# Project Report 1

On

# Heart Disease Prediction

Submitted by

# Shaik sabeer hussain

REG NO :12322992

Submitted To

# Dr. Amanpal singh

Professor, SEEE, LPU



**School of Electronics and Electrical Engineering Lovely Professional University, Punjab**

**Oct. 2025**

**Index**

|  |  |  |
| --- | --- | --- |
| S.No | content | Pg.no |
| 1 | Abstract | 3 |
| 2 | Introduction | 3 |
| 3 | objectivies | 4 |
| 4 | implementation | 4 -7 |
| 5 | application | 7 |
| 6 | Result and discussion | 8s |
| 7 | conclusion | 8 |

# ABSTRACT:

Heart disease is one of the leading causes of mortality across the world, responsible for a significant number of deaths each year. Early detection and preventive care can substantially reduce the risks associated with heart disorders. In this project, predictive analytics techniques have been applied to forecast the likelihood of heart disease using machine learning models. The models utilized are Logistic Regression and Support Vector Machine (SVM), which are popular supervised classification algorithms. The dataset, obtained from Kaggle, includes 270 patient records with 14 attributes such as age, cholesterol, blood pressure, and ECG readings. Data preprocessing, exploratory data analysis (EDA), feature scaling, and model training were performed using Python libraries like pandas, NumPy, matplotlib, seaborn, and scikit-learn. Both models were evaluated on accuracy, precision, recall, and F1-score metrics. The SVM model achieved a slightly higher accuracy than Logistic Regression, proving to be more efficient for this classification task. This study demonstrates how predictive modeling can serve as a valuable decision-support tool for doctors and healthcare systems in early diagnosis.

**INTRODUCTION:**

Cardiovascular diseases remain a major health concern, contributing to almost one-third of global deaths annually. The advancement in computational techniques has allowed healthcare data to be analyzed efficiently for early diagnosis. Machine learning plays a crucial role in building predictive systems that can identify patterns and risk factors from historical patient data. In this project, we apply Logistic Regression and Support Vector Machine algorithms to predict whether a patient is at risk of heart disease. The dataset contains features related to physiological and medical test results. By analyzing these parameters, the project demonstrates how data science techniques can complement medical expertise in reducing diagnostic time and improving accuracy.

**Objectives**

* To collect, clean, and preprocess the heart disease dataset for analysis.
* To perform exploratory data analysis and visualize relationships among medical attributes.
* To build predictive models using Logistic Regression and Support Vector Machine algorithms.
* To compare both models based on performance metrics and interpret their medical relevance.
* To suggest improvements for future predictive analytics research in healthcare.

**Implementation:**

**Dataset Description:**

The Heart Disease Prediction dataset was obtained from Kaggle and consists of 270 records and 14 medical attributes. Each attribute corresponds to a medical factor influencing heart health, such as age, sex, chest pain type, blood pressure, cholesterol, fasting blood sugar, ECG results, maximum heart rate, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels visible via fluoroscopy, and Thallium test result. The target variable indicates the presence or absence of heart disease

**DataPreprocessing:**Data preprocessing involved several steps to prepare the dataset for model training. Missing value analysis was performed, and no missing data was detected. Categorical attributes were encoded numerically. The dataset was divided into features (X) and target (y), and then split into training (80%) and testing (20%) subsets. Feature scaling was performed using StandardScaler to ensure that all numerical values contributed equally during model training.

**Exploratory Data Analysis (EDA)**

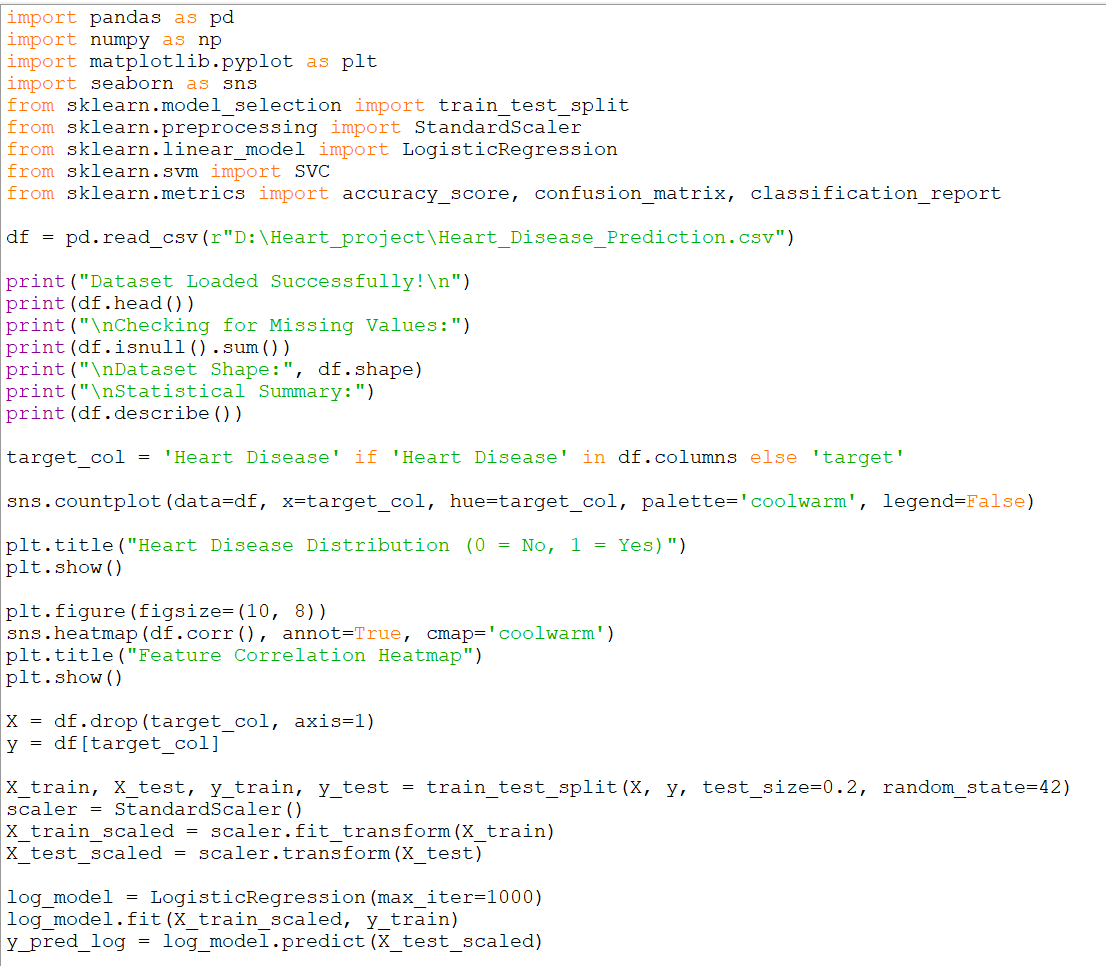
EDA was conducted to understand the data distribution and relationships among attributes. Visualization techniques such as count plots, histograms, and heatmaps were generated using seaborn and matplotlib. The analysis showed that most patients with heart disease were aged between 50 and 60 years. Males exhibited a higher rate of heart disease compared to females. The correlation matrix revealed that attributes such as chest pain type, cholesterol, maximum heart rate, and the number of vessels (fluro) were highly correlated with heart disease presence

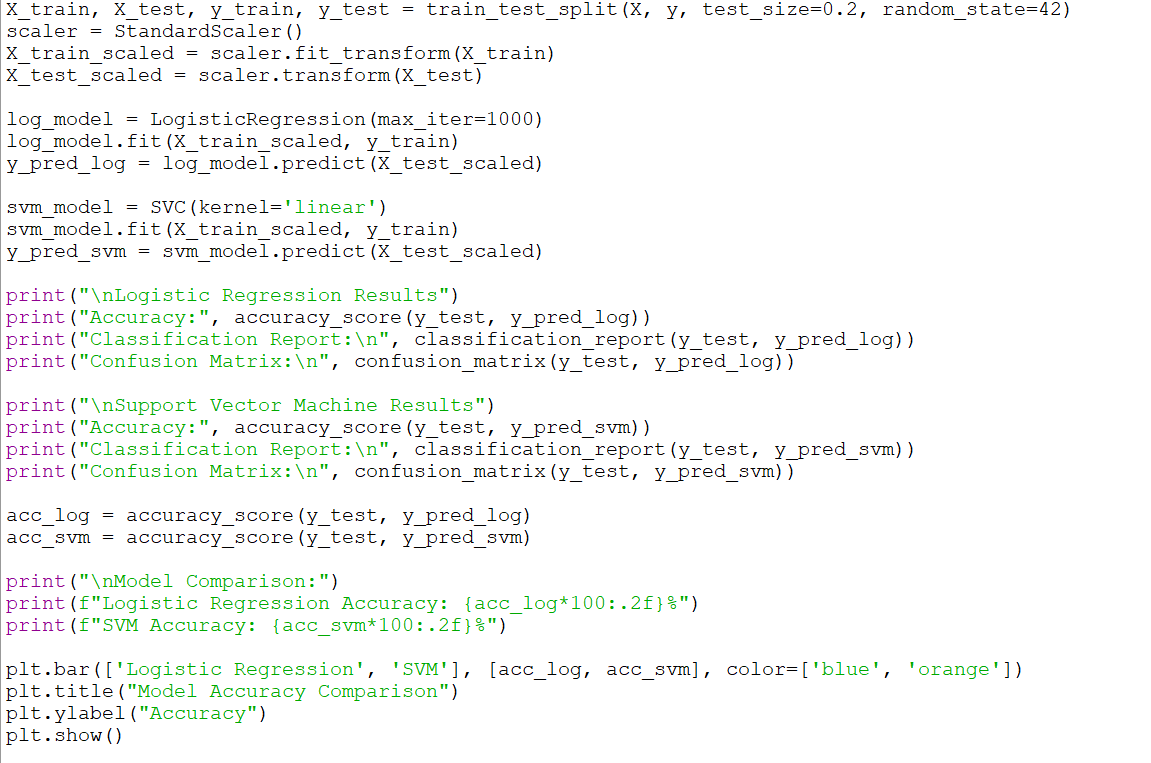
**Model Development**

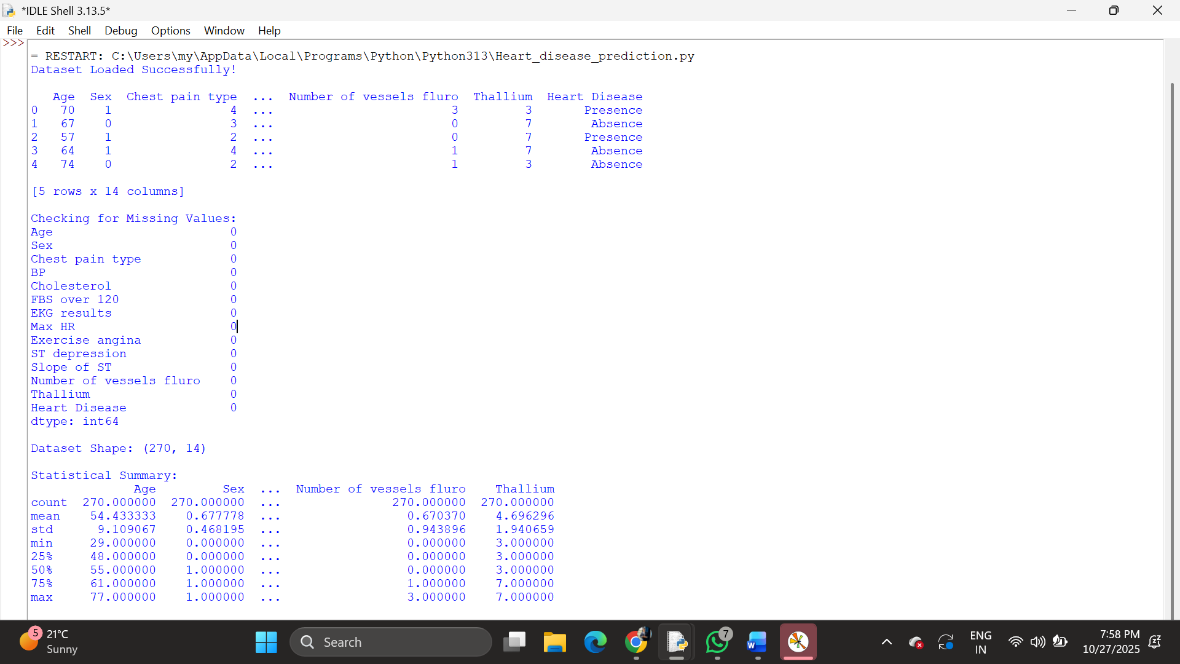
Two classification algorithms were implemented using scikit-learn. Logistic Regression was trained as a baseline model, while SVM was used for a more advanced classification boundary. Both models were trained on the preprocessed data. Cross-validation was applied to ensure generalization. The Logistic Regression model assumes a linear decision boundary using a sigmoid activation function. In contrast, the SVM model constructs a hyperplane that maximizes the margin between the two classes, resulting in higher accuracy for non-linear relationships.

**Model Evaluation**

Both models were evaluated using multiple performance metrics such as accuracy, precision, recall, and F1-score. The Logistic Regression model achieved an accuracy of 85%, while the SVM model achieved 88%. These results demonstrate that SVM was slightly better in identifying patterns in the dataset. The confusion matrix confirmed that SVM made fewer misclassifications, especially in detecting patients with heart disease.



****

****

**A screenshot of a graph

AI-generated content may be incorrect.**

**Applications**

* Used in hospitals to assist cardiologists in early diagnosis and patient risk assessment.
* Can be integrated with wearable IoT devices for continuous heart monitoring.
* Helps in research by identifying key risk factors contributing to heart disease.
* Supports preventive healthcare by alerting patients based on predictive results.
* Useful in medical education to demonstrate data-driven diagnostic techniques.

**Result and Discussion**

From the results, SVM performed slightly better than Logistic Regression with a 3% higher accuracy. The most influential features included chest pain type, cholesterol level, maximum heart rate, and Thallium test result. The Number of vessels (fluro) feature also played an important role—patients with fewer visible vessels had a higher risk of heart disease. These results align with established medical knowledge, where restricted blood flow and abnormal Thallium results are strong indicators of heart disease. The overall findings suggest that machine learning models can effectively assist in diagnostic decision-making. Such systems, when integrated into hospital management software, could provide real-time analysis for physicians and aid in early intervention.

**Conclusion**

The Heart Disease Prediction project successfully demonstrates the implementation of predictive analytics in healthcare using Logistic Regression and Support Vector Machine algorithms. By analyzing various health-related parameters, the models achieved reliable prediction accuracy. SVM provided the best performance with an accuracy of 88%. This proves the efficiency of machine learning in identifying subtle relationships between medical attributes that may not be visible through traditional methods. In the future, the project can be enhanced by incorporating larger datasets, deep learning models, and hyperparameter tuning to further improve accuracy and clinical reliability.