MENTAL DISORDER CLASSIFICATION

Group 10

- Chathurangi Nethmini – 16012
- Kusara Udayana – 16073
- Amindu Yohan - 16385

Abstract

This project looks at the symptoms that help tell apart Bipolar Type 1, Bipolar Type 2, and Depression using data from 120 patients. We found that signs like mood swings, suicidal thoughts, euphoria, overthinking, and sexual activity were different in people with mental disorders compared to healthy people. These results match what doctors already know from real-life medical experience.

We tried methods like FAMD and Fisher's Discriminant Analysis to see if the data forms clear groups. But they didn't. So, we built machine learning models like SVM, Random Forest, Naive Bayes, and KNN to classify patients based on their symptoms. Among them, the Random Forest model was selected as the best model.

This work shows that using data and models can help doctors diagnose mental disorders more quickly and correctly, especially when symptoms are hard to notice. The goal is not to replace doctors but to support them in making better decisions.

Table of Contents

List of Figures

List of Tables

Introduction

In today's world, mental health is gaining the attention it truly deserves. Accurate diagnosis of mental disorders is crucial **-** not only for effective treatment but also for improving the quality of life of millions affected globally. Yet, diagnosing complex conditions like **bipolar type I**, **bipolar type II** and **Depression** can be challenging, often relying heavily on behavioral symptoms that are subtle and overlapping. This project dives into patient symptom data to explore the patterns that distinguish different mental health conditions. By carefully analyzing key symptoms and their relationships, we aim to provide clearer insights that can support psychiatrists in making more informed and timely diagnoses. Finally, we aim to build a classification model that can predict the correct mental health condition based on symptoms. This is especially important because mental disorders often go unnoticed and misunderstood, unlike physical illnesses. Our results suggest that data-driven classification tools can help support early, consistent and accurate diagnosis-enhancing, not replacing, clinical decisions.

Description of the Question

1. **Identifying Key Symptoms in Diagnosed vs. Healthy Individuals**

Identify which symptoms best distinguish individuals with Depression, Bipolar I, or Bipolar II from those without any diagnosis - aligning findings with clinical knowledge.

2. **Building a Symptom-Based Classification model**

Build a classification model that predicts the correct disorder type based on a patient's reported symptoms that aid early detection and clinical decisions.

Description of the Dataset

This project was conducted using a publicly available dataset found on Kaggle, titled "**Mental Disorder Classification (kaggle.com)**". The dataset contains **120 unique patient records** and includes **19 variables**, out of which **17 are essential psychological symptoms**, and one variable represents the **expert diagnosis** (target variable). The description of each variable used in this project is presented in the table below.

Variable Name	Variable Type	Description
Patient Number	Identifier	Unique identifier/index number of each patient.
Sadness	Ordinal (More often > Usual > Sometimes > Seldom)	Frequency of experiencing sadness.
Euphoric	Ordinal (More often > Usual > Sometimes > Seldom)	Frequency of experiencing elevated or euphoric mood.
Exhausted	Ordinal (More often > Usual > Sometimes > Seldom)	Frequency of feeling emotionally or physically exhausted.
Sleep Disorder	Ordinal (More often > Usual > Sometimes > Seldom)	Frequency of sleep-related disturbances.
Mood Swing	Nominal (Binary: Yes/No)	Presence of intense and frequent mood fluctuations.
Suicidal Thoughts	Nominal (Binary: Yes/No)	Presence of suicidal ideation.
Anorexia	Nominal (Binary: Yes/No)	Presence of disinterest in eating or food.
Authority Respect	Nominal (Binary: Yes/No)	Tendency to respect authority or follow rules.
Try-Explanation	Nominal (Binary: Yes/No)	Effort made to explain one's behavior or thoughts.
Aggressive Response	Nominal (Binary: Yes/No)	Tendency to react aggressively.
Ignore & Move-On	Nominal (Binary: Yes/No)	Ability to let go of issues and move forward.
Nervous Breakdown	Nominal (Binary: Yes/No)	Frequency or severity of emotional breakdowns.
Admit Mistakes	Nominal (Binary: Yes/No)	Willingness to acknowledge personal errors.
Overthinking	Nominal (Binary: Yes/No)	Tendency to overanalyze situations.
Sexual Activity	Ordinal (1–10)	Level of interest or activity in sexual behavior.
Concentration	Ordinal (1–10)	Ability to focus and stay attentive.
Optimism	Ordinal (1–10)	Degree of positive outlook or hopefulness.
Expert Diagnose	Nominal (Target Variable)	Final psychiatric diagnosis (Bipolar Type-1, Bipolar Type-2, Depression, Normal).

Table 1- Description of the Dataset

Data Pre-processing

The dataset was checked for duplicates and missing values. No duplicate records or missing values were found. Using the Isolation Forest algorithm, 1 outlier was identified**.** However, no data points were removed.

Feature Engineering

We cleaned the variables *Optimism*, *Sexual Activity*, and *Concentration* by removing the text "from 10" and keeping only the numeric part. This made them usable as ordinal values for analysis and visualization.

Important results of the descriptive analysis

Bipolar Disorder

Figure 1- Distribution of mood swing among Normal vs Bipolar Type 1

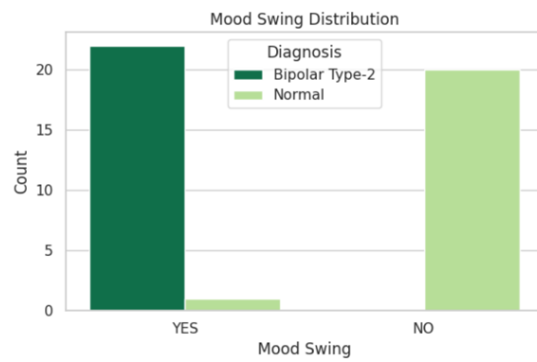


Figure 2 - Distribution of mood swing among Normal vs Bipolar Type 2

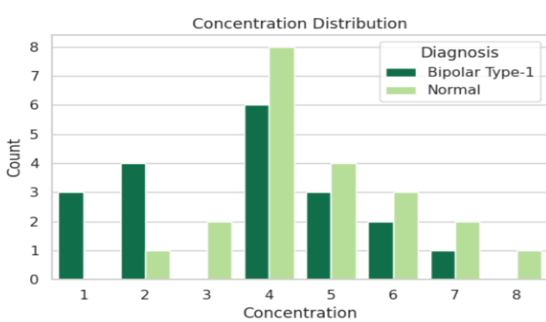


Figure 4 - Distribution of Concentration among Normal vs Bipolar Type 1

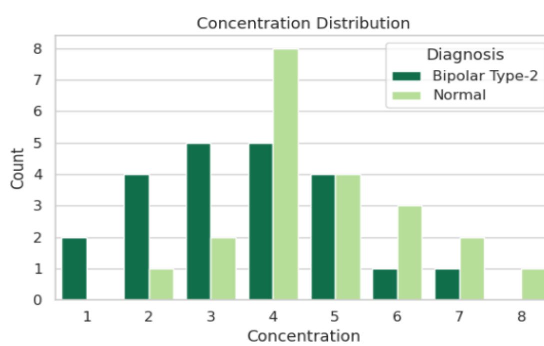


Figure 3 - Distribution of Concentration among Normal vs Bipolar Type 2

Both Bipolar Type 1 and Type 2 individuals reported frequent mood swings far more often than those without a diagnosis. The Chi-square test of independence (p-value < 0.0001) also indicates a significant difference in mood swings between individuals with Bipolar 1 or 2 and those considered normal. Due to these rapid mood changes, individuals with Bipolar Type 1 or Type 2 often struggle to concentrate like those without **disorder**. As a result, there is a significant difference in concentration levels between people with bipolar disorder and those considered normal.

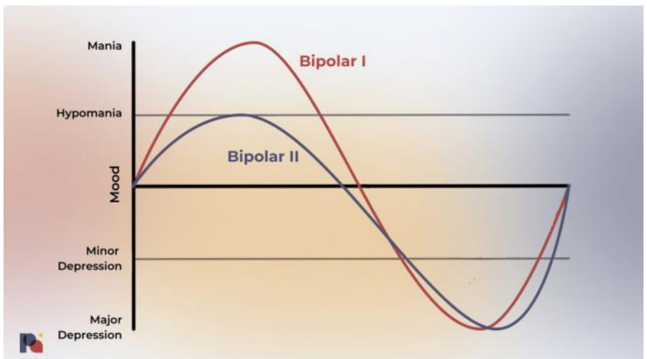


Figure 5 - Mood swing for Bipolar 1 &2

According to medical field Mood swings are different in Bipolar Type 1 and Type 2. Both types have depressive episodes, where the person feels very sad and low. But the difference is in the high mood phase: Bipolar Type 1 has mania, which is a very high energy state. Bipolar Type 2 has hypomania, which is similar but less intense. This difference is very important. Because of the manic episode, Bipolar Type 1 shows more severe symptoms than Type 2. That's why Bipolar Type 1 can sometimes be more serious and harder to manage.

The manic phase in Bipolar Type 1 causes euphoria. But not just normal happiness. It's a very extreme and intense feeling. According to the Mann-Whitney test, people with Bipolar Type 1 feel more euphoria than those with Bipolar Type 2, and the p-value is 0.009, which means this difference is statistically significant. During manic phases, some people with Bipolar Type 1 show something called unrealistic optimism. According to clinical mania often involves impulsivity and heightened libido, leading to riskier or more frequent sexual activity. Our analysis confirms this: Bipolar Type 1 individuals had significantly higher sexual activity than Type 2, with a Mann-Whitney p-value = 0.0001. (Bipolar 1 Disorder and Bipolar 2 Disorder: What Are the Differences?, 2023)

Depression

1. Sadness

WHO lists persistent sadness as a core depression symptom. Our chart showed higher sadness in depressed individuals, but the Mann–Whitney U test (U = 235.5, p = 0.9182) found no significant difference - possibly due to small or skewed groups. Despite this, sadness remains clinically vital and often goes unrecognized.

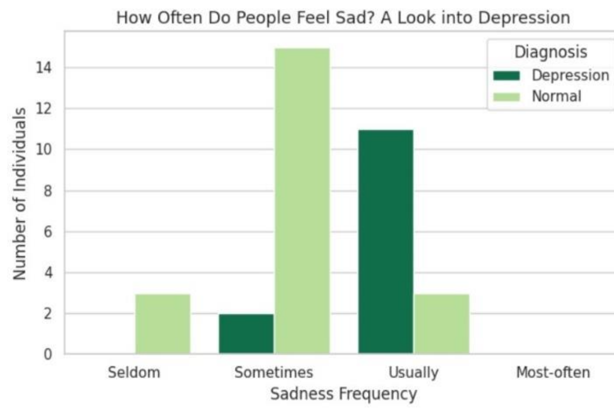


Figure 6 - Distribution of Sadness among Normal vs

2. Suicidal Thoughts

Suicidal thoughts were far more common in depressed individuals. A Chi-square test ($\chi^2 = 18.02$, $p < 0.001$) confirmed this as highly significant, aligning with clinical evidence and highlighting the need for early awareness and support.

3. Optimism

Optimism was significantly lower in those with depression (U = 75.5, p = 0.0001). This matches clinical signs of hopelessness, showing how depression can manifest not just in sadness but in a loss of outlook. (Gupta, 2023)

FAMD and Fisher Discriminant Analysis

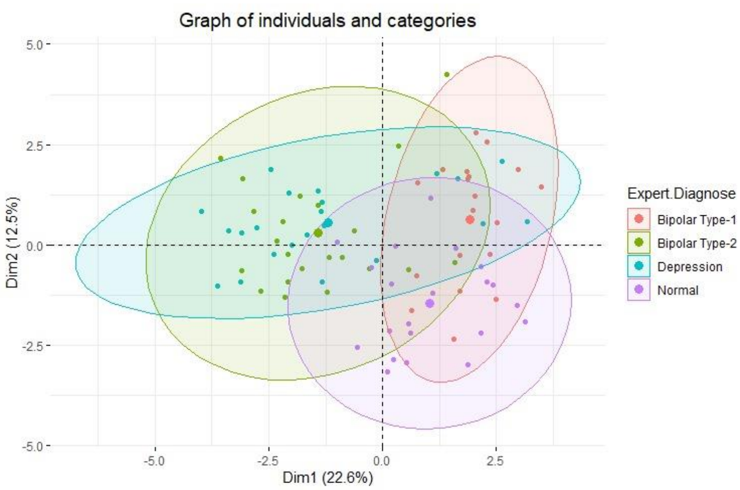


Figure 7 - FAMD graph of individuals and categories

After applying FAMD to the dataset, treating scaled variables as continuous, we extracted 8 dimensions. We visualized the two dimensions with the highest explained variance and marked all observations. To identify potential clusters, we circled the regions corresponding to each diagnosis. However, the overlapping areas suggest that there is no clear separation between depression, bipolar type 1, bipolar type 2, and normal classes.

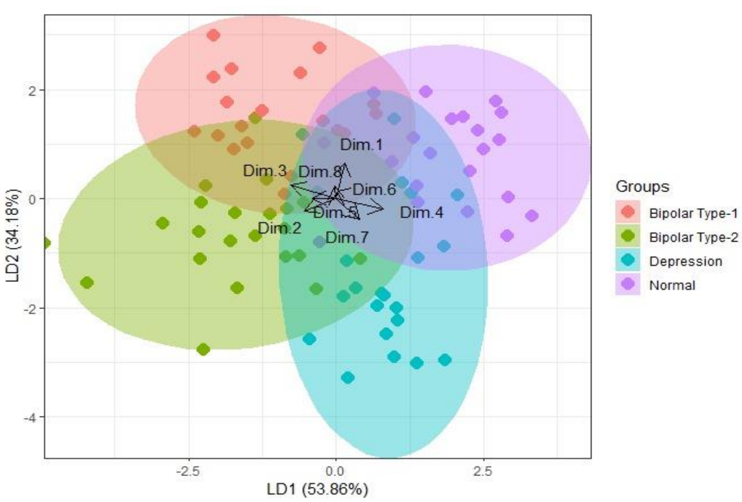


Figure 8 - FDA graph

Using the 8 FAMD components, we applied Fisher Discriminant Analysis and plotted two new components to investigate possible class separation. The result still showed no distinct groupings, indicating that the four classes overlap significantly in the reduced feature space.

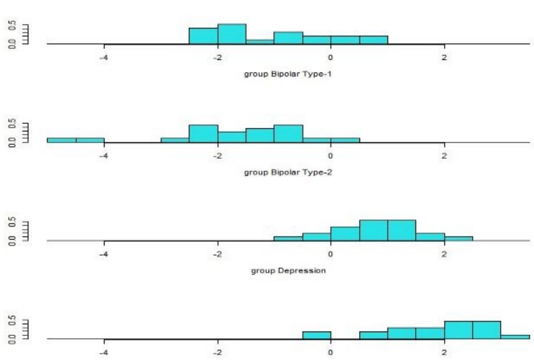


Figure 9 - Distribution of four classes along first FAMD

We also plotted the distribution of all four classes along the first FAMD component. The distributions were largely overlapping, reinforcing that no clear boundary exists between classes. Furthermore, since the data violated multivariate normality, we could not apply Linear Discriminant Analysis (LDA) to define a proper decision boundary.

Important results of the advanced analysis

- Since no clear decision boundary was observed in the data, we selected multiple classification models to improve accuracy: Support Vector Machine (SVM), Random Forest, Naive Bayes, and K-Nearest Neighbors (KNN).
- For SVM, key hyperparameters include the regularization parameter (C), the kernel type (e.g., linear, RBF), and kernel-specific parameters such as gamma.
- In Random Forest, we tuned the maximum tree depth, number of leaf nodes, and number of estimators.
- For Naive Bayes, although it has fewer tunable parameters, we considered different distributional assumptions (e.g., Gaussian vs. multinomial) and smoothing parameters (alpha).
- In KNN, we optimized the number of neighbors (k) and the distance metric (e.g., Euclidean, Manhattan).
- All models were tuned using 5-fold cross-validation to identify the best-performing hyperparameter combinations and avoid overfitting. The table below summarizes the performance of each model.

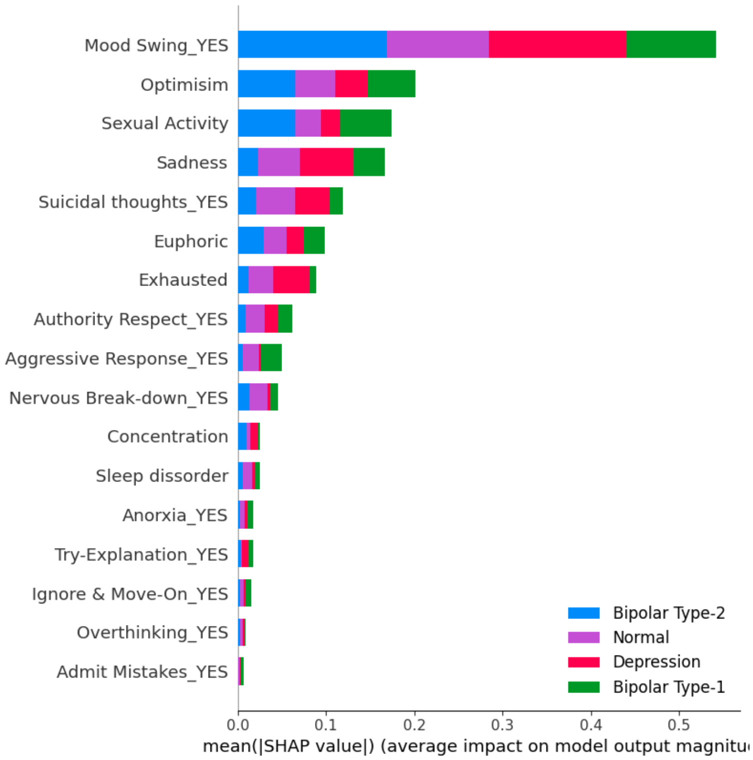
Model		Accuracy	Precision	Recall	F1
SVM	Train	0.976			
	Test	0.86	0.87	0.86	0.86
Random Forest	Train	0.964			
	Test	0.86	0.86	0.86	0.86
Naive Bayes	Train	0.845			
	Test	0.67	0.78	0.67	0.60
KNN	Train	1.00			
	Test	0.78	0.81	0.78	0.77

Table 2 - Model performance

Although SVM achieved high testing accuracy, the large gap between training and testing accuracy indicated potential overfitting. To address this, we applied bagging with SVM, which reduced the training accuracy to 85.56%, helping to balance the model. (bagging vs Boosting in Machine Learning, n.d.)

However, after comparing all models, we selected Random Forest as the best-performing classifier, as it showed a minimal gap between training and testing accuracy, indicating better generalization and stability compared to the other models.

Feature importance and Model Refinement



We used SHAP (SHapley Additive exPlanations) to interpret our Random Forest model and identify which symptoms contribute the most to predicting each mental health condition.

The graph shows the average absolute SHAP value for each feature across all predictions, broken down by class:

Figure 10 - Feature importance graph

- **Mood Swing** had the highest impact on predictions across all classes, especially for Bipolar disorders.
- **Optimism**, **Sexual Activity**, and **Sadness** also showed strong influence in distinguishing between Depression, Bipolar types, and Normal cases.
- Other impactful symptoms include **Suicidal Thoughts**, **Euphoric states**, and **Exhaustion**.

Based on this analysis, we refit our Random Forest model using only the top 7 features Mood Swing, Optimism, Sexual Activity, Sadness, Suicidal Thoughts, Euphoric and Exhausted.This reduced model achieved nearly the same accuracy, with improved interpretability and efficiency.

Issues you encountered and proposed solutions

1. **Issue**: Our dataset included only nominal and ordinal variables, which made FAMD (Factor Analysis of Mixed Data) unsuitable.

   Solution: We converted ordinal variables to continuous (1–10 scale) to apply FAMD meaningfully.

2. **Issue**: Fisher's Discriminant Analysis (FDA) showed that diagnostic groups were not well-separated, meaning no clear decision boundaries.

   Solution: This indicated the need for more complex or non-linear models to better capture underlying patterns, and we tried SVM, Random Forests and Naïve Bayes.

Discussion and conclusions

This study explored how symptom-level data can help distinguish Depression, Bipolar Type 1, Bipolar Type 2, and healthy individuals. While clustering techniques like FAMD and Fisher Discriminant Analysis showed overlapping groups, individual symptom patterns aligned well with clinical expectations, such as mood swings and euphoria in Bipolar Type 1, and low optimism and suicidal thoughts in Depression. Among the models tested, Random Forest offered the best balance of accuracy and generalization, highlighting its potential to support early diagnosis. Overall, this project demonstrates how combining clinical insight with data-driven methods can enhance mental health assessment and decision-making.

Appendix

ChathurangiAkmeemana/Mental_Disorder_Classification (github.com)

References

bagging vs Boosting in Machine Learning. (n.d.). Retrieved from www.geeksforgeeks.org: https://www.geeksforgeeks.org/machine-learning/bagging-vs-boosting-in-machine-learning/

Bipolar 1 Disorder and Bipolar 2 Disorder: What Are the Differences? (2023, 08 30). Retrieved from https://www.healthline.com/: https://www.healthline.com/

Gupta, S. (2023, 12 16). *Major Depressive Disorder: Symptoms, Causes, and Treatment*. Retrieved from https://www.verywellmind.com: https://www.verywellmind.com/major-depressive-disorder-symptoms-causes-and-treatment-5270926