

Heart Disease Prediction Project

Project Overview

This project aims to analyze a heart disease dataset and build predictive models to determine whether a patient is likely to have heart disease. The report outlines data cleaning, exploratory data analysis, machine learning model results, and actionable insights.

Dataset Overview

The dataset includes key medical attributes such as:

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure
- Serum Cholesterol
- Fasting Blood Sugar
- Resting ECG
- Max Heart Rate Achieved
- Exercise-induced angina
- ST depression
- Major Vessels colored by fluoroscopy
- Target (Heart disease present or not)

Key Insights From EDA

- Higher age groups (50+) show increased risk.
- Male patients exhibit higher probability of heart disease.
- High cholesterol levels correlate with disease presence.
- Patients with exercise-induced angina show strong risk signals.
- Abnormal ECG patterns indicate higher risk.
- Lower maximum heart rate achieved often corresponds to disease.

Machine Learning Model Results

Multiple models were tested including:

1. Logistic Regression

- Moderate performance due to linear boundaries.
- Most useful for interpretability.

2. Decision Tree

- Captures non-linear patterns.
- Performs better on small datasets.

3. Random Forest

- Best overall performance.
- Handles variance and improves stability

Evaluation Metrics

- Accuracy Score
- Precision, Recall, F1 Score
- Confusion Matrix

Recommendations

- Collect more balanced data for improved generalization.
- Include lifestyle factors (smoking, physical activity, diet).
- Use advanced models such as XGBoost, CatBoost.
- Deploy the model into a healthcare dashboard.
- Automate data monitoring with Power BI or Streamlit.

Conclusion

The heart disease prediction model is a powerful tool that can assist healthcare professionals in identifying high-risk patients. By analyzing key medical indicators, the model can support early detection and preventive care strategies.