PREDICTING MENTAL ILLNESS OF WORKING PROFESSIONALS USING MACHINE LEARNING

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING(AI&ML)

Submitted By

DEVULAPALLY LAVANYA	**20UK1A6601**
VADDEPELLI GANESH	**20UK1A6602**
CHEMMALA AMULYA	**20UK1A6642**
BARIGALA VAMSHI PRABHATH	**20UK1A6637**

Under the guidance of

Dr .K. SHARMILA REDDY

Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) –

506005

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING(AI&ML)

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION
INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled "**PREDICTING MENTAL ILLNESS OF WORKING PROFESSIONALS USINF MACHINE LEARNING**" is being submitted by **DEVULAPALLI LAVANYA(20UK1A6601),VADDEPELLI GANESH(20UK1A6602),CHEMMALA AMULYA(20UK1A6642),BARIGALA VAMSHI PRABHHATH(20UK1A6637)** in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023- 2024.

Project Guide
Dr. K Sharmila Reddy
(Professor)

HOD
Dr. K Sharmila Reddy
(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr . P. PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr . K. SHARMILA**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, For their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

DEVULAPELLY LAVANYA (20UK1A6601)
VADDEPELLI GANESH (20UK1A6602)
CHEMMALA AMULYA (20UK1A6642)
BARIGALA VAMSHI PRABHATH (20UK1A6637)

ABSTRACT

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 benefits 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. Organizations that incorporate mental health awareness help to create a healthy and productive work environment that reduces the stigma associated with mental illness, increases the organizations' mental health literacy, and teaches the skills to safely and responsibly respond to a co-worker's mental health concern. 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.

We will be using classification algorithms such as Logistic Regression, KNN, Decision tree, Random Forest, AdaBoost, Gradient Boost, and Boost. We will train and test the data with these algorithms. From this, the best model is selected and saved in pkl format. We will also be deploying our model locally using Flask.

TABLE OF CONTENTS: -

1.INTRODUCTION

1.1. OVERVIEW

The project aims to utilize machine learning techniques to predict mental illness among working professionals. It involves collecting and analysing various data points to develop predictive models that can identify potential mental health issues among employees in a professional setting. Factors such as work-related stress, lifestyle habits, behavioural patterns, and other relevant data will be considered to create a system that can forecast and potentially prevent mental health challenges in the workplace.

1.2. PURPOSE

The primary purpose of this project is to proactively address mental health concerns in the working environment. By leveraging machine learning algorithms on collected data, the goal is to develop a system that can detect early signs of mental health issues among employees. This predictive model could assist employers in taking preventive measures, offering support, or interventions to improve the mental well-being of their staff. Ultimately, the project aims to contribute to a healthier and more supportive work environment by identifying potential mental health challenges and enabling timely intervention.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

An existing problem in the context of predicting mental illness among working professionals using machine learning is the lack of proactive measures to address mental health issues in the workplace.

- An existing problem in the context of predicting mental illness among working professionals using machine learning is the lack of proactive measures to address mental health issues in the workplace. Currently, many organizations primarily focus on reactive approaches, addressing mental health problems only after they become evident or severe. This often leads to increased absenteeism, reduced productivity, and high turnover rates, impacting both employees' well-being and a company's overall performance.

- Additionally, there might be a stigma associated with seeking help or discussing mental health problems in many work environments, leading to underreporting and concealment of issues. This lack of open dialogue and data-driven early intervention strategies makes it challenging to identify, address, and support employees dealing with mental health challenges effectively.

- Therefore, the existing problem lies in the absence of predictive tools or models that can leverage machine learning to detect potential mental health concerns in professionals early on, allowing for proactive measures and support systems to be put in place before these issues escalate.

2.2 PROPOSED SOLUTION

- **Data Collection:** Gather diverse data sources such as employee demographics, work-related factors (e.g., job role, workload, working hours), behavioural patterns, health history, and potentially external factors like social media activities or wellness habits.
- **Feature Selection and Engineering**: Identify relevant features and variables that might influence mental health. This could involve pre-processing and transforming data to ensure its suitability for the machine learning model.

- **Model Development:** Utilize machine learning algorithms, such as classification models (e.g., logistic regression, decision trees, or neural networks), to build a predictive model. Train the model on historical data to predict potential mental health concerns based on the collected features.

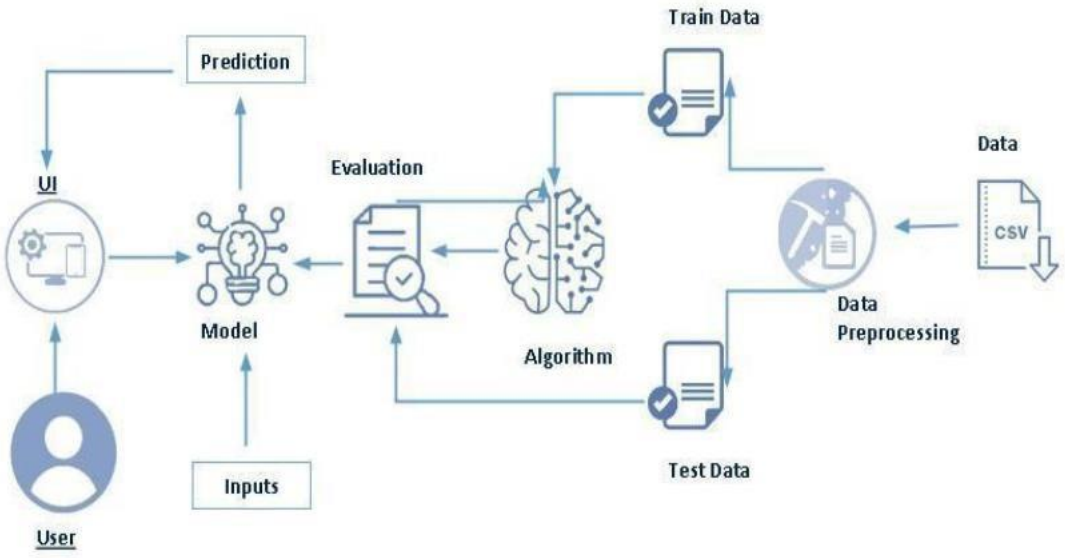
- **Validation and Testing:** Validate the model's performance using different evaluation metrics and test it with new, unseen data to ensure its accuracy and reliability in predicting mental health issues.

- **Implementation and Intervention:** Integrate the developed model into the workplace system. Regularly monitor employee data, and when the model predicts potential issues, trigger interventions. These interventions could include offering mental health resources, support, or tailored wellness programs to the employees.

- **Continuous Improvement:** Regularly update and refine the model based on feedback and new data, ensuring that it stays effective and adaptive to changing workplace dynamics.
- This solution aims to create a proactive approach by leveraging machine learning to detect potential mental health concerns among working professionals, thereby allowing for early interventions and support mechanisms to maintain a healthier and more supportive work environment.

3.THEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM



3.2. SOFTWARE DESIGNING

The following is the Software required to complete this project:

- **API Integration:** Interface to collect data from various sources like HR databases, wellness surveys, and health records.
- **Data Preprocessing**: Tools for cleaning, transforming, and structuring the data for analysis.
- Machine Learning Model Development:
- **Model Building:** Employing algorithms like logistic regression, decision trees, or neural networks to develop predictive models.
- Training Pipeline: System for model training and refinement using historical data.
- Model Integration and Deployment:
- **API/Service Integration:** Incorporating the trained model into the software infrastructure.
- **Real-time Prediction:** Capability to generate predictions as new data comes in.

User Interface (UI) Components:

- **Dashboard:** Visual representation of predictions and trends for HR managers and relevant stakeholders.
- **Interactive Visualization:** Graphs, charts, and summary reports for easy interpretation of model outputs.
- **Notification System:** Alerts for potential mental health concerns and suggested interventions.
- **Security and Compliance Measures:**
- Data Encryption and Compliance: Ensuring data security and compliance with relevant regulations (e.g., GDPR, HIPAA).
- **Access Controls:** User authentication and role-based access to protect sensitive data.
- **User Feedback Loop:** Mechanism for users to provide feedback on the model's predictions and suggested interventions.
- **Reporting Tools:** Generating comprehensive reports on model performance and efficacy of interventions.
- **Automated Learning:** Mechanisms to continuously update and refine the model based on new data and user feedback.
- **Version Control:** Managing different iterations of the model to track changes and improvements.
- **Resource Management:** Optimizing computing resources for model training and real-time predictions.
- **System Scalability:** Designing the software to handle larger datasets efficiently.

4.EXPERIMENTAL INVESTIGATION

Machine Learning Approaches in Predicting Depression and Anxiety:

- A machine learning algorithm is developed to predict the clinical remission from a 12-week course of citalopram . Data are collected from the 1949 patients that experience depression of level 1. A total of 25 variables from the data set are selected to make a better prediction outcome. Then, the gradient boosting method is being deployed for the prediction because of its characteristics that combine the weak predictive models when built. An accuracy of 64.6% is obtained by using the gradient boosting method. In order to identify depression and anxiety at an early age, a model has been proposed by Ahmed et al. .
- The model involves psychological testing, and machine learning algorithms such as convolutional neural network, support vector machine, linear discriminant analysis, and K-nearest neighbour have been used to classify the intensity level of the anxiety and depression, which consists of two data sets. Based on the results obtained, the convolutional neural network achieved the highest accuracy of 96% for anxiety and 96.8% for depression.
- The support vector machine showed a great result and was able to obtain an accuracy of 95% for anxiety and 95.8% for depression. Besides that, the linear discriminant analysis reached the accuracy of 93% for anxiety and 87.9% for depression. Meanwhile, the K-nearest neighbour obtained the lowest accuracy among the models with 70.96% for anxiety and 81.82% for depression.
- Hence, the convolutional neural network can be a helpful model to assist psychologists and counsellors for making the treatments efficient. In the research paper by Sau and Bhakta, they developed a predictive model for diagnosing the anxiety and depression among elderly patients with machine learning technology. Elderly patients have different sociodemographic factors and factors related to health. The data set involved

510 geriatric patients and tested with a tenfold cross-validation method. Then, ten classifiers as shown in Table 1 were selected to predict the anxiety and depression in elderly patients. The metrics of each classifier were evaluated and summarized.

- According to Table 1, the highest prediction was obtained by random forest with 89.0%. Then, the J48 accuracy score was 87.8% followed by random subspace with an accuracy of 87.5%. Random tree showed the prediction accuracy with 85.1%; meanwhile, the Bayesian network achieved an accuracy of 79.8%. Next, the naive Bayes and multilayer perceptron achieved the accuracy of 79.6% and 77.8%, respectively. Sequential minimal optimisation and K-star achieved the same accuracy, which is 75.3%. Finally, logistic regression showed the lowest accuracy prediction of 72.4%. In research conducted byKatsis et al., a system based on physiological signals for the assessment of affective states in anxiety patients has been proposed. The system is proposed to predict the affective state of an individual according to five predefined classes, which are neutral, relaxed, startled, apprehensive, and very apprehensive. The authors use machine learning algorithms in this research such as artificial neural networks, random forest, neurofuzzy systems, and support vector machine. The neuro-fuzzy system can obtain the highest accuracy with a score of 84.3% followed by random forest with an accuracy of 80.83%. Meanwhile, the support vector machine and artificial neural networks achieved the accuracies of 78.5% and 77.33%, respectively.
- A research paper by Sau and Bhakta shows the prediction of depression and anxiety among seafarers. Seafarers are easily exposed to mental health problems, which typically are depression and anxiety. Hence, machine learning technology has been useful in predicting and diagnosing them for early treatments.
- The authors were able to obtain a data set of 470 seafarers who were interviewed. In this research conducted by them, features including age, educational qualification, marital status, job profile, type of family, duration of service, existence or nonexistence

of heart disease, body mass index, hypertension, and diabetes have been selected to predict the outcome.

- Five classifiers, which are CatBoost, random forest, logistic regression, naive Bayes, and support vector machine, were chosen on the training data set with 10-fold crossvalidation. In order to determine the strength of the machine learning algorithms, the data set with 56 instances are deployed on the trained model. For the training set, the results indicate that the boosting algorithms method CatBoost performs best on this training data set with an accuracy of 82.6%. Random forest has achieved a satisfying accuracy score of 81.2%; meanwhile, logistic regression obtained an accuracy.
- **Table 1**: Evaluation metrics of ten classifiers in predicting the anxiety and depression among elderly patients.

Classifiers	Evaluation metrics (%)		
	Accuracy	F-measure	AUC
Bayesian network	79.8	79.7	88.9
Naive Bayes	79.6	79.4	85.3
Logistic regression	72.4	72.2	81.1
Multiple layer perceptron	77.8	77.8	85.0
Sequential minimal optimisation	75.3	74.6	75.9
K-star	75.3	75.3	81.4
Random subspace	87.5	87.5	91.7
J48	87.8	87.8	86.0
Random forest	89.0	89.0	94.3
Random tree	85.1	85.1	85.0

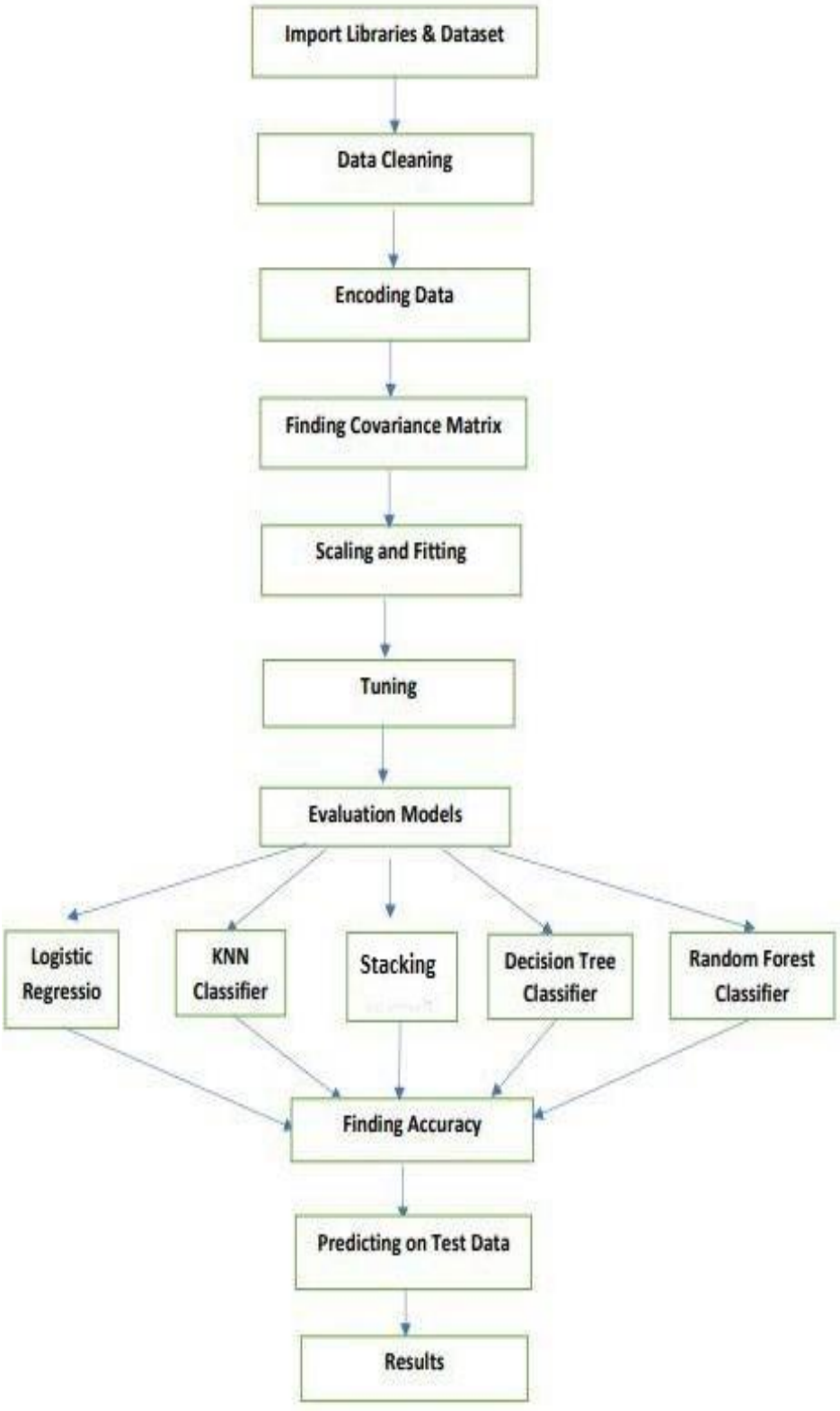
score of 77.8%. The support vector machine and naive Bayes obtained 76.1% and 75.8%, respectively. For the test data set, the CatBoost algorithm has performed better than the other machine learning algorithms with a predictive accuracy of 89.3%.

- Meanwhile, logistic regression has performed very well with the predictive accuracy of 87.5%. Besides, the support vector machine and naive Bayes score the accuracy with

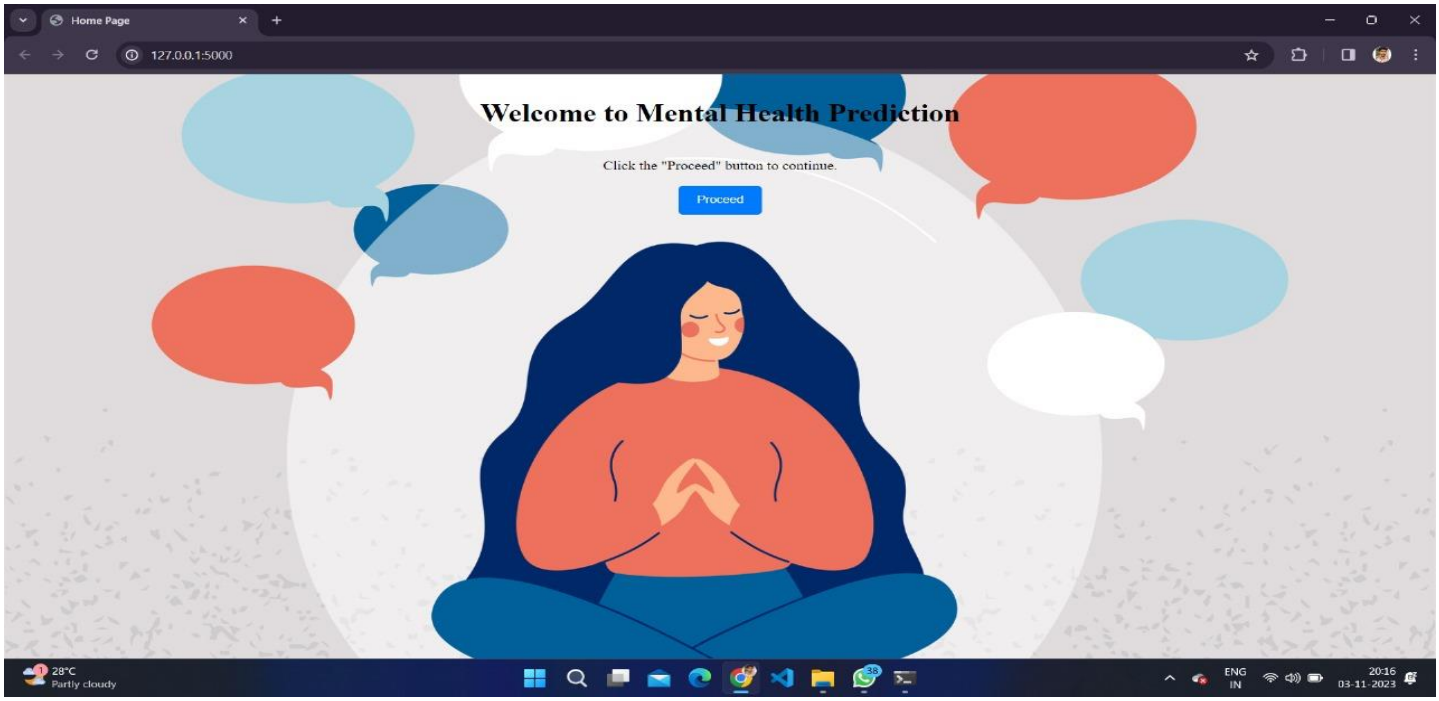
the same percentage, which is 82.1%. The random forest shows the lowest accuracy percentage score of 78.6% for the test data set. Hilbert et al. used machine learning approaches to separate the complicated subjects from healthy ones and distinguish generalized anxiety disorders from major depression without generalized anxiety disorder .

- For the data set, they used the multimodal behavioural data from a sample of generalized anxiety disorders, healthy persons, and major depression. They applied a binary support vector machine and found out that the prediction of generalized anxiety disorders was difficult when using the clinical questionnaire data. Meanwhile, the input involves the in conclusion of cortisol and grey matter volume can reach accuracies of 90.10% and 67.46% for the classification of case and disorder, respectively.
- A study has been conducted to detect depression from text and audio by Jerry and others. The study aims to collect the data and improve the analysis from the features of text and voice. The mean of F1-score is analyzed and recorded to determine the best performance among the machine learning algorithms. Tables 2 and 3 show the performance of the machine learning algorithms in detecting depression in text and audio features, respectively.

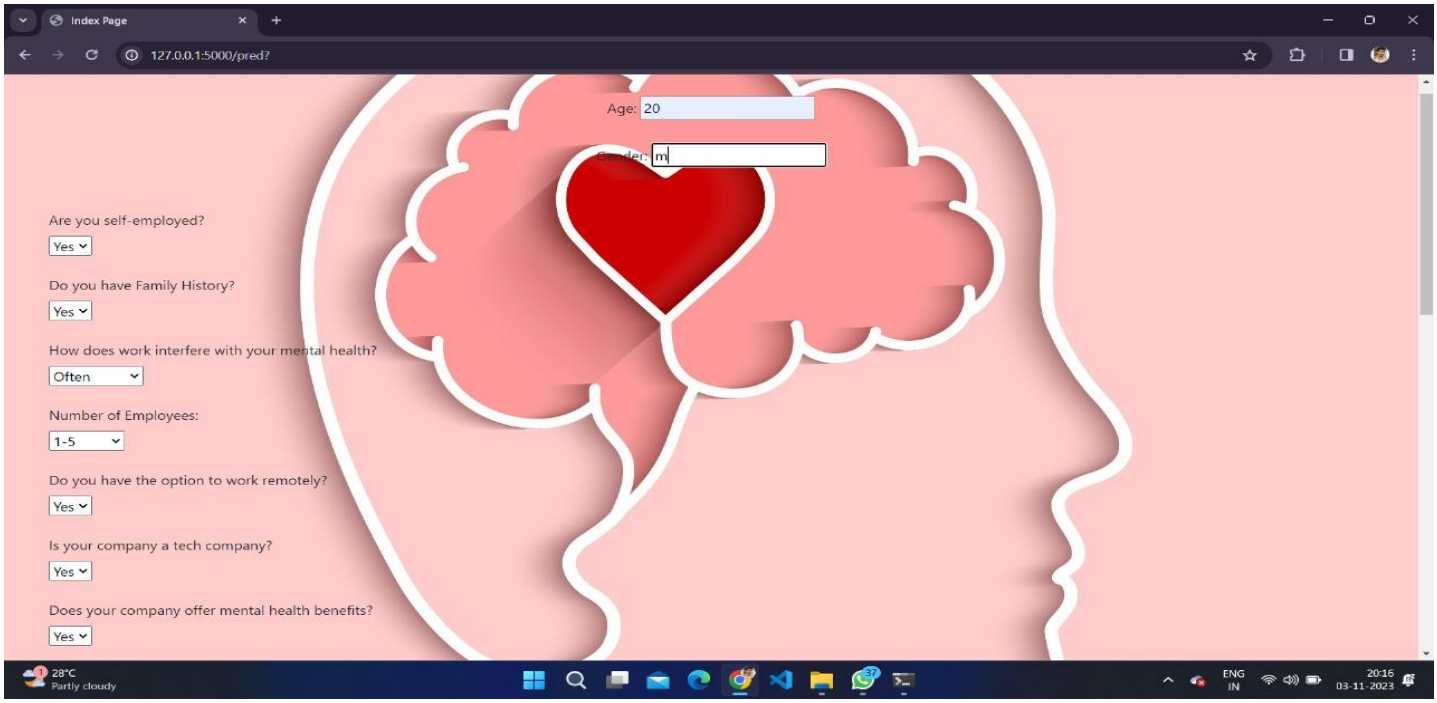
5.FLOWCHART


Import Libraries & Dataset
Data Cleaning
Encoding Data
Finding Covariance Matrix
Scaling and Fitting
Tuning
Evaluation Models
Logistic Regressio
KNN Classifier
Stacking
Decision Tree Classifier
Random Forest Classifier
Finding Accuracy
Predicting on Test Data
Results

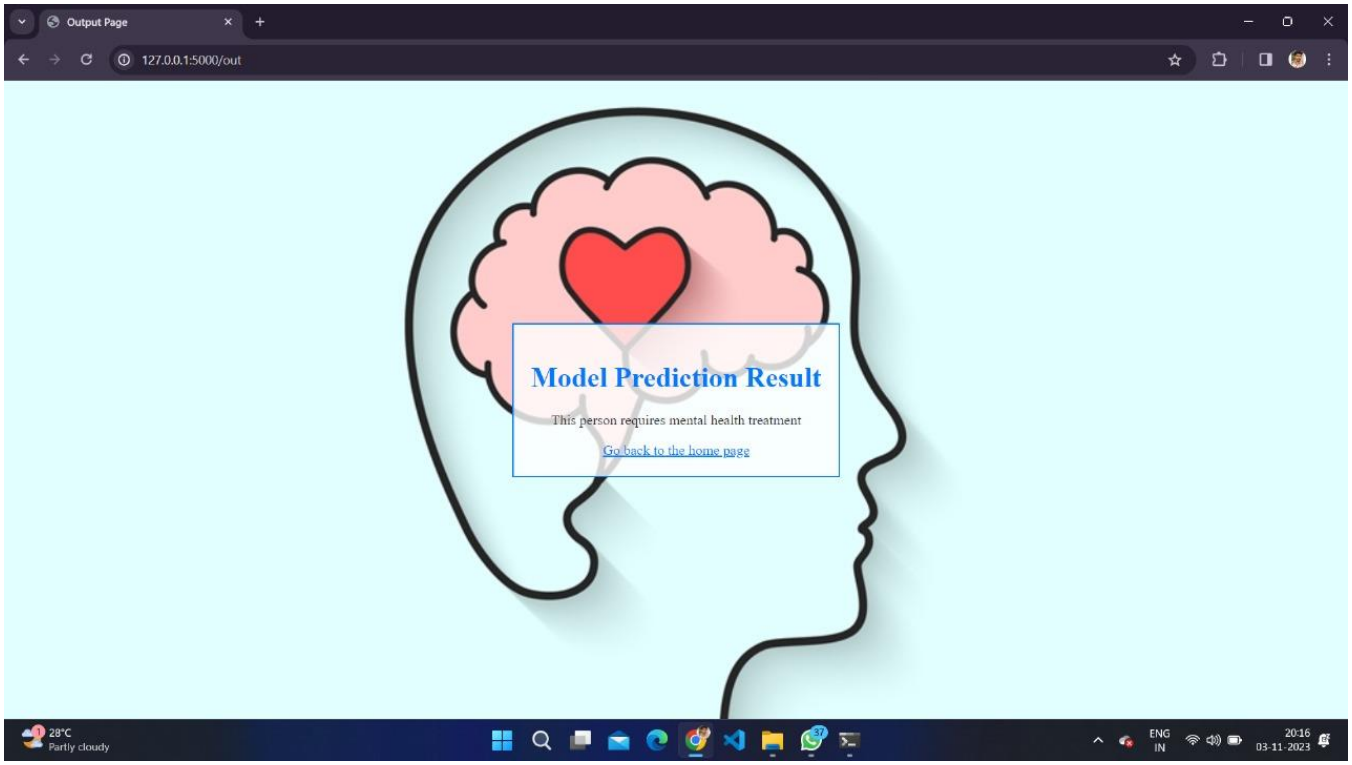
6.RESULTS



INDEX PAGE



OUTPUT


Output Page
127.0.0.1:5000/out
Model Prediction Result
This person requires mental health treatment
Go back to the home page
28°C
Partly cloudy
ENG
IN
20:16
03-11-2023

7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

1. **Early Intervention:** Detecting potential mental health issues early can facilitate timely interventions and support, preventing severe problems from developing.
2. **Proactive Workplace Culture:** Encourages a supportive work environment focused on employee well-being.
3. **Resource Allocation:** Helps organizations allocate resources more efficiently by targeting support where it's most needed.
4. **Reduction in Stigma:** Normalizes discussions about mental health and seeks to remove the stigma associated with seeking help in the workplace.
5. **Improved Productivity:** Addressing mental health concerns can potentially enhance employee productivity and overall work performance.

DISADVANTAGES:

1. **Data Privacy Concerns:** Handling sensitive employee data could raise privacy and ethical concerns.
2. **Algorithm Bias:** The model might inadvertently be biased or inaccurate due to the nature of the data collected or limitations within the model itself.
3. **Over-Reliance on Technology:** There's a risk of solely relying on machine learning models without considering human expertise or factors beyond the model's scope.
4. **Lack of Subjectivity:** Models might not account for individual differences and subjective experiences related to mental health.
5. **Misinterpretation of Results:** Incorrect interpretation of the model's predictions could lead to inappropriate interventions or decisions.

8.APPLICATIONS

- **Corporate Settings:** In corporate environments, this application helps HR departments and managers proactively identify and address mental health concerns among employees, fostering a supportive workplace culture.
- **Healthcare Services:** Mental health practitioners and clinics can leverage such models to predict and prevent mental health issues, providing early interventions and tailored support to patients.
- **Educational Institutions:** Universities and schools can utilize similar models to identify potential mental health challenges among students and provide appropriate counselling or support.
- **Government and Public Services:** Public health agencies can use predictive models to assess the mental health status of the working population, aiding in policy-making and resource allocation for mental health programs.
- **Wellness and Mental Health Apps:** Developers of mental health apps or wellness platforms can incorporate predictive models to offer personalized support and guidance based on user data.
- **Insurance and Healthcare Providers:** Insurance companies can use these models to personalize mental health coverage and interventions for their clients.

9.CONCLUSION

- The utilization of machine learning models to predict mental illness among working professionals offers a proactive and innovative approach to address mental health concerns in the workplace. By leveraging diverse data sources and predictive algorithms, this method can potentially identify early signs of mental health issues, enabling timely interventions and support for employees.

- The benefits of such an approach include early detection, fostering a supportive workplace culture, optimizing resource allocation, and potentially improving overall productivity. However, there are notable challenges, including data privacy concerns, the risk of algorithm bias, and the importance of not over-relying solely on technology without considering individual experiences and human expertise.

- Despite these challenges, the applications of predictive models for mental health in professional settings extend across various sectors, including corporate environments, healthcare services, education, public policy, wellness apps, insurance, and research.

10.FUTURE SCOPE

- **Enhanced Predictive Models:** Continuous advancements in machine learning algorithms, such as deep learning and natural language processing, can lead to more accurate and nuanced predictive models. These learning and natural language processing, can lead to more accurate and nuanced predictive models. These
- models can potentially account for a wider array of factors influencing mental health and offer more precise predictions.

- **Personalized Interventions:** Future developments may focus on tailoring interventions and support systems based on individualized predictions. This could involve adaptive systems that learn from user feedback and behaviour, offering more personalized and effective support for each individual.

- **Ethical and Regulatory Standards:** With a growing emphasis on data privacy and ethics, the future scope involves developing and implementing stringent ethical guidelines and regulatory standards to ensure the responsible handling of sensitive employee data.
- Integration with Wearable Technology: Integrating predictive models with wearable technology or health-tracking devices can provide real-time data for mental health predictions, potentially offering more comprehensive and continuous monitoring of an individual's well-being.

- **Global Implementation and Public Health Initiatives:** The application of these models on a broader scale can assist in public health initiatives and policies, offering insights into mental health at a societal level, influencing organizational strategies, and shaping government policies.

- **Education and Awareness:** Increasing awareness about mental health and destigmatizing seeking help in professional settings will be a crucial part of the future scope. These models can aid in promoting a culture of openness and support for mental health in the workplace.

- **Interdisciplinary Research and Collaboration:** Future advancements may involve collaboration between data scientists, mental health professionals, sociologists, and ethicists to develop more holistic models that consider the multi-faceted nature of mental health.

- **Longitudinal Studies and Long-Term Impact Assessment:** Conducting longitudinal studies to assess the long-term impact of implementing predictive models on mental health in the workplace will be crucial in understanding their efficacy and refining strategies.

- The future scope for predicting mental illness among working professionals using machine learning involves advancements in technology, ethical considerations, personalized interventions, and widespread application, aiming to create healthier and more supportive work environments while continuously refining and improving these predictive systems.

11.BIBILOGRAPHY

[1] G. Miner, L. Miner, M. Goldstein et al., Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research, Academic Press, Inc., Cambridge, MA, USA, 1st edition, 2016.

[2] M. Hamilton, "Development of a rating scale for primary depressive illness," British Journal of Social and Clinical Psychology, vol. 6, no. 4, pp. 278–296, 1967.

[3] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, Reston, VA, USA, 2013.

[4] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, 2015.

[5] F. Dabek and J. J. Caban, "A neural network based model for predicting psychological conditions," in Brain Informatics and Health, pp. 252–261, Springer International Publishing, Berlin, Germany, 2015.

[6] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, "Review of machine learning algorithms for diagnosing mental illness," Psychiatry Investigation, vol. 16, no. 4, pp. 262–269, 2019. Applied Computational Intelligence and Soft Computing 17

[7] E. G. Pintelas, T. Kotsilieris, I. E. Livieris, and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," in Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion—DSAI 2018,)essaloniki, Greece, 2018.

[8] R. B. Rutledge, A. M. Chekroud, and Q. J. Huys, "Machine learning and big data in psychiatry: toward clinical applications," Current Opinion in Neurobiology, vol. 55, pp. 152–159, 2019.

[9] 130 Executive Board, “Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social workers at the country level,” 2012, https:// apps.who.int/iris/handle/10665/23741.

[10] M. Hadzic, M. Chen, and T. S. Dillon, “Towards the mental health ontology,” in Proceedings of the 2008 IEEE International Conference on Bioinformatics and Biomedicine, Philadelphia, PA, USA, 2008.

[11] J. Diederich, A. Al-Ajmi, and P. Yellowlees, “Ex-ray: data mining and mental health,” Applied Soft Computing, vol. 7, no. 3, pp. 923–928, 2007.

[12] P. D. McGorry, A. R. Yung, L. J. Phillips et al., “Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms,” Archives of General Psychiatry, vol. 59, no. 10, pp. 921–928, 2002.

[13] T. I. Mueller, A. C. Leon, M. B. Keller et al., “Recurrence after recovery from major depressive disorder during 15 years of observational follow-up,” American Journal of Psychiatry, vol. 156, no. 7, pp. 1000–1006, 1999.

[14] L. L. Judd, P. J. Schettler, and H. S. Akiskal, “)e prevalence, clinical relevance, and public health significance of subthreshold depressions,” Psychiatric Clinics of North America, vol. 25, no. 4, pp. 685–698, 2002.

[15] J. Angst and R. Sellaro, “Historical perspectives and natural history of bipolar disorder,” Biological Psychiatry, vol. 48, no. 6, pp. 445–457, 2000.

[16] M. Olfson, R. C. Kessler, P. A. Berglund, and E. Lin, “Psychiatric disorder onset and first treatment contact in the United States and ontario,” American Journal of Psychiatry, vol. 155, no. 10, pp. 1415–1422, 1998.

[17] M. J. Zaki, J. T. L. Wang, and H. T. T. Toivonen, “Data mining in bioinformatics,” ACM SIGKDD Explorations Newsletter, vol. 5, no. 2, pp. 198-199, 2003.

11.APPENDIX

Model building:

1)Dataset
2)Google colab and VS code Application Building
    1.HTML file (index file, predict file)
    2.CSS file
    3.models in pickle format

SOURCE CODE:

HOME.HTML

```
<!DOCTYPE html>
<html>
<head>
  <title>Home Page</title>
  <style>
    body {
      background-image:
url("https://healthcare.utah.edu/sites/g/files/zrelqx136/files/media/images/2023/GettyImages-
1446012687-mental%20health.jpg");
      background-size: cover; /* This ensures the background image covers the entire page */
      background-repeat: no-repeat;
      background-attachment: fixed; /* This keeps the background fixed while scrolling */
      background-position: center center;
    }

    h1, p {
      color: black; /* Set the text color to white for better visibility on the background */
    }

    /* Center the content vertically and horizontally */
    body {
      display: flex;
      flex-direction: column;
      align-items: center;
```

```
            justify-content: center;
            /* height: 100vh; This makes the body take up the full viewport height */
        }

        button {
            padding: 10px 20px;
            background-color: #007bff; /* Blue button background color */
            color: white; /* Button text color */
            border: none;
            border-radius: 5px;
            cursor: pointer;
        }
    </style>
</head>
<body>
    <h1>Welcome to Mental Health Prediction</h1>
    <p>Click the "Proceed" button to continue.</p>
    <form action="/pred" method="GET" enctype="multipart/form-data">
        <button type="submit">Proceed</button>
    </form>
</body>
</html>
```

INDEX.HTML

```
<!DOCTYPE html>
<html>


<head>
    <title>Index Page</title>
     <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css" integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z"
crossorigin="anonymous" />
```

```
    <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js" integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
crossorigin="anonymous"></script>
    <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
integrity="sha384-
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
crossorigin="anonymous"></script>
    <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
integrity="sha384-B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
crossorigin="anonymous"></script>
    <style>
        body {
            background-image: url("https://healthwaymedical.com/wp-content/uploads/2020/09/mental-
health-and-heart1.jpg");
            background-size: cover; /* This ensures the background image covers the entire page */
            background-repeat: no-repeat;
            background-attachment: fixed; /* This keeps the background fixed while scrolling */
            background-position: center center;
        }
    </style>
</head>

<body>

    <div  class="background-image">

        <form action="/out" method="POST" enctype="multipart/form-data">
            <div class="m-5 text-center">
            <label for="age">Age:</label>
            <input type="text" name="age" id="age" /><br><br>
            <label for="gender">Gender:</label>
            <input type="text" name="gender" id="gender" /><br>
```

```
</div>
<div class="pl-5">
   <label for="self_employed">Are you self-employed?</label><br />
<select id="self_employed" name="self_employed">
   <option value="Yes">Yes</option>
   <option value="No">No</option>
</select><br><br/>
<label for="family_history">Do you have Family History?</label><br />
<select id="family_history" name="family_history">
   <option value="Yes">Yes</option>
   <option value="No">No</option>
</select>
<br /><br/>

<label for="work_interfere">How does work interfere with your mental health?</label><br
/>
<select id="work_interfere" name="work_interfere">
   <option value="Often">Often</option>
   <option value="Sometimes">Sometimes</option>
   <option value="Rarely">Rarely</option>
   <option value="Never">Never</option>
</select><br><br/>
<label for="no_employees">Number of Employees:</label><br />
<select id="no_employees" name="no_employees">
   <option value="1-5">1-5</option>
   <option value="6-25">6-25</option>
   <option value="26-100">26-100</option>
   <option value="100-500">100-500</option>

</select><br><br/>
<label for="remote_work">Do you have the option to work remotely?</label><br />
<select id="remote_work" name="remote_work">
```

```
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="tech_company">Is your company a tech company?</label><br />
        <select id="tech_company" name="tech_company">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="benefits">Does your company offer mental health benefits?</label><br />
        <select id="benefits" name="benefits">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="care_options">Are you offered care options by your company?</label><br />
        <select id="care_options" name="care_options">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="seek_help">Are you willing to seek help for mental health issues?</label><br
/>
        <select id="seek_help" name="seek_help">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="anonymity">Can you be anonymous when seeking help for mental health
issues?</label><br />
        <select id="anonymity" name="anonymity">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="leave">How easy is it for you to take medical leave for mental
health?</label><br />
        <select id="leave" name="leave">
            <option value="Very easy">Very easy</option>
```

```
            <option value="Somewhat easy">Somewhat easy</option>
            <option value="Don't know">Don't know</option>
            <option value="Very difficult">Very difficult</option>
        </select><br><br/>
        <label for="mental_health_consequence">Do you think discussing a mental health issue would have negative consequences for your career?</label><br />
        <select id="mental_health_consequence" name="mental_health_consequence">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="phys_health_consequence">Do you think discussing a physical health issue would have negative consequences for your career?</label><br />
        <select id="phys_health_consequence" name="phys_health_consequence">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="coworkers">Would you be willing to discuss a mental health issue with your coworkers?</label><br />
        <select id="coworkers" name="coworkers">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>
        <label for="supervisor">Would you be willing to discuss a mental health issue with your direct supervisor(s)?</label><br />
        <select id="supervisor" name="supervisor">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
        </select><br><br/>

        <label for="wellness_program">Does your company offer a wellness program?</label><br />
        <select id="wellness_program" name="wellness_program">
            <option value="Yes">Yes</option>
            <option value="No">No</option>
```

```
</select><br><br/>
<label for="mental_health_interview">Would you bring up a mental health issue with a potential employer in an interview?</label><br />
<select id="mental_health_interview" name="mental_health_interview">
  <option value="Yes">Yes</option>
  <option value="No">No</option>
</select><br><br/>
<label for="phys_health_interview">Would you bring up a physical health issue with a potential employer in an interview?</label><br />
<select id="phys_health_interview" name="phys_health_interview">
  <option value="Yes">Yes</option>
  <option value="No">No</option>
</select><br><br/>
<label for="mental_vs_physical">Do you feel that your employer takes mental health as seriously as physical health?</label><br />
<select id="mental_vs_physical" name="mental_vs_physical">
  <option value="Yes">Yes</option>
  <option value="No">No</option>
</select><br><br/>
<label for="obs_consequence">Have you observed any negative consequences for discussing a mental health issue in your workplace?</label><br />
<select id="obs_consequence" name="obs_consequence">
  <option value="Yes">Yes</option>
  <option value="No">No</option>
</select><br><br/>
</div>


<div class="text-center">
<button class="btn btn-primary" type="submit">Predict</button>
</div>
  </form>
</div>
```

```
</body>

</html>
<!DOCTYPE html>
<html>
<head>
    <title>Output Page</title>
    <style>
        body {
            background-image: url("https://vidhilegalpolicy.in/wp-content/uploads/2021/10/mental-health-wellness-during-covid-19.jpg");
            background-size: cover;
            background-repeat: no-repeat;
            background-attachment: fixed;
            background-position: center center;
            margin: 0; /* Remove default margin to ensure full coverage */
            display: flex;
            justify-content: center;
            align-items: center;
            height: 100vh;
        }

        .result-box {
            background-color: rgba(255, 255, 255, 0.8); /* Semi-transparent white background */
            border: 2px solid #007bff; /* Blue border */
            padding: 20px;
            text-align: center;
        }

        h1 {
            color: #007bff; /* Blue text color for the header */
        }
```

```
        p {
            color: #333; /* Dark text color for the prediction result */
        }

        a {
            color: #007bff; /* Blue text color for the link */
        }
    </style>
</head>
<body>
    <div class="result-box">
        <h1>Model Prediction Result</h1>
        <p>{{ y }}</p>
        <a href="/">Go back to the home page</a>
    </div>
</body>
</html>
```

APP.PY

```
from flask import Flask, render_template, request
import pickle
import joblib
import pandas as pd

app = Flask(_name_, template_folder='template')
# Assuming that you have your model training code in a variable named 'clf'
model = pickle.load(open("C:\\Users\\Ganesh\\Desktop\\min\\flask\\model.pkl", "rb"))

ct = joblib.load("C:\\Users\\Ganesh\\Desktop\\min\\flask\\feature_values")
```

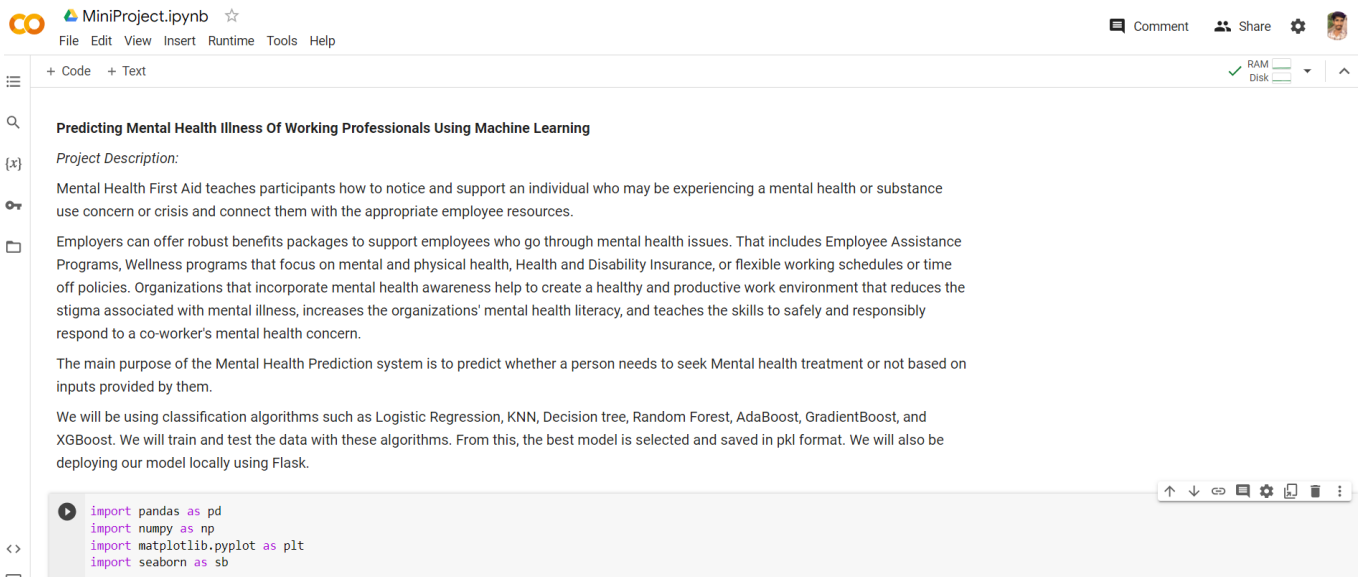
```
gender_mapping = {
    "Male": ["Male", "male", "M", "m", "Male", "Cis Male", "Man", "cis male", "Mail", "Male-ish", "Male (CIS)", "Cis Man", "msle", "Malr", "Mal", "maile", "Make"],
    "Female": ["Female", "female", "F", "f", "Woman", "Female", "femail", "Cis Female", "cis-female/femme", "Femake", "Female (cis)", "woman"],
    "Non-Binary": ["Female (trans)", "queer/she/they", "non-binary", "fluid", "queer", "Androgyne", "Trans-female", "male leaning androgynous", "Agender", "A little about you", "Nah", "All", "ostensibly male, unsure what that really means", "Genderqueer", "Enby", "p", "Neuter", "something kinda male?", "Guy (-ish) ^_^", "Trans Woman"]
}

@app.route('/')
def home():
    return render_template("home.html")

@app.route('/pred')
def predict():
    return render_template("index.html")

@app.route('/out', methods=["POST"])
def output():

    age = request.form["age"]
    gender = request.form["gender"]
    self_employed = request.form["self_employed"]
    family_history = request.form["family_history"]
    work_interfere = request.form["work_interfere"]
    no_employees = request.form["no_employees"]
    remote_work = request.form["remote_work"]
    tech_company = request.form["tech_company"]
    benefits = request.form["benefits"]
    care_options = request.form["care_options"]
    wellness_program = request.form["wellness_program"]
    seek_help = request.form["seek_help"]
```

```
    anonymity = request.form["anonymity"]
    leave = request.form["leave"]
    mental_health_consequence = request.form["mental_health_consequence"]
    phys_health_consequence = request.form["phys_health_consequence"]
    coworkers = request.form["coworkers"]
    supervisor = request.form["supervisor"]
    mental_health_interview = request.form["mental_health_interview"]
    phys_health_interview = request.form["phys_health_interview"]
    mental_vs_physical = request.form["mental_vs_physical"]
    obs_consequence = request.form["obs_consequence"]
    for mapped_gender, aliases in gender_mapping.items():
        if gender.lower() in [alias.lower() for alias in aliases]:
            gender = mapped_gender
            break



    data = [
        [age, gender, self_employed, family_history, work_interfere, no_employees, remote_work,
tech_company, benefits,
         care_options, wellness_program, seek_help, anonymity, leave, mental_health_consequence,
         phys_health_consequence, coworkers, supervisor, mental_health_interview,
phys_health_interview,
         mental_vs_physical, obs_consequence]]


    feature_cols = ['Age', 'Gender', 'self_employed', 'family_history', 'work_interfere', 'no_employees',
'remote_work',
                    'tech_company', 'benefits', 'care_options', 'wellness_program', 'seek_help', 'anonymity',
'leave',
                    'mental_health_consequence', 'phys_health_consequence', 'coworkers', 'supervisor',
                    'mental_health_interview', 'phys_health_interview', 'mental_vs_physical',
'obs_consequence']



    pred = model.predict(ct.transform(pd.DataFrame(data, columns=feature_cols)))
```

```
    pred = pred[0]

    if pred:
        return render_template("output.html", y="This person requires mental health treatment")
    else:
        return render_template("output.html", y="This person doesn't require mental health treatment")

if _name_ == '_main_':
    app.run(debug=True)
```

CODE SNIPPETS

Model Building:



\

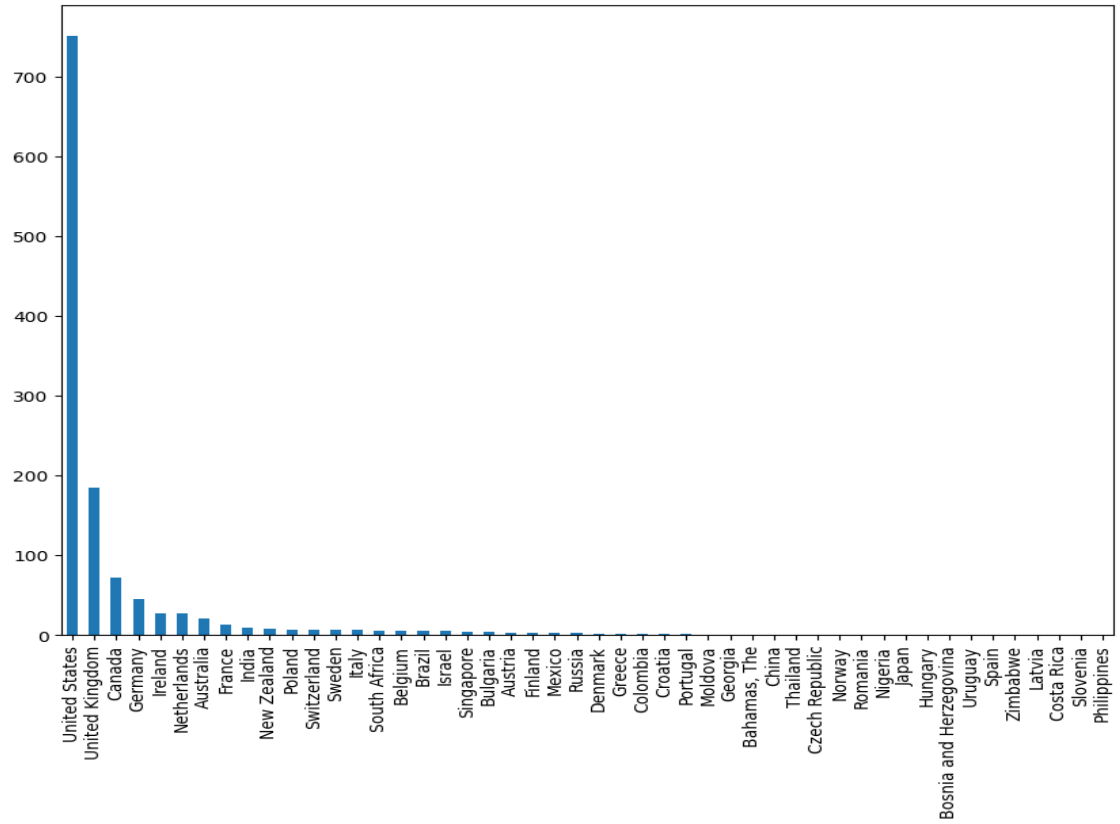
```
data.head()
```

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence	phys_health_consequence	cow
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	No	No	
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	Maybe	No	
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	No	No	
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Yes	Yes	
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	No	No	

5 rows × 27 columns

```
[ ] data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
 #   Column                     Non-Null Count  Dtype 
---  ------                     --------------  ----- 
 0   Timestamp                  1259 non-null   object
 1   Age                        1259 non-null   int64 
 2   Gender                     1259 non-null   object
 3   Country                    1259 non-null   object
 4   state                      744 non-null    object
 5   self_employed              1241 non-null   object
 6   family_history             1259 non-null   object
 7   treatment                  1259 non-null   object
 8   work_interfere             995 non-null    object
 9   no_employees               1259 non-null   object
 10  remote_work                1259 non-null   object
 11  tech_company               1259 non-null   object
 12  benefits                   1259 non-null   object
 13  care_options               1259 non-null   object
 14  wellness_program           1259 non-null   object
 15  seek_help                  1259 non-null   object
 16  anonymity                  1259 non-null   object
 17  leave                      1259 non-null   object
 18  mental_health_consequence  1259 non-null   object
 19  phys_health_consequence    1259 non-null   object
 20  coworkers                  1259 non-null   object
 21  supervisor                 1259 non-null   object
 22  mental_health_interview    1259 non-null   object
 23  phys_health_interview      1259 non-null   object
 24  mental_vs_physical         1259 non-null   object
 25  obs_consequence            1259 non-null   object
```



```
data.isnull().sum()
```

```
Age                           0
Gender                        0
self_employed                18
family_history                0
treatment                     0
work_interfere              264
no_employees                  0
remote_work                   0
tech_company                  0
benefits                      0
care_options                  0
wellness_program              0
seek_help                     0
anonymity                     0
leave                         0
mental_health_consequence     0
phys_health_consequence       0
coworkers                     0
supervisor                    0
mental_health_interview       0
phys_health_interview         0
mental_vs_physical            0
obs_consequence               0
dtype: int64
```

```
[ ] data['self_employed'].value_counts()
```

```
No     1095
Yes     146
Name: self_employed, dtype: int64
```

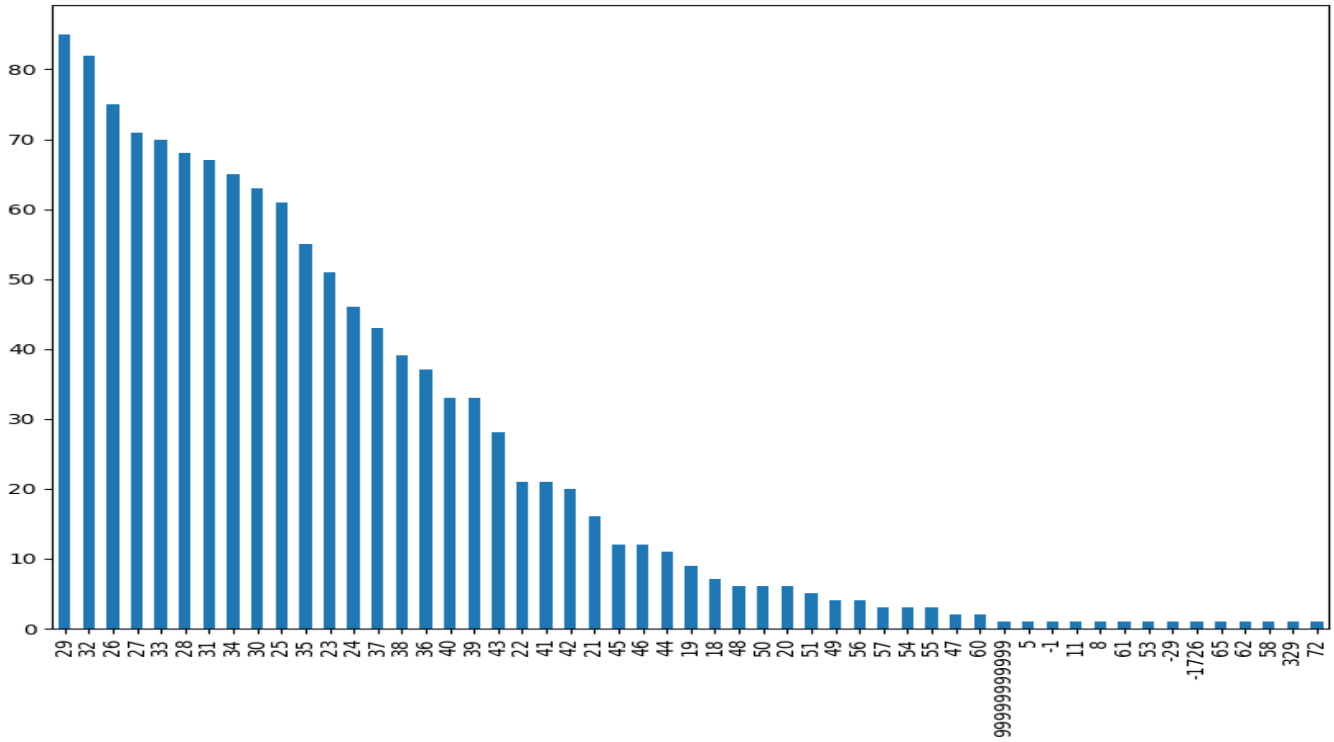
```
data['self_employed'].fillna('No',inplace=True)
```

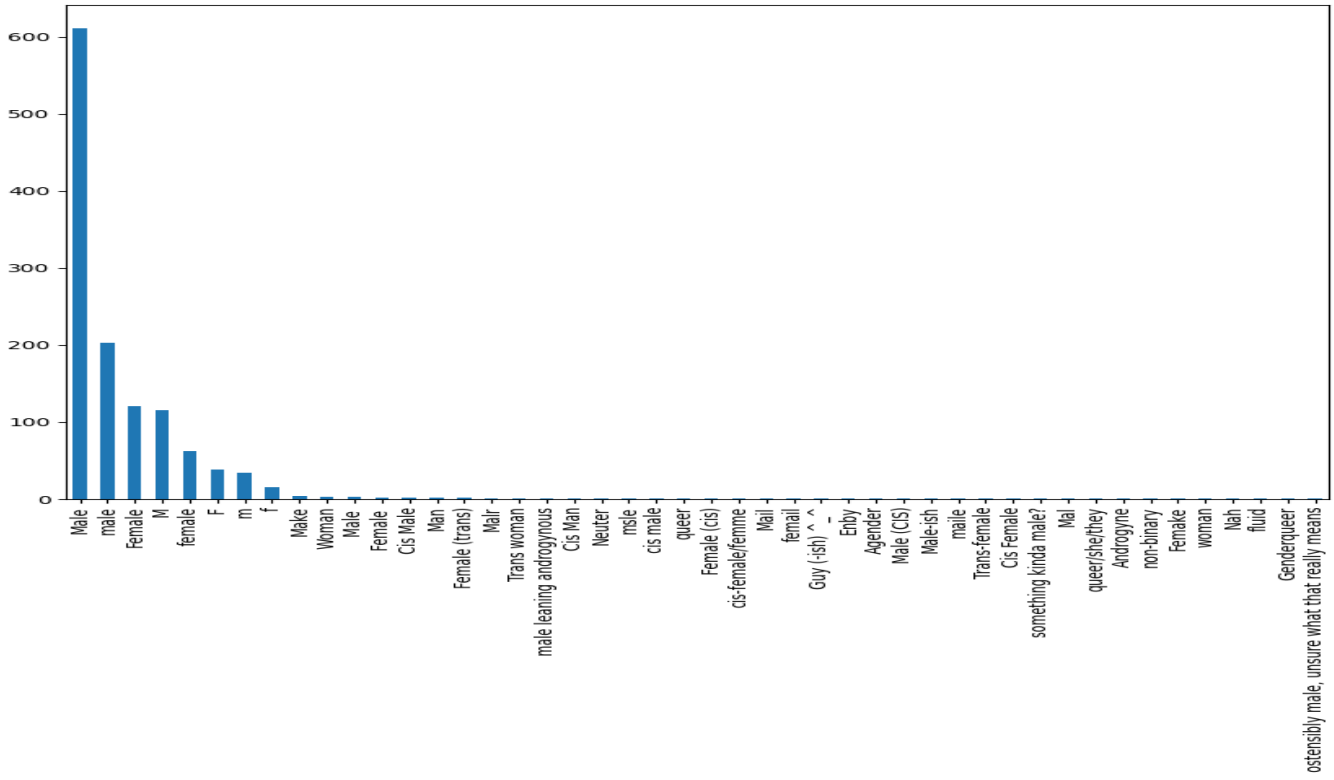
+ Code + Text

```
[ ] data['work_interfere'].value_counts()
```

```
Sometimes    465
Never        213
Rarely       173
Often        144
Name: work_interfere, dtype: int64
```

```
[ ] data['work_interfere'].fillna('N/A',inplace=True)
```

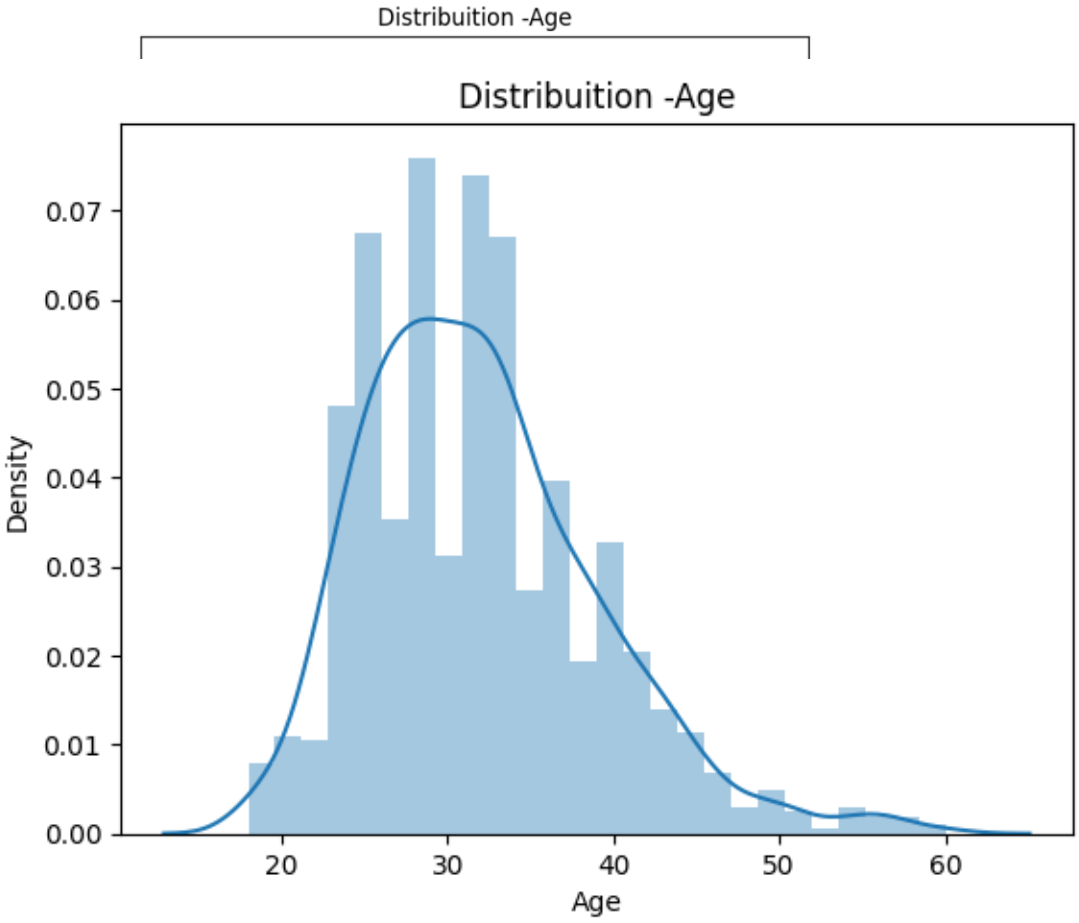

0
10
20
30
40
50
60
70
80
29
32
26
27
33
28
31
34
30
25
35
23
24
37
38
36
40
39
43
22
41
42
21
45
46
44
19
18
48
50
20
51
49
56
57
54
55
47
60
99999999999
5
-1
11
8
61
53
-29
-1726
65
62
58
329
72

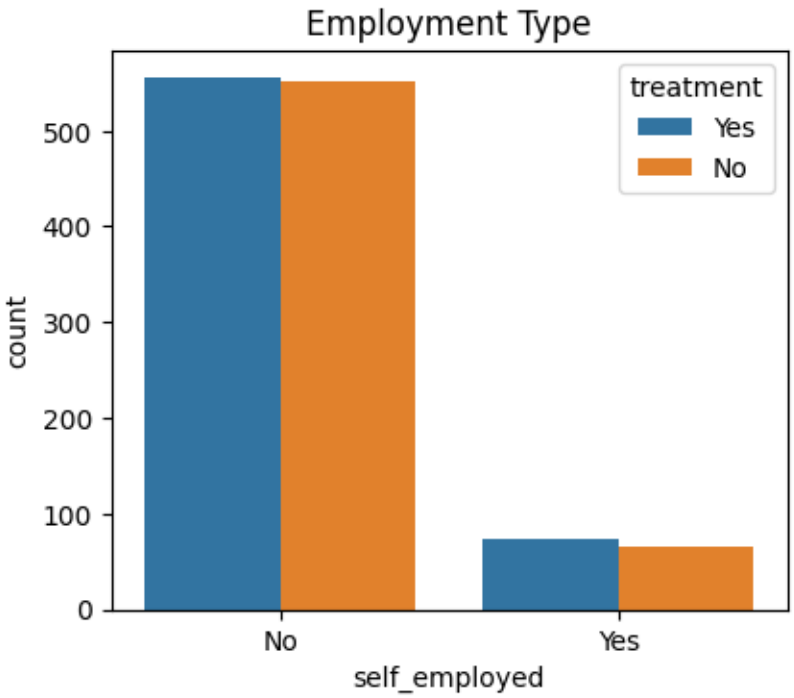

0
100
200
300
400
500
600
Male
male
Female
M
female
F
m
f
Make
Woman
Male
Female
Cis Male
Man
Female (trans)
Malr
Trans woman
male leaning androgynous
Cis Man
Neuter
msle
cis male
queer
Female (cis)
cis-female/femme
Mail
femail
Guy (-ish) ^_^
Enby
Agender
Male (CIS)
Male-ish
maile
Trans-female
Cis Female
something kinda male?
Mal
queer/she/they
Androgyne
non-binary
Femake
woman
Nah
fluid
Genderqueer
ostensibly male, unsure what that really means

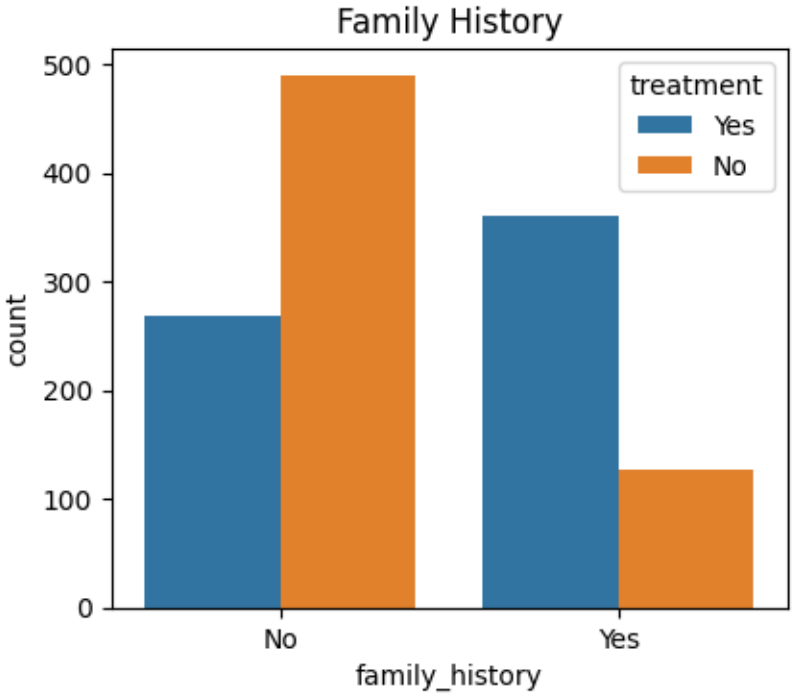
```
[ ] data['Gender'].replace(['Male','male','M','m','Male','Cis Male','Man','cis male','Mail','Male-ish','Male (CIS)','Cis Man','ms
    data['Gender'].replace(['Female','female','F','f','Woman','Female','femail','Cis Female','cis-female/femme','Femake','Female
    data["Gender"].replace(['Female (trans)','queer/she/they','non-binary','fluid','queer','Androgyne','Trans-female','male leani
```

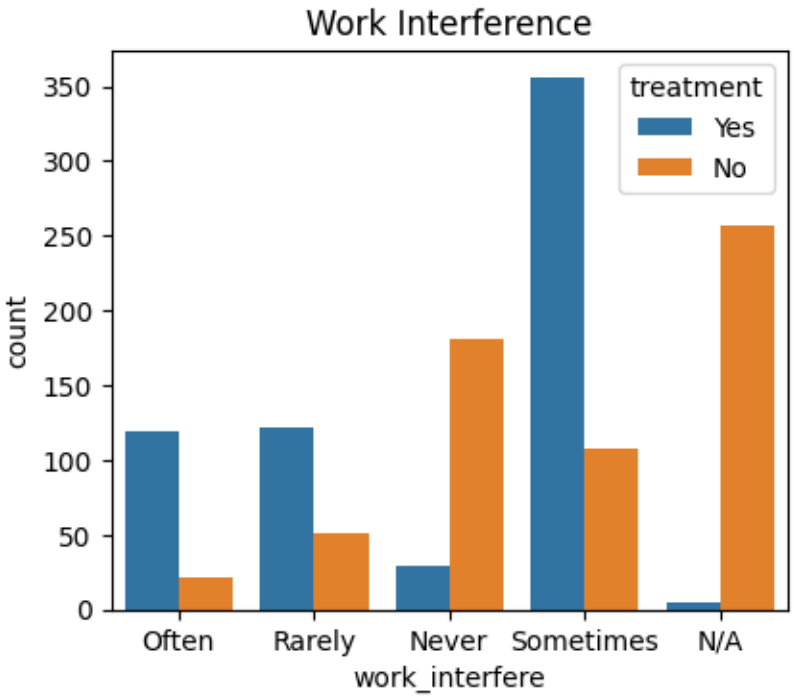
```
sb.distplot(data["Age"])
plt.title("Distribuition -Age")
plt.xlabel("Age")
```

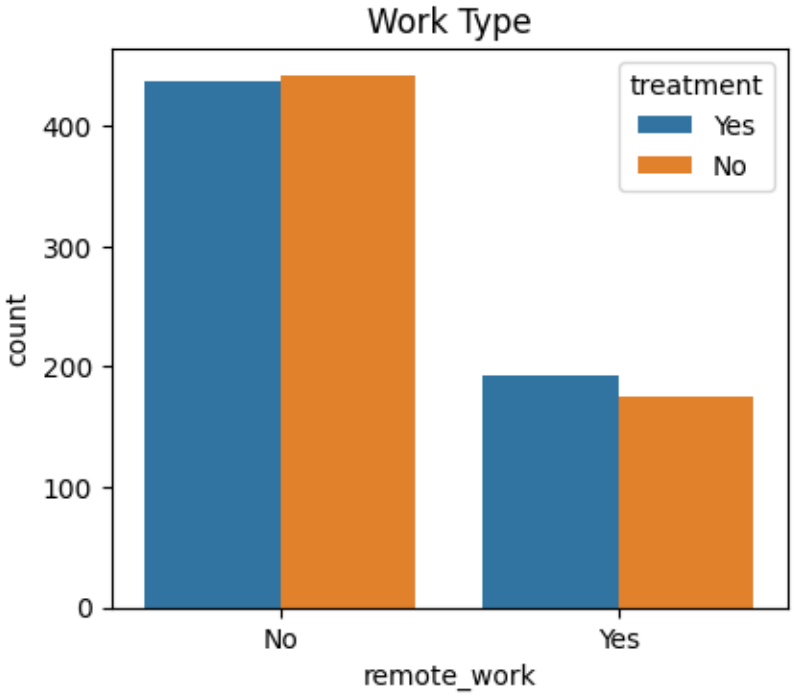
```
<ipython-input-18-ffe2a170eb87>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sb.distplot(data["Age"])
Text(0.5, 0, 'Age')
```

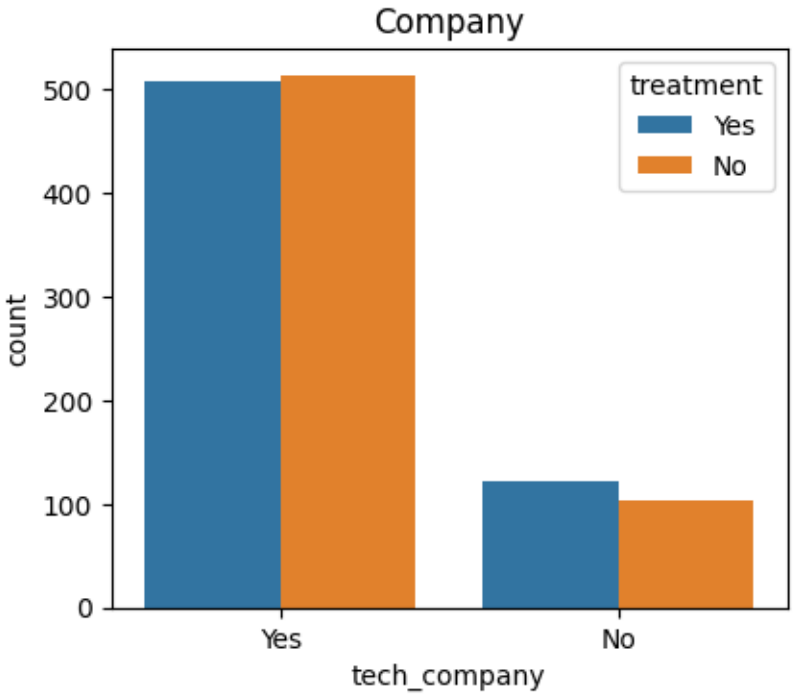


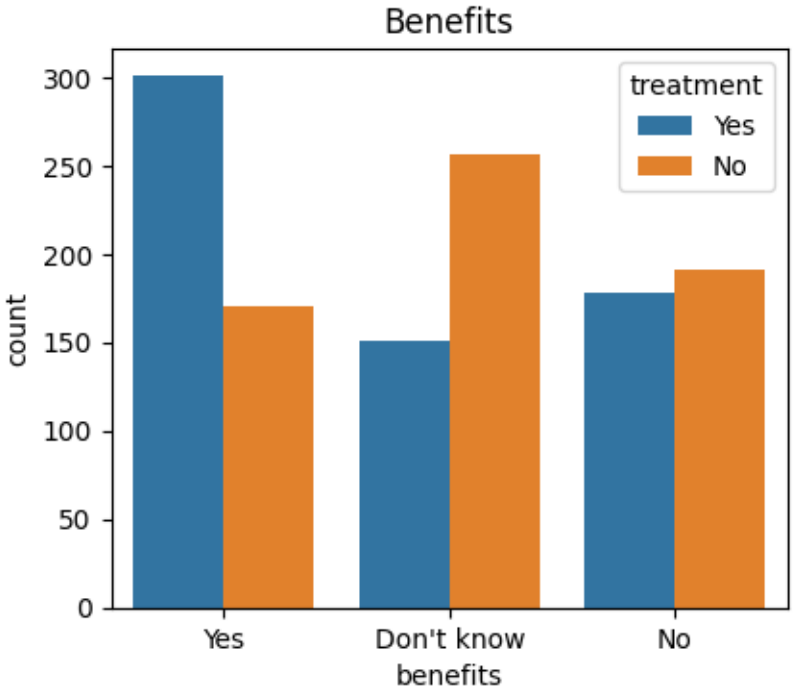

Employment Type
treatment
Yes
No
500
400
300
200
100
0
count
No
Yes
self_employed

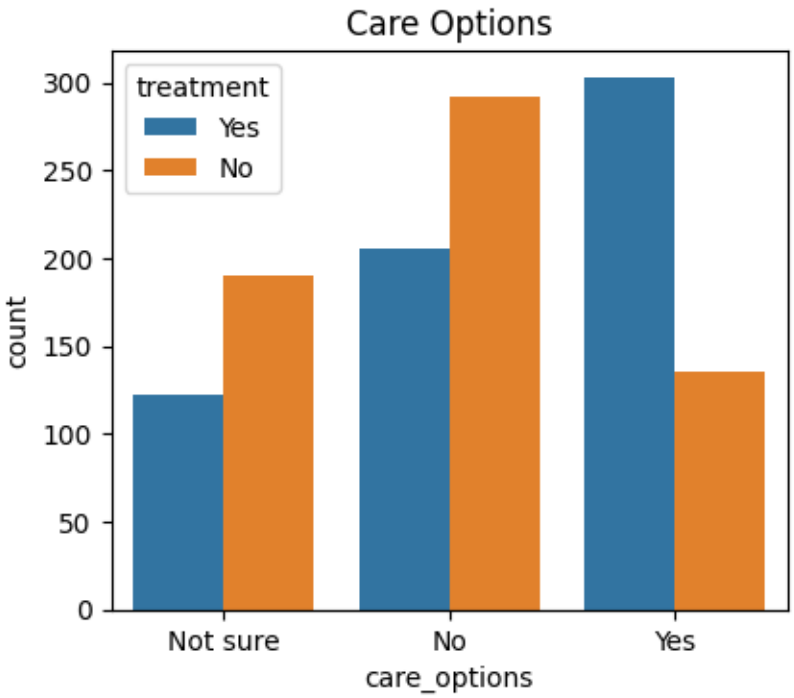

Family History
treatment
Yes
No
500
400
300
200
100
0
count
No
Yes
family_history

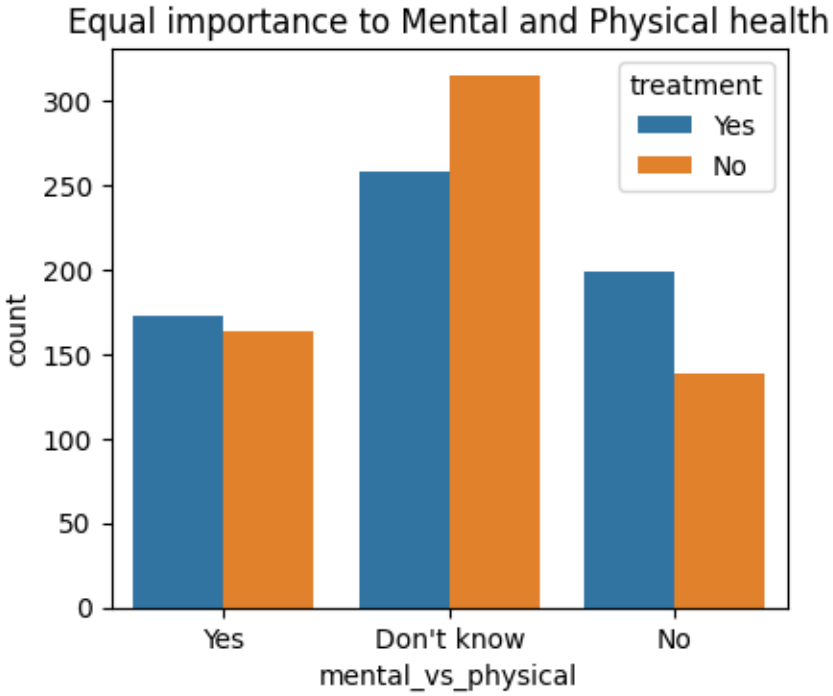

Work Interference
treatment
Yes
No
count
0
50
100
150
200
250
300
350
Often
Rarely
Never
Sometimes
N/A
work_interfere

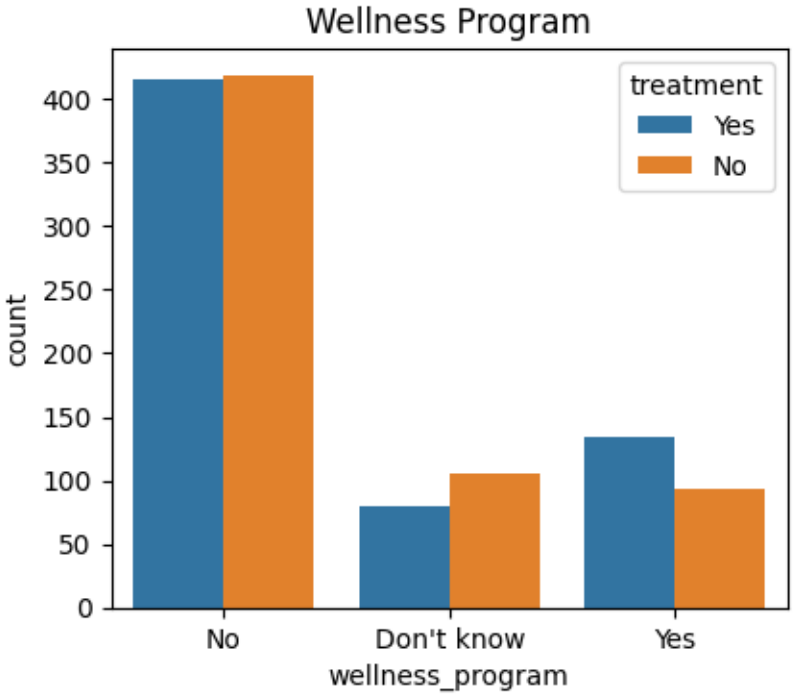

Work Type
treatment
Yes
No
count
0
100
200
300
400
No
Yes
remote_work

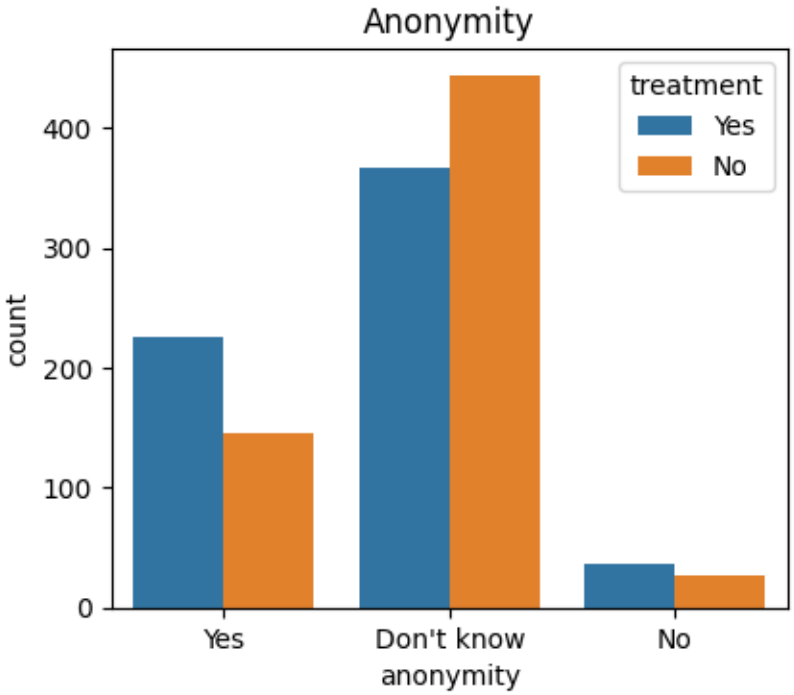

Company
treatment
Yes
No
500
400
300
200
100
0
count
Yes
No
tech_company

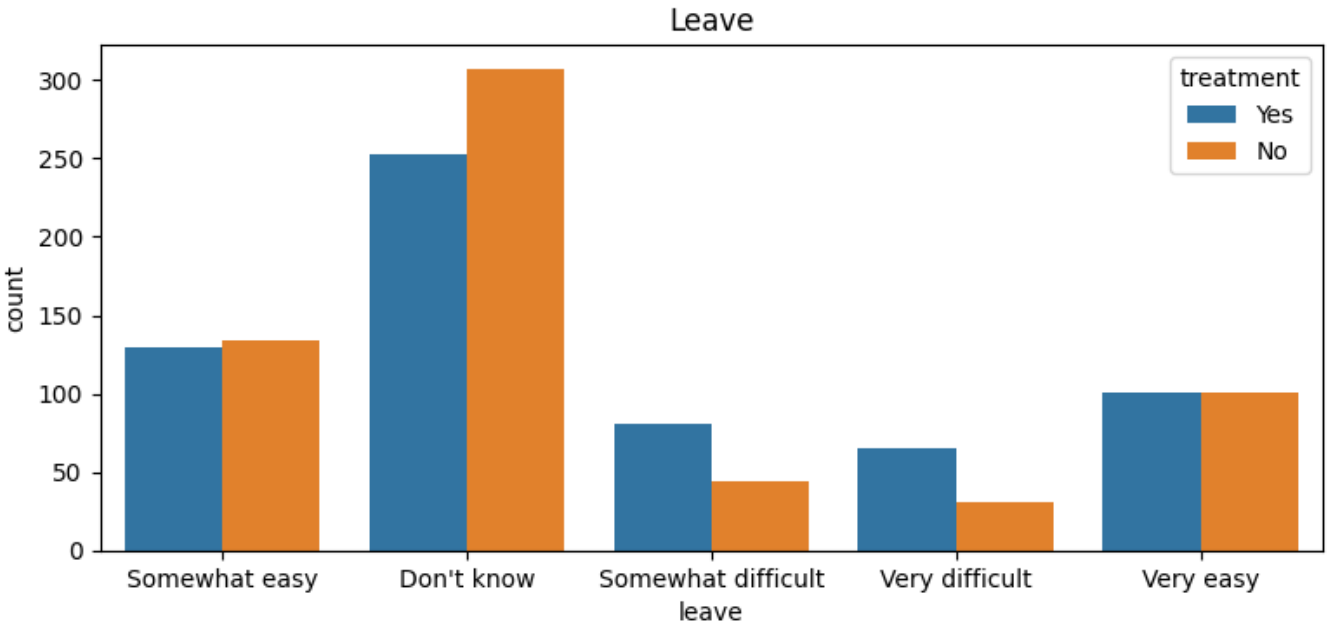

Benefits
treatment
Yes
No
300
250
200
150
100
50
0
count
Yes
Don't know
No
benefits

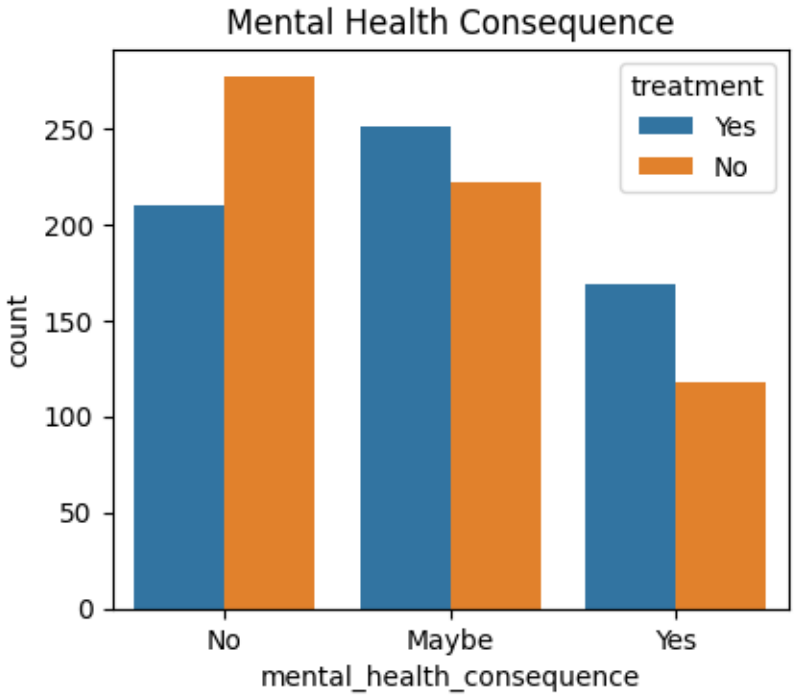

Care Options
treatment
Yes
No
300
250
200
150
100
50
0
count
Not sure
No
Yes
care_options

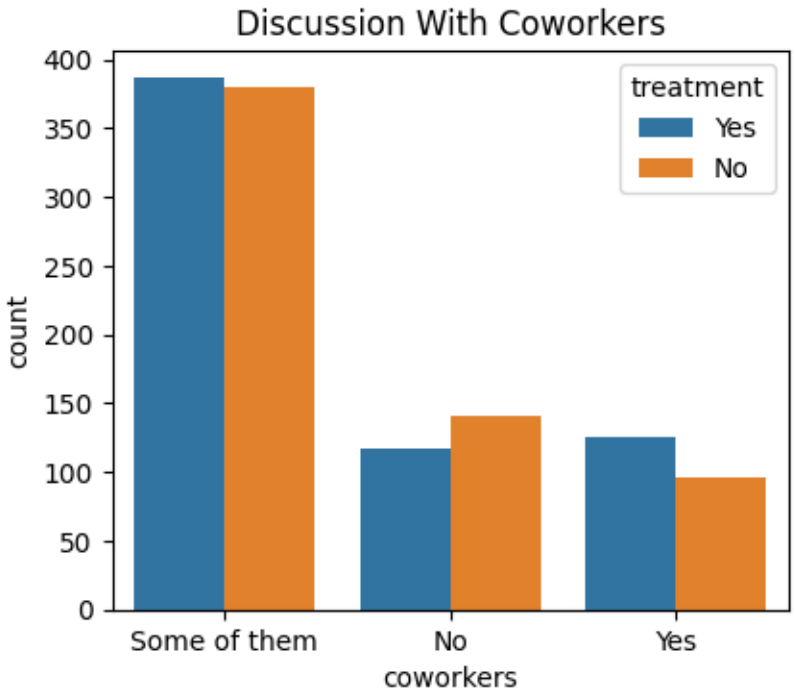

Equal importance to Mental and Physical health
treatment
Yes
No
300
250
200
150
100
50
0
count
Yes
Don't know
No
mental_vs_physical

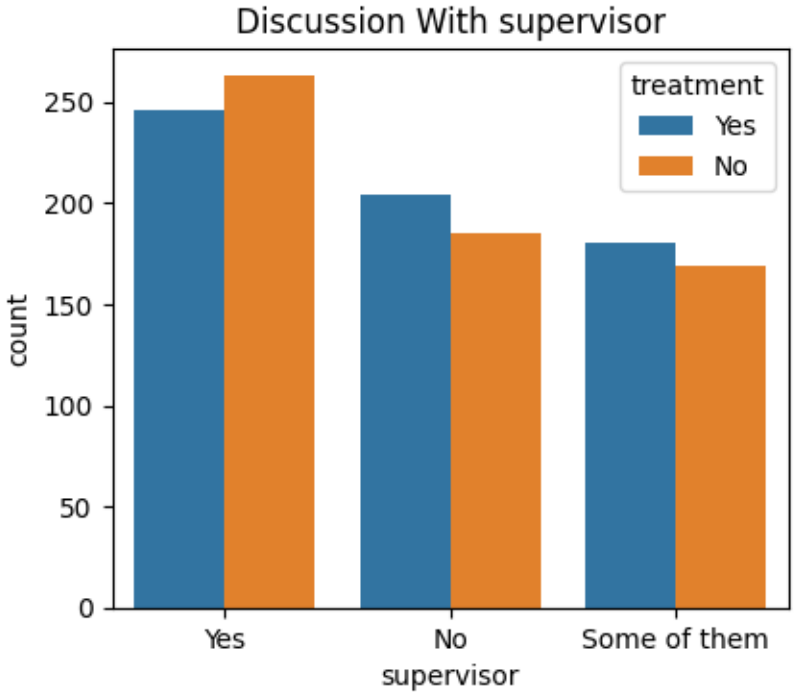

Wellness Program
treatment
Yes
No
count
0
50
100
150
200
250
300
350
400
No
Don't know
Yes
wellness_program

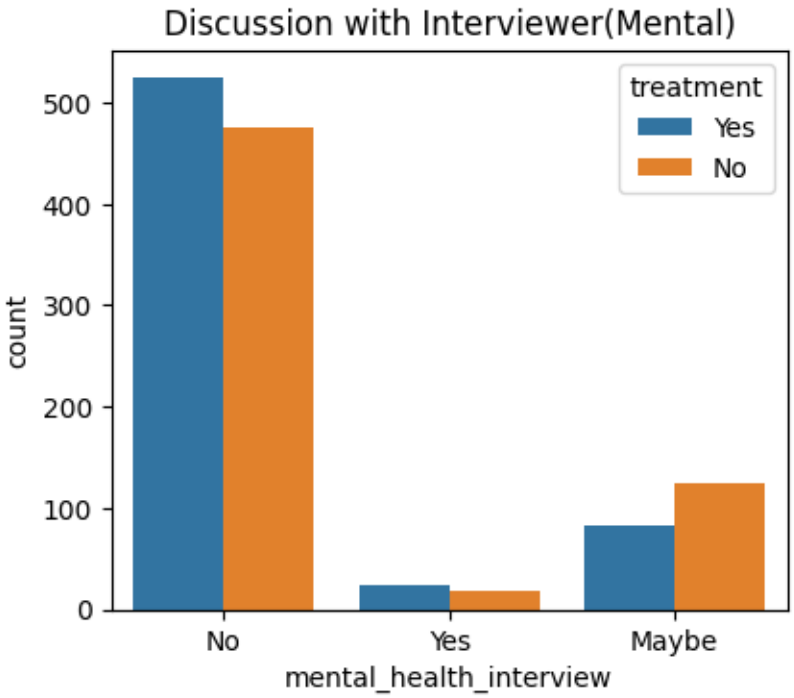

Anonymity
treatment
Yes
No
count
0
100
200
300
400
Yes
Don't know
No
anonymity

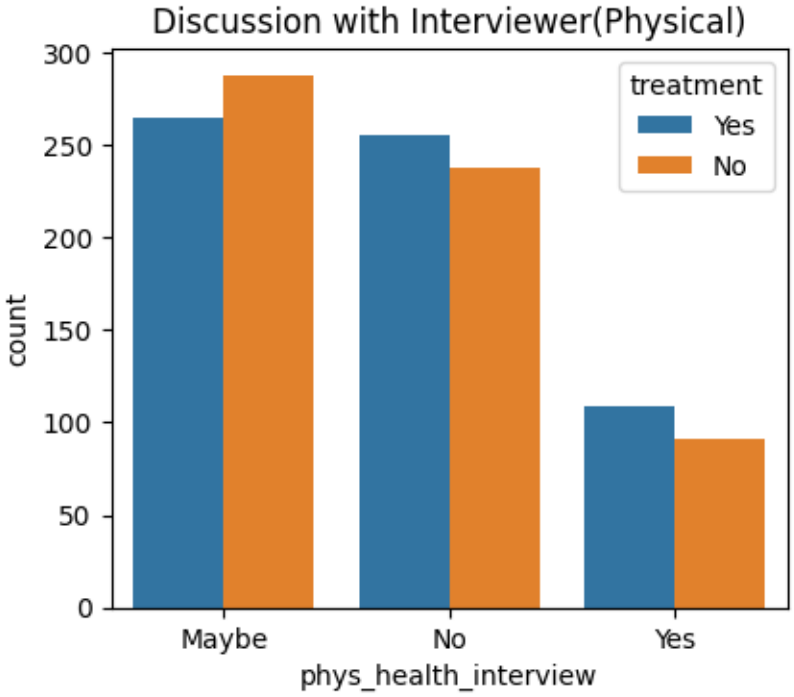

Leave
treatment
Yes
No
count
300
250
200
150
100
50
0
Somewhat easy
Don't know
Somewhat difficult
Very difficult
Very easy
leave

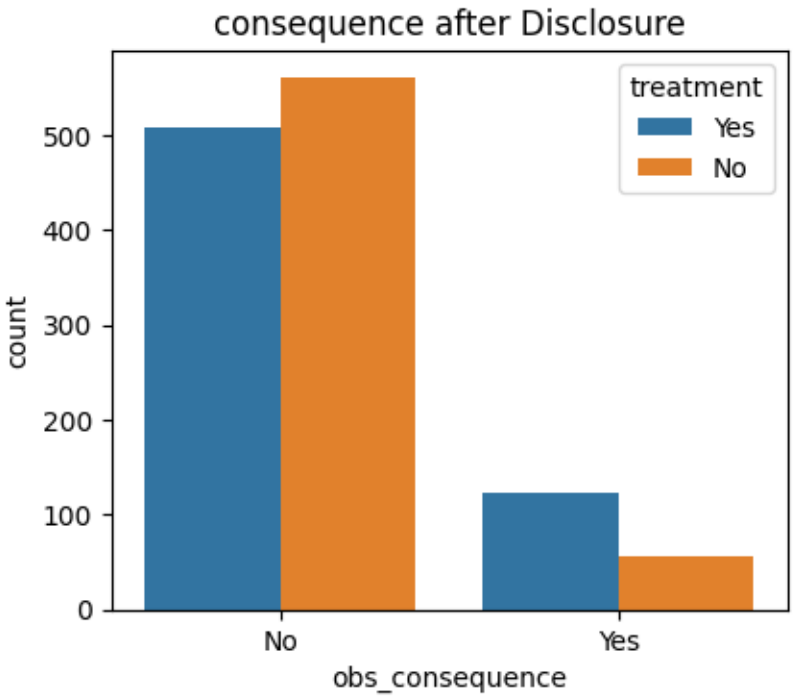

Mental Health Consequence
treatment
Yes
No
count
250
200
150
100
50
0
No
Maybe
Yes
mental_health_consequence

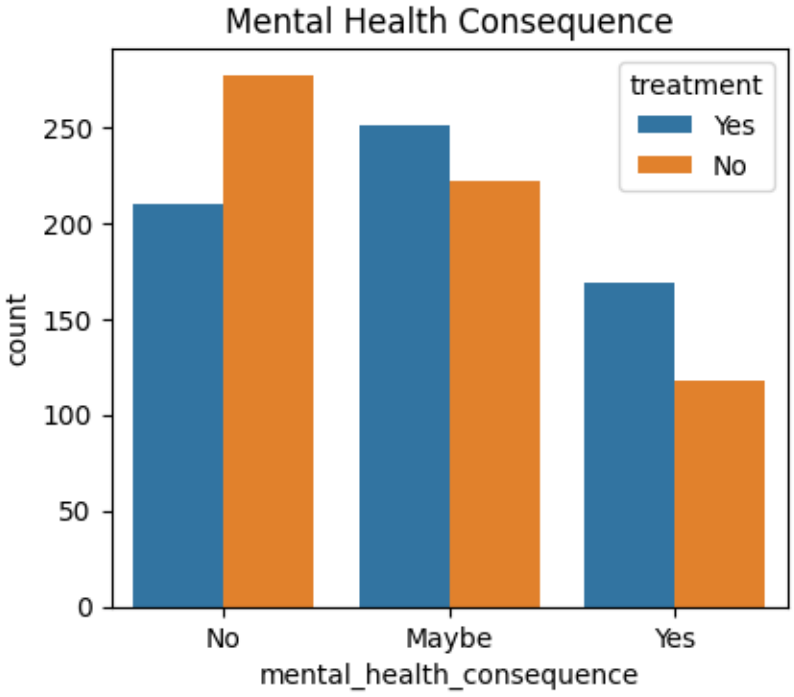

Discussion With Coworkers
treatment
Yes
No
count
400
350
300
250
200
150
100
50
0
Some of them
No
Yes
coworkers

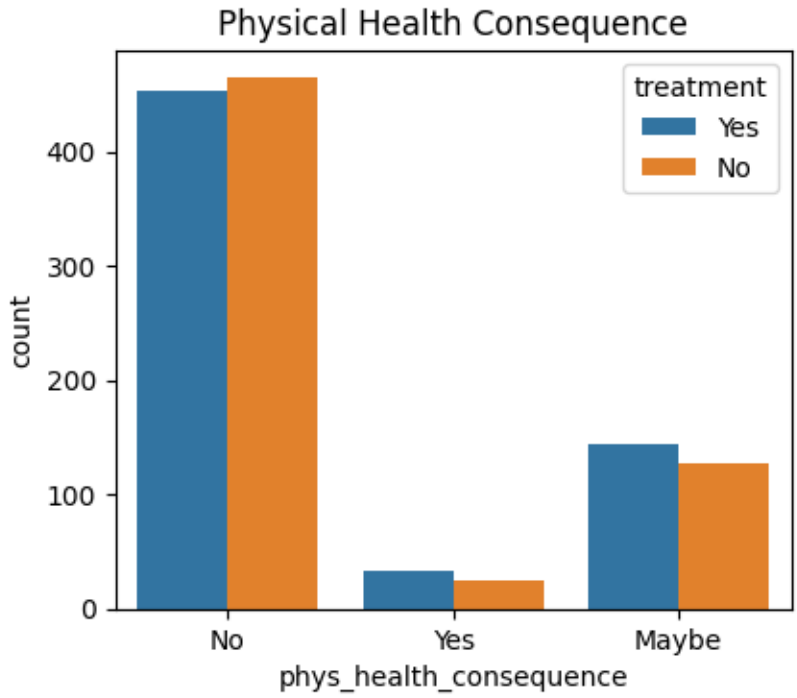

Discussion With supervisor
treatment
Yes
No
count
250
200
150
100
50
0
Yes
No
Some of them
supervisor

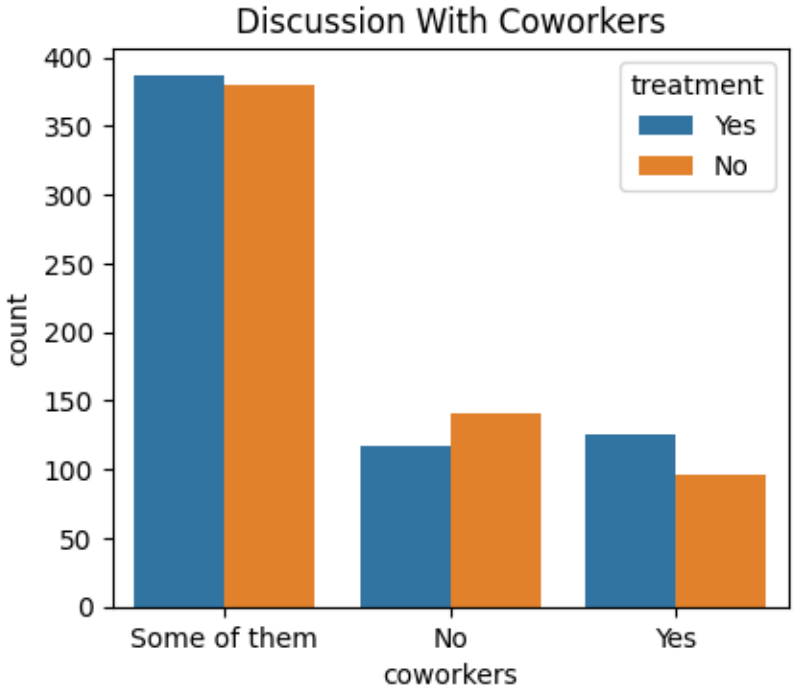

Discussion with Interviewer(Mental)
treatment
Yes
No
count
500
400
300
200
100
0
No
Yes
Maybe
mental_health_interview

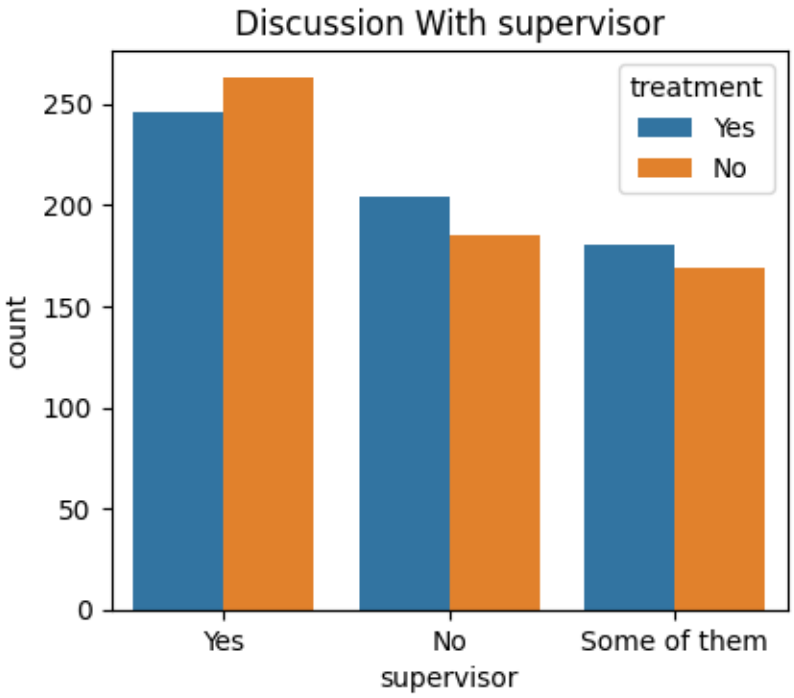

Discussion with Interviewer(Physical)
treatment
Yes
No
count
300
250
200
150
100
50
0
Maybe
No
Yes
phys_health_interview

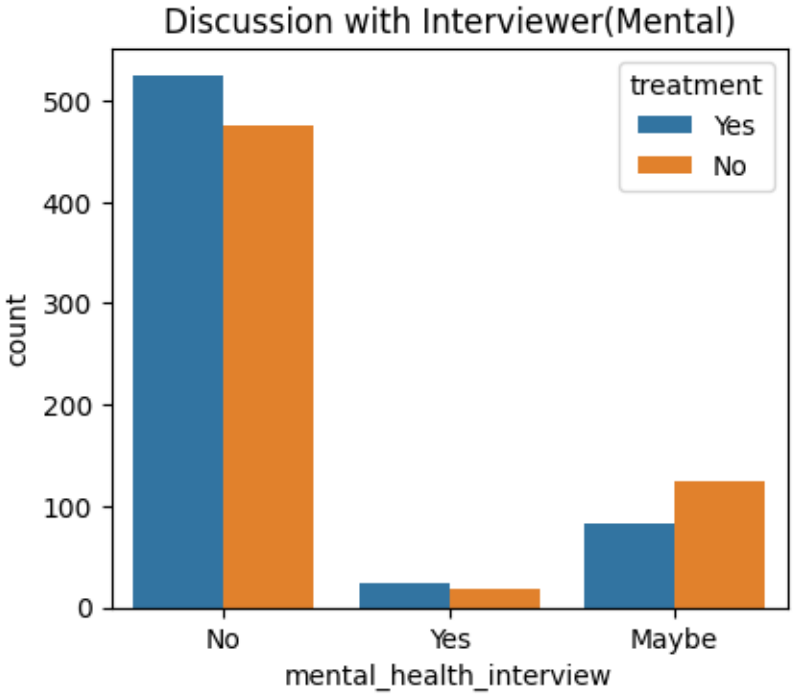

consequence after Disclosure
treatment
Yes
No
500
400
300
200
100
0
count
No
Yes
obs_consequence

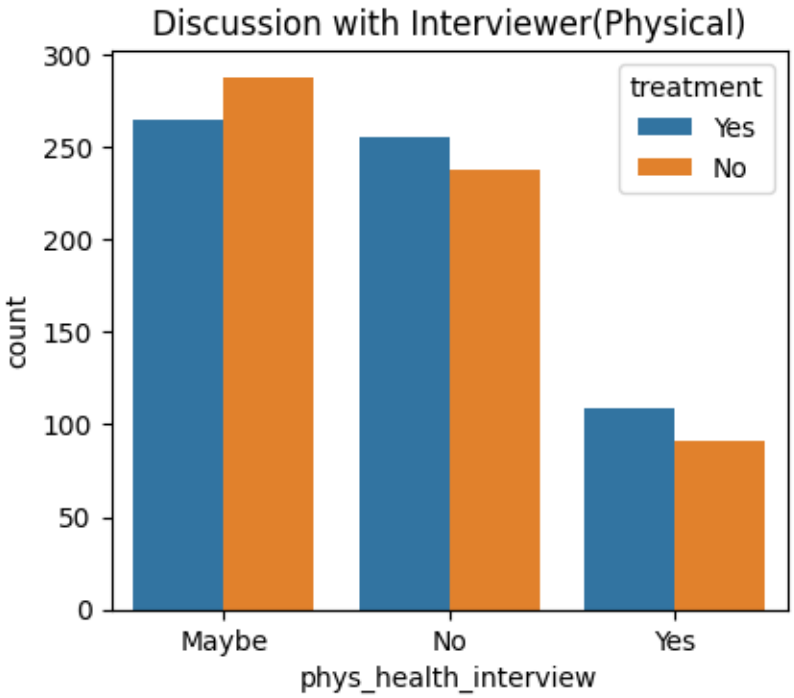

Mental Health Consequence
treatment
Yes
No
250
200
150
100
50
0
count
No
Maybe
Yes
mental_health_consequence

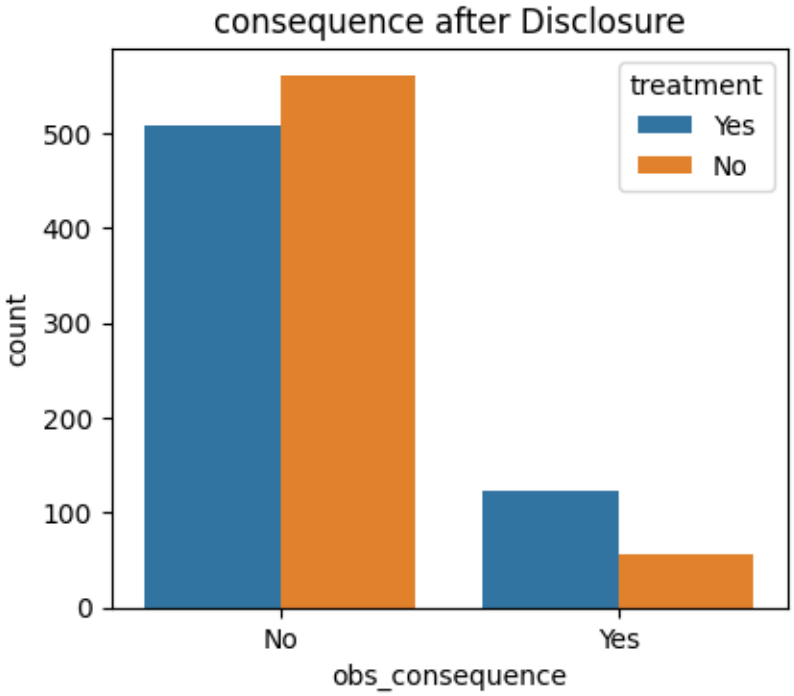

Physical Health Consequence
treatment
Yes
No
400
300
200
100
0
count
No
Yes
Maybe
phys_health_consequence


Discussion With Coworkers
treatment
Yes
No
400
350
300
250
200
150
100
50
0
count
Some of them
No
Yes
coworkers


Discussion With supervisor
treatment
Yes
No
250
200
150
100
50
0
count
Yes
No
Some of them
supervisor


Discussion with Interviewer(Mental)
treatment
Yes
No
500
400
300
200
100
0
count
No
Yes
Maybe
mental_health_interview


Discussion with Interviewer(Physical)
treatment
Yes
No
count
300
250
200
150
100
50
0
Maybe
No
Yes
phys_health_interview


consequence after Disclosure
treatment
Yes
No
count
500
400
300
200
100
0
No
Yes
obs_consequence

```
data.describe(include='all')
```

	Age	Gender	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	benefits	...	anonymity	leave	mental_health_consequence
count	1247.000000	1247	1247	1247	1247	1247	1247	1247	1247	1247	...	1247	1247	1247
unique	NaN	7	2	2	2	5	6	2	2	3	...	3	5	3
top	NaN	Male	No	No	Yes	Sometimes	6-25	No	Yes	Yes	...	Don't know	Don't know	No
freq	NaN	980	1107	759	630	463	288	879	1023	471	...	811	560	487
mean	31.971131	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
std	7.052598	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
min	18.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
25%	27.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
50%	31.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
75%	36.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
max	60.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

11 rows × 23 columns

```
[ ] x=data.drop('treatment',axis=1)
    y=data['treatment']
```

```
[ ] from sklearn.compose import ColumnTransformer
    from sklearn.preprocessing import LabelEncoder,OrdinalEncoder
```

```
[ ] x=data.drop('treatment',axis=1)
    y=data['treatment']
```

```
[ ] ct=ColumnTransformer([('oe',OrdinalEncoder(),['Gender','self_employed','family_history','work_interfere','no_employees','remote_work','tech_company','benefits','care_options','wellnes
```

```
[ ] x=ct.fit_transform(x)
```

```
[ ] data.shape

    (1247, 23)
```

```
[ ] le=LabelEncoder()
    y=le.fit_transform(y)
```

```
[ ] import joblib
    joblib.dump(ct,'feature_values')

    ['feature_values']
```

```
[ ] from sklearn.model_selection import train_test_split
    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=49)
```

```
[ ] x_train.shape,x_test.shape,y_train.shape,y_test.shape

    ((872, 22), (375, 22), (872,), (375,))
```

```
[ ] pip install xgboost

    Requirement already satisfied: xgboost in /usr/local/lib/python3.10/dist-packages (2.0.1)
    Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from xgboost) (1.23.5)
    Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from xgboost) (1.11.3)
```

```
[ ] from sklearn.linear_model import LogisticRegression

    from sklearn.tree import DecisionTreeClassifier

    from sklearn.neighbors import KNeighborsClassifier

    from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier

    from xgboost.sklearn import XGBClassifier

    from sklearn.metrics import accuracy_score, roc_curve, confusion_matrix, classification_report, auc
```

```
model_dict = {}


model_dict['Logistic regression']= LogisticRegression (solver='liblinear', random_state=49)

model_dict['KNN Classifier'] = KNeighborsClassifier()

model_dict['Decision Tree Classifier'] = DecisionTreeClassifier(random_state=49)
model_dict['Random Forest Classifier'] = RandomForestClassifier(random_state=49)

model_dict['AdaBoost Classifier'] = AdaBoostClassifier(random_state=49)

model_dict['Gradient Boosting Classifier'] = GradientBoostingClassifier(random_state=49)

model_dict['XGB Classifier'] =XGBClassifier(random_state=49)
```

```
def model_test(x_train, x_test, y_train, y_test, model, model_name):
    model.fit(x_train,y_train)
    y_pred=model.predict(x_test)
    accuracy=accuracy_score(y_test,y_pred)
    print('==================================={}=================='.format(model_name))
    print('score is :{}'.format(accuracy))
    print()
```

```
[ ] for model_name,model in model_dict.items():
        model_test(x_train,x_test,y_train,y_test,model,model_name)
```

```
==================================Logistic regression==================
score is :0.8506666666666667

==================================KNN Classifier==================
score is :0.792

==================================Decision Tree Classifier==================
score is :0.7946666666666666

==================================Random Forest Classifier==================
score is :0.848

==================================AdaBoost Classifier==================
score is :0.864

==================================Gradient Boosting Classifier==================
score is :0.832

==================================XGB Classifier==================
score is :0.8293333333333334
```

```
[ ] abc = AdaBoostClassifier(random_state=99)

    abc.fit(x_train,y_train)

    pred_abc = abc.predict(x_test)

    print('Accuracy of AdaBoost=' ,accuracy_score (y_test, pred_abc))
```

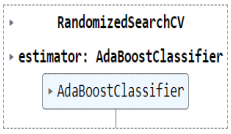
```
Accuracy of AdaBoost= 0.864
```

```
from sklearn.model_selection import RandomizedSearchCV

params_abc = {'n_estimators': [int (x) for x in np.linspace(start = 1, stop = 50, num= 15)],

'learning_rate': [(0.97 + x/100) for x in range(0, 8)],}



abc_random = RandomizedSearchCV(random_state=49, estimator=abc, param_distributions = params_abc,n_iter =50, cv=5,n_jobs=-1)
```

```
[ ] params_abc
```

```
{'n_estimators': [1, 4, 8, 11, 15, 18, 22, 25, 29, 32, 36, 39, 43, 46, 50],
 'learning_rate': [0.97, 0.98, 0.99, 1.0, 1.01, 1.02, 1.03, 1.04]}
```

```
[ ] abc_random.fit(x_train,y_train)
```

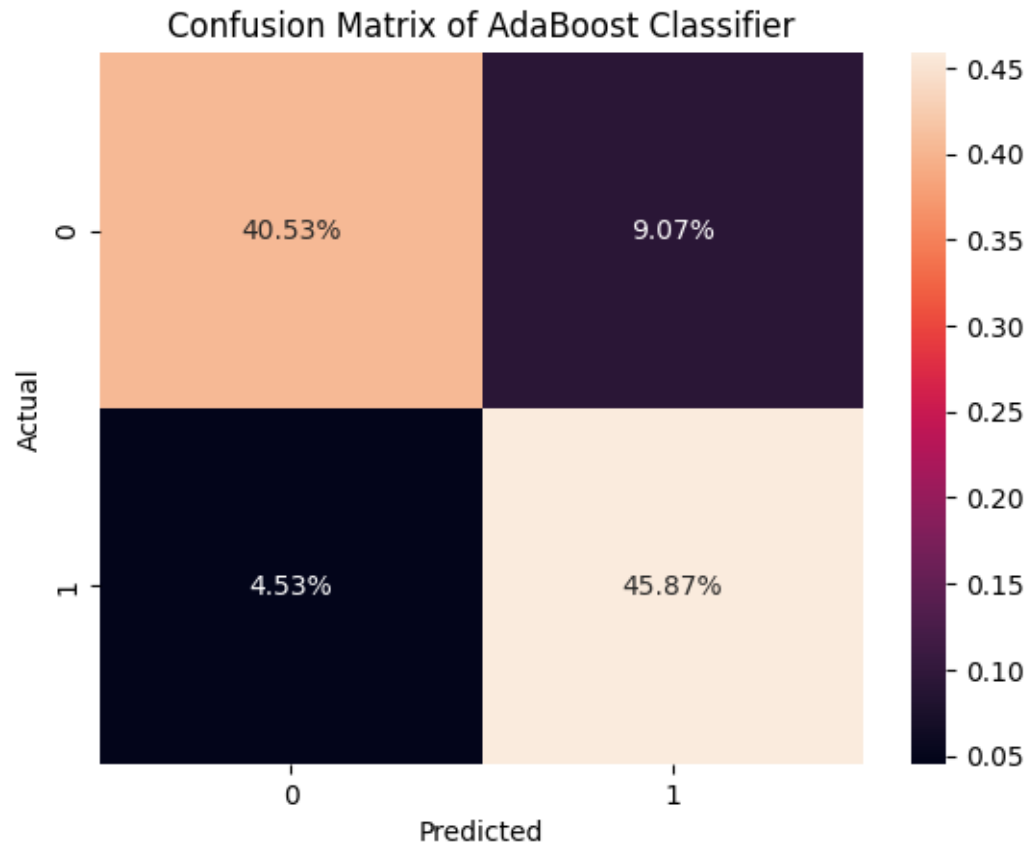


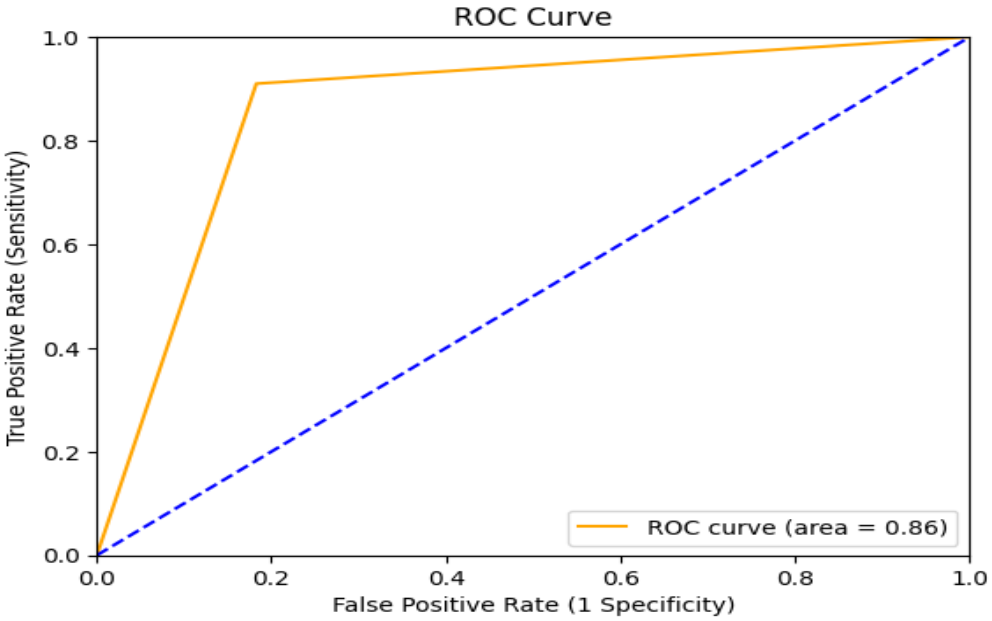

```
[ ] abc_random.best_params_
```

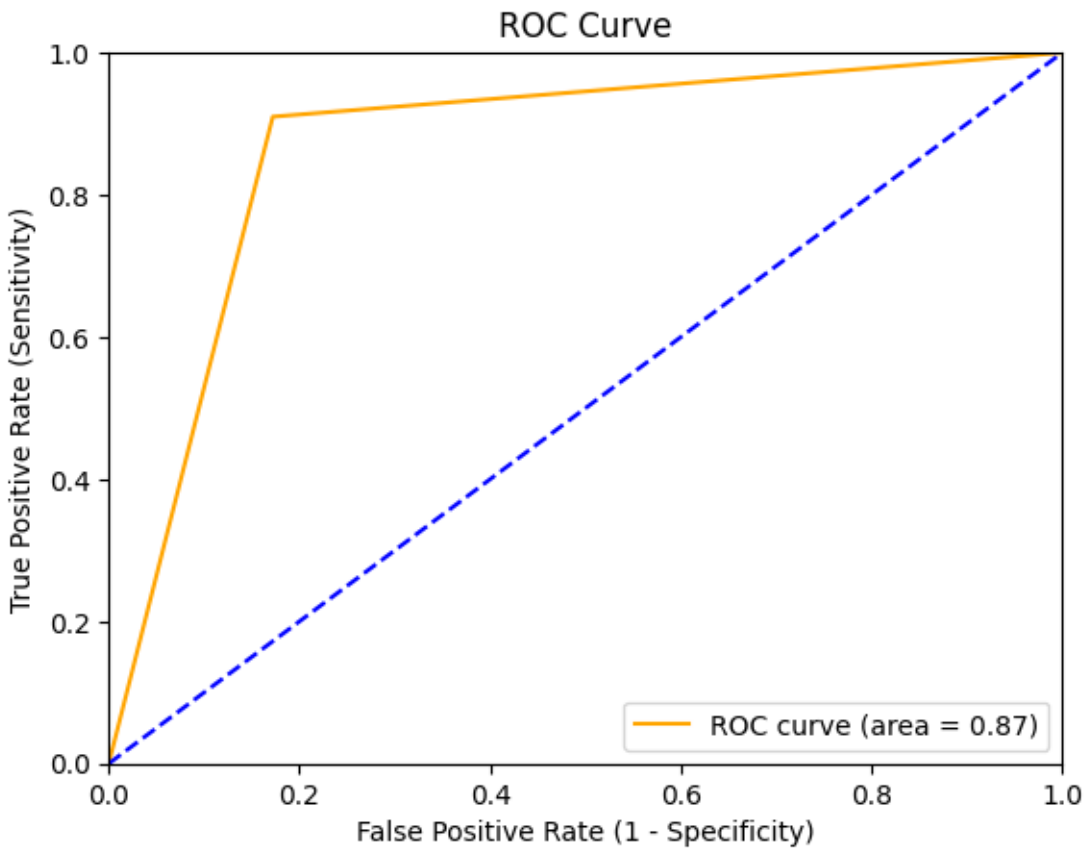
```
{'n_estimators': 11, 'learning_rate': 1.02}
```

```
abc_tuned = AdaBoostClassifier (random_state=49,n_estimators=11, learning_rate=1.02)
abc_tuned.fit(x_train, y_train)

pred_abc_tuned = abc_tuned.predict(x_test)

print('Accuracy of AdaBoost (tuned)=', accuracy_score (y_test, pred_abc_tuned))
```

```
Accuracy of AdaBoost (tuned)= 0.8693333333333333
```


Confusion Matrix of AdaBoost Classifier
40.53%
9.07%
4.53%
45.87%
Actual
Predicted
0
1
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05


ROC Curve
True Positive Rate (Sensitivity)
False Positive Rate (1 Specificity)
ROC curve (area = 0.86)
0.0
0.2
0.4
0.6
0.8
1.0




```
print(classification_report(y_test,pred_abc))
```

```
              precision    recall  f1-score   support

           0       0.90      0.82      0.86       186
           1       0.83      0.91      0.87       189

    accuracy                           0.86       375
   macro avg       0.87      0.86      0.86       375
weighted avg       0.87      0.86      0.86       375
```

```
print(classification_report(y_test,pred_abc_tuned))
```

```
              precision    recall  f1-score   support

           0       0.90      0.83      0.86       186
           1       0.84      0.91      0.88       189

    accuracy                           0.87       375
   macro avg       0.87      0.87      0.87       375
weighted avg       0.87      0.87      0.87       375
```

```
[ ] import pickle
    pickle.dump(abc_tuned,open('model.pkl','wb'))
```