Mental Health Prediction

Project Report

**Table of Contents**

**Introduction**

* Background
* Objective
* Dataset

**Data Wrangling**

* Data Collection
* Data Preprocessing
* Handling Missing Values
* Feature Engineering

**Exploratory Data Analysis (EDA)**

* Correlation Analysis
* Data Distribution

**Model Development**

* Feature Selection
* Model Selection
* Model Training
* Model Validation

**Model Evaluation**

* Confusion Matrix
* Classification Report
* Model Comparison

**Conclusion**

* Summary of Findings
* Limitations
* Future Work

**1. Introduction**

**Background**

Mental health is a critical aspect of an individual's well-being, and early detection of mental health conditions can significantly impact treatment outcomes. This project aims to develop predictive models for mental health based on various factors to assist in identifying individuals who may be at risk.

**Objective**

The primary objective of this project is to predict mental health conditions using machine learning techniques. Specifically, we will develop three different classification models: Logistic Regression, Decision Tree Classification, and Random Forest Classification. We will evaluate these models using confusion matrices and classification reports to assess their performance.

**Dataset**

The dataset used for this project contains various features related to individuals' demographics, lifestyle, and other relevant factors, along with a binary target variable indicating whether an individual has a mental health condition (1) or not (0).

**2. Data Wrangling**

**Data Collection**

The dataset was obtained from Kaggle (link: https://www.kaggle.com/code/jagannathrk/mental-health-prediction/input), and it consists of 334 samples and 31 features.

**Data Preprocessing**

Data Cleaning: Removed duplicates and irrelevant columns.

Data Encoding: Encoded categorical variables using one-hot encoding.

Data Splitting: Split the dataset into training and testing sets (70% training and 30% testing).

**Handling Missing Values**

Missing values were handled through methods such as imputation using mean values where appropriate.

**Feature Engineering**

Created new features or transformed existing ones to improve model performance and interpretability.

**3. Exploratory Data Analysis (EDA)**

**Heatmaps to show feature correlations.**

**Correlation Analysis**

Analyzed correlations between features and the target variable to identify potential predictors.

**Data Distribution**

Visualized the distribution of the target variable (mental health condition) to understand class imbalance.

**4. Model Development**

**Feature Selection**

Used feature importance techniques to select relevant features for modeling.

**Model Selection**

Three classification models were chosen:

* Logistic Regression
* Decision Tree Classification
* Random Forest Classification

**Model Training**

Each model was trained using the training dataset.

**Model Validation**

Model performance was assessed using the testing dataset.

**5. Model Evaluation**

**Confusion Matrix**

Confusion matrices were generated for each model to calculate metrics such as accuracy, precision, recall, and F1-score.

**Classification Report**

A classification report for each model was created, providing a detailed breakdown of performance metrics for both positive and negative classes.

**Model Comparison**

Models were compared based on their evaluation metrics to identify the best-performing model.

**6. Conclusion**

**Summary of Findings**

Identified key features associated with mental health conditions. Evaluated three models (Logistic Regression, Decision Tree, Random Forest) for predicting mental health conditions. Presented confusion matrices and classification reports for model evaluation.

**Limitations**

* Limited dataset scope and availability.
* Model performance may vary based on different datasets.

**Future Work**

* Explore additional features and data sources.
* Investigate techniques for addressing class imbalance.
* Fine-tune models for improved performance.
* Deploy the best-performing model for real-time predictions and intervention.