Uncovering Global Mental Health Trends: Patterns and Predictions

ORIE 4741

Authors: Nicole Choi (nhc35), Grace Wang (gzw2)

Github Link: https://github.com/hyuseee/Global-Mental-Health-Analysis/tree/main

Abstract:

Aligned with the World Federation for Mental Health (WFMH), this research aims to contribute to combating mental disorders and facilitating recovery globally. Our study investigates the impacts of global events on mental health, including natural disasters and wars, and their associations with disorders such as depression, bipolar disorder, and anxiety. Our primary objective is to deepen our understanding of the global mental health landscape, uncovering trends across nations and identifying regional disparities. By addressing key questions such as overall prevalence trends, cross-country variations, and patterns in response to global events, we aim to inform future interventions and policies effectively. Employing data analysis tools such as visualization, feature engineering, linear regression, and bagging, we seek to advance mental health awareness, prevent disorders, and promote recovery-focused interventions worldwide.

Table of Contents

1. Introduction

Aligned with the World Federation for Mental Health's mission, our research aims to combat mental disorders and facilitate recovery globally. Using the multiple datasets from Kaggle, we investigate how global events impact mental health, including depression, bipolar disorder, and anxiety. By analyzing trends across nations and identifying regional disparities, we aim to inform effective interventions and policies. Leveraging data analysis tools, we seek to advance mental health awareness and promote recovery-focused interventions worldwide.

2. Data

2.1 Dataset Characteristics

1. Global Mental Health Disorders [1] [4.88 MB]

This large data set comprises significant insights into the prevalence of 7 mental health conditions: Schizophrenia, Bipolar Disorder, Eating Disorders, Anxiety Disorders, Drug Use Disorders, Depression, and Alcohol Use Disorders. Each row of the dataset represents the percentage of citizens with the listed disorders for a country or territory in some year between 1990 to 2017. In this dataset, there are 233 unique countries and territories that we have analyzed.

2. Country Mapping [2] [19.7 KB]

This dataset consists of all countries and territories and their respective ISO codes, regions, subregions, and additional identification codes. This dataset allows us to group countries with neighboring countries to determine geographical trends in mental health disorders at a more global scale rather than individual.

3. Natural Disasters [3] [4.7 MB]

This dataset was used to discover trends and correlations between global events and mental health disorders. This dataset logs all natural disasters ranging from 1900 to 2021; however, we only used values that are dated past 1990. Each row represents a single disaster that has occurred, and gives information such as location, disaster type, total number of people affected, total deaths, and more.

2.2 Data Cleaning and Feature Engineering

In order to effectively use our three datasets to produce answers to our overarching questions, we used Pandas to merge the datasets together into one dataframe. Due to the initial dataset containing extremely minimal information, we merged the Country Mapping dataset into our working dataset to gain access to information such as what regions, subregions, and Continents a country belongs to. This way, we were able to identify trends in geographical locations in respect to mental health disorders.

To determine correlations and responses to global events, we incorporated data from the Natural Disasters dataset. Out of the many features provided in this dataset, we determined that the most informative ones would simply be the country/territory, year, disaster type, number of deaths, and number of people affected. Thus, we aggregated these values per country per year, and appended these new columns to our working dataset.

3. Analysis

3.1 Methods, Approaches and Supporting Questions

To approach this broad goal of finding trends and patterns in mental disorders at a global scale, we created several supporting questions to help us understand the raw data and apply it to find solutions to address this issue. First, we wanted to understand the general trajectories of these different mental health disorders at a global scale and highlight subregions in which each disorder is the most prevalent. To achieve this, we created multiple **data visualizations** through **feature engineering** that aggregated and averaged the overall percentages of each disorder. Then, we generated additional graphs to depict the top subregions for which these disorders were the most prevalent.

Next, we wanted to be able to predict the changes in disorders for all countries. To do this, we created a script that uses **linear regression** to generate a model that predicts the level of some given disorder in a given country for a given year (e.g. predict the rate of Schizophrenia in Turkey in 2040). Additionally, we created a script that would return the top *n* countries with the highest predicted rate of increase for a given disorder based on a linear regression model.

Another question we seeked to answer was how do major catastrophes affect disorders. To achieve this, we used a **random forest regressor** model and the **mean squared error** to determine and understand the feature importances for such events on mental health.

Lastly, we aimed to determine the correlation and impact of different mental health disorders on each other. To do this, we created a correlation matrix with heatmap to show one disorder may be linked to another.

3.2 Data Visualizations

3.2.1 Global Trajectory of Mental Disorders

Understanding the global trajectory of mental disorders is essential for assessing the impact of global events and enhancing predictive modeling efforts. We refrained from plotting all the disorders on a single graph due to the varying magnitudes of change associated with each disorder. We wanted to avoid potentially distorting the accuracy of trends observed for individual disorders.

From 1990 to 2017, schizophrenia, bipolar, eating, anxiety, and drug use disorders show a relatively increasing linear trend as the years go by (Figs. 1, 2, 3, 4, 5). As for depression, there is a surprising drastic non linear decrease in global prevalence percentage over time (Fig. 6). Alcohol use shows an inconsistent non linear pattern that has its ups and down but overall increasing in the recent years (Fig. 7).

3.2.2 Top Subregions with Predominant Disorder Prevalence

We conducted a thorough examination using the subregions dataset to extract insights regarding the prevalence of mental health disorders. The outcomes are summarized in the table below:

Table 1: Top 5 Subregions of Each Mental Health Disorder (based on Figs 8 - 14)

DISORDER	1	2	3	4	5
Schizophrenia	Australia & New Zealand	Northern America	Western Europe	Eastern Asia	Northern Europe
Bipolar	Australia & New Zealand	Western Europe	Northern Europe	Latin America & the Caribbean	Western Europe
Eating	Australia & New Zealand	Western Europe	Northern America	Northern Europe	Southern Europe
Anxiety	Australia & New Zealand	Western Europe	Northern America	Northern Europe	Northern Europe

Drug Use	Australia & New Zealand	Northern America	Northern Africa	Western Asia	Western Family
Depression	Northern America	Australia & New Zealand	Northern Europe	Northern Africa	Western Europe
Alcohol Use	Eastern Europe	Northern Europe	Central Asia	Northern America	Latin America & the Caribbean

Our analysis revealed that the Australia and New Zealand subregion consistently rank among the top five regions most often with the highest percentages of various mental health disorders, as depicted in Figures 8 to 14.

3.3 Linear Regression

3.3.1 Predictive Modeling of Disorder Trends and Trajectories

We developed a predictive method that takes into account a country, a specific mental health disorder, and a designated time frame to forecast the prevalence of that disorder in the given country. This method trains a linear regression model on the country's historical data for the specified year, which is then used to predict the disorder rate for the desired time frame. This model is able to identify trends in various countries, enabling policymakers and other international organizations to pinpoint areas that require additional aid and resources to address the rising prevalence of mental health disorders. Furthermore, it is able to give insight into how different disorders are projected to grow or decrease in following years, which may be valuable to mental health clinics and other medical professionals in the field. This method can inform data driven decisions to allocate resources more effectively and improve mental health outcomes globally.

Fig 15: Eating Disorder Prediction in Taiwan

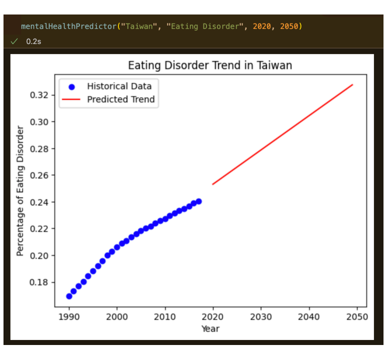


Fig 16: Anxiety Prediction in Uruguay

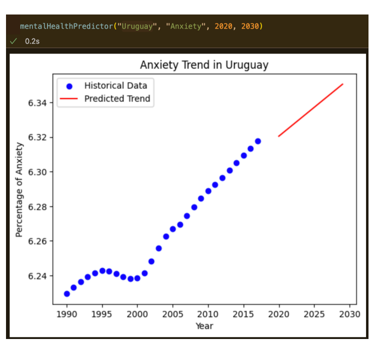
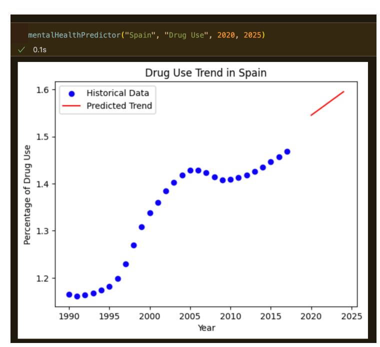


Fig 17: Drug Abuse Prediction in Spain



3.3.2 Forecasting Countries with Greatest Disorders Growth

We also developed a script that enables the identification of countries that are likely to experience a significant surge in a specific mental health disorder. This script takes as input a target disorder and a desired number of countries (*n*), and returns the top *n* countries that are predicted to have the highest increase in the given disorder. To achieve this, the script runs a linear regression model on *all* countries for the specified disorder, calculates the percentage increase, and then ranks the countries accordingly. The output provides valuable insights into the countries that are expected to experience alarmingly large increases in the disorder, thereby allowing policymakers and related organizations to allocate resources more effectively to the most critical locations and respond proactively to emerging mental health crises. From the examples below, we are able to learn that Afghanistan and Libya have the highest predicted increase in drug use over the next year. Additionally, Equatorial Guinea and Myanmar seem to also be on the rise for Schizophrenia, and Mongolia and Kazakhstan have concerning rising levels of Alcoholism. With this script, we are able to easily determine which countries need immediate resources and attention to address these rising levels of mental disorders.

Fig 18: Top 10 Countries with Greatest Drug Use Increase

```
get_top_countries_with_highest_percent_change("Drug Use", 10)
```

✓ 0.7s

	PercentChange
Afghanistan	1.428612
Libya	1.241115
United Arab Emirates	1.196421
Lebanon	1.057106
Chile	1.056258
United States	1.045779
Tunisia	0.980588
Estonia	0.973190
Sweden	0.918665
Malta	0.900683

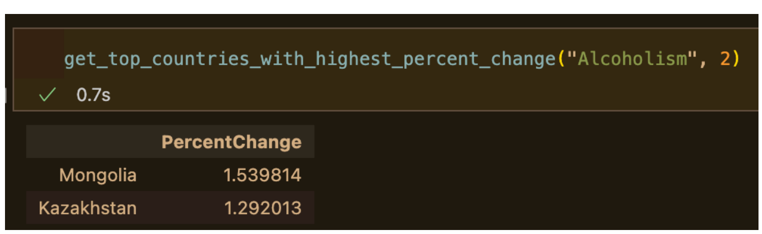
Fig 19: Top 5 Countries with Greatest Schizophrenia Increase

```
get_top_countries_with_highest_percent_change("Schizophrenia", 5)
```

✓ 0.7s

	PercentChange
Equatorial Guinea	0.716820
Myanmar	0.454037
Maldives	0.380717
Laos	0.349308
Malaysia	0.345127

Fig 20: Top 2 Countries with Greatest Alcoholism Increase



```
get_top_countries_with_highest_percent_change("Alcoholism", 2)
```

✓ 0.7s

	PercentChange
Mongolia	1.539814
Kazakhstan	1.292013

3.4 Random Forest Regression: MSE and Feature Contributions

3.4.1 Quantifying the Impact of Global Events on Mental Health Disorder Rates

Fig 21: MSE and Feature Importance Output

```
Model for Schizophrenia: MSE = 0.00
Model for Bipolar: MSE = 0.02
Model for Eating Disorder: MSE = 0.03
Model for Anxiety: MSE = 1.32
Model for Drug Use: MSE = 0.16
Model for Depression: MSE = 0.41
Model for Alcoholism: MSE = 0.58
Feature Importances for Schizophrenia:
  Number of Disasters: 0.41
  Total Affected: 0.29
  Total Deaths: 0.30

Feature Importances for Bipolar:
  Number of Disasters: 0.27
  Total Affected: 0.26
  Total Deaths: 0.47

Feature Importances for Eating Disorder:
  Number of Disasters: 0.34
  Total Affected: 0.34
  Total Deaths: 0.32

Feature Importances for Anxiety:
  Number of Disasters: 0.33
  Total Affected: 0.33
  Total Deaths: 0.34

Feature Importances for Drug Use:
  Number of Disasters: 0.33
  Total Affected: 0.32
  Total Deaths: 0.35

Feature Importances for Depression:
  Number of Disasters: 0.32
  Total Affected: 0.31
  Total Deaths: 0.37

Feature Importances for Alcoholism:
  Number of Disasters: 0.31
  Total Affected: 0.33
  Total Deaths: 0.35
```

In order to determine the impact of global events on mental health, we implemented a random forest regressor to predict the rate of various mental health disorders based on disaster related features. We chose to implement a random forest regressor to improve our original linear regressor and reduce variance in our output and model. The output of running this is shown on the left (Figs 22 to 28 were created based on the outputs for better interpretability). The mean squared error (MSE) represents the average squared difference between our predicted and actual value for the disorder. Based on our outputs, the MSE values are relatively low for most disorders, indicating that the model is performing well. However, the MSE value for Anxiety (1.32) is higher than the others, suggesting that the model may not be performing as well for this specific disorder. The listed feature importances reflect the relative importance of each feature in predicting the disorder. For each disorder, the feature importance values are normalized to sum up to 1, and features with higher values reflect a higher impact on the prevalence of the disorder. In our output, we can see that the number of disasters is an extremely important feature for all disorders, the total affected and total deaths are also important, but their importance varies across disorders. With this information, we now know that all three features are important factors in predicting mental health disorders (except for anxiety), and we use this information to predict increases in specific mental health disorders based on the magnitude and types of natural disasters occuring in different countries. For example, we can infer that deaths caused by natural disasters are closely linked to bipolar disorder and depression, while Schizophrenia is closely linked to the number of disasters.

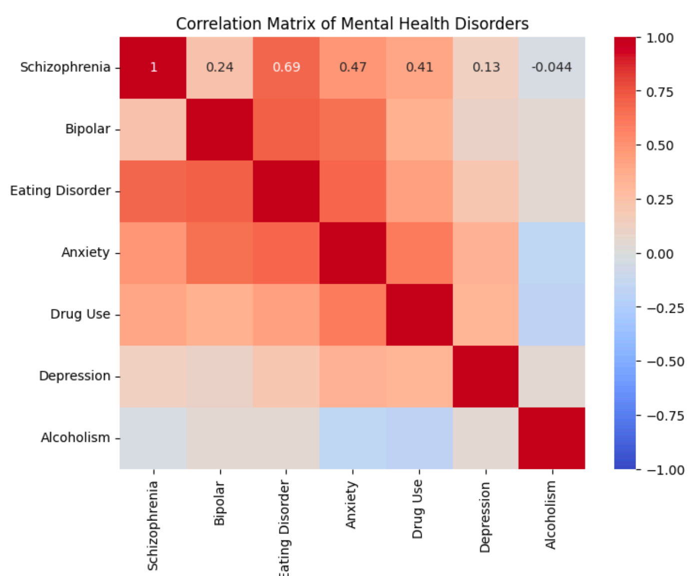
3.5 Correlations Between Disorders

Mental health disorders often co-occur with each other, which means that knowing the correlations helps us understand the complex interplay between them. For instance, understanding if depression often coexists substance abuse allows for more comprehensive treatment approaches. Existing correlations can guide further research to investigate the underlying mechanisms that link certain disorders or explore common risk factors contributing to their co-occurrence.

3.5.1 Correlation Matrix of Mental Health Disorders

In our analysis, we explored the Correlation Matrix of Mental Health Disorders to gain insights into the interconnectedness of the mental health conditions from the dataset. There seems to be higher degrees of correlation between eating disorders and schizophrenia, bipolar, and anxiety disorders. Using this information, organizations specializing in mental health can develop targeted interventions that take into account the relationships between disorders, adopting a more holistic approach to treatment that addresses underlying factors, leading to more effective treatment outcomes, improved patient care, and better resource allocation.

Figure 29: Correlation Matrix of Mental Health Disorders



4. Ethical Aspects

Upon reflecting on our project, we believe that there are no characteristics in our project that could classify our mission or methodologies as a Weapon of Math Destruction. Through our analysis and findings, we are only able to highlight correlations between mental health disorders, their geographical locations, and natural disasters. While our algorithms and methodologies are scalable, there is nothing that can be done with the learned information that could potentially cause harm. Our sole aim is to help people identify, predict, and take action when cases of mental health disorder arise, and these models we created have no potential to cause destruction or danger upon anyone. Even if our model were to incorrectly label a country as “high risk” of some disorder, it would actually benefit that country, as mental health organizations would prioritize health care and deliver proper resources to that country to address this disorder.

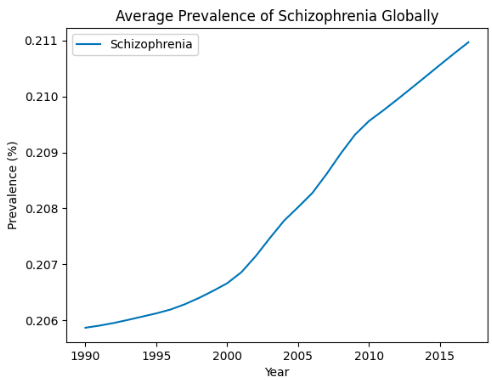
5. Conclusion

In conclusion, we were able to leverage different global datasets to investigate the trajectories and trends of different countries and regions. Using data visualization and different predictive modeling techniques, we were able to identify different geographical regions for different disorders and forecasted future trends. The impact of global events on mental health was further quantified with the help of machine learning models. The insights gained from this study can help enterprises, nonprofit organizations, and even governments form targeted interventions, create better resource allocation strategies, and make more informed policy decisions to help combat mental health challenges globally. We are extremely confident in our results, as many of the correlations and trends we uncovered have either scientific reasoning or logical reasoning that proves our discoveries. However, we believe that there is definitely more exploration to be done and patterns to be uncovered that can further advance our understanding of mental health and its relationships. Hopefully, these tools and algorithms can be implemented in the future by organizations, like the World Federation for Mental Health (WFMH), to bring accessible mental health aid to those in need.

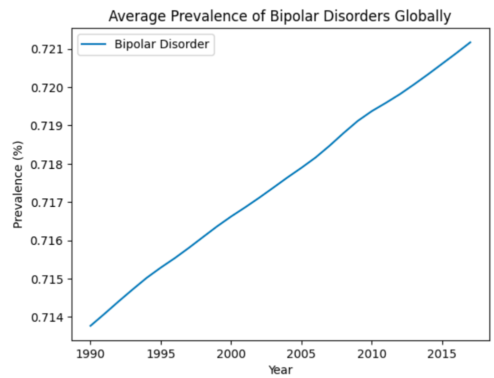
6. Appendix

Figures 1 - 7: Average Prevalence of Mental Health Disorders Globally from 1990 to 2017

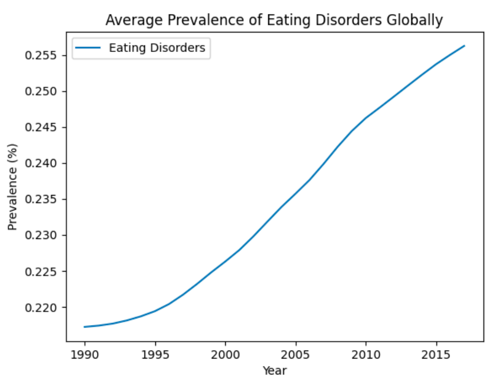
1. Schizophrenia



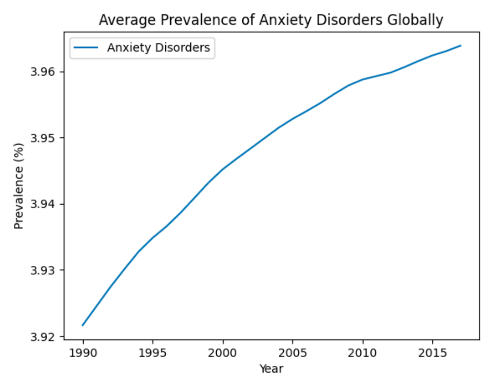
2. Bipolar Disorder



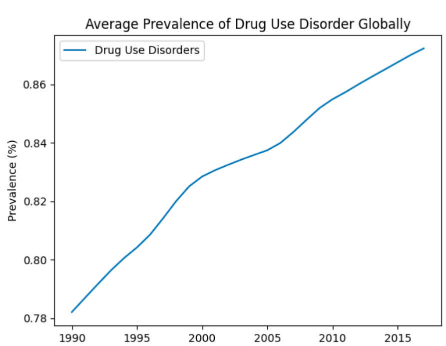
3. Eating Disorder



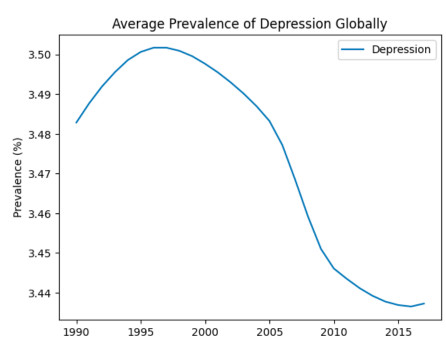
4. Anxiety Disorder



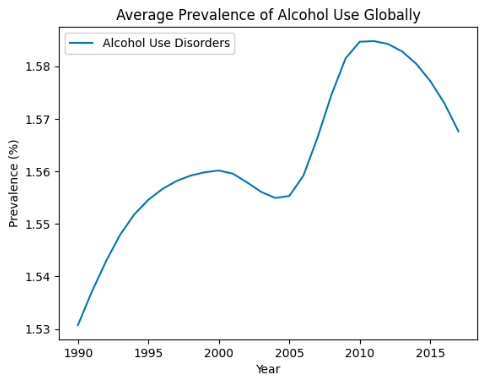
5. Drug Use Disorder



6. Depression

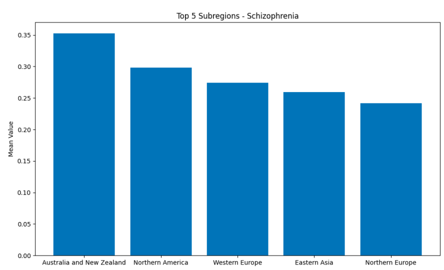


7. Alcohol Use Disorder

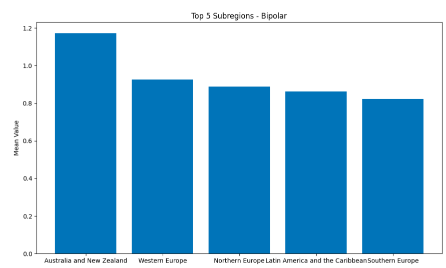


Figures 8 - 14: Top 5 Subregions for Mental Health Disorders

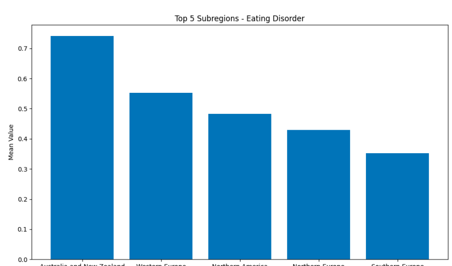
8. Schizophrenia



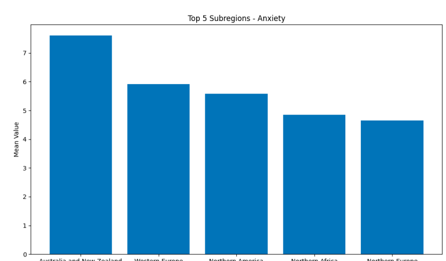
9. Bipolar Disorder



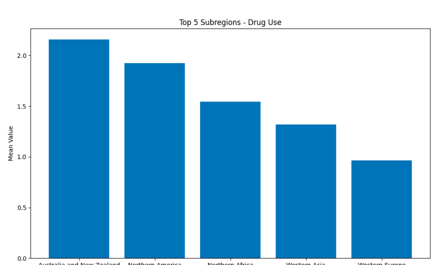
10. Eating Disorder



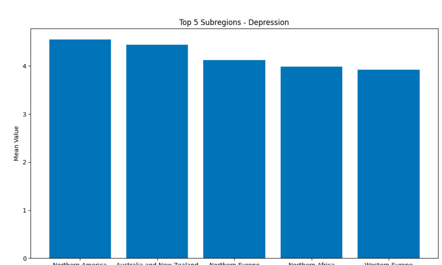
11. Anxiety Disorder



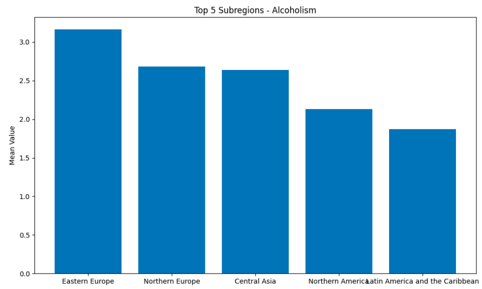
12. Drug Use Disorder



13. Depression

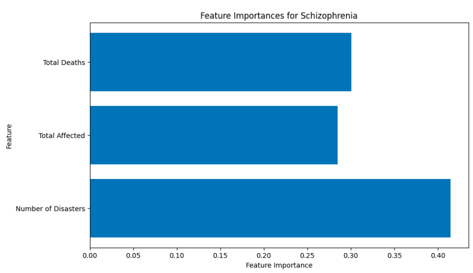


14. Alcohol Use Disorder

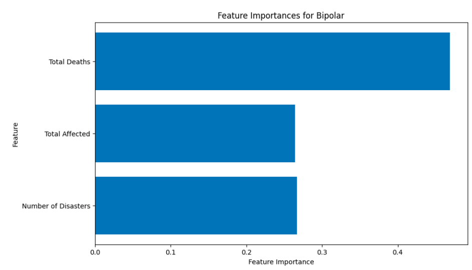


Figures 19 - 25: Feature Importance for Mental Health Disorders

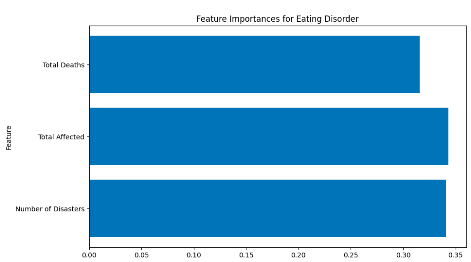
19. Schizophrenia



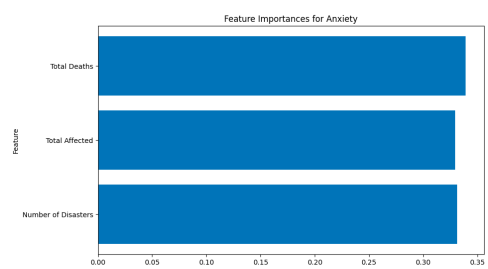
20. Bipolar Disorder



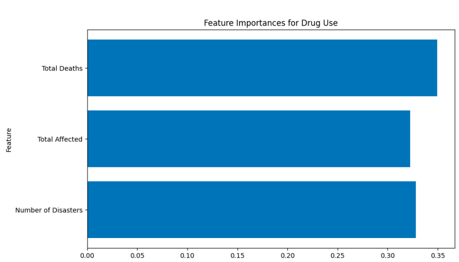
21. Eating Disorder



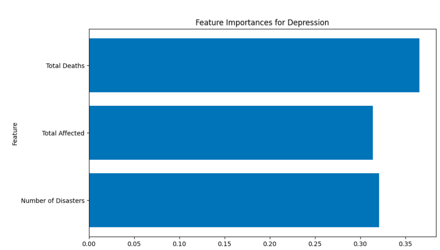
22. Anxiety Disorder



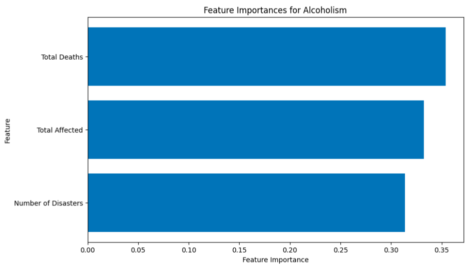
23. Drug Use Disorder



24. Depression



25. Alcohol Use Disorder



7. Citations

[1] The Devastator. (2022). Global Mental Health Disorders [Data set]. Kaggle. https://www.kaggle.com/datasets/thedevastator/global-mental-health-disorders

[2]Olteanu, A. (2022). Country Mapping (ISO, Continent, Region) [Data set]. Kaggle. https://www.kaggle.com/datasets/andradaolteanu/country-mapping-iso-continent-region

[3]Dinçer, B. R. (2023). All Natural Disasters 1900-2021 - EoSDIS [Data set]. Kaggle. https://www.kaggle.com/datasets/brsdincer/all-natural-disasters-19002021-eosdis/data