Mental Health Prediction System Report

1. Dataset Preprocessing Steps

The dataset preprocessing was handled within model_testing.ipynb, ensuring clean and structured data for:

- Handling Missing Values: Missing entries were filled using appropriate imputation techniques.
- Feature Encoding: Categorical variables (e.g., age range, gender, CGPA, and scholarship) were encoded
- Normalization: Numeric features were normalized to ensure the model's efficiency.
- Train-Test Split: The dataset was divided into training and testing sets for validation.

2. Model Selection Rationale

The model chosen for prediction is a Random Forest Classifier, due to:

- Robustness to Overfitting: It generalizes well on unseen data.
- Feature Importance Interpretation: It allows understanding which factors contribute most to mental health - High Accuracy: The model demonstrated strong performance in validation tests.

Additionally, a T5-based LLM (google/flan-t5-base) was used to generate detailed explanations for the pred

3. How to Run the Inference Script

The inference script is contained in predict_mental_health_UI.py. Follow these steps to run it:

- Ensure Dependencies Are Installed: pip install streamlit numpy joblib torch transformers
- Run the Script Using Streamlit:

streamlit run predict_mental_health_UI.py

- Interact with the UI: Enter required details, and the system will predict the mental health condition based

4. UI Usage Instructions

- Input Details: Select your age range, gender, CGPA, and scholarship status.
- Rate Symptoms: Adjust sliders (0-5) for various mental health symptoms.
- Prediction Output: Click the "Predict Mental Health Condition" button to get the result.
- LLM-Generated Explanation: The system provides a detailed text explanation of the diagnosed condition.

This system provides an interactive way to assess and understand mental health conditions based on user