

Heart Disease Prediction using Machine Learning

Final Project Presentation

Problem Statement

- Heart disease is a leading cause of death globally.
- Goal: Predict heart disease using clinical data for early intervention.
- Machine learning can support diagnosis and improve outcomes.

Data Description

- Source: UCI Heart Disease Dataset
- 303 records with 14 attributes.
- Key Features: Age, Sex, Chest Pain, Cholesterol, Max Heart Rate, etc.
- Target: 1 = Disease, 0 = No Disease.

Key Insights from EDA

- - Slight class imbalance, ~55% have heart disease.
- - Males show higher prevalence.
- - Lower max heart rate and higher ST depression correlate with disease.
- - Chest pain type is a strong predictor.

Data Cleaning & Preparation

- - Handled missing values with mean imputation.
- - Encoded categorical features.
- - Scaled numerical data using StandardScaler.
- - Split into 80% training and 20% testing.

Top Questions & Insights

- 1. What is the distribution of heart disease?
~55% positive cases.
- 2. How does gender relate to heart disease?
Males higher.
- 3. How does max heart rate relate to disease?
Lower rate = higher risk.

Model Experimentation

- Tested Models:
 - - Logistic Regression
 - - KNN
 - - Decision Tree
 - - Random Forest
 - - SVM
 - - XGBoost
- Evaluation: Accuracy, Precision, Recall, F1-

Model Comparison

- Baseline: Logistic Regression (~84%)
- Top Performing Model: XGBoost (~90%)
- XGBoost provided highest accuracy and robustness.

Final Model & Business Relevance

- - Final Model: XGBoost
- - Justification: High accuracy, handles complex data patterns.
- Business Relevance:
 - - Supports early diagnosis
 - - Can be integrated into hospital systems
 - - Prioritizes at-risk patients for screening