Heart Disease Analysis Project Report

Abstract

This project focuses on analyzing heart disease data using Machine Learning techniques to predict the likelihood of heart disease in patients. Various classification algorithms were implemented and evaluated to determine the best-performing model.

1. Introduction

Heart disease is one of the leading causes of death worldwide. Early detection and prediction can significantly reduce mortality rates. This project uses data science techniques to analyze medical attributes and predict heart disease risk.

2. Dataset Description

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure
- Cholesterol
- Fasting Blood Sugar
- Resting ECG
- Maximum Heart Rate
- Exercise Induced Angina
- ST Depression
- Number of Major Vessels
- Thalassemia
- Target Variable

3. Methodology

The project followed these steps: Data preprocessing, Exploratory Data Analysis (EDA), feature scaling, model training, and performance evaluation.

4. Machine Learning Models Used

- Logistic Regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors (KNN)

5. Model Performance

Model	Accuracy
Logistic Regression	82%
Decision Tree	78%
Random Forest	87%
KNN	80%

6. Conclusion

Random Forest achieved the highest accuracy of 87%, making it the best-performing model. The analysis shows that age, cholesterol level, and chest pain type are significant factors in predicting heart disease. Machine Learning can assist medical professionals in early diagnosis.