Final Project Proposal

Mental Health & Depressive Disorder Risk Analysis & Prediction Using Lifestyle and Medical Data

- ## Project Description

This project aims to analyze the impact of medical and lifestyle factors on the risk of depressive disorder and other mental health issues, and to build a predictive model that can estimate the likelihood of developing depression based on individual behavioral and health indicators.

However, it's important to acknowledge that mental health is influenced by a wide range of psychological, emotional, and social factors that extend far beyond the measurable lifestyle features included in this dataset.
Therefore, while the model may not achieve very high accuracy, its predictions still highlight significant patterns and useful indicators that can contribute to understanding mental health risks and guiding further clinical evaluation

- ## Group Members & Roles
  - ✓ Karim Mohamed (Team Leader)
  - ✓ Naira Zaidan
  - ✓ Ahmed Zaghloul
  - ✓ Mohamed Hisham
  - ✓ Hossam Eldeen Ahmed
  - ✓ Aya Elsaeed

As a team, we decided to collaborate equally across all stages of the project, from data collection and preparation to model development and final deployment. Each member will actively participate in every phase, with specific sub-tasks assigned within each stage (such as data cleaning, feature engineering, model training, evaluation, and deployment).
All decisions and progress will be discussed collectively to ensure accuracy, knowledge sharing, and a consistent workflow throughout the project.

The Team Leader, Karim Mohamed, will coordinate the workflow, ensure task distribution, and oversee progress tracking and communication between team members.

- **Objectives**

1. To analyze general health and lifestyle data to identify patterns and correlations related to mental health risk.
2. To clean, preprocess, and prepare a large dataset for accurate and efficient analysis
3. To develop and evaluate machine learning models that can predict the likelihood of Depressive Disorder based on individual health indicators.
4. To visualize key insights and model results through an interactive dashboard that supports better understanding and decision-making.
5. To deploy the predictive model as an accessible and scalable solution for practical use in healthcare analytics.

- **Tools & Technologies**

  - **Data Processing & Analysis :** Python (Pandas, NumPy).
  - **ML & DL:** Scikit-learn, TensorFlow.
  - **Visualization Tools:** Matplotlib, Seaborn, Plotly, Power BI & Dash (optional for summary reports).
  - **Version Control & Collaboration:** Git & GitHub.
  - **Deployment:** Flask or FastAPI.

- **KPIs (Key Performance Indicators)**

1) Data Quality:
   - Percentage of missing values handled: ≥ 98%
   - Data accuracy after preprocessing: ≥ 95%
2) Model Performance
   - Model accuracy / F1-Score: 72%–77%
   - Model prediction speed (latency): < 200 milliseconds
   - Error rate (False Positive / False Negative): < 15% / < 12%
3) Deployment & Scalability
   - API uptime: ≥ 99%
   - Response time per request: ≤ 300 milliseconds
   - (If applicable) Real-time processing speed: Not applicable / To be evaluated post-deployment
4) Business Impact & Practical Use
   - Reduction in manual effort: ≥ 70%
   - Expected cost savings: ≥ 40%
   - User satisfaction: ≥ 90%

- **Milestones & Deadlines**

Milestone	Description	Deadline
Data Collection & Understanding	Collect the dataset, explore its structure, and define key health and lifestyle features related to heart disease.	Sep 20 – Oct 1
Data Cleaning & Preprocessing	Handle missing values, remove duplicates, encode categorical variables, and prepare data for modeling.	Oct 11 – Oct 15
Exploratory Data Analysis (EDA)	Analyze distributions, visualize trends, and identify correlations between variables.	Oct 16 – Oct 20
Model Development & Training	Build and train machine learning models (e.g., Logistic Regression, Random Forest, XGBoost) to predict Depressive Disorder risk.	Oct 22 – Oct 27
Model Evaluation & Optimization	Assess model performance using metrics like Accuracy, Precision, Recall, and F1-score; tune hyperparameters.	Oct 28 – Nov 5
Dashboard Design & Visualization	Develop an interactive dashboard using Dash/Plotly to present insights and predictions.	Nov 7 – Nov 10
Deployment & Final Report	Deploy the model, finalize documentation, and prepare the project presentation.	Nov 12 – Nov 19