Case Study Title

Shree LR Tiwari College Of Engineering

Gururaj Krishna Sharma
Pooja Bhuvnesh Sharma
Ronit Sharma
Satyam Pandey
Suraj Sah

Problem Statement

Heart Disease Prediction – Project Overview

Heart disease is one of the leading causes of death worldwide, and early detection plays a crucial role in reducing mortality rates. This project focuses on predicting the presence of heart disease using **Machine Learning techniques** based on patient health data.
The system uses a structured dataset containing medical attributes such as age, gender, blood pressure, cholesterol level, heart rate, and other clinical parameters. These features are analyzed to identify patterns associated with heart disease.
Various **machine learning algorithms** are trained and evaluated to build an accurate prediction model. The model learns from historical patient data and classifies whether a person is likely to have heart disease or not.

The objective of this project is to:
1. Assist healthcare professionals in early diagnosis
2. Reduce manual errors in medical analysis
3. Provide a fast and cost-effective decision-support system

Overall, this project demonstrates how machine learning can be effectively applied in the healthcare domain to improve predictive accuracy and support preventive care.

Dataset Overview(Optional)

Dataset Description
The dataset used for this project is a **Heart Disease Dataset** sourced from publicly available medical repositories (commonly derived from the **UCI Machine Learning Repository**). It contains real-world patient health records used for analyzing and predicting the presence of heart disease.
Dataset Size:
The dataset consists of multiple patient records (typically around **300+ instances**) with each record representing one individual.
Type of Data:
Structured, tabular medical data containing numerical and categorical attributes.

Key Features:
Age
Sex
Chest pain type
Resting blood pressure
Serum cholesterol level
Fasting blood sugar
Resting electrocardiographic results
Maximum heart rate achieved
Exercise-induced angina
ST depression (oldpeak)
Number of major vessels
Thalassemia

Target Variable:
Presence or absence of heart disease (binary classification: **0 – No Disease, 1 – Disease**)
This dataset is widely used for machine learning and data mining research in healthcare due to its relevance, reliability, and balanced feature set, making it suitable for building and evaluating predictive models.

Source : heart_disease_full_range_dataset.xlsx

Methodology

- **Approach**

Data Collection
A publicly available heart disease dataset was collected, containing medical and demographic information of patients.

Data Preprocessing
Handled missing or inconsistent values ,Converted categorical variables into numerical form
Normalized/scaled numerical features where required

Exploratory Data Analysis (EDA)
Analyzed relationships between features and heart disease , Visualized distributions and correlations to understand key risk factors
Feature Selection
Important attributes influencing heart disease were selected to improve model performance and reduce noise.

Model Training
The dataset was split into **training and testing sets**. Multiple machine learning algorithms were trained using the training data.
Model Evaluation
Models were evaluated using performance metrics such as **accuracy, precision, recall, and confusion matrix**.
Prediction
The best-performing model was used to predict whether a patient is likely to have heart disease.

Conclusion

- **Summary**:

This project presented a machine learning–based approach for **heart disease prediction** using clinical and demographic patient data. The dataset was preprocessed, analyzed, and used to train multiple classification models. Various machine learning algorithms were evaluated to identify the most accurate and reliable model.

Future Work:

Use larger and more diverse datasets
Apply advanced models to improve accuracy
Deploy the system as a web or mobile application

References

1. **Heart Disease Dataset (UCI Machine Learning Repository)**
   🔗 *UCI Heart Disease Dataset* — https://archive.ics.uci.edu/dataset/45/heart

2. Machine Learning Research Article (PubMed)
🔗 *A novel ML approach for cardiovascular disease prediction (PubMed)* — https://pubmed.ncbi.nlm.nih.gov/38992943/

3. Heart Disease Prediction Paper (IJRASET)
🔗 *Heart Disease Prediction Using ML Algorithms* — https://www.ijraset.com/research-paper/heart-disease-prediction-using-ml-algorithms

4. Heart Disease Risk Prediction (MDPI)
🔗 *Heart Disease Risk Prediction Using ML Classifiers* — https://www.mdpi.com/2076-3417/11/18/8352

5. WHO Cardiovascular Diseases Fact Sheet (Global Health Info)
🔗 *World Health Organization – Cardiovascular diseases (CVDs)* — https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29

6. scikit-learn (Machine Learning Library Documentation)
🔗 *scikit-learn: Machine Learning in Python* — https://scikit-learn.org/

7. GitHub Implementations (Example Projects)
🔗 *UCI Heart Disease Dataset GitHub Repo* — https://github.com/nmiuddin/UCI-Heart-Disease-Dataset
🔗 *Heart Disease Detection ML Project (GitHub)* — https://github.com/NavjotDS/Heart-Disease-Detection

Thank You