Name : Bhushan Bharat Gavali

Heart Disease Detection – Project Report

1. Introduction

The goal of this project is to develop a machine learning model that predicts whether a patient has heart disease, based on various clinical and lifestyle factors. The task is framed as a binary classification problem.

2. Dataset & Exploratory Data Analysis (EDA)

The dataset was imported and explored. Exploratory Data Analysis (EDA) was performed to understand patterns and correlations between features and the target variable. Checked data distributions, missing values, and class imbalance.

3. Models Implemented

Three machine learning algorithms were tested:
1. Logistic Regression – a linear classification algorithm.
2. K-Nearest Neighbors (KNN) – a distance-based classifier.
3. Random Forest Classifier – an ensemble of decision trees.

4. Model Training & Evaluation

The dataset was split into training and testing sets. Models were trained and evaluated using metrics such as accuracy, precision, recall, and AUC-ROC.
Initial results showed Random Forest achieving almost perfect performance (accuracy ~1.0). Hyperparameter tuning was applied to improve KNN and Logistic Regression.

5. Hyperparameter Tuning

Hyperparameter tuning was performed as follows:
- Logistic Regression: tuned regularization and solver parameters.
- KNN: optimized the number of neighbors and distance metrics.
- Random Forest: already performing very well, with adjustments to n_estimators and max_depth.

6. Key Results

Based on the notebook outputs:
- Logistic Regression: Accuracy ~97.6%, Precision ~0.811, Recall ~0.894, AUC-ROC ~0.899.

- KNN: Precision ~0.989, Recall ~0.893, AUC-ROC ~0.998.
- Random Forest: Precision improved from 0.976 to 0.989, Recall from 0.811 to 0.893.

7. Conclusion & Future Work

Random Forest and KNN performed extremely well, making them strong candidates for deployment.
Future improvements include:
- Testing with larger datasets.
- Deploying as a web API for integration into healthcare systems.
- Exploring explainable AI techniques to make predictions interpretable.