**Mental Health Prediction System - Summary Report**

## 1. Introduction

This report presents a detailed analysis of the Mental Health Prediction System. The system predicts depression and anxiety severity levels using machine learning models and provides explanations with suggested coping mechanisms.

## 2. Dataset Insights

### 2.1 Data Description

The dataset includes various attributes related to mental health, such as:

- **Demographics**: Age, gender, BMI
- **Mental Health Scores**: PHQ-9 (depression severity), GAD-7 (anxiety severity), Epworth Sleepiness Score
- **Mental Health Diagnoses & Treatments**: Depression and anxiety diagnoses, treatments, and symptoms

### 2.2 Data Preprocessing

- **Handled missing values** by filling numerical columns with median values and categorical columns with mode.
- **Encoded categorical variables** (gender, BMI category, severity levels) using Label Encoding.
- **Feature Engineering**:
    - Created interaction term: bmi_who_interaction = bmi * who_bmi
    - Combined mental health scores: total_mental_health_score = phq_score + gad_score

### 2.3 Correlation Analysis

- **PHQ-9 Score, GAD-7 Score, and Anxiousness** were highly correlated with depression and anxiety severity.
- **BMI and sleepiness** had a moderate impact on mental health severity.

## 3. Model Development & Performance

### 3.1 Machine Learning Models Used

- Random Forest Classifier
- MLP Neural Network
- Logistic Regression

| 3.2 Model Evaluation Metrics | | | | | |
|---|---|---|---|---|---|
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **ROC-AUC** |
| **Random Forest (Depression)** | 0.97 | 0.97 | 0.97 | 0.97 | 0.99 |
| **Random Forest (Anxiety)** | 0.98 | 0.97 | 0.98 | 0.98 | 0.99 |
| **MLP Neural Network (Depression)** | 0.78 | 0.78 | 0.78 | 0.78 | 0.95 |
| **MLP Neural Network (Anxiety)** | 0.73 | 0.81 | 0.73 | 0.76 | 0.91 |
| **Logistic Regression (Depression)** | 0.86 | 0.87 | 0.86 | 0.86 | 0.97 |
| **Logistic Regression (Anxiety)** | 0.89 | 0.90 | 0.89 | 0.89 | 0.97 |

### 3.3 Feature Importance Analysis (SHAP)

- **PHQ-9 Score and GAD-7 Score** were the most important predictors for depression and anxiety.
- **Anxiousness and Suicidal tendencies** had a strong influence on classification.
- **Epworth Sleepiness Score and BMI interaction terms** played a moderate role in predictions.

## 4. Explanation & Coping Mechanisms

### 4.1 LLM-based Explanations

Using an open-source LLM, the system generates:

- **Personalized explanations** for the predicted mental health severity.
- **Coping strategies** tailored to the severity level.

### 4.2 Suggested Coping Mechanisms

| Severity Level | Coping Strategies |
|---|---|
| **Mild** | Mindfulness, journaling, light exercise |
| **Moderate** | Support groups, counseling, stress management |

| Severity Level | Coping Strategies |
| --- | --- |
| **Severe** | Professional intervention, medication, therapy |

### 5. Findings & Recommendations

- **Random Forest outperforms other models**, making it the best choice for predictions.
- **PHQ-9 and GAD-7 Scores are the most critical features** in determining severity levels.
- **LLM integration provides human-like explanations**, enhancing user engagement.
- **Future work** can involve fine-tuning a transformer model for mental health explanations.

---

📌 **Developed by: Sai Chiranthan H M**