Machine Learning based analysis and timely diagnosis of Mental Health Issues

Mayank Pujara

Date: 11/01/2024

Abstract: The world is seeing a growing number of mental health issues, expected to continue rising. Various factors—physiological, environmental, and biological are significant in causing these conditions. Failing to address mental health problems adequately could lead to a substantial global disease burden, given that about one in five adults worldwide already grapples with mental disorders. Using artificial intelligence methods helps create risk models to predict an individual's likelihood of developing mental illness, offering improved screening tools for early detection. These days, artificially intelligent virtual beings who can converse with caregiver seekers and suggest treatments are also being used in place of therapists.

Overview:

1. Problem Statement:

Developing a Machine Learning driven system capable of analyzing multifaceted data to detect and diagnose mental health issues promptly. The system should leverage various data sources to accurately identify potential mental health concerns in individuals. The primary aim is to create a reliable and scalable solution that assists in early detection, providing insights for timely intervention and support, thus mitigating the adverse impacts of untreated mental health conditions.

2. Market/Customer/Business Need Assessment

Globally, mental health issues have gotten worse in recent years. Mental health conditions now cause 1 in 5 years lived with disability. Around 20% of the world's children and adolescents have a mental health condition, with suicide the second leading cause of death among 15-29-year-olds. Approximately one in five people in post-conflict settings have a mental health condition.

Mental health issues can take many different forms, such as addiction, bipolar illness, depression, anxiety, and schizophrenia. The stigma attached to mental illness prevents many people from seeking treatment, even though mental health issues are becoming more common. The reason for this is because some people lack access to sufficient mental health treatments, while others have insufficient funds or qualified personnel. This has led to a significant gap in the provision of mental health treatment, which calls for creative methods to address the issue. The product’s aim is to make use of Machine Learning and NLP techniques to detect and diagnose mental issues promptly with an intention to reduce the gap between diagnosis and treatment.

3. Target Specifications and Characterization

- **College Students/Teenagers:** Adolescents and college students frequently deal with pressures associated with hormonal changes, social and academic challenges. The Machine Learning based system can provide private, easily available mental health help, assisting with coping mechanisms, anxiety reduction, and stress management.
- **Working Adults/Corporate Crowd:** Such health support platforms can help professionals in high-stress work situations manage work-life balance, burnout, and workplace stress. This is easily accessible and discrete, making it suitable for hectic schedules.
- **Individuals with Stigmatized Mental Health Concerns:** Some people may be reluctant to seek out standard mental health care because they fear stigma. The system can offer a more covert and private way to get help for problems like trauma, addiction, or specific phobias. It can provide a judgment-free environment that is safe for conversations and guidance.
- **Mental Health Professionals:** They can make use of such systems for initial assessments, continuous monitoring of patients, exploring new treatment options, and expanding mental health education.

4. External Search (information sources)

Statistical Data:

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

Relevant Papers:

- A. Almaleh, "Machine Learning-Based Forecasting of Mental Health Issues Among Employees in the Workplace," 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), BALI, Indonesia, 2023, pp. 118-124, doi: 10.1109/IAICT59002.2023.10205620.
- K. Yadav and Y. Hasija, "Artificial Intelligence and Technological Development in Behavioral and Mental Healthcare," 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022, pp. 1-6, doi: 10.1109/ICONAT53423.2022.9726100.
- Prof. Priti Kakde, Aditya Armarkar, Aman Kathale, Aryant Kshirsagar, Akshay Lokhande, Aditya Taitkar, "Sentimental Analysis Using Machine Learning for Mental Healthcare Management System", Ijraset Journal For Research in Applied Science and Engineering Technology (IJRASET), doi: 10.22214/ijraset.2023.53450
- Jetli Chung, Jason Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges", Applied Computational Intelligence and Soft Computing, vol. 2022, Article ID 9970363, 19 pages, 2022. https://doi.org/10.1155/2022/9970363

Data Source:

The dataset which I have referred for this project is the Mental Health in Tech Survey on Kaggle:

Dataset Link

This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace.

Overview of the dataset:

In [5]:
```
#Loading the dataset
data = pd.read_csv(r'Mental Health Tech Survey Dataset.csv')
data.head(10)
```

Out[5]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	No
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	Maybe
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	No
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Yes
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	No
5	2014-08-27 11:31:22	33	Male	United States	TN	NaN	Yes	No	Sometimes	6-25	...	Don't know	No
6	2014-08-27 11:31:50	35	Female	United States	MI	NaN	Yes	Yes	Sometimes	1-5	...	Somewhat difficult	Maybe
7	2014-08-27 11:32:05	39	M	Canada	NaN	NaN	No	No	Never	1-5	...	Don't know	No
8	2014-08-27 11:32:39	42	Female	United States	IL	NaN	Yes	Yes	Sometimes	100-500	...	Very difficult	Maybe
9	2014-08-27 11:32:43	23	Male	Canada	NaN	NaN	No	No	Never	26-100	...	Don't know	No

10 rows × 27 columns

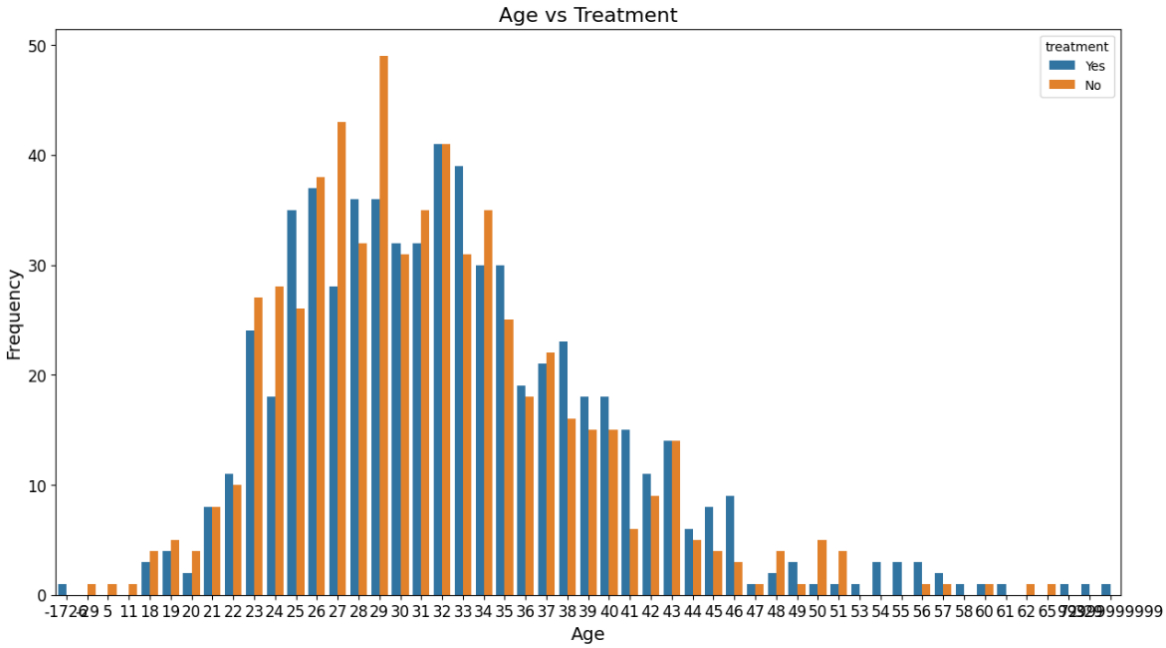
In [6]:
```
data.info()
```

```
<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
 26  comments                   164 non-null    object
dtypes: int64(1), object(26)
memory usage: 265.7+ KB
```

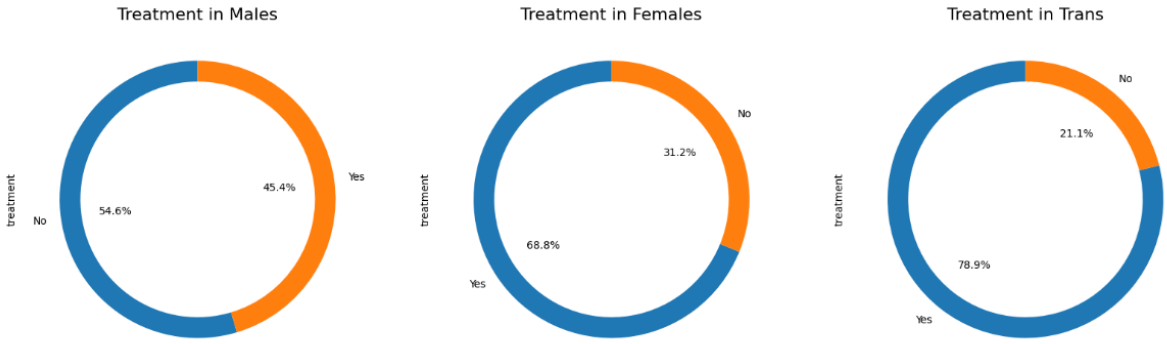
In [8]:
```
data.shape
```

Out[8]: (1259, 27)

Benchmarking:



Individuals over the age of 30 are addressing their mental health concerns



Individuals identifying as Trans and Females exhibit a higher tendency to seek treatment for mental health issues in comparison to Males.

5. Applicable Patents

- Fayyaz Memon (2021). Diagnostic and assessment system for mental illness based on collecting and analyzing multifactorial data using machine learning and artificial intelligence algorithms. *U.S. Patent Application No. 202203281844A1*. Washington, DC: U.S. Patent and Trademark Office.

  The system consolidates inputs from genetic reports, imaging results, neurological tests, and clinical information. It compares the data from current research to develop a score using machine learning algorithms and data analysis techniques, including linear and logistics regression, decision trees, Naive Bayes, and ensemble methods.

6. Applicable Regulations

Regulations regarding the use of AI in mental health diagnostics can vary significantly between countries and regions.

- *Data Privacy and Security Regulations:* Many countries have stringent regulations governing the collection, storage, and usage of personal health data, such as the General Data Protection Regulation (GDPR) in the European Union or the Health Insurance Portability and Accountability Act (HIPAA) in the United States.
- *Medical Device Regulations:* AI-powered tools used for diagnosing or treating health conditions may fall under the purview of medical devices. The Medical Device Rules, 2017, introduced by the Central Drugs Standard Control Organization (CDSCO), govern the registration and regulation of medical devices in India.
- *Ethical Guidelines:* Ethical considerations are vital, particularly in mental health. Many countries have ethical guidelines or institutional review boards (IRBs) that oversee research involving human subjects, ensuring that the use of AI in mental health remains ethical and respects patients' rights.

7. Applicable Constraints:

- **Data Privacy and Security:** Ensure robust measures to safeguard sensitive personal mental health data.
- **Transparency:** Transparent model will build trust by explaining how and why a conclusion or prediction was made.
- **Expert Involvement and Clinical Evaluation:** Collaboration with mental health professionals is vital for model validation, interpretation of results and aligns with the clinical expertise
- **Informed Consent and Awareness:** Users interacting with the system should be informed about the limitations, risks, and potential outcomes of using such technology

8. Business Model:

The business model will be based on a sophisticated subscription-based platform, which will redefine mental wellness services.

Tiered Subscription Plans:

- **Basic Plan:** Access to limited features such as mental health assessments, stress management tips, and general resources.
- **Premium Plan:** Full access to advanced diagnostics, personalized treatment recommendations, 24/7 chat support with AI counselors, progress tracking and direct involvement of mental health experts and doctors.

Monetization Points:

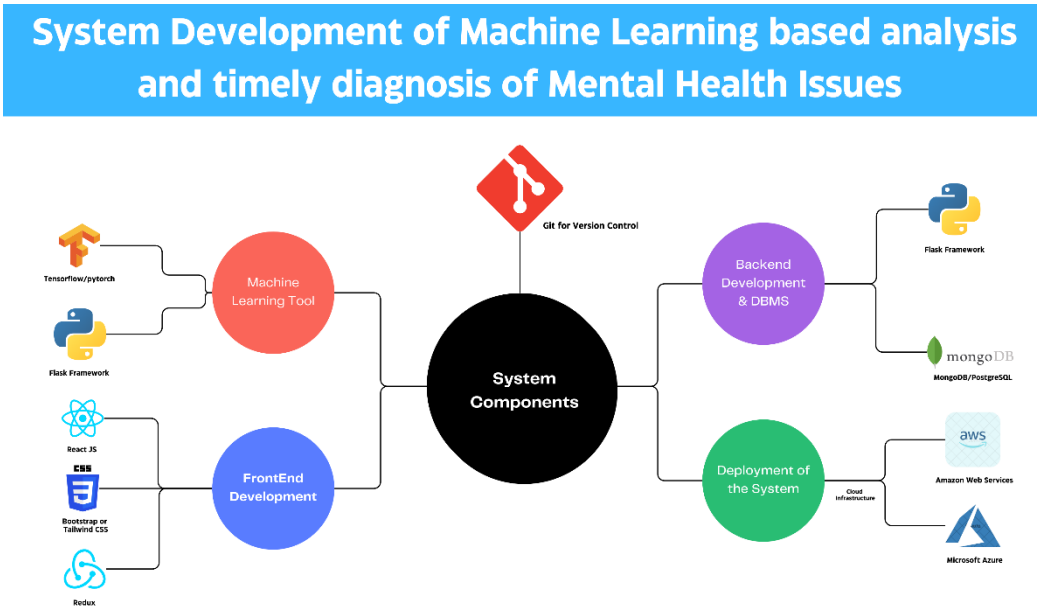
- **Monthly/Annual Subscriptions:** Provide several durations for your subscriptions, with lower prices for longer terms.
- **Consultation Add-ons:** Fees apply for prearranged online consultations with licensed mental health practitioners or therapists that are arranged through the site.
- **Corporate/Institutional Packages:** Personalized programs designed for companies, educational institutions, or medical facilities that give group subscriptions for mental health services for staff members or students.

9. Concept Generation:

This project on the Machine Learning-driven mental health diagnostic tool centers on revolutionizing mental health care through accessible and personalized support. The tool incorporates a machine learning-based assessment that integrates biometric data and analyzes user inputs to provide thorough evaluations that will be created from the ground up. Various machine learning models can be applied in the development of this tool, with each serving specific purposes. For instance, Support Vector Machines (SVMs) are proficient in classification tasks. Decision trees provide interpretability, which makes it possible to comprehend the model's decision-making process clearly. The main purpose is to ensure a comprehensive and accurate approach to diagnosing mental health conditions

10. Concept Development:

A basic overview of how the overall system can be developed is as follows:



1. Identify sources for diverse data followed by cleaning and preprocessing of the collected data to ensure quality for machine learning model training.
   - *PostgreSQL or MongoDB for structured and unstructured data storage.*
   - *Pandas and NumPy for data cleaning, transformation, and manipulation.*

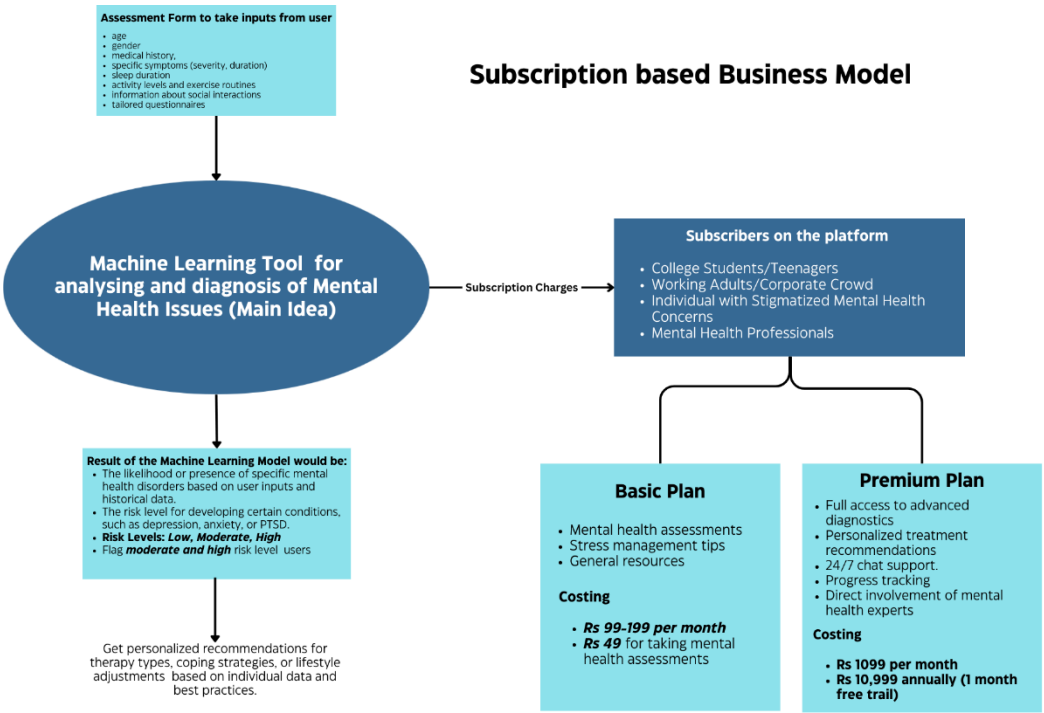
2. Selection of appropriate machine learning models for mental health diagnostics, followed by training the suitable models using preprocessed data to develop the tool.
   - *Scikit-learn, TensorFlow, or PyTorch for implementing various ML models.*
   - *Flask or FastAPI for deploying ML models as REST APIs.*

3. Creation of intuitive and accessible interfaces for users so that they can interact with the system easily.
   - *React.js for creation of interfaces*
   - *Bootstrap or Tailwind CSS for designing*
   - *Redux for state management.*

4. Building APIs for seamless communication between frontend and backend components and connecting the machine learning algorithms with backend APIs to enable real-time predictions and support generation.
    - *python for backend logic and API development*
    - *Flask for creating RESTful APIs and server-side logic*
    - *MongoDB or PostgreSQL to establish a secure and scalable database infrastructure for storage and retrieval.*

5. Other tools to be utilized:
    - *Git* for version control.
    - Amazon Web Services (AWS) or Microsoft Azure for scalable cloud hosting.

11.Final Product Prototype:



12.Product Details:

Working of the system:

- The users once registered on our platform will start with filling out an assessment form by providing personal information related to age, gender, medical history, specific symptoms (severity, duration), sleep duration and few other details.
- The user-provided data will then be sent to the backend component that will integrate and provide the user's data to the machine learning model to process inputs and predict the likelihood mental health conditions and assess the risk levels associated with the condition.
- The Predictions generated by machine learning models are then delivered back to the user, categorizing their mental health status or likelihood of certain conditions.

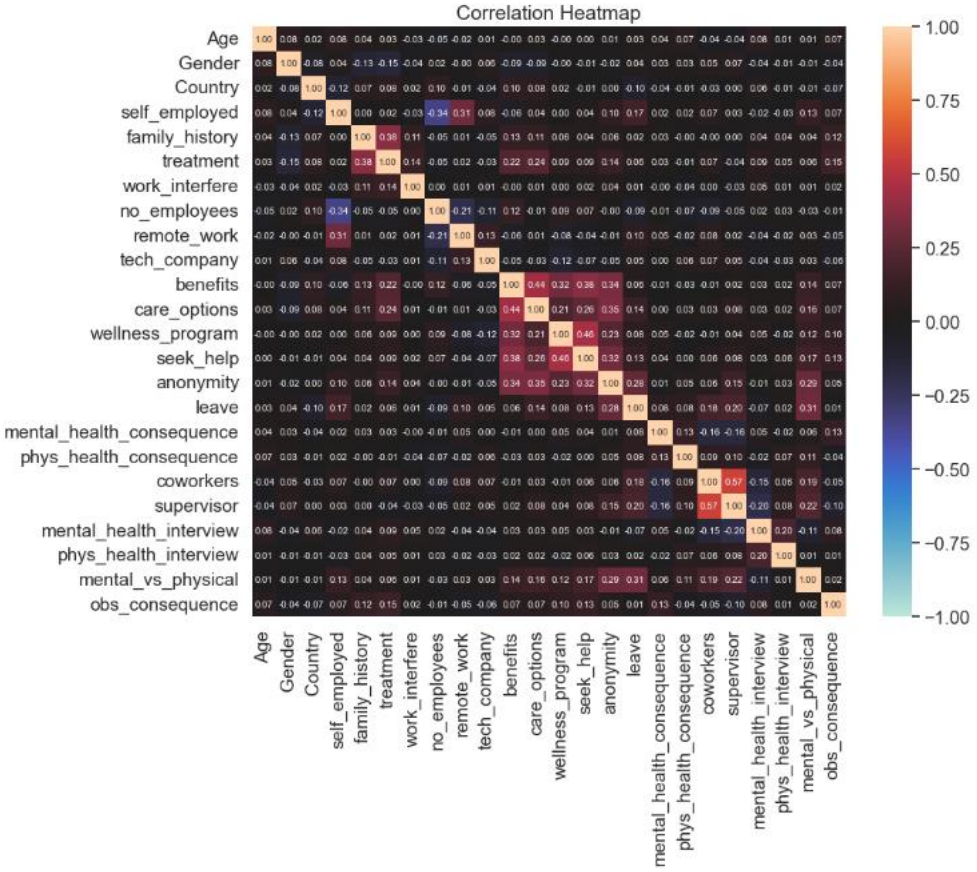
Data Sources:

The main source of data for the execution of the platform is the information provided by users through the assessment forms. For the purpose of training the model, the datasets related to mental-health data can be used as mentioned above.

Team Required to Develop:

- Frontend Developer who is proficient in React.js for building the user interface and assessment form.
- Backend Developer skilled in Flask for creating API endpoints and handling data processing.
- Machine Learning Engineer having experience in machine learning model development and integration process.
- Database Administrator familiar with managing Database Management Systems like MongoDB/PostgreSQL.

13.Code Implementation:



Above, we find the correlation between all the columns of the dataset.

- We start with processing and cleaning the data (handling missing values and removing duplicate values)
- This is followed by splitting the dataset into training dataset (70%) and testing dataset (30%).

For prediction purpose, making use of the *treatment* column

```
In [24]: X = data.drop(columns= ['treatment'])
         X.shape
Out[24]: (1257, 23)

In [25]: y= data['treatment']
         y.shape
Out[25]: (1257,)
```

Splitting the data into training and testing sets

```
In [26]: 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)
```

- Model Implementation

Model Implementation using *K Nearest Neigbour Classifier*

```
In [27]: classifier = KNN(n_neighbors= 8)
         classifier.fit(X_train,y_train)
         y_= classifier.predict(X_test.values)
         print('Accuracy of the KNN Model is:', accuracy_score(y_true=y_test, y_pred=y_)*100)
Accuracy of the KNN Model is: 68.51851851851852

C:\Python310\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but KNeighborsClassifier
was fitted with feature names
  warnings.warn(
```

Model Implementation using *Random Forest Classifier*

```
In [28]: steps_rfc = [('Scaler', StandardScaler()),('clf', RFC(n_estimators = 100))]
         clf_rfc = Pipeline(steps=steps_rfc)
         clf_rfc.fit(X_train, y_train)
         y_pred_rfc = clf_rfc.predict(X_test)
         print('Accuracy of the Random Forest Classifier is:', accuracy_score(y_true=y_test, y_pred=y_pred_rfc)*100)
Accuracy of the Random Forest Classifier is: 73.80952380952381
```

Model Implementation using *Support Vector Classification Model*

```
In [29]: steps_svc = [('Scaler', StandardScaler()),
                     ('clf', SVC())]
         clf_svc = Pipeline(steps = steps_svc)
         clf_svc.fit(X_train, y_train)
         y_pred_svc = clf_svc.predict(X_test)
         print('Accuracy of the Support Vector Classifier is:', accuracy_score(y_true=y_test, y_pred=y_pred_svc)*100)
Accuracy of the Support Vector Classifier is: 70.63492063492063
```

Model Implementation using *Decision Tree Classifier*

```
In [30]: classifier = DT()
         classifier.fit(X_train,y_train)
         accuracy = classifier.score(X_test,y_test)
         print('Accuracy of the Decision Tree Classifier is:', accuracy*100)
Accuracy of the Decision Tree Classifier is: 64.28571428571429
```

14. Results and Conclusion:

I made used of the various Machine Learning classifiers for training the model and the results were as follows:

Sr. No.	Model	Accuracy of the model
1.	K Nearest Neigbour Classifier	68.5185 %
2.	Random Forest Classifier	73.8095 %
3.	Support Vector Classification Model	70.6349 %
4.	Decision Tree Classifier	64.2857 %

Several machine learning models were used in the execution of our mental health prediction project, and their predictive accuracy using the dataset was assessed. Among the models tested, **the Random Forest Classifier** had the highest success rate ***(73.81%)***, making it the most accurate model. These results demonstrate how well machine learning approaches capture intricate correlations in the dataset, offering insightful predictions about mental health. The Random Forest model has proven to be a successful technique in improving pre-diagnosis screening in mental health examinations.

This initiative demonstrates how important it is to use advanced analytics to support continuous efforts to comprehend and treat mental health issues, opening the door for future intervention options that will be even more successful.

Source Code: *https://github.com/mayankpujara/Prediction-of-Mental-Health-Issues*