Early Prediction of Hypertension Risk

- ## Project Description

  This project aims to develop a robust machine learning **classification system** for the early prediction of **hypertension risk**(high/low) using a comprehensive dataset of health, demographics, and lifestyle factors such as- **age, gender, BMI, and blood pressure**. The solution will involve data preprocessing, feature analysis, and the comparison of multiple classification algorithms (Random Forest, Logistic Regression) to determine the most accurate model. **Finally,** deploying the best performing model as a high availability low-latency prediction API for potential integration into clinical or public health screening tools.

- ## Group Members & Roles

Name	Role
Abdelrhman Elyamny Moawad	Team Leader, Model Design & Coordination
Ali Alsabahi Ibrahim Metwaly	Data Preprocessing & Features Engineering
Habiba Mohamed Elsayed Elsharkawy	Model Training & Evaluation
Ali Ahmed Abdalhamid Dandsh	API Development &Deployment
Omar Abdelaziz Amer Kassem	Data Visualization
Habiba Abdelmalek El-Sayed	Data Preprocessing & Features Engineering

- **Objectives**

  -To build a predictive model capable of classifying hypertension risk (High or Low).

  -To analyze the impact of key health and lifestyle features on hypertension.

  -To compare multiple classification algorithms for performance and reliability.

  -To deploy the final model as an API for real-time hypertension risk prediction.

  -To ensure scalability, accuracy, and clinical relevance of the system.

- **Tools & Technologies**

  Programming Languages: Python

  Data Processing Libraries: Pandas, NumPy, Scikit-Learn

  Modeling Frameworks: Scikit-learn (Logistic Regression, Random Forest, Decision Tree, XGBOOST)

  API: Streamlit app

  Environment: Jupyter Notebook / VS Code

  Version Control: Git/GitHub

- **Milestones & Deadlines**

Milestone	Description	Deadline
Milestone1	Data collection, Exploration & Preprocessing	5/10/2025
Milestone2	Data Analysis & Visualization	15/10/2025

Milestone3	Predictive Model Development & Optimization	20/10/2025
Milestone4	MLOPs, Deployment & Monitoring	
Milestone5	Final Documentation & Presentation	

- ## KPIs (Key Performance Indicators)

  1. ## Data Quality

     Percentage of missing values handled: 100%

     Data accuracy after preprocessing: 95%

     Dataset diversity (representation of different categories): 95%

  2. ## Model Performance

     Accuracy: 71.8%

     Precision: 42%

  3. ## Deployment & Scalability

     The model can be integrated into clinical systems for initial screening. Accuracy is 71.8% but resampling or class weighting reduced performance due to data issues. Improved data quality is required before large-scale deployment.

  4. ## Business Impact & Practical Use

     The model provides an initial tool for early hypertension screening. It can assist doctors in decision-making but requires better data for reliable use. With improved data quality, it can reduce workload and support early detection.