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	Deployment
Omar Abdelaziz Amer Kassem	Data Visualization
Habiba Abdelmalek El-Sayed	Data Preprocessing & Features Engineering

- ## Objectives

1. To build a predictive model capable of classifying hypertension risk (High or Low).
2. To analyze the impact of key health and lifestyle features on hypertension.
3. To compare multiple classification algorithms for performance and reliability.
4. To deploy the final model as an API for real-time hypertension risk prediction.
5. To ensure scalability, accuracy, and clinical relevance of the system.

- ## Tools & Technologies

Data Processing & Analysis: Python (Pandas, NumPy).
ML & DL: Scikit-learn, Tensorflow
Visualization Tools: Seaborn, Matplotlib, Sciby
Deployment: Streamlit
Version Control: Git/GitHub

- ## 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.

- ## Milestones & Deadlines

Milestone	Description	Deadline
Data Collection & Understanding	Collecting data, understanding its structure, and performing initial exploration	5/10 – 20/10
Preprocessing	Cleaning data, handling missing values, and transforming data into the proper format	5/10 – 20/10
Visualization	Creating charts and graphs to understand relationships between variables	20/10 – 1/11
Exploratory Data Analysis	Performing exploratory analysis to identify patterns, trends, and distributions	20/10 – 1/11

Predictive Model Development & Optimization	Building predictive models and tuning them for best performance	20/10 – 1/11
Modeling, MLOps & Monitoring	Implementing models in a production-like environment and setting up monitoring	1/11 – 20/11
Deployment	Deploying the final model and ensuring it works in the target environment	1/11 – 20/11
Final Documentation & Presentation	Preparing the final report and presenting results to stakeholders	20/11 – 25/11