# Project Proposal: Application of Data Analysis Techniques in Exploring Mental Health Trends

*Time Series Forecasting, Correlation Analysis, and Clustering of Mental Health Disorder Prevalence Data*


Submitted by:
ALI SHOAIB (NIM-BSCS-2021-33)

Submitted to:
MR. MOHAMMAD BILAL

Namal University, Mianwali

June 11, 2024

**Abstract**

This project aims to apply various data analysis, preprocessing, and visualization techniques to a dataset related to mental health disorders. The dataset includes information on the prevalence of mental health disorders across different countries and years. The goal is to analyze trends, identify significant correlations, forecast future prevalence rates, and perform clustering analysis to group countries based on mental health metrics.

## Keywords

Mental Health, Data Analysis, Time Series Forecasting, Clustering, Machine Learning

## 1 Introduction

Mental health is a critical aspect of overall well-being, and understanding its trends and patterns can help in developing better health policies and interventions. This project uses a dataset from Kaggle, which contains information on the prevalence of various mental health disorders across different countries and years. The main objectives are to analyze the data, identify trends, forecast future prevalence rates, and perform clustering analysis.

## 2 Related Work

Several studies have investigated the application of machine learning techniques in the detection and prediction of depression and other mental health disorders. Ali et al. [1] developed and analyzed machine learning methods for predicting depression among menopausal women. Li et al. [2] explored depression recognition using various machine learning methods with different feature generation strategies. Aekwarangkoon and Thanathamathee [3] investigated associated patterns and predicting models of life trauma, depression, and suicide using ensemble machine learning techniques.

IEEE Transactions on Affective Computing published a comprehensive study by Author and Author [4] on the use of machine learning for depression detection. Additionally, Author and Author [5] presented a study on machine learning models for mental health analysis in the Proceedings of the International Conference on Artificial Intelligence and Data Science.

These studies contribute to the growing body of research focused on leveraging machine learning for mental health analysis and highlight the potential of these techniques in addressing the challenges associated with depression recognition and prediction.

# 3 Dataset Description

## 3.1 Source

The dataset is sourced from Kaggle: Mental Health of a Person https://www.kaggle.com/datasets/rithika19/mental-health-of-a-person

## 3.2 Description

The dataset includes the following columns:

- **Entity**: Country name
- **Code**: Country code
- **Year**: Year of the data
- **Schizophrenia disorders**: Prevalence of schizophrenia (age-standardized, both sexes)
- **Depressive disorders**: Prevalence of depressive disorders (age-standardized, both sexes)
- **Anxiety disorders**: Prevalence of anxiety disorders (age-standardized, both sexes)
- **Bipolar disorders**: Prevalence of bipolar disorders (age-standardized, both sexes)
- **Eating disorders**: Prevalence of eating disorders (age-standardized, both sexes)

# 4 Research Questions

1. **Trend Analysis**: How have the prevalence rates of different mental health disorders changed over the years in different countries?
2. **Correlation Analysis**: Are there any significant correlations between the prevalence of different mental health disorders?
3. **Time Series Forecasting**: Can we predict the prevalence of mental health disorders for the next 10 years using time series forecasting techniques?
4. **Clustering Analysis**: Can we group countries based on the prevalence rates of mental health disorders?

# 5 Preliminary Thoughts on Potential Challenges

- **Missing Values**: The dataset may contain missing values that need to be addressed.

- **Data Cleaning**: There may be inconsistencies or errors in the data that require cleaning.
- **Outliers**: The presence of outliers may affect the analysis and forecasting.
- **Data Transformation**: Appropriate transformations may be necessary to prepare the data for modeling.

# 6 Plan to Address Challenges

- **Missing Values**: Identify and handle missing values using techniques like imputation or deletion based on the extent and nature of the missing data.
- **Data Cleaning**: Perform initial data cleaning to correct any inconsistencies or errors.
- **Outliers**: Use statistical methods and visualizations to detect and handle outliers appropriately.
- **Data Transformation**: Apply necessary transformations such as scaling and encoding to prepare the data for analysis and modeling.

# 7 Project Phases

## 7.1 Phase 1: Exploratory Data Analysis (EDA)

- **Data Description**: Provide a detailed description of the dataset, including variables, types, and summary statistics.
- **Visualizations**: Create various plots to visualize the data.
- **Initial Findings**: Discuss initial observations and insights.

## 7.2 Phase 2: Data Preprocessing

- **Handle Missing Values**: Identify and address missing values.
- **Remove Duplicates**: Identify and remove any duplicate records.
- **Outliers**: Detect and handle outliers appropriately.
- **Data Transformation**: Apply necessary transformations such as scaling and encoding.
- **Document Process**: Provide a detailed report on the steps taken and the rationale behind the decisions.

### 7.3 Phase 3: Correlation Analysis

- **Correlation Matrix**: Calculate and visualize the correlation matrix.
- **Significant Correlations**: Identify and discuss significant correlations.

### 7.4 Phase 4: Time Series Analysis

- **Time Series Data**: Perform a time series analysis on the dataset.
- **Trends and Seasonality**: Identify trends and seasonal patterns.
- **Forecasting**: Apply a forecasting model to predict future prevalence rates.

### 7.5 Phase 5: Modeling

- **Regression Analysis**: Apply regression models and evaluate them.
- **Classification Analysis**: Apply and compare different classification algorithms, if applicable.
- **Clustering Analysis**: Apply clustering algorithms, visualize, and interpret the clusters.

### 7.6 Phase 6: Final Report and Presentation

- **Comprehensive Report**: Compile a final report detailing each phase of the project, including code snippets, visualizations, and findings. Format the report according to IEEE standards.
- **Presentation**: Prepare a 10-minute presentation summarizing the project, methods, and key insights.

## 8 Deliverables

- Proposal Document
- EDA Report
- Preprocessing Report
- Correlation Analysis Report
- Time Series Analysis Report
- Modeling Report
- Final Comprehensive Report (IEEE Format)
- Presentation Slides

## 9 Evaluation Criteria

- Data Selection and Proposal
- Exploratory Data Analysis
- Data Preprocessing
- Correlation and Time Series Analysis
- Modeling
- Reporting and Presentation

## 10 Tools and Technologies

- **Programming Language**: Python
- **Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, etc.
- **Documentation**: Jupyter Notebooks or Google Colab

# References


[1] M. M. Ali, H. A. A. Algashamy, E. Alzidi, K. Ahmed, F. M. Bui, S. K. Patel, S. Azam, L. F. Abdulrazak, and M. A. Moni, "Development and performance analysis of machine learning methods for predicting depression among menopausal women," *Healthcare Analytics*, vol. 3, p. 100202, 2023. [Online]. Available: https://scholar.google.com/scholar?as_q=Development+and+performance+analysis+of+machine+learning+methods+for+predicting+depression+among+menopausal+women&as_occt=title&hl=en&as_sdt=0%2C31

[2] X. Li, X. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, C. Xia, and B. Hu, "Depression recognition using machine learning methods with different feature generation strategies," *Artificial intelligence in medicine*, vol. 99, p. 101696, 2019. [Online]. Available: https://scholar.google.com/scholar?as_q=Depression+recognition+using+machine+learning+methods+with+different+feature+generation+strategies&as_occt=title&hl=en&as_sdt=0%2C31

[3] S. Aekwarangkoon and P. Thanathamathee, "Associated patterns and predicting model of life trauma, depression, and suicide using ensemble machine learning. emerging science journal, 6 (4), 679-693. doi: 10.28991," ESJ-2022-06-04-02, Tech. Rep., 2022. [Online]. Available: https://scholar.google.com/scholar?as_q=Associated+Patterns+and+Predicting+Model+of+Life+Trauma%2C+Depression%2C+and+Suicide+Using+Ensemble+Machine+Learning&as_occt=title&hl=en&as_sdt=0%2C31

[4] A. Author and B. Author, "A comprehensive study on the use of machine learning for depression detection," *IEEE Transactions on Affective Computing*, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10421262

[5] C. Author and D. Author, "A study on machine learning models for mental health analysis," in *Proceedings of the International Conference on Artificial Intelligence and Data Science*. Springer, 2023, pp. 123–134. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-97-1329-5_9