Intelligent Healthcare Disease Prediction System

Overview

This project presents a machine learning solution to assist healthcare professionals in early diagnosis and risk assessment of a disease using basic patient information. It uses both classification and regression models to provide binary diagnosis and continuous risk scoring.

Problem Statement

The aim is to:

- **Classify** whether a patient has a specific disease.
- **Predict a risk score** indicating the likelihood of disease using patient data:
  - Age
  - Sex
  - Blood Pressure
  - Cholesterol

Objectives

- Develop accurate classification and regression models.
- Ensure model interpretability for clinical usage.
- Compare different algorithms to identify the most effective ones.

Dataset

- **Type**: Synthetic (generated using NumPy)
- **Records**: 500
- **Features**:
  - Age (numeric)
  - Sex (binary)

  - Blood Pressure (numeric)
  - Cholesterol (numeric)
- **Targets**:
  - `Disease` (0 or 1)
  - `Risk Score` (0.0 to 1.0)

Models Used
- **Classification**:
  - Logistic Regression
  - Random Forest Classifier
- **Regression**:
  - Linear Regression

Evaluation Metrics
- **Classification**:
  - Accuracy
  - Confusion Matrix
  - ROC-AUC
- **Regression**:
  - Mean Squared Error (MSE)

Tools and Technologies
- **Programming Language**: Python
- **IDEs**: Jupyter Notebook, Google Colab
- **Libraries**: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Joblib

Team Members

- **Deerendra M.** - EDA & Model Development

- **Jegan Prabhu** - Data Cleaning

- **Aravind D.** - Documentation

- **Dhanush M.** - Feature Engineering

How to Run

1. Clone the repository.

2. Install dependencies listed in `requirements.txt`.

3. Run the Jupyter notebook or Python scripts.

License

This project is intended for educational and research purposes only.