

Self-Analysis Mental Health Model - Project Documentation

Submitted by: Sushant Chivale

The Self-Analysis Mental Health Model : This is to predict possible mental health conditions based on user-provided symptoms. This project includes data preprocessing, model selection, an inference script, and a Streamlit-based UI for user interaction.

So I begin with the dataset preprocessing:

Dataset Preprocessing

Data Collection

- For this I had taken the open source dataset from the Kaggle : **Depression and anxiety data**,

About Dataset :

<https://www.kaggle.com/datasets/shahzadahmad0402/depression-and-anxiety-data/data>

- Basically this dataset (depression_anxiety_data.csv) is about the Depression. Symptoms such as the sensation of guilt or worthlessness, decreased energy, difficulties concentrating, suicidality, and psychotic symptoms may be present. The number, severity, and length of symptoms, as well as the impact on one's ability to function socially and at work, are all factors that go into determining the severity of a depression. The dataset is collected from undergraduate students of University of lahore. There are 787 participants contributed to the dataset.

Below are the features of the dataset :

- id
- school_year
- age
- gender
- bmi
- who_bmi
- phq_score
- depression_severity

- depressiveness
- suicidal
- depression_diagnosis
- depression_treatment
- gad_score
- anxiety_severity
- anxiousness
- anxiety_diagnosis
- anxiety_treatment
- epworth_score
- sleepiness

I made a prediction of the label **"Suicidal"** .

Data Cleaning & Preprocessing

- **Handling Missing Values:** Removed or imputed missing data.
 - **Feature Encoding:** Converted categorical symptom descriptions into numerical format using One-Hot Encoding or Label Encoding.
 - **Text Processing:** If symptoms were text-based, applied NLP techniques like tokenization and vectorization.
 - **Normalization:** Scaled numerical features using Min-Max Scaling or Standardization.
 - **Splitting Data:** Divided into training (80%) and testing (20%) sets.
-

Model Selection Rationale

I selected **Random Forest** and **XGBoost** for predicting mental health conditions based on user symptoms due to the following reasons:

1. Random Forest

- A robust ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.
- Handles missing values well and provides feature importance insights.
- Works effectively with high-dimensional data and is resistant to noise.

2. XGBoost

- An optimized gradient boosting algorithm known for its efficiency and high performance.
- Handles both classification and regression tasks effectively.
- Incorporates regularization (L1 & L2) to prevent overfitting.
- Faster training speed due to parallel computation and tree pruning.

Both models were evaluated based on **accuracy, precision, recall** to ensure optimal performance for mental health condition predictions.

How to Run the Inference Script

Prerequisites

Ensure the following dependencies are installed:

```
pip install -r requirements.txt
```

UI/CLI Usage Instructions

Streamlit App

To launch the Streamlit UI for an interactive experience:

“streamlit run app.py”

- Users can enter their symptoms via a form.
- The app processes inputs and displays predictions along with possible resources.

Features of the UI

- **User-Friendly Interface:** Simple input fields for symptom entry.
- **Instant Results:** Predicts mental health conditions in real-time.
- **Additional Resources:** Provides links to mental health support materials.

Or you go directly on the below website as I deployed this on Streamlit Community Cloud

<https://self-analysismentalhealthmodelarogoi-2xratdau99pfcmkqrru9h4.streamlit.app/>

Conclusion

Basically, this project offers a simple yet effective way for users to assess potential mental health conditions based on symptoms. Future improvements include expanding the dataset, integrating more advanced NLP techniques, and deploying the model as a web service.
