Heart Disease Predictor

Introduction

Heart disease remains one of the leading causes of death worldwide. Predicting cardiovascular conditions early can help in taking preventive measures and saving lives. This dataset provides patient health records with multiple attributes related to lifestyle, medical history, and physiological conditions, which can be analyzed to predict the presence of heart disease.

Context

Cardiovascular diseases are influenced by factors like age, gender, lifestyle habits, cholesterol levels, blood pressure, and more. Machine learning models can help doctors and healthcare professionals assess the risk of heart disease in patients.

This dataset can be used for:

- Predictive modeling of heart disease likelihood
- Statistical and exploratory data analysis (EDA)
- Building classification models for healthcare applications
- Understanding correlations between risk factors and heart disease outcomes

Content

The dataset (heart.csv) contains patient medical information with the following columns:

- age → Age of the patient
- sex → Gender (1 = male, 0 = female)
- cp → Chest pain type (categorical)
- trestbps → Resting blood pressure (mm Hg)
- chol → Serum cholesterol (mg/dl)
- fbs → Fasting blood sugar (>120 mg/dl, 1 = true; 0 = false)
- restecg → Resting electrocardiographic results
- thalach → Maximum heart rate achieved
- exang → Exercise-induced angina (1 = yes; 0 = no)
- oldpeak → ST depression induced by exercise relative to rest
- slope → Slope of the peak exercise ST segment
- ca → Number of major vessels (0–3) colored by fluoroscopy
- thal → Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)
- target → Presence of heart disease (1 = disease, 0 = no disease)

Acknowledgements

This dataset is a commonly used benchmark in medical data mining and machine learning research. It is derived from the Cleveland Heart Disease dataset, originally made available through the UCI Machine Learning Repository. Many research papers and Kaggle notebooks have used this dataset for classification problems related to heart disease.

Source Reference:
- UCI Machine Learning Repository: Heart Disease Dataset
(https://archive.ics.uci.edu/ml/datasets/heart+disease)