Heart Disease Prediction using Logistic Regression

Student Name: VINITH K M

Internship / Course: Machine Learning / Data Science

Project Title: Heart Disease Prediction using Logistic Regression

Tools Used: Python, Pandas, Scikit-learn, Matplotlib, Seaborn

Date: 22-01-2026

Introduction

Heart disease is one of the leading causes of death worldwide. Early detection of heart disease can help in providing timely treatment and reduce the risk of severe complications. With the availability of medical data, machine learning techniques can be used to predict the presence of heart disease efficiently.

This project focuses on building a **Heart Disease Prediction system** using the **Logistic Regression** algorithm. The model predicts whether a person has heart disease based on various medical attributes.

Dataset Overview

- **Dataset Name:** Heart Disease Dataset
- **Number of Records:** 1025
- **Number of Features:** 13
- **Target Variable:** target
    - 0 → No Heart Disease
    - 1 → Heart Disease

The dataset contains medical attributes such as age, cholesterol level, blood pressure, heart rate, and other clinical parameters.

Feature Description

Some important features used in this dataset include:

- **age:** Age of the patient
- **sex:** Gender (1 = male, 0 = female)
- **cp:** Chest pain type

- **trestbps:** Resting blood pressure
- **chol:** Serum cholesterol
- **thalach:** Maximum heart rate achieved
- **exang:** Exercise-induced angina
- **oldpeak:** ST depression induced by exercise
- **target:** Presence of heart disease

Methodology

The following steps were performed in this project:

1. Loaded the dataset using the Pandas library
2. Explored the dataset and checked for missing values
3. Separated features and target variable
4. Split the dataset into **training (80%)** and **testing (20%)** sets
5. Trained a **Logistic Regression** model
6. Made predictions on test data
7. Evaluated the model using performance metrics

Model Used

Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for binary classification problems. It predicts the probability of a target class using a logistic (sigmoid) function.

This model is suitable for this task as the target variable has two classes: presence or absence of heart disease.

Model Evaluation

The performance of the model was evaluated using the following metrics:

- **Accuracy:** 79.51%
- **Precision:** 75.63%
- **Recall:** 87.38%

Confusion Matrix

Actual \ Predicted	0	1
0	73	29
1	13	90

Interpretation:

- The model correctly predicts most heart disease cases.
- High recall indicates that the model is effective in identifying patients with heart disease.

Results and Discussion

The Logistic Regression model achieved good performance with an accuracy of nearly 80%. The recall score is high, which is important in medical applications where missing a positive case can be dangerous.

Some false positives exist, but overall, the model provides reliable predictions.

Conclusion

This project successfully demonstrates the use of machine learning for predicting heart disease. Logistic Regression proved to be an effective baseline model for this classification problem.

The model can be further improved by:

- Applying feature scaling
- Trying advanced models like Random Forest or XGBoost
- Performing hyperparameter tuning

Tools and Technologies Used

- Python
- Pandas
- Scikit-learn
- Matplotlib
- Seaborn
- Google Collab
- GitHub