**Exploring and Modeling**

**on Mental Health Data**

**: Insights from Kaggle Data Analysis**

Prepared for: Prof. Divya Pandove

(TA: Bhumika Mallikarjunhorapet)

Prepared by: Team 7

Haeyeon Jeong, Sai Rachana Kandikattu

Abirham Getie, Yonathan Shimelis

Dec 17, 2024

Introduction1

Research Methods and Results2

*Task 1. Data Acquisition and Data Preprocessing 2*

*Task 2. Exploratory Data Analysis and Statistical Testing* 5

*Task 3. Modeling* 7

*Task 4. Model Evaluation* 8

Results12

Conclusions12

Recommendations12

References 14

References 14

Appendix 14

**Introduction**

Mental health challenges are becoming more visible across different sectors, affecting individuals in diverse ways. Understanding the factors that contribute to mental stress is essential for developing effective support systems. This research will explore mental health data using various classification models to identify key factors influencing mental health outcomes. Specifically, we will examine how variables such as mood swings, family history, treatment status, occupations, and time spent indoors affect mental health. By analyzing these factors through data models, we hope to uncover patterns that can help improve mental health management and prevention strategies.

This study seeks to address the following key question:

* How can we develop models to assess mental stress, and which factors are most influential in predicting mental health outcomes?

**Research Methods**

To achieve these objectives, we will perform the following tasks:

***Task 1. Data Acquisition and Preparation***

The dataset used in this study is available on Kaggle at this ([link](https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset)). It contains 292,364 observations with 17 variables.

***Task 2. Exploratory Data Analysis and Data Visualization***

We will conduct exploratory data analysis to identify key trends, relationships, and distributions in the dataset.

***Task 3. Modeling and Evaluation***

Various classification models will be employed, including Logistic Regression, SVM, KNN, and Random Forest, to assess the impact of identified factors on mental health outcomes.

**Results**

**Conclusions**

**Recommendations**

**Reference**

Git Repository: [**github**](https://github.com/sairachanak/6103-Team7.git)