

Department of Computer Science & Engineering

**Title: Artificial Intelligence System & Expert System Lab Course Code: CSE 404**

**Project Name: Heart Disease Prediction Using Logistic Regression**

**Submission Date: 27/04/2025**

**Submitted To: Submitted By:**

**Noor Mairukh Khan Arnob Name: Humaira Ahmed Proma**

**Lecturer, Reg no: 21201178**

**Department of CSE, UAP Sec: B2**

**Problem Title:**

Predicting Heart Disease Using Logistic Regression

**Problem Description:**

Heart disease is one of the leading causes of mortality worldwide. Early detection and prediction of heart disease can significantly improve patient outcomes by enabling timely medical intervention. This project aims to build a machine learning model to predict the presence of heart disease based on various clinical and lifestyle features. The dataset used is the processed Cleveland dataset from the UCI Machine Learning Repository, containing 14 attributes related to heart health.

The objective is to build a classification model that predicts whether a patient has heart   
 disease based on clinical features. Logistic regression is used because:

* It models the probability of binary outcomes
* Provides interpretable coefficients
* Efficient for medical diagnostic tasks

**Key Metrics:**

**Accuracy:** Overall prediction correctness

**Precision:** Reliability of positive predictions

**Recall:** Ability to detect actual cases

**Tools & Language Used:**

* Programming Language: Python
* Tools: Google Colab Notebook
* Libraries: Scikit-learn, Pandas, Matplotlib

**Methodology:**

**Data Loading and Initial Exploration:**

1. The dataset was loaded from the UCI repository with predefined column names.
2. Initial exploration included checking the first few rows and dataset information to understand the structure and data types.

**Data Preprocessing:**

1. Rows with missing values were dropped to ensure data quality.
2. To simplify the classification task, the target variable was converted to binary (0 = no disease, 1 = disease).
3. To evaluate model performance, the dataset was split into training (80%) and testing (20%) sets.

**Feature Scaling:**

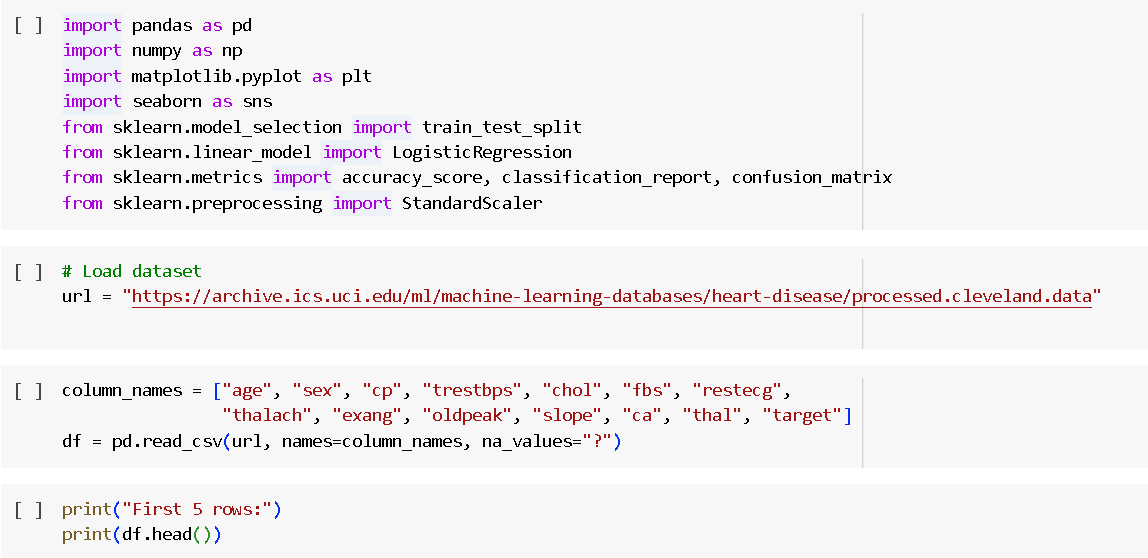
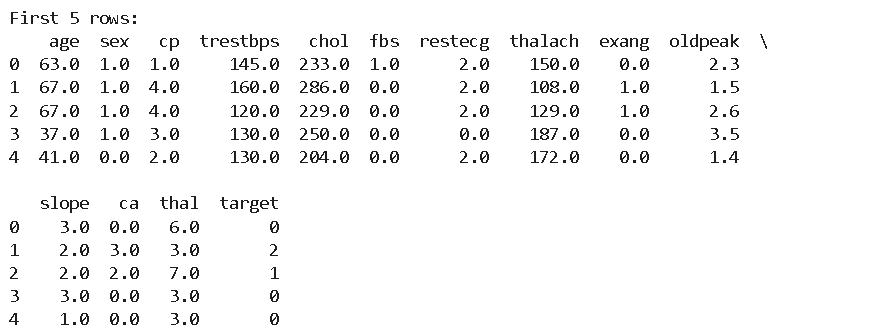
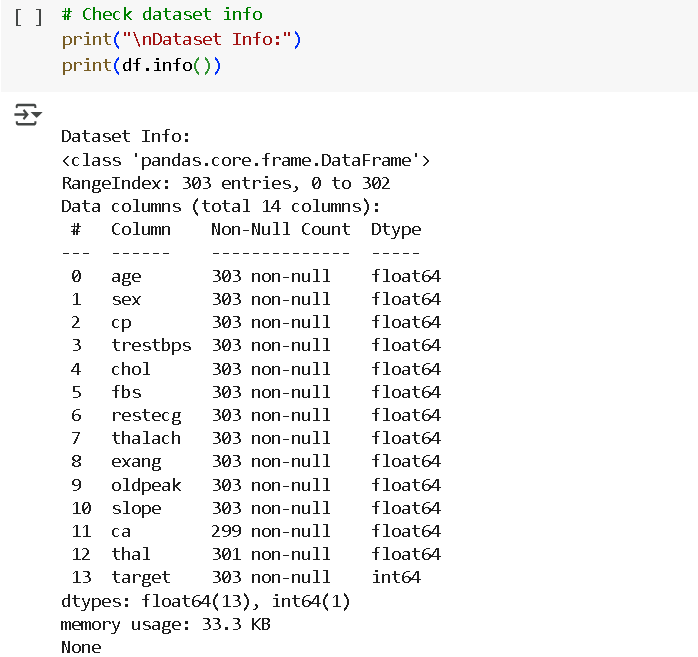
1. Features were standardized using StandardScaler to ensure all features contribute equally to the model.

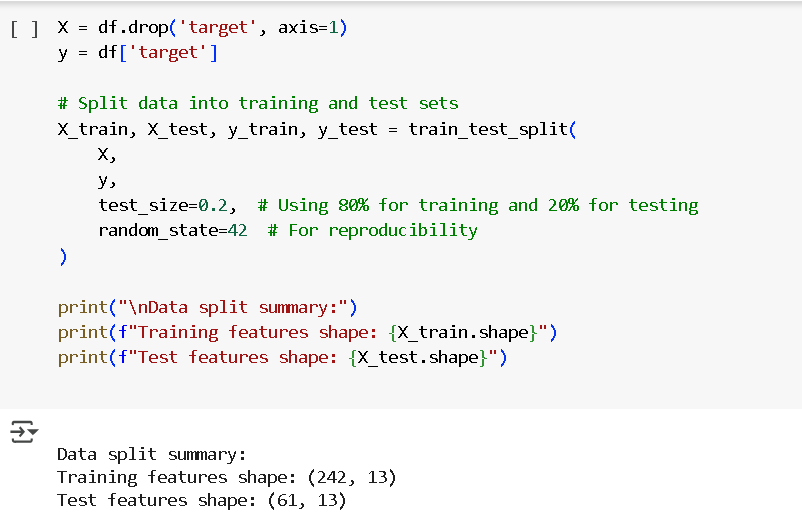
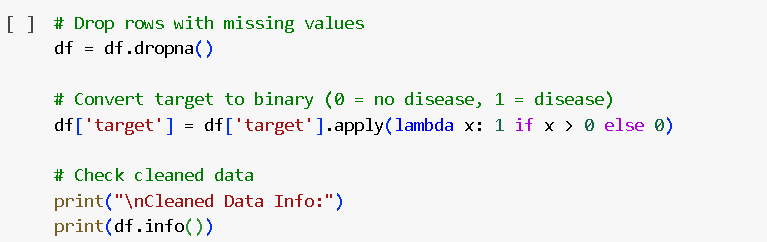
**Model Building:**

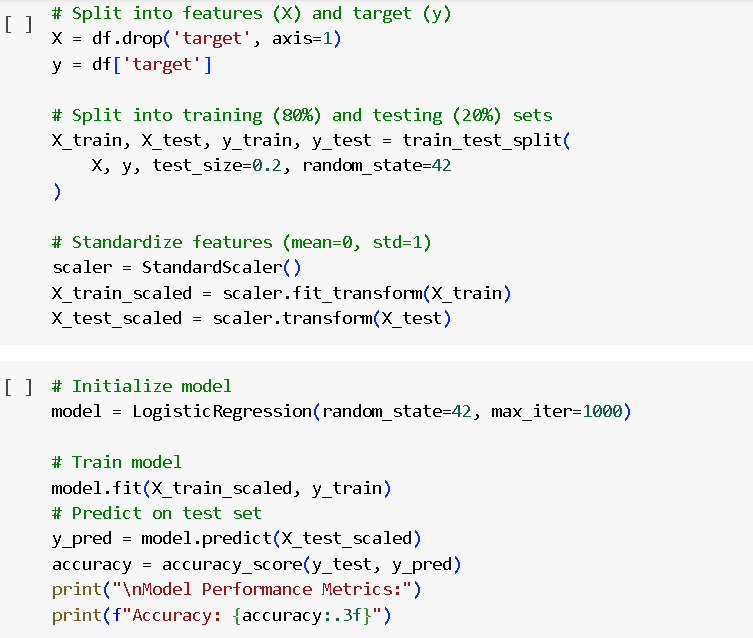
1. A Logistic Regression model was chosen due to its interpretability and effectiveness for binary classification tasks.
2. The model was trained on the scaled training data and evaluated on the test set.

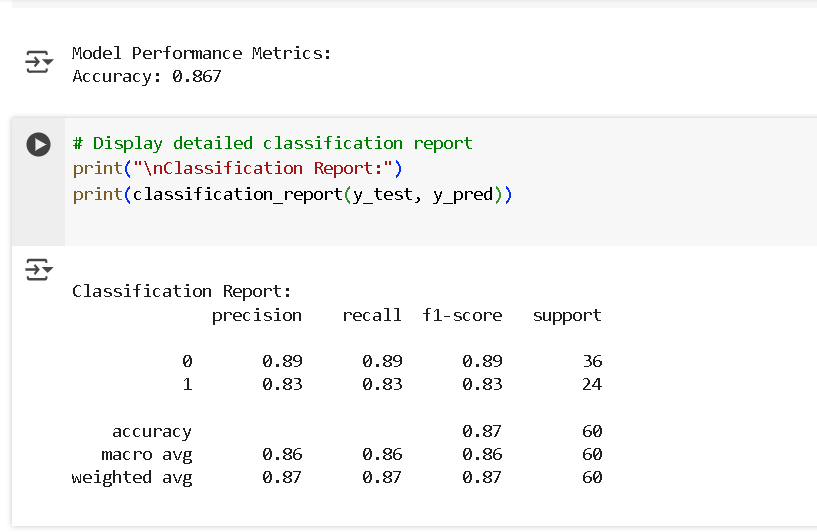
**Model Evaluation:**

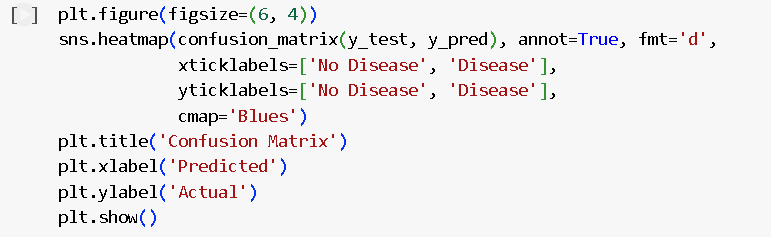
1. Accuracy: The model achieved an accuracy of 86.7% on the test set.
2. Classification Report: Provided precision, recall, and F1-score for both classes (0 and 1).
3. Confusion Matrix: Visualized the model's performance in terms of true positives, true negatives, false positives, and false negatives.
4. Feature Importance: Analyzed the coefficients of the Logistic Regression model to understand which features most influenced the predictions.

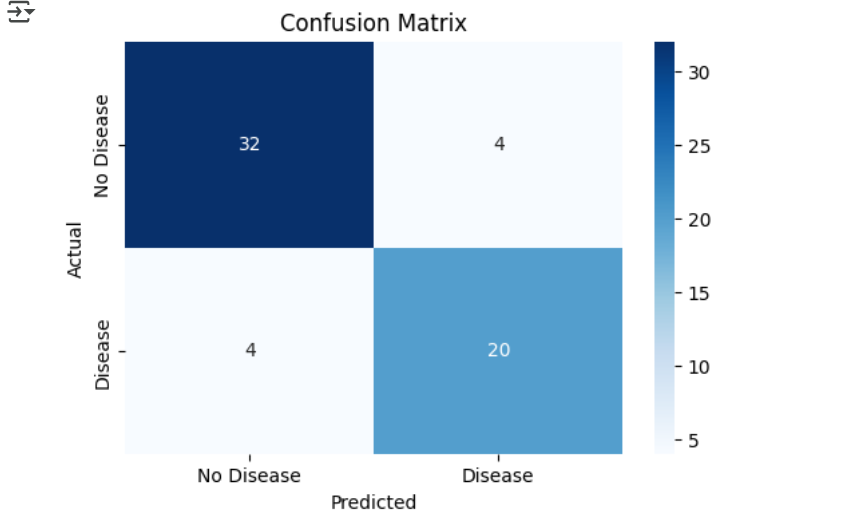
**Source Code & Output:**

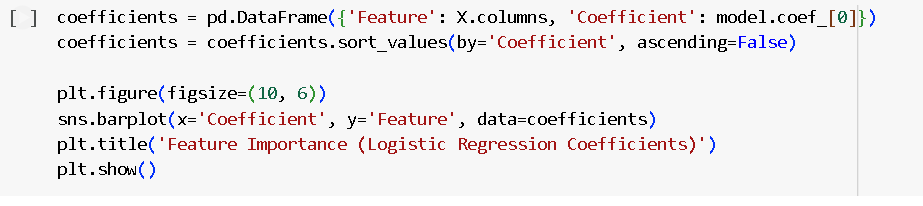
****

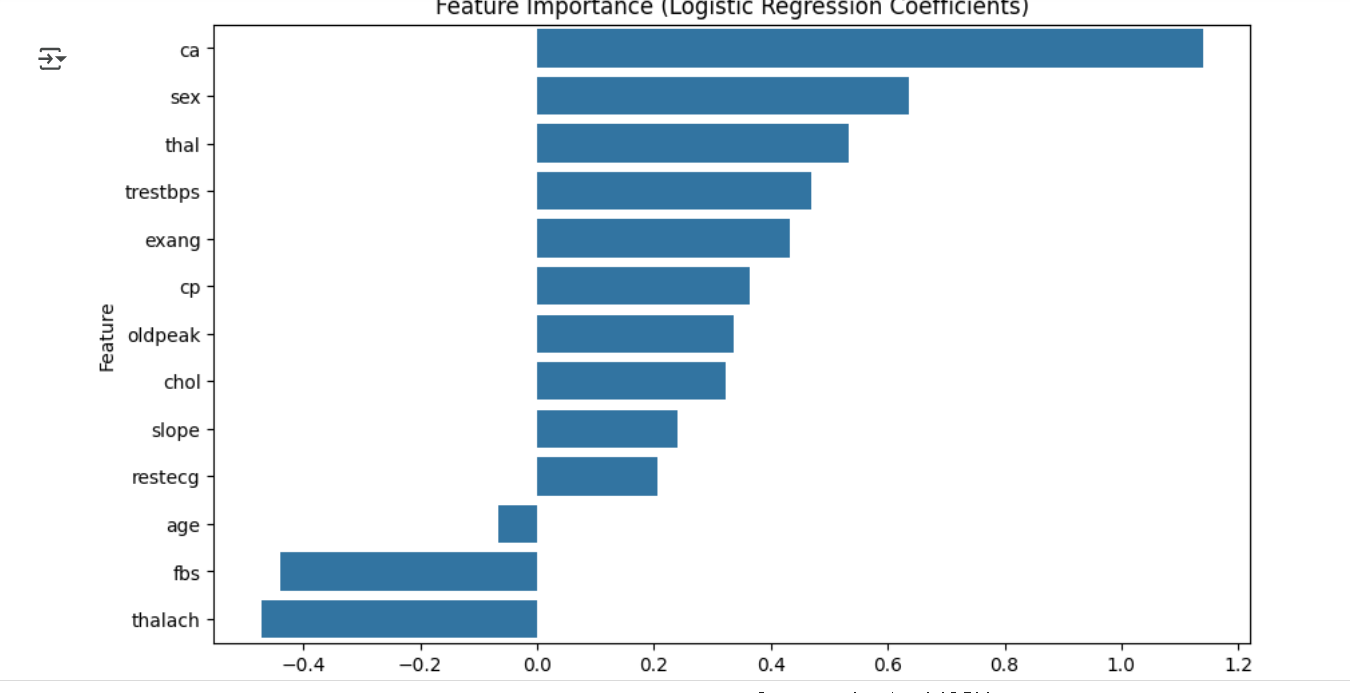












**Conclusion:**

The Logistic Regression model demonstrated strong predictive performance for heart disease detection. The analysis highlighted the importance of features like maximum heart rate and chest pain type in predicting heart disease. Future work could explore more complex models, additional feature engineering, and cross-validation to further improve performance. This model can serve as a valuable tool for early heart disease detection, aiding healthcare professionals in making informed decisions.