

Heart Disease Prediction – Summary Report

1. Objective

The objective of this project is to analyze heart disease data and develop machine learning models to predict the presence of heart disease in patients. The process includes exploratory data analysis, data preprocessing, model training, and evaluation.

2. Dataset Overview

The dataset contains patient health information such as age, gender, chest pain type, blood pressure, cholesterol levels, and other medical indicators. The target variable represents the presence or absence of heart disease.

3. Exploratory Data Analysis (EDA)

- 1 Analyzed distributions of numerical and categorical health features.
- 2 Identified strong relationships between chest pain type, age, and heart disease.
- 3 Observed higher risk patterns in patients with abnormal clinical measurements.
- 4 Detected correlations among key medical indicators.

4. Data Preprocessing

- 1 Handled missing values and verified data consistency.
- 2 Encoded categorical variables into numerical form.
- 3 Scaled numerical features to improve model performance.
- 4 Split data into training and testing sets.

5. Model Development

- 1 Logistic Regression used as a baseline classification model.
- 2 K-Nearest Neighbors evaluated for pattern-based prediction.
- 3 Decision Tree captured non-linear relationships in the data.
- 4 Random Forest provided improved accuracy and robustness.

6. Model Evaluation

Models were evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Comparative evaluation helped identify the most effective model for heart disease prediction.

7. Conclusion

The analysis shows that machine learning models can effectively assist in early heart disease detection. Accurate predictions can support healthcare professionals in decision-making and risk

assessment.