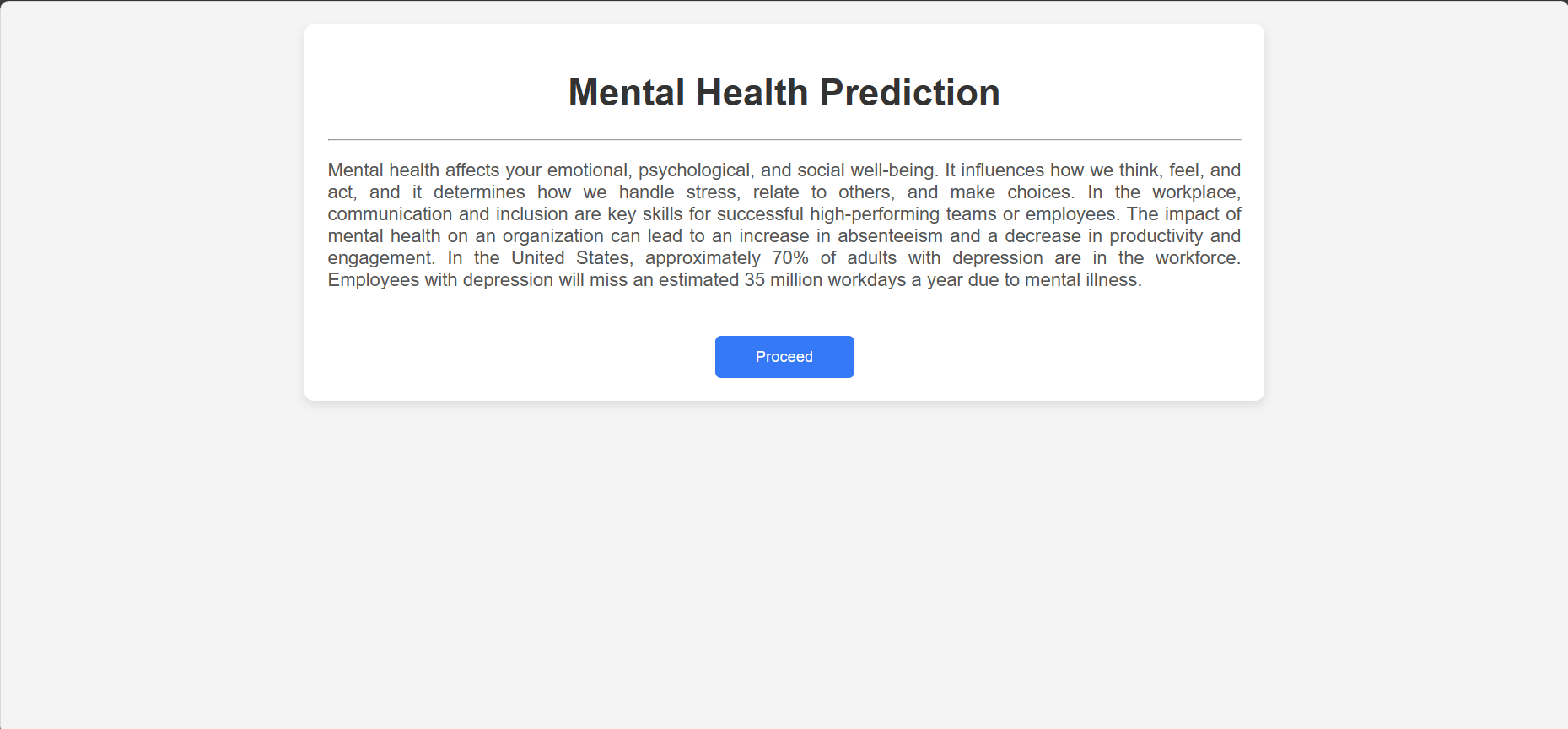
Classification Report:

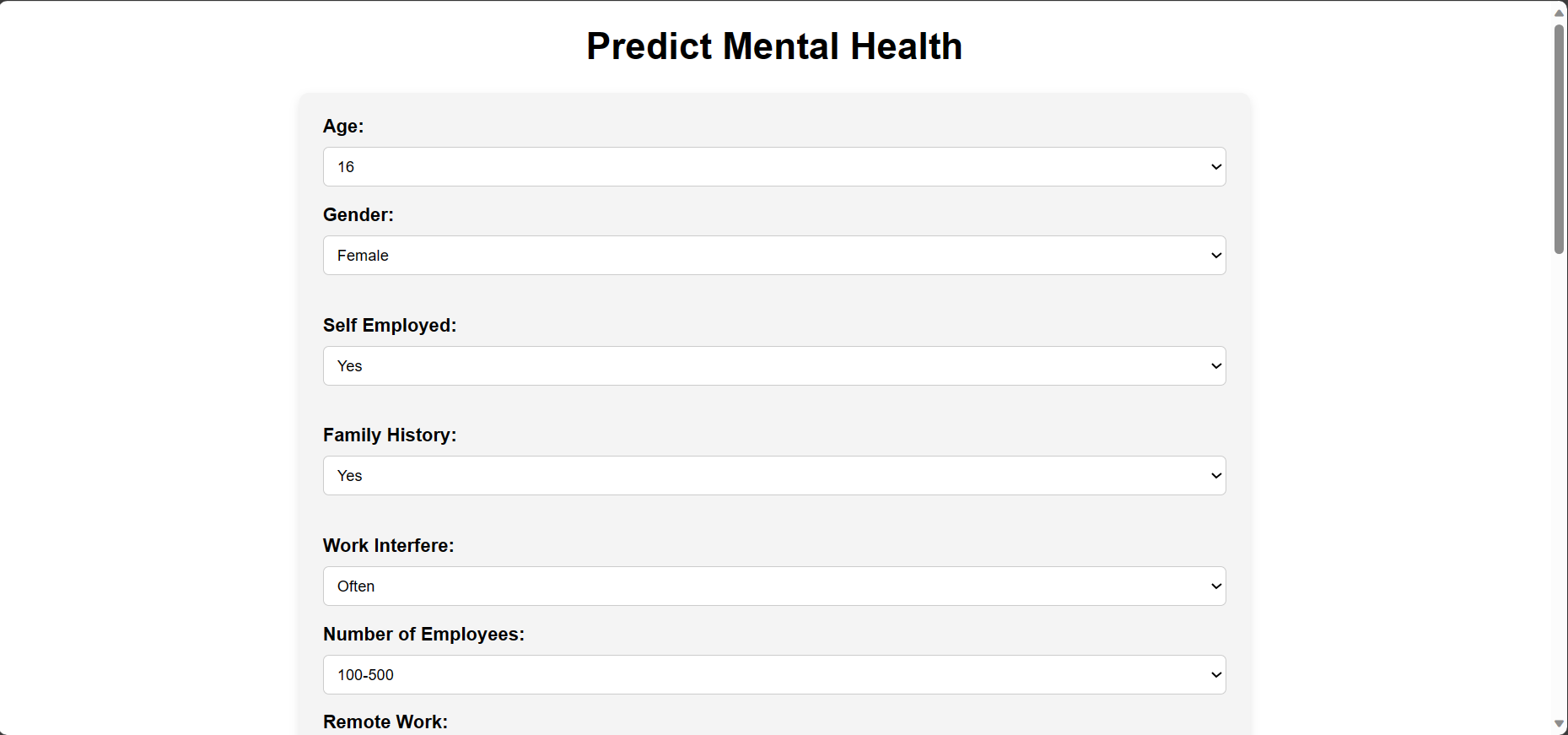
**9. RESULTS**

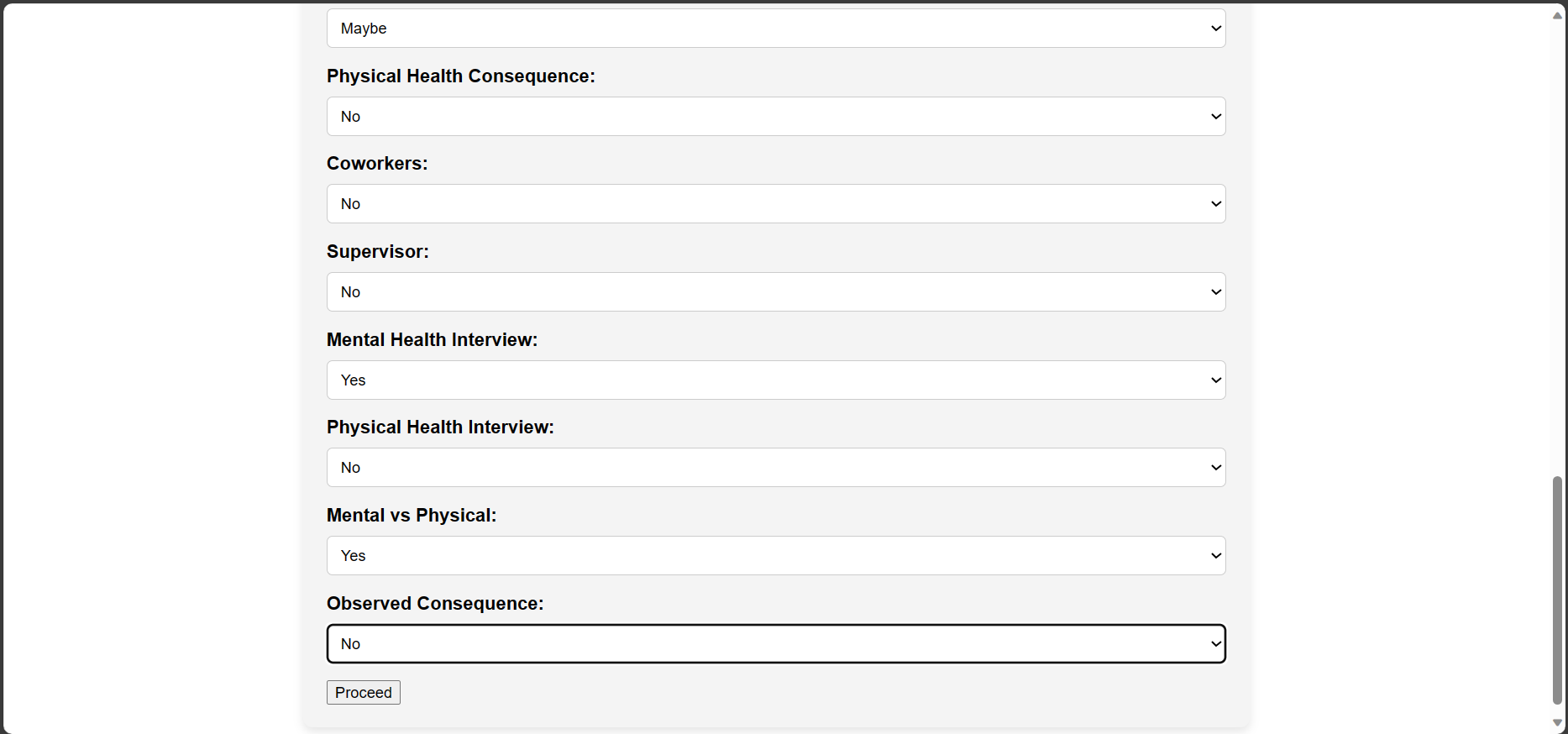
9.1 Output Screenshots

**Home page:**

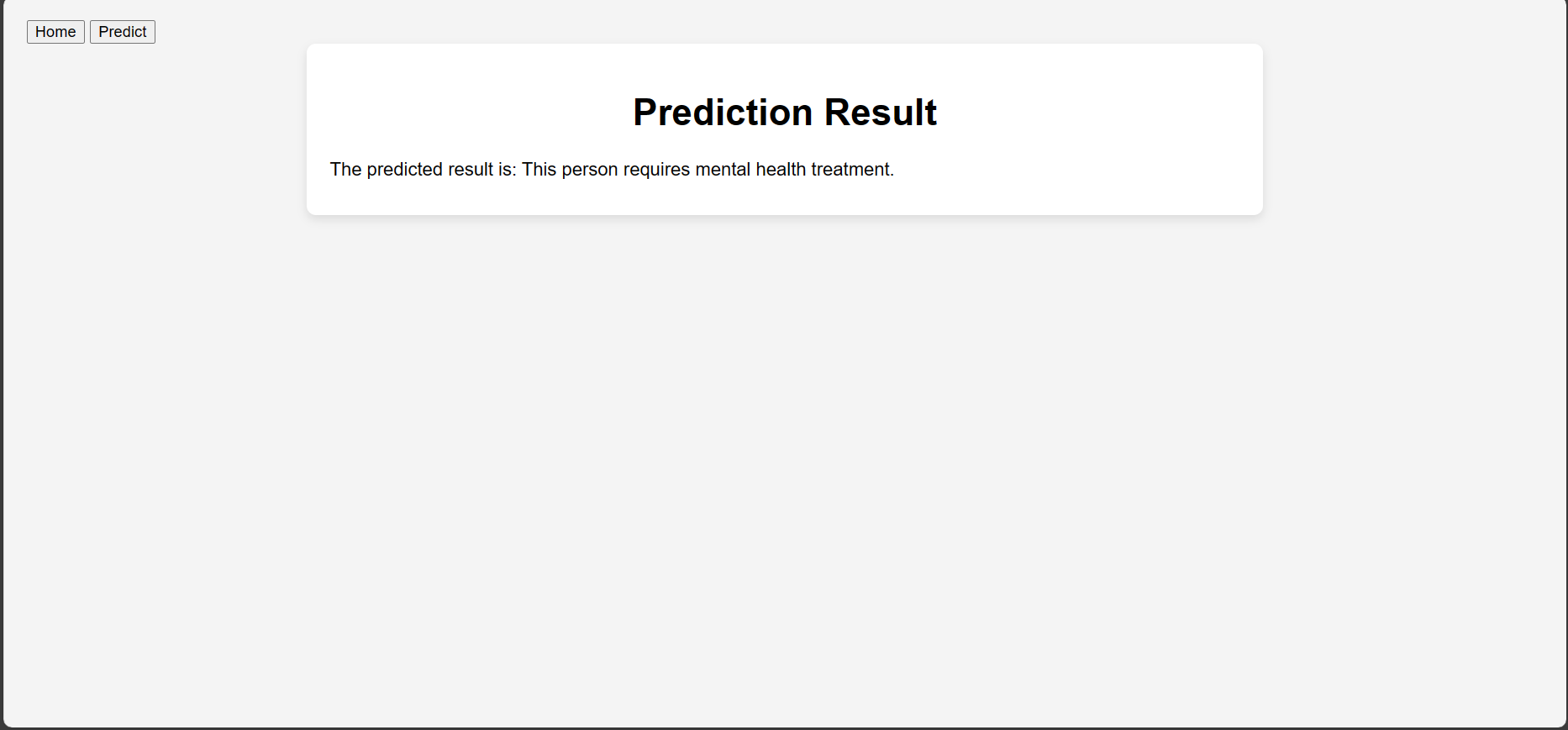
****

**Prediction input form:-**

****

****

**Result page:-**

****

**10. ADVANTAGES & DISADVANTAGES**

10.1 Advantages:

The mental illness prediction system brings significant benefits to users and organizations. It prioritizes mental health well-being by providing a user-friendly platform for individuals to assess their mental state. The high prediction accuracy of the integrated algorithm enhances the system's reliability. Simplicity in design ensures easy implication, making it accessible to a broader audience. Importantly, the system guarantees user anonymity, fostering a safe space for individuals to engage with mental health assessments.

10.2 Disadvantages:

While the system addresses key aspects, it does have limitations. Its simplicity, while an advantage, may be considered a disadvantage for those seeking more in-depth mental health assessments. The system is limited in its scope, offering a basic analysis without delving into nuanced mental health conditions. Additionally, the reliance on self-reported data may introduce subjective biases, impacting the precision of predictions.

**11. CONCLUSION**

In conclusion, the mental illness prediction project establishes a foundational approach to addressing the well-being of working professionals. By prioritizing simplicity, accuracy, and user anonymity, the system provides a valuable tool for self-assessment. While recognizing its limitations, the project lays the groundwork for future enhancements and collaborations to refine mental health prediction strategies.

**12. FUTURE SCOPE**

The future scope of the project involves continuous improvement and expansion. Enhancements could include the integration of more sophisticated algorithms for a broader range of mental health conditions. Collaborations with mental health professionals and organizations could lead to the development of comprehensive support systems. Additionally, exploring avenues for data validation and incorporating additional features, such as real-time monitoring, would contribute to the ongoing evolution of the mental illness prediction system. The project aims to adapt and grow, staying responsive to emerging technologies and evolving understanding of mental health.

**13. APPENDIX**

A. Algorithm Development

A.1 Model Selection

The predictive algorithm employs a machine learning model based on [insert model name], chosen for its ability to handle complex patterns in multidimensional data.

### A.2 Training and Validation

1. Training Set: A subset of the data used to train the algorithm.
2. Validation Set: Independent subset used to assess model performance during training.
3. Hyperparameter Tuning: Optimization of model parameters for improved accuracy.

### A.3 Performance Metrics

The algorithm's performance is assessed using the following metrics:

1. Accuracy: Overall correctness of predictions.
2. Precision: Proportion of true positive predictions among all positive predictions.
3. Recall: Proportion of actual positives correctly predicted.
4. F1 Score: Balance between precision and recall.

## B. Ethical Considerations

### B.1 Bias Mitigation

The algorithm is designed to minimize bias by:

1. Fair Feature Selection: Ensuring the inclusion of diverse demographic variables.
2. Regular Model Audits: Periodic assessments to identify and rectify biased patterns.

### B.2 Privacy Measures

1. Data Encryption: Secure transmission and storage of sensitive information.
2. Consent Protocols: Implementation of transparent and explicit consent procedures.

Source Code

GitHub & Project Demo Link