

**Mental Disorder Classification for Enhanced Behavioral Health Screening**



June 1, 2025

Katherine grey

Bellevue University DSC 680

## Executive Summary

The escalating need for efficient online behavioral health screening tools is addressed by this white paper, which details a machine learning-driven model designed to predict potential mental disorders. This system aims to assist therapists in initial diagnosis and guide patients towards appropriate care. Utilizing a dataset of 120 psychology patients characterized by 17 key symptoms, the project focuses on classifying individuals into categories such as Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, or other disorders. The primary goal is to enhance the therapy intake process—improving speed, accuracy, and ultimately the experience for both patients and providers—by streamlining initial assessments.

## 1. Business Problem

The shift towards online behavioral health services, significantly propelled by recent global events, highlights a critical need for efficient patient screening. Traditional intake processes are often inefficient, hindering the swift matching of patients with suitable therapists. There is a pressing demand for a predictive model that can quickly analyze patient-reported symptoms and information to forecast potential mental disorders. Such a tool would not only empower practitioners with data-driven insights for diagnosis but also suggest preliminary treatment pathways, thereby optimizing the entire therapy lifecycle from initial selection to ongoing treatment.

## 2. Background/History

The landscape of online medical services, including behavioral health, was already on an upward trajectory before the recent global health crisis, which then triggered an exponential surge in demand. This growth has fundamentally altered the interaction dynamics between therapists and patients. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative forces in this sector, significantly enhancing both patient experience and provider operational efficiency. This project harnesses the power of ML to construct a robust mental health assessment prediction tool based on a defined set of symptomatic criteria, moving beyond traditional intake methods.

## 3. Data Explanation

The dataset, sourced from Kaggle, comprises information from 120 psychology patients, each characterized by 17 essential symptoms. This rich feature set is designed to capture a comprehensive picture of mental well-being for classification into one of four disorder categories: Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, or another unspecified disorder.

The symptoms included are levels of: Sadness, Exhaustion, Euphoria, Sleep Disorder, Mood Swings, Suicidal Ideations, Anorexia, Anxiety, Try-explaining (difficulty explaining oneself), Nervous Breakdown, Ignore/Move-on (coping mechanism), Admitting Mistakes, Overthinking, Aggressive Response, Optimism, Sexual Activity, and Concentration.

A graph of a patient's thoughts

AI-generated content may be incorrect.

Data Preparation:

The Mental\_Classifier.ipynb notebook indicates standard data preprocessing steps such as dropping unnecessary columns (Id and Timestamp) and renaming columns for clarity (Patient\_ID). Missing values are handled by dropping rows with NaNs, ensuring data integrity for model training. Categorical features Where discretized for modeling building. Features were then selected bases on correlation matrix and a new data frame with these features created and split 70/30.

A chart with numbers and a red line

AI-generated content may be incorrect.

## 4. Methods

The project employs a supervised machine learning approach to classify mental disorders. The general methodology includes:

* **Data Loading and Initial Inspection**: Reading the CSV data into a Pandas DataFrame and performing initial schema and data checks.
* **Feature Preprocessing**:
  + **Categorical Feature Indexing**: Converting string-based categorical features (e.g., 'Gender', 'Major') into numerical indices using StringIndexer.
  + **Label Indexing**: Transforming the target categorical variable (e.g., 'StressLevel' or the final disorder classification) into a numerical index.
* **Data Splitting**: Dividing the processed dataset into training and testing sets (typically 70% for training and 30% for testing) to evaluate model performance on unseen data.
* **Model Selection and Training**: A variety of classification algorithms can be considered, with the notebook specifically highlighting the **Random Forest Classifier** as a robust choice. The model is trained on the prepared training data.
* **Prediction and Evaluation**: The trained model is used to make predictions on the test set, and its performance is evaluated using appropriate classification metrics such as accuracy.

The choice of **Random Forest Classifier** is suitable for this problem due to its ability to handle high-dimensional data, mitigate overfitting, and provide good predictive performance.

## 5. Analysis

The Mental\_Classifier.ipynb notebook provides insights into the model's performance, particularly through a table illustrating the accuracy of various classification models. This table is a critical component of the analysis, allowing for a comparative understanding of model effectiveness.

**Model Performance Comparison:**

The notebook presents a summary table comparing the performance of different classification models. Although the exact values are in a styler object, the intention is to show how various models (e.g., Logistic Regression, Decision Tree, Random Forest) perform on the given dataset. This comparative analysis is crucial for selecting the most appropriate model for deployment.

Example Visualization (Conceptual - based on notebook output description):

A table or plot showing the "Graphical representation of models" would display the accuracy (or other metrics) of different models, often with a visual highlight (e.g., dark blue, red, purple) to indicate the best-performing model across rows, columns, or the entire table. Such a visualization helps quickly identify which algorithm is most effective for the task.

The analysis would delve into the strengths and weaknesses of each model based on these metrics, providing a rationale for selecting the Random Forest Classifier as a primary candidate due to its balance of accuracy and robustness. The notebook also implicitly suggests that feature engineering (like Try-explaining which is created from Try Explaining and Explanation Difficult) and handling of categorical data are crucial steps before feeding data to the model.

K-NN

A diagram of a blue and yellow grid

AI-generated content may be incorrect.

Gaussian Bias

A diagram of a blue and yellow grid

AI-generated content may be incorrect.

MLP Classifier;

A blue and green squares with white numbers

AI-generated content may be incorrect.

## 6. Conclusion

This project successfully demonstrates the feasibility of developing an AI-powered mental disorder classifier using machine learning techniques on a symptomatic dataset. By leveraging Pythons SKLearn library of machine learning models, the solution provides a scalable and efficient means to process patient data and predict potential mental disorders. The classification model serves as a foundational component for streamlining the initial assessment phase of behavioral health services, promising improved efficiency for practitioners and better directional guidance for patients. The system's ability to classify individuals into specific disorder categories offers a data-driven approach to initial diagnostic support.

## 7. Assumptions

* The provided dataset accurately reflects the symptoms and corresponding mental disorders of the patient population it represents.
* The 17 identified symptoms are sufficient and comprehensive enough to predict the specified mental disorders with reasonable accuracy.
* Patient self-reported data is reliable and accurate.
* The four disorder categories (Mania Bipolar, Depressive Bipolar, Major Depressive, Other) are mutually exclusive or can be appropriately distinguished by the model.
* The kgrandom.csv file structure and content align with the notebook's expectations for column names and data types.

## 8. Limitations

* **Dataset Size**: The dataset of 120 patients is relatively small, which can limit the model's ability to generalize to a larger, more diverse population and may lead to overfitting.
* **Feature Scope**: While 17 symptoms are included, other crucial factors influencing mental health might be missing, potentially impacting the model's comprehensiveness.
* **Imbalance**: There is no explicit mention of handling class imbalance, which might affect the model's performance, particularly for less frequent disorder types.
* **Static Data**: The model is trained on a static dataset. Real-world symptoms can evolve, requiring continuous model retraining.
* **Lack of External Validation**: The model's performance is evaluated on a test set from the same dataset; external validation on completely new and diverse data would be beneficial.

## 9. Challenges

* **Data Quality and Consistency**: Ensuring the accuracy and consistency of self-reported symptom data can be challenging.
* **Interpreting Symptoms**: The subjective nature of some symptoms (e.g., "Sadness," "Try-explaining") can lead to variability in reporting.
* **Ethical Deployment**: Deploying a model in a sensitive domain like mental health necessitates careful consideration of ethical implications, including bias and privacy.
* **Model Explainability**: For clinical adoption, understanding why a model makes a particular prediction (explainability) is often as important as the prediction itself.
* **Generalization**: Building a model that performs well across diverse demographics and cultural backgrounds remains a significant challenge.

## 10. Future Uses/Additional Applications

Beyond the initial classification, the developed model lays the groundwork for several advanced applications:

* **Personalized Therapist Matching**: The model's output could be integrated into a system that matches patients with therapists specializing in their predicted disorder type, considering additional criteria like therapist gender, insurance compatibility, and therapeutic approach.
* **Treatment Path Recommendations**: The system could evolve to suggest evidence-based treatment options tailored to the predicted disorder, serving as a starting point for therapists.
* **Early Intervention Systems**: By identifying potential disorders early, the model could support proactive interventions, potentially preventing the escalation of conditions.
* **Mental Health Trends Analysis**: Aggregated and anonymized data processed by the model could provide insights into prevalent mental health trends, aiding public health initiatives.
* **Progress Monitoring**: With repeated assessments, the model could potentially track changes in symptom levels and assess the effectiveness of ongoing treatments.

## 11. Recommendations

The following recommendations are crucial for the continued development and robustness of the mental disorder classifier:

* **Acquire Larger and More Diverse Datasets**: Training the model on a substantially larger and more demographically diverse dataset is paramount to improve generalization and accuracy.
* **Explore Broader Classification Architectures**: Investigate advanced machine learning or deep learning architectures capable of simultaneously analyzing multiple disorder families or employing an iterative process to narrow down the top 3-5 most probable outcomes.
* **Integrate Professional Diagnoses for Feedback**: Establish a robust feedback loop where "professional" diagnoses from therapists can be continuously fed back into the training set, enhancing the model's real-world accuracy over time.
* **Develop Explainable AI (XAI) Components**: Incorporate XAI techniques to provide insights into why the model makes specific predictions, fostering trust and aiding clinical interpretation.
* **Address Class Imbalance**: Implement techniques to handle potential class imbalance in the dataset, ensuring the model performs well across all disorder categories.

## 12. Implementation Plan

The implementation of this AI-powered mental health screening tool would follow a systematic pipeline:

1. **Data Ingestion Pipeline**: Establish automated processes to securely collect and ingest self-reported patient information from online intake forms or questionnaires.
2. **Feature Engineering Module**: Develop a robust module to perform necessary data cleaning, transformation, categorical encoding, and feature assembly on the incoming raw patient data.
3. **Model Inference Engine**: Integrate the trained machine learning model (e.g., Random Forest Classifier) into a scalable inference engine that can process new patient data and generate disorder predictions in real-time or near real-time.
4. **Recommendation and Notification System**: Create a system to deliver the disorder predictions and, potentially, initial therapist recommendations to both the patient (e.g., via a secure portal) and the intake staff or assigned therapist.
5. **Continuous Feedback Loop and Model Retraining**: Implement a mechanism where the "professional" diagnosis made by a qualified therapist after an initial assessment can be captured and securely fed back into the model's training data. This continuous learning loop is vital for iterative model improvement and increasing accuracy over time.

## 13. Ethical Assessment

The development and deployment of any medical or psychological AI tool demand rigorous ethical consideration, with paramount concerns centered on patient privacy and data accuracy.

* **Patient Privacy and Anonymization**: All demographic information not directly correlated with the data necessary for the model (e.g., patient names, precise addresses) will be meticulously scrubbed and anonymized to ensure the highest level of patient confidentiality. Adherence to data protection regulations (e.g., HIPAA in the US, GDPR in Europe) is non-negotiable.
* **Regulatory Compliance**: The entire system will be designed and operated in strict compliance with all relevant medical, privacy, and data security regulations to ensure legal and ethical operation.
* **Bias Detection and Mitigation**: Continuous monitoring and auditing for algorithmic bias within the model's predictions are essential. Efforts will be made to ensure the model provides equitable and fair assessments across diverse patient populations, avoiding discriminatory outcomes based on demographic or other protected characteristics.
* **Augmentation, Not Replacement**: It is critical to communicate that the model's output is intended as a **guidance tool for healthcare professionals**, not a definitive diagnostic instrument. The AI provides a prediction to assist, but it does not replace the nuanced judgment and professional diagnosis of a licensed therapist or medical practitioner.
* **Transparency and Consent**: Patients will be fully informed about the use of AI in their initial screening process, and explicit consent will be obtained for the use of their anonymized data for model improvement.

**14. Question and Answers**

* What where some of the issues ?
  + Overall this was a fairly straightforward model. Being as such there was no real significant issues.
* What if anything would you do better or change?
  + Well given the limited time and scope not much but if there was more time I would like to try building a model that can screen for more than just one mental illness. Patient psychology isn’t usually just one thing but an interplay of many factors that can lead to episodic disorders and co-morbidities that make saying someone is having depression and only that a bit limiting. Depression might just be a symptom of something deeper or not related. Having been misdiagnosed myself, it can be very frustrating.