

? Heart Disease Prediction Using Machine Learning

1. Introduction

This project aims to build a machine learning model that can predict the likelihood of heart disease in patients based on various medical and demographic attributes. By using patient data such as age, sex, chest pain type, cholesterol levels, and more, we can train a model to help doctors make early and accurate diagnoses.

2. Dataset Description

The dataset contains several important features related to cardiovascular health, including:

- Age
- Sex
- Blood Pressure
- Cholesterol
- Chest Pain Type
- Fasting Blood Sugar
- Resting ECG
- Maximum Heart Rate
- ST Depression
- Target (0 = No heart disease, 1 = Heart disease)

3. Steps Involved in the Project

The following steps were taken to build this project:

1. Data Cleaning and Preprocessing
2. Exploratory Data Analysis (EDA)
3. Feature Engineering and Scaling
4. Model Training (Logistic Regression, Decision Tree, Random Forest)
5. Model Evaluation using Accuracy, Precision, Recall, F1 Score, and ROC-AUC
6. Saving the Model using Joblib
7. Optional Deployment using Streamlit

4. Model Accuracy and Selection

We trained and evaluated three machine learning models:

- Logistic Regression: Accuracy ~83%
- Decision Tree Classifier: Accuracy ~81%
- Random Forest Classifier: Accuracy ~86%

The Random Forest model performed the best with an accuracy of approximately 86%. It also gave the best precision, recall, and F1-score among the three models.

5. Model Insights and Findings

From feature importance analysis, we observed that chest pain type, maximum heart rate, and cholesterol were strong indicators of heart disease. The model was able to learn patterns in the data that helped distinguish between patients with and without heart disease.

6. Real-world Application

In the real world, this model can be integrated into healthcare systems to provide instant heart disease risk analysis. Doctors and healthcare providers can use it in hospitals or clinics to make early assessments based on patient records.

Additionally, it can be deployed in mobile health apps where patients enter their medical stats and instantly get predictions.

7. Conclusion

The project demonstrates the effective use of machine learning in the healthcare sector. Using a Random Forest model, we achieved 86% accuracy in predicting heart disease. This project proves that predictive analytics can assist doctors and patients in making faster, data-driven health decisions.