HEART FAILURE PREDICTION

USING MACHINE LEARNING ALGORITHMS

A Course Project report submitted in partial fulfilment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

By

K. GREESHMA 2203A51052

D. HASINI 2203A51044

V. VARSHA 2203A51132

CH. SRIPUJA 2203A51108

under the guidance of

DR. MAMATA PANDEY

Associate professor, Department of CSE.



CERTIFICATE

This is to certify that the Project entitled “HEART FAILURE PREDICTION USING USING MACHINE LEARNING ALGORITHMS” is the bona-fide work carried out by K. Greeshma, D. Hasini, V. Varsha, Ch. Sripuja as a Course Project for the partial fulfillment to award the degree BACHELOR OF TECHNOLOGY in ARTIFICAL INTELLIGENCE AND MACHINE LEARNING during the academic year 2023-2024 under my guidance and Supervision.

Dr. MAMATA PANDEY Dr. M. SHESHIKALA

Ass Prof. (CSE) Ass Prof.&associate

S R University, S R University,

Ananthasagar, Warangal. Ananthasagar, Warangal.

**ABSTRACT**

Objective: Heart failure (HF) is a critical and escalating public health issue with significant morbidity and mortality rates worldwide. Early prediction of heart failure can lead to timely intervention, management, and potentially improved clinical outcomes for at-risk patients. This study aims to develop and validate a machine learning model for the early prediction of heart failure based on a comprehensive set of patient parameters.

Methods: We utilized a dataset comprising N patients (where N represents the number of patients in your dataset) with X features (e.g., age, blood pressure, biomarkers, medical history). Data was split into training (70%) and test (30%) subsets. Various machine learning algorithms, including Random Forest, Gradient Boosting Machine, and Deep Neural Networks, were trained and evaluated based on their predictive performance. The primary outcome was the occurrence of heart failure within a specified follow-up period.

Conclusion: Machine learning models, particularly Gradient Boosting Machines, show promising results in the early prediction of heart failure. Such models can augment traditional diagnostic methods, potentially enabling healthcare providers to identify at-risk patients and initiate early interventions. Future work should focus on external validation of these models in diverse populations and integration into clinical workflows.

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to all those who contributed to the successful completion of the project on Heart failure prediction using machine learning algorithms. I am immensely thankful to my supervisor for their unwavering guidance and support throughout this endeavor. I would also like to extend my appreciation to my colleagues and friends for their invaluable input and encouragement. Furthermore, I am indebted to the academic community and researchers whose prior work laid the foundation for this project. Last but not least, I acknowledge the funding and resources provided by our institution, without which this research would not have been possible.

TABLE OF CONTENT

|  |  |  |
| --- | --- | --- |
| Chapter No. | Title | Page No |
| 1 | Introduction |  |
|  | 1.1. Problem Statement |  |
|  | 1.2. Existing system |  |
|  | 1.3. Proposed system |  |
|  | 1.4. Objectives |  |
|  | 1.5. Architecture |  |
| 2 | Literature survey |  |
|  | 2.1. Document the survey done by you |  |
| 3 | Data pre-processing |  |
|  | 3.1. Flow chart |  |
|  | Figure 3.1. Flow chart |  |
|  | 3.2. Dataset description |  |
|  | Figure 3.2.1 dataset |  |
|  | 3.3. Data cleaning |  |
|  | 3.4. Data Visualization |  |
|  | Figure 3.4.1-3.4.9 Graphs |  |
| 4 | Methodology |  |
|  | 4.1Multiple linear regression |  |
|  | Figure 4.1.1. Multi Linear Regression |  |
|  | Model architecture |  |
|  | Software description Figure 4.1.2 Graph |  |
|  | 4.2 Ridge Regression |  |
|  | 4.2.1 Graph |  |
|  | Model architecture |  |
|  | Software description |  |
|  | 4.3 Support Vector Regression |  |
|  | Figure 4.3.1 Graph |  |
|  | Model architecture |  |
|  | Software description |  |
|  | Table 4.3.2. Results |  |
| 5 | Conclusion and future scope |  |
| 6 | References |  |
|  |  |  |

CHAPTER 1

INTRODUCTION

Heart failure (HF) stands as one of the leading causes of morbidity, mortality, and health-related expenditures worldwide. As an escalating global health concern, heart failure results in substantial social and economic burdens, accounting for numerous hospital admissions and long-term care requirements. This progressive condition sees affected individuals experiencing reduced heart pump efficiency, leading to decreased blood flow to vital organs and various clinical manifestations such as fatigue.

The prognostic landscape of heart failure is heterogeneous. While some patients can live with the condition for years without experiencing significant deterioration, others may face rapid health declines. The variability in patient trajectories underscores the necessity for precise and early predictive models. Early prediction facilitates the identification of at-risk individuals, enabling timely interventions, personalized treatment, and, consequently, better outcomes.

Historically, the prediction of heart failure relied on clinical intuition and traditional risk factor assessments, such as the presence of coronary artery disease, hypertension, or diabetes. However, these methods have limitations in specificity and sensitivity. With the exponential growth of electronic health records (EHRs) and advancements in data science, the healthcare industry has shifted towards harnessing the power of machine learning (ML) and artificial intelligence (AI) to refine the predictive accuracy for a multitude of conditions, including heart failure.

1

Problem Statement

I will make a project for Heart failure prediction. The problem statement is heart failure, with its growing prevalence and consequential socioeconomic impact, is a pressing global health concern. Despite advances in clinical care and management, mortality rates associated with heart failure remain significant. A substantial challenge in managing this condition is the timely identification of patients at risk, which would enable proactive interventions and possibly alter disease progression trajectories. This model is user-friendly, it helps the user to get the rainfall rate at the specified location excluding noisy reading.

* 1. Existing System

Heart failure prediction of a certain persons has been studied extensively. Works related include regression model which helps with better precision than some simple multiple regression. Support Vector Machines (SVMs) are typically used for classification tasks, where the goal is to separate data into different classes or categories. They do this by finding a hyperplane that maximizes the margin between classes. SVMs can also be used for regression tasks, and this variant is known as Support Vector Regression (SVR).. To support advanced decision tree algorithm; basic indicators such as mean, variance and standard deviation are required.

* 1. Proposed System

Optimization of the data set, data cleaning, Random Forest regression are key metrics in this project. The classifiers are evaluated in terms of accuracy, precision, Decision Tree Regression, Logistic Regression and RFR. This produces expansion of e-commerce business and increased volume of products as well as improved website usability to navigate the desired product quickly and easily. RFR is a supervised learning algorithm that works well with classification and regression tasks. The algorithm divides the data into small parts and identify the patterns that can be used for making an accurate prediction. The learning strategy behind Random Forest Regression is to ensemble methods(bagging); by constructing a multitude of decision trees at training time and outputting mean of prediction of individual trees.

1.3 Objective:

The objective of predicting heart failure is to enhance clinical outcomes, patient quality of life, and healthcare efficiency through the early and accurate identification of individuals at risk of developing heart failure. Specific objectives can be detailed as follows:

Identify individuals at an elevated risk of heart failure at the earliest possible stage, even before pronounced clinical manifestations appear. This would allow for prompt interventions that could delay or prevent the onset of the disease. Harness the detailed insights provided by predictive models to tailor medical interventions, lifestyle recommendations, and follow-up strategies to individual patient profiles, thereby optimizing care pathway .Arm healthcare professionals with data-driven insights that complement their clinical expertise, facilitating more informed decision-making regarding diagnostics, treatments, and patient counselling . By pinpointing those at higher risk, healthcare systems can better allocate resources, such as specialized monitoring and therapeutic interventions, to the individuals who need them most. Allow for the stratification of patients based on their risk profile, enabling healthcare providers to determine the frequency and intensity of monitoring required. High-risk patients can be monitored more closely, while low-risk patients can avoid unnecessary medical interventions.

* 1. Architecture

There isn't a traditional neural network architecture like those used in deep learning. Instead, the machine learning architecture revolves around linear regression and support vector regression (SVR) models, which are not neural networks but are common supervised learning algorithms in machine learning. Let's break down the architecture. In this architecture, the primary machine learning models used for heart failure prediction are linear regression and SVR. These models learn the relationships between the meteorological features (input) and the target variable 'ANNUAL' (output) in order to make predictions. While there's no deep neural network architecture in this specific code, deep learning architectures are used for more complex tasks when dealing with large datasets and non-linear patterns. However, linear regression and SVR models are well-suited for tasks like rainfall prediction, where there may not be complex, non-linear relationships in the data.

Requirement Specification

Software Requirements

✓ OS : Windows 10 or Higher Versions

✓ Platform : Google collab

✓ Program Language : Python

CHAPTER 2

LITERATURE SURVEY

2.1 Document the survey done by you

Machine learning is helpful for a variety of situations. The prediction of dependent variable values from independent variables is one of the uses of this methodology. The power consumption bill prediction is a data mining field since it contains extensive data resources that are challenging to manually handle. We will have data and with the help of that data inputs that we have we predict the bill of power consumed. Multiple Linear Regression is a statistical technique that is used for predictive modeling and understanding the relationships between multiple independent variables and a single dependent variable.

* Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. [1]
* In this modern era people are very busy and working hard in order to satisfying their materialistic needs and not able to spend time for themselves which leads to physical stress and mental disorder[2]
* It’s important to know the symptoms of heart failure. [3]
* Heart failure, also known as congestive heart failure, is a condition that develops when your heart doesn’t pump enough blood for your body’s needs. [4]
* centre around various topics in the heart failure community[5]
* This study sought to review the literature for risk prediction models in patients with heart failure and to identify the most consistently reported independent predictors of risk across models.[6]
* study sought to review the literature for risk prediction models in patients with heart failure and to identify the most consistently reported independent predictors of risk across models.[7]
* Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand.[8]
* Cardiovascular diseases kill approximately 17 million people globally every year, and they mainly exhibit as myocardial infarctions and heart failures.[9]
* The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time.[10]

 leads to Kaggle, which is a popular online platform for data science and machine learning competitions, datasets, and collaborative data science projects. Machine learning method scan none the less be helpful in solving this challenge and in anticipating danger early. Some of the techniques used for such prediction problems are multiple linear regression algorithms. The existing research have less inputs. We can predict the bill by taking few more data and we used different methods to improve the accuracy in the prediction of bill but they are not much accurate than our proposed system with the comparison accuracy rates when comparative to existing system to the proposed system. In proposed system my model showed accuracy rates with the Linear regression with 80.73% accuracy.

CHAPTER 3

DATA PREPROCESSING

Data preprocessing in rice classification using machine learning is a critical step to ensure the accuracy and effectiveness of the model. Initially, raw rice grain images are collected and may exhibit variations in lighting, orientation, and resolution. These images undergo several preprocessing steps to standardize the data. First, resizing and normalization are applied to ensure all images have the same dimensions and pixel values. Additionally, techniques like contrast enhancement and noise reduction are employed to improve image quality and reduce unwanted variations. Colour spaces may be adjusted to enhance feature extraction, and histogram equalization can be applied for uniform illumination. To mitigate overfitting, data augmentation techniques such as rotation, flipping, and cropping can be used to increase the diversity of the training dataset. Furthermore, image segmentation may be utilized to isolate rice grains from background noise. Overall, data preprocessing in rice classification is indispensable for creating a robust and accurate machine learning model by preparing the input data for effective feature extraction and classification, ultimately contributing to more reliable and consistent rice grain categorization.

3.1 Flowchart



Figure 3.1 flowchart

5

This is an analysis of rainfall prediction problem using different Machine learning models. It predicts the rainfall in the given area with the configurations given to it.

3.2 Dataset Description

For this project, I have obtained my dataset from csv file. This dataset contains 18185 rows of data

and 12 columns (features) that we could focus onto build our model i.e., I have used

12 attributes to predict the bill of power consumed.

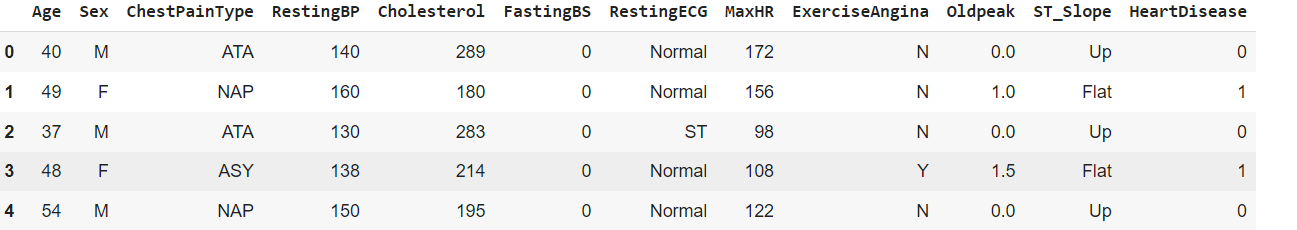


Figure 3.2.1 data set

3.3. DATA CLEANING:

**Data cleaning** is one of the important parts of machine learning. It plays a significant part in building a model. It surely isn’t the fanciest part of machine learning and at the same time, there aren’t any hidden tricks or secrets to uncover. However, the success or failure of a project relies on proper data cleaning. Professional data scientists usually invest a very large portion of their time in this step because of the belief that **“Better data beats fancier algorithms”**. If we have a well-cleaned dataset, there are chances that we can get achieve good results with simple algorithms also, which can prove very beneficial at times especially in terms of computation when the dataset size is large. Obviously, different types of data will require different types of cleaning. However, this systematic approach can always serve as a good starting point.

3.4 DATA VISUALIZING:

Data visualization algorithms create images from raw data and display hidden correlations so that humans can process the information more effectively. It is also an important evaluation metrics for learning, since the ultimate goal of artificial intelligence is to create a machine that can understand and respond to data even better than a human could. Data visualization is a crucial aspect of machine learning that enables analysts to understand and make sense of data patterns, relationships, and trends. Through data visualization, insights and patterns in data can be easily interpreted and communicated to a wider audience, making it a critical component of machine learning.we will discuss the significance of data visualization in machine learning, its various types, and how it is used in the field. General types of data visualization are charts, tables, graphs, maps, The below are the graphs of our data set which are plotted between each feature and target variables in the dataset.

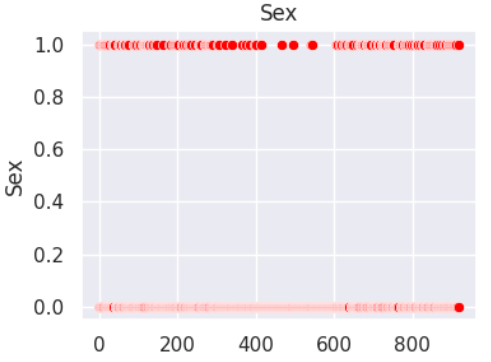
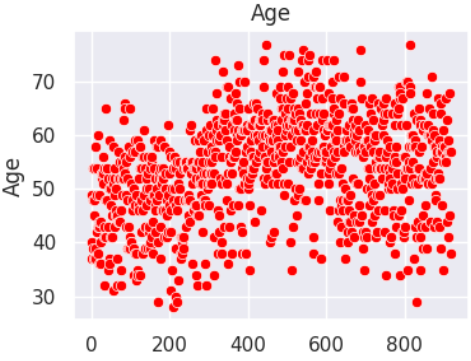


Figure 3.4.1. Figure 3.4.2.

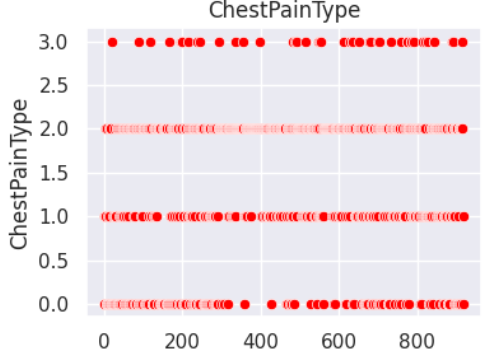
 

Figure 3.4.3. Figure 3.4.3.

CHAPTER 4

METHODOLOGY

Enough methods are performed on the data to evaluate the data set and gather knowledge about the data. Let's perform some Machine Learning model and Experimentation to create a model that helps us to achieve the goal I stated in the problem definition. For this study, the raw data were collected from the regional meteorological station at Bahir Dar City, Ethiopia. Ten data features such as year, month, date, evaporation, sunshine, maximum temperature, minimum temperature, humidity, wind speed, and rainfall were included. The meteorology station records the values of the environmental variable every day for each year directly from the devices in the station. Then, the data were recorded in the Microsoft Excel file tabular format. The year and the days of the month were arranged in the row of tables related to environmental variables in the column of the table. The raw data recorded at the station for 114years (1901–2015) were used for the study. In this we talk about the various machine learning algorithms used for the project. They are linear regression, Support Vector Machine(SVM).

4. LOGISTIC REGRESSION:

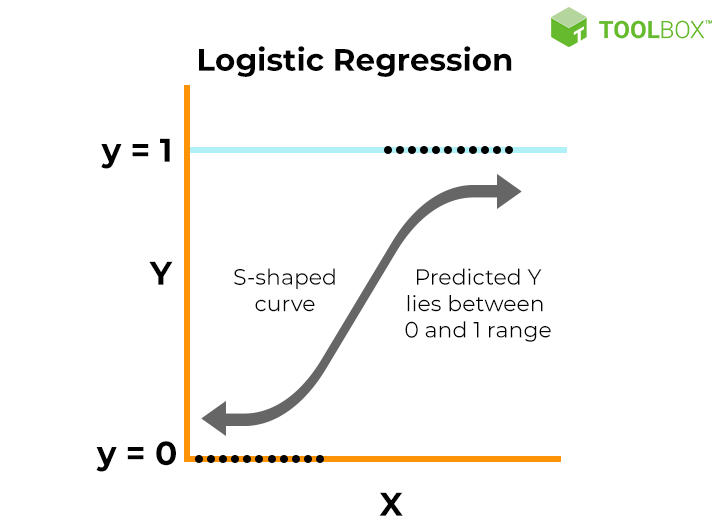
Logistic regression is a statistical method used for modeling ,the probability of a binary outcome, which means it's ideal for classification problems. It's a type of generalized linear model (GLM) that's widely used in fields such as machine learning, statistics, and social sciences for various applications, including medical diagnosis, spam detection, and credit scoring.

Sigmoid function:

P(Y=1∣X)=1+e−(b0+b1X1+b2X2+…+bnXn)1

Here, $P(Y=1|X)$ is the probability of the target variable being 1, and $X\_1, X\_2, lots, X n$ are the independent variables. $b\_0, b\_1, b\_2, \dots, b n$ are the coefficients that need to be estimated .Logistic regression is a simple yet powerful algorithm for solving binary classification problems. It's widely used due to its interpretability and ease of implementation. However, in cases where the relationship between independent variables and the target is not linear, other

**Figure 4.1.1:Logistic Regression**



Logistic regression can be expressed mathematically as:

P(Y=1∣X)=1+e−(b0+b1X)1

Here,

• $P(Y=1|X)$ represents the probability of the dependent variable Y being 1 given the value of the independent variable X.

• $b\_0$ is the intercept term, representing the value of the log-odds when X is 0.

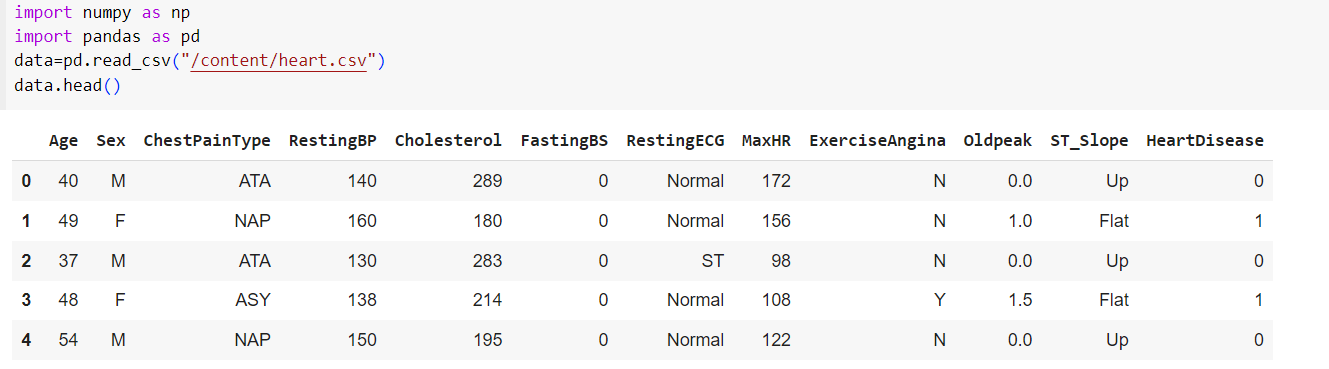
• $b\_1$ is the coefficient of the independent variable X, indicating how a one-unit change in X affects the log-odds of Y being 1.

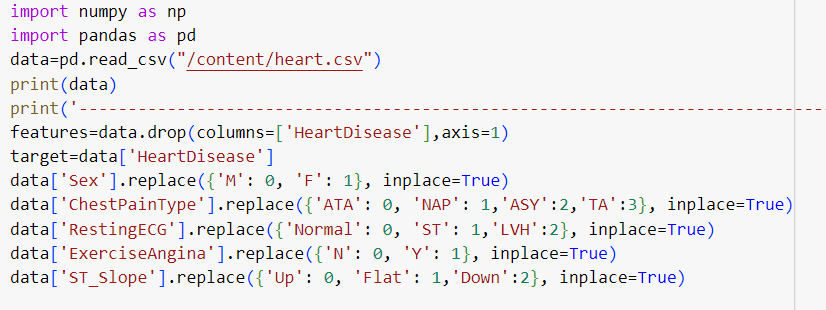
• $e$ is the base of the natural logarithm, approximately equal to 2.718

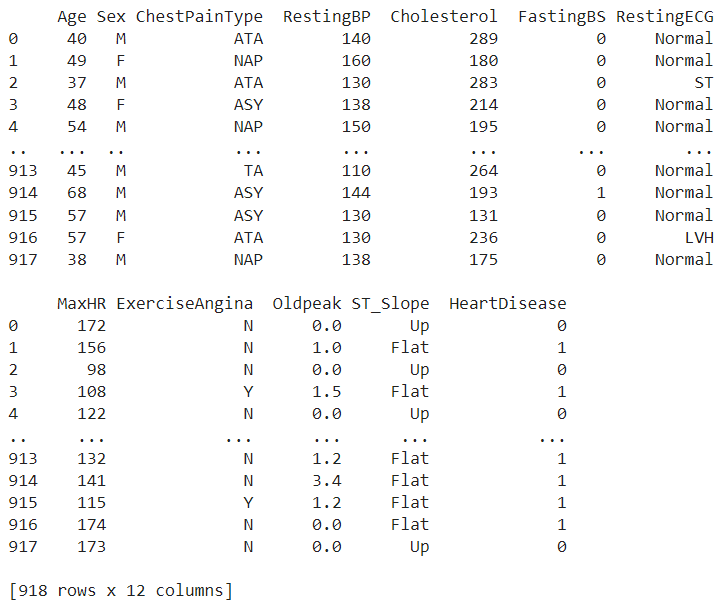
ADVANTAGES AND EXAMPLES OF **LOGISTIC** REGRESSION:

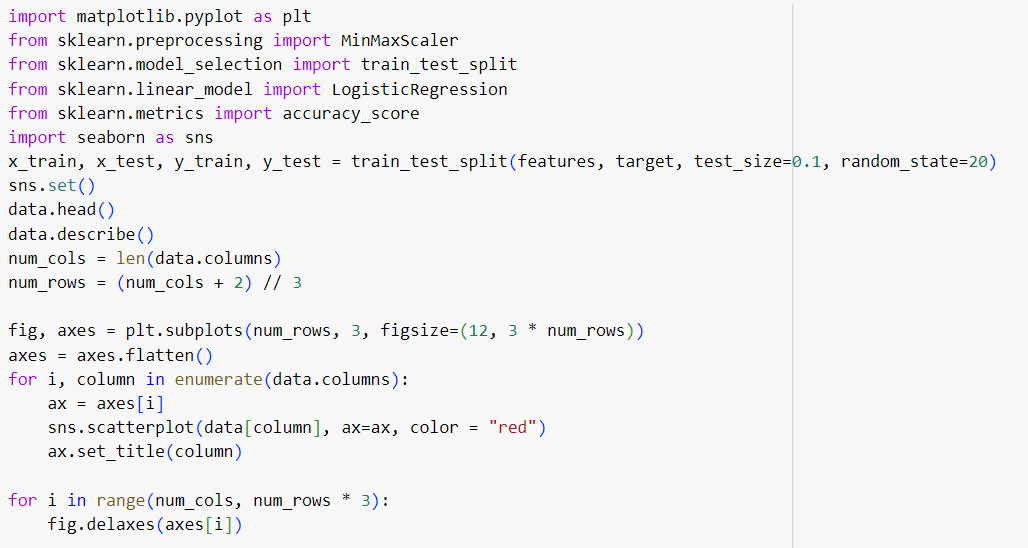
1. The main advantage of logistic regression is that it is much easier to set up and train than other machine learning and AI applications. Another advantage is that it is one of the most efficient algorithms when the different outcomes or distinctions represented by the data are linearly separable
2. Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable.

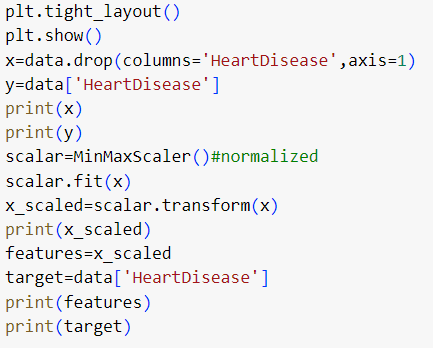
**CODE:**

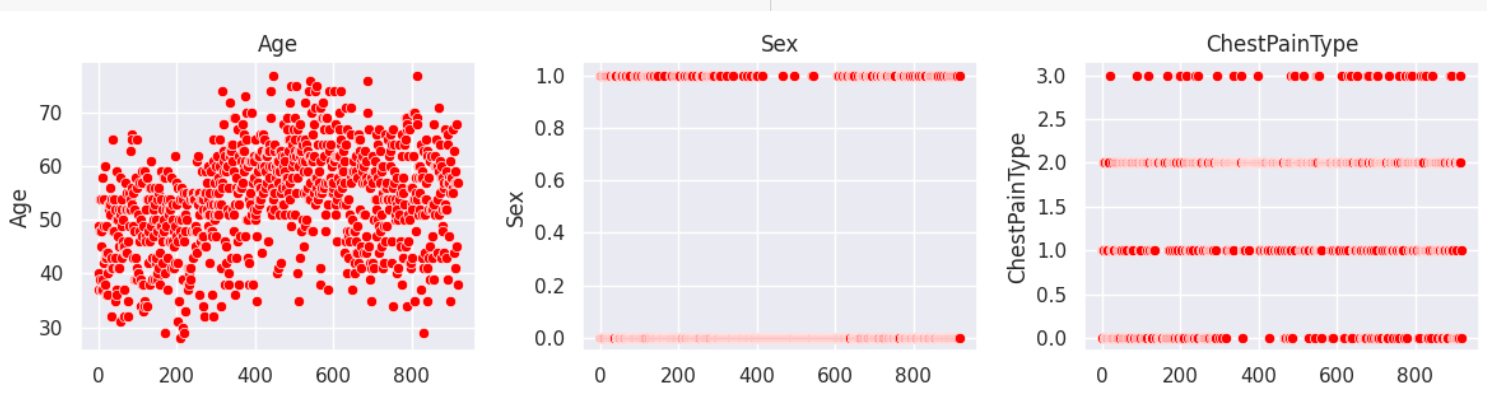


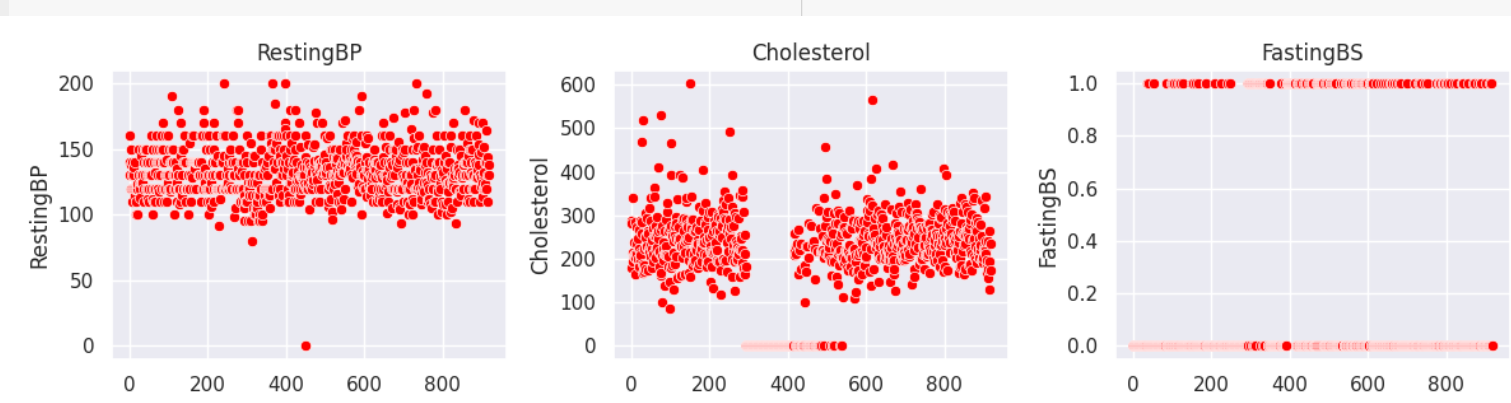


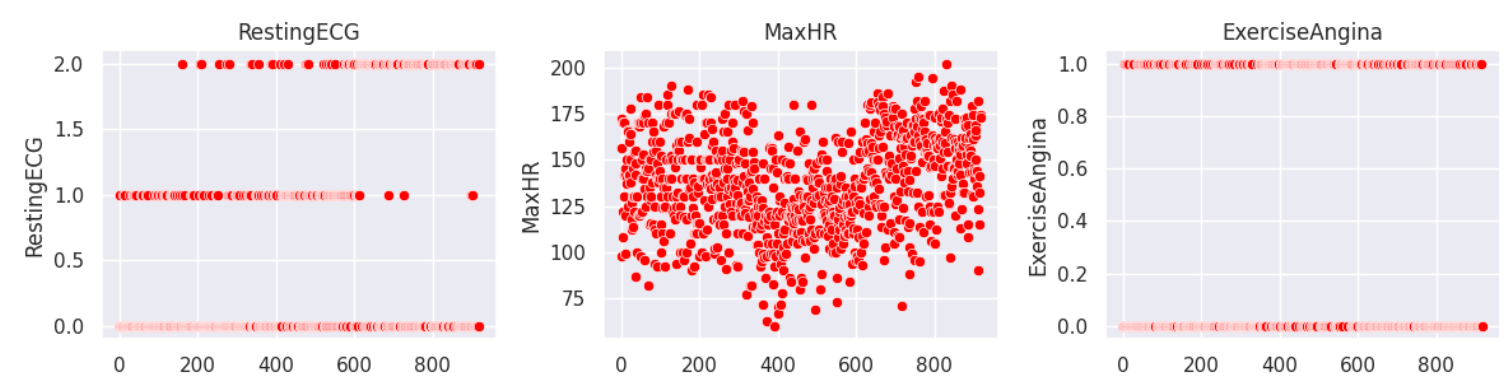


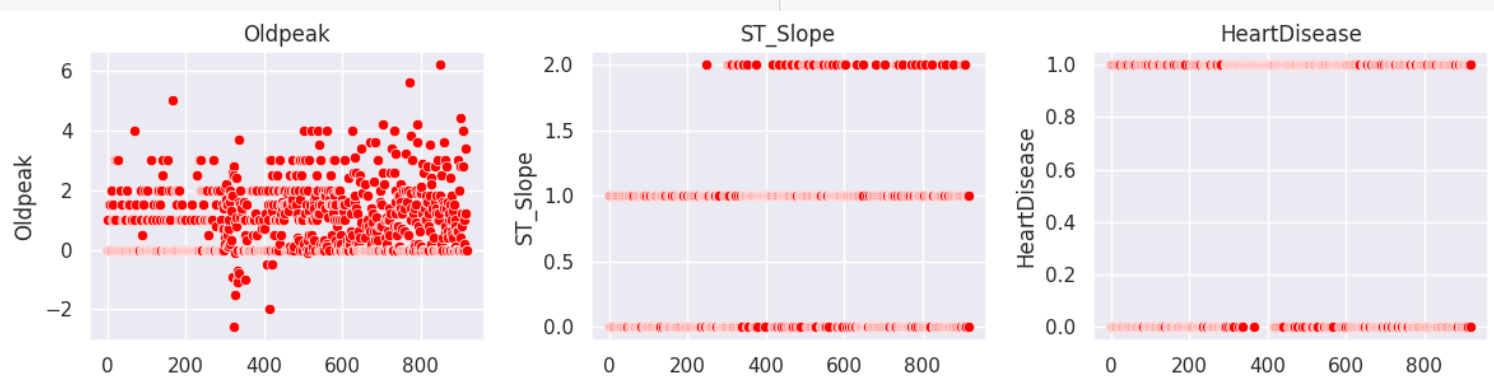


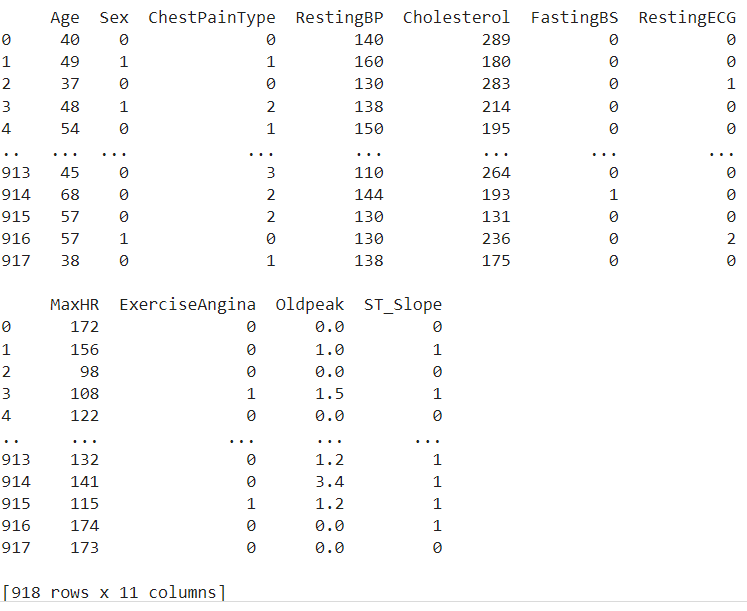


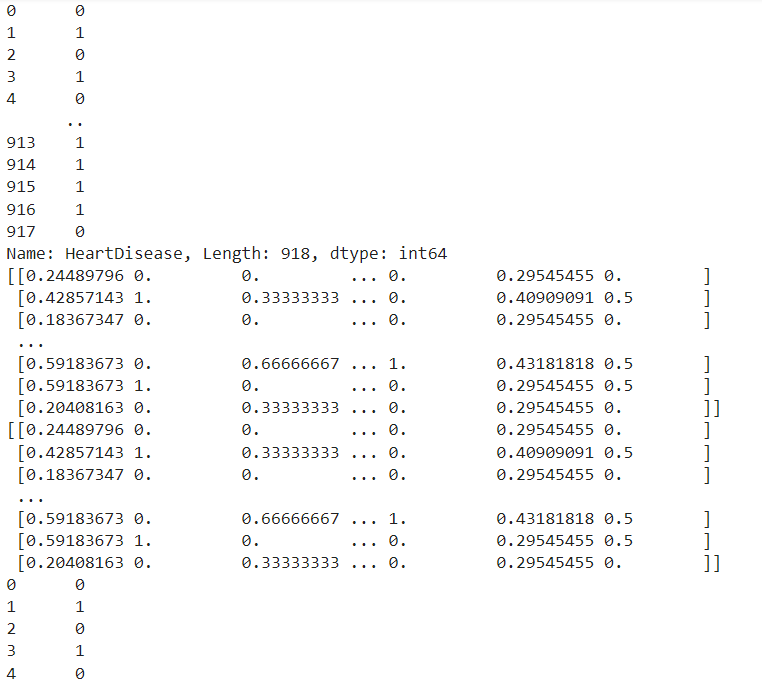


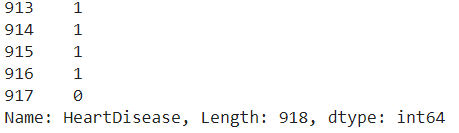


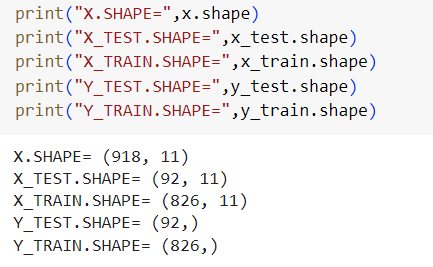


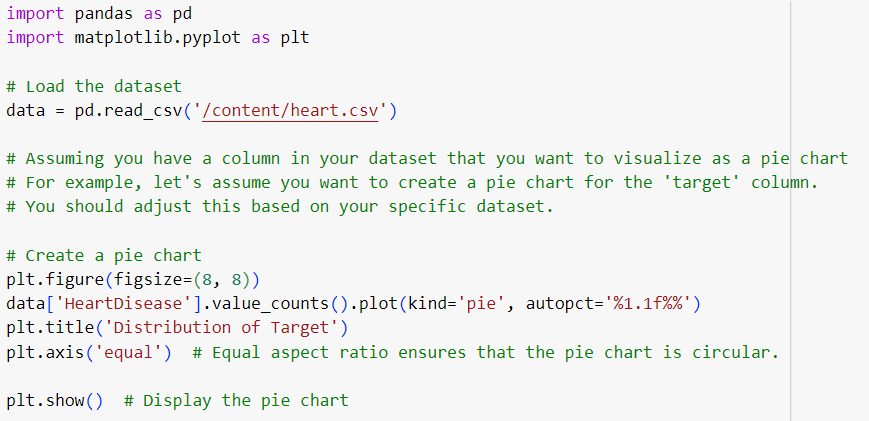


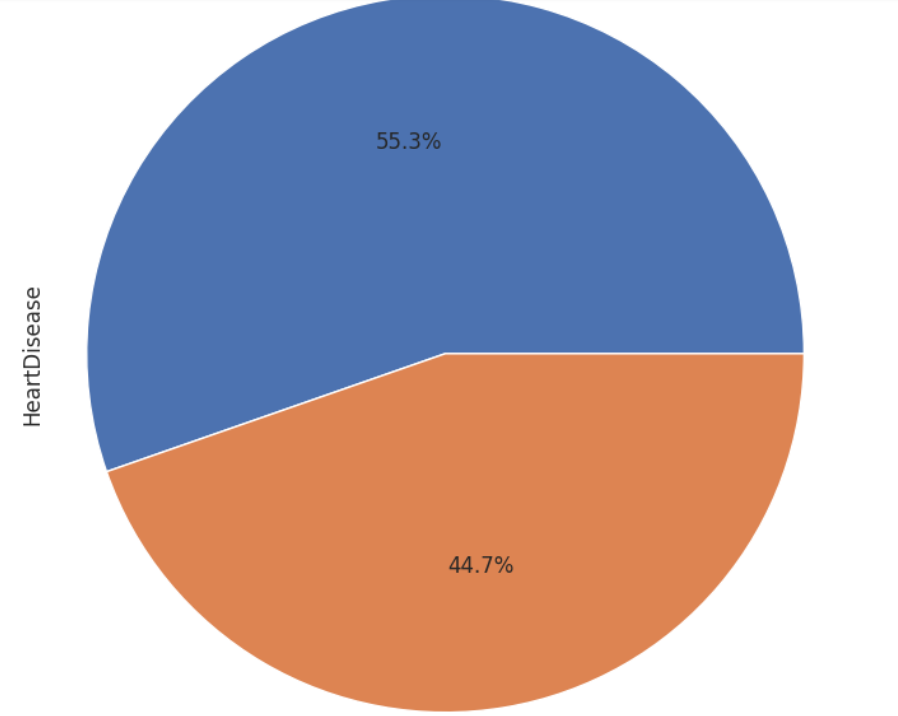


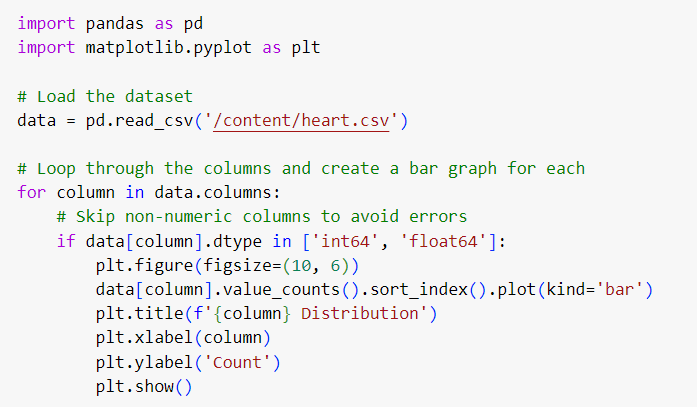


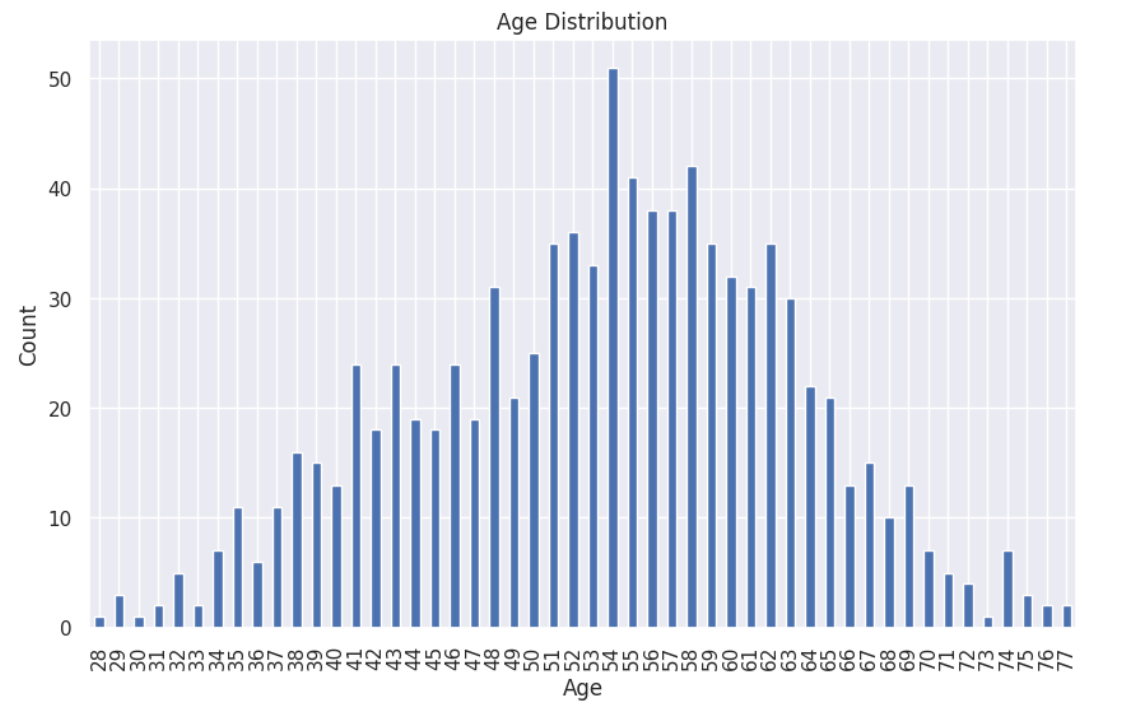


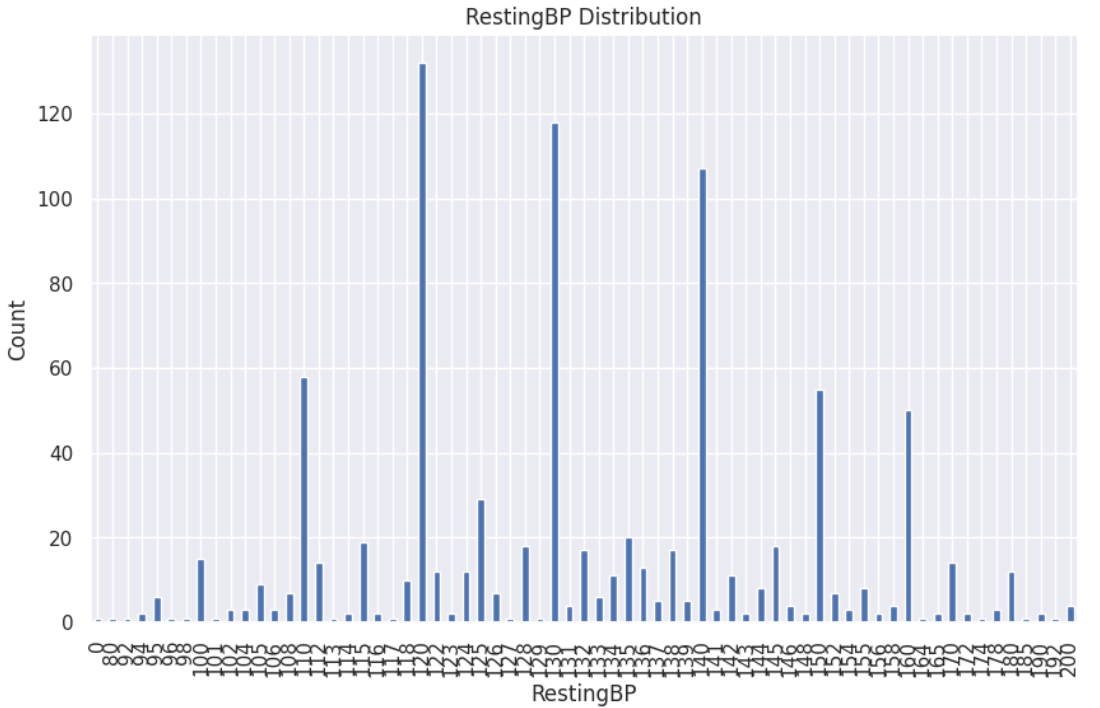


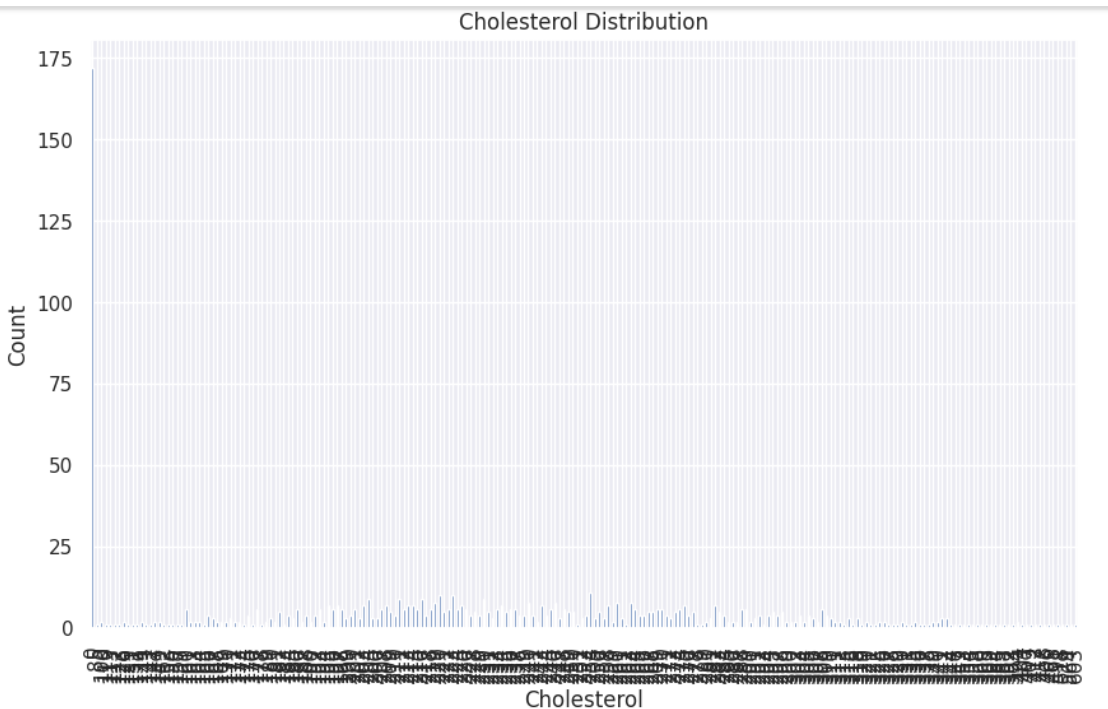


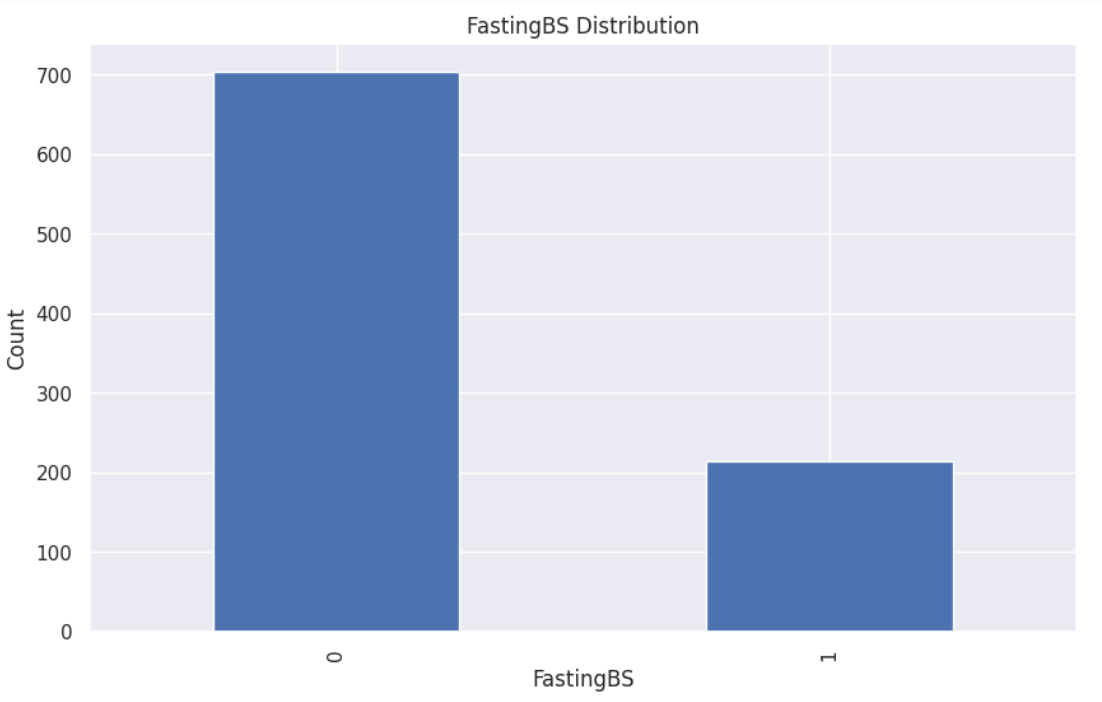


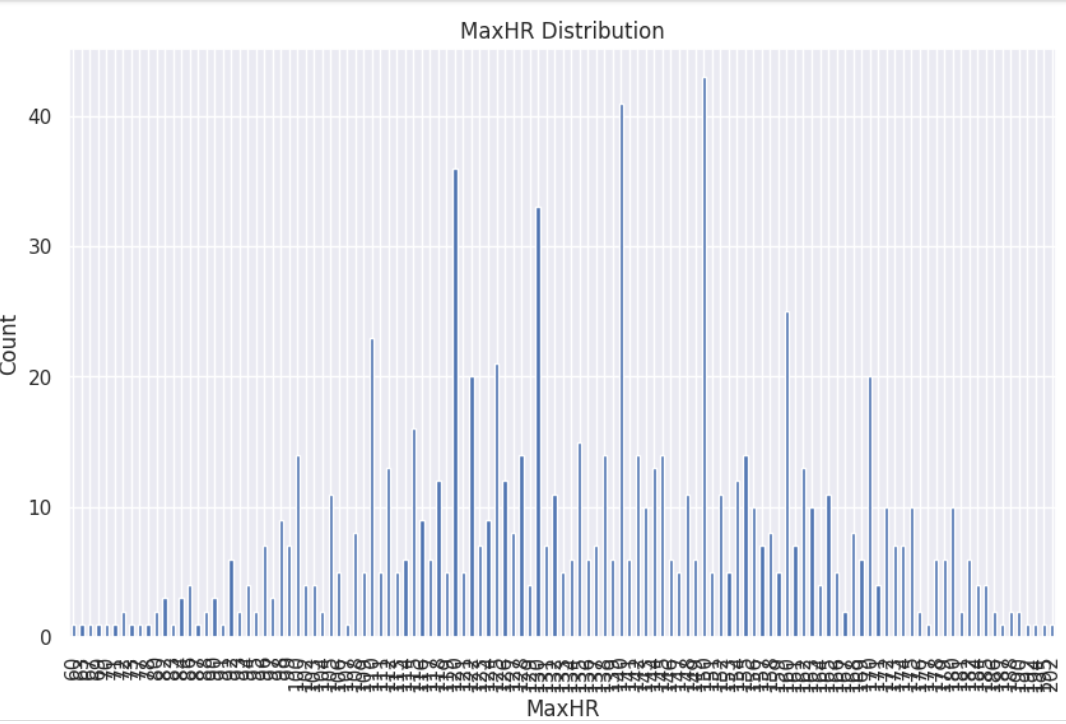


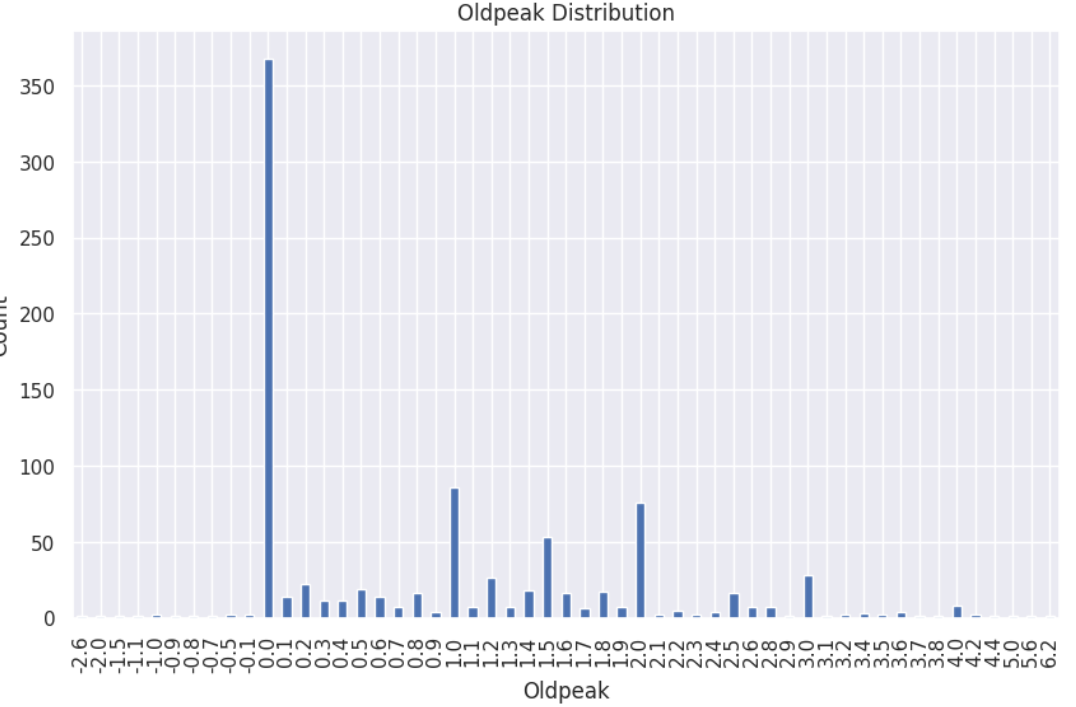


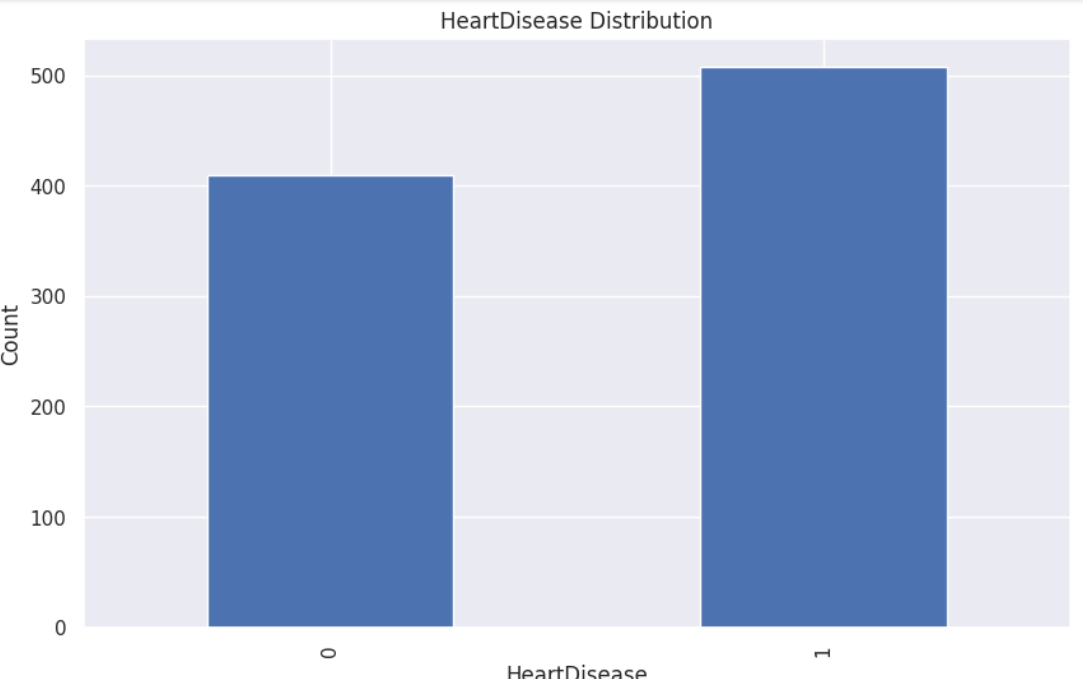


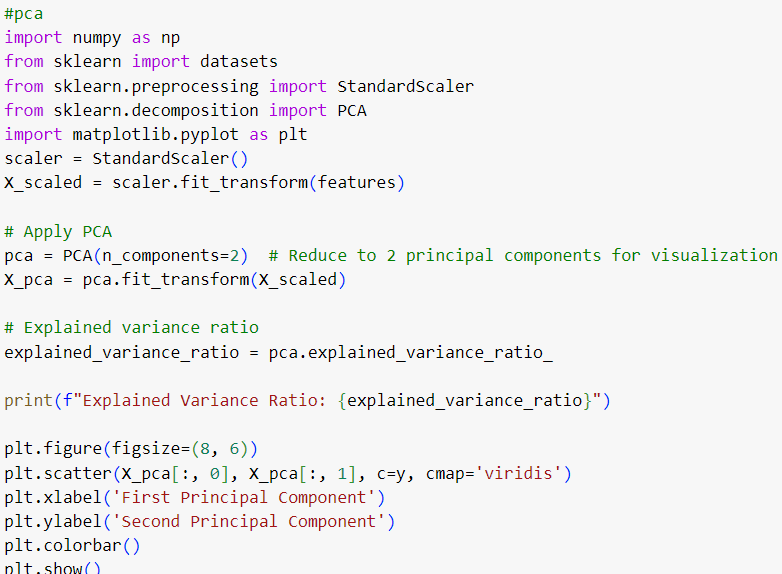


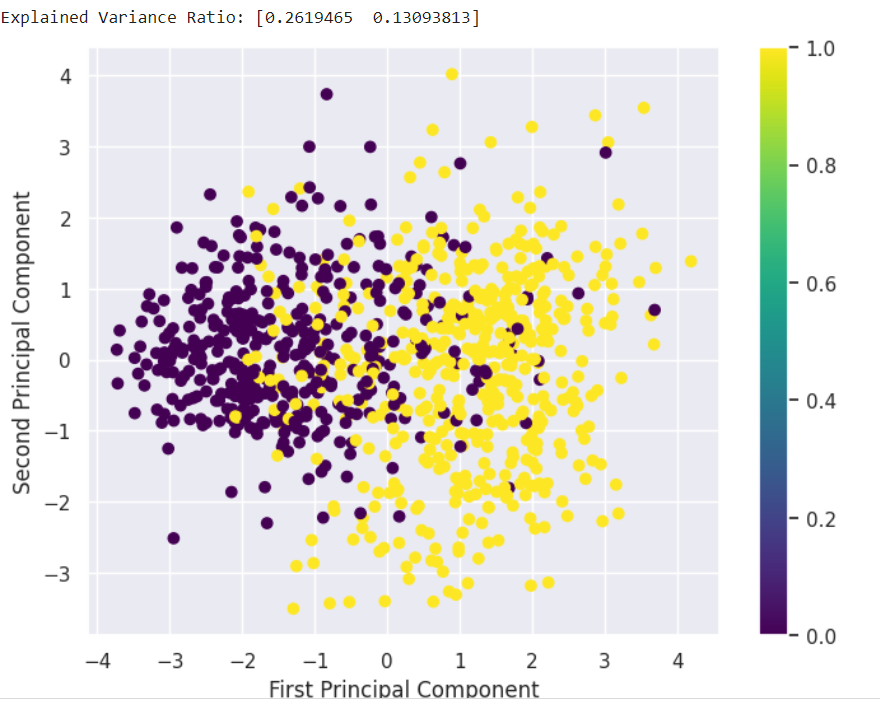


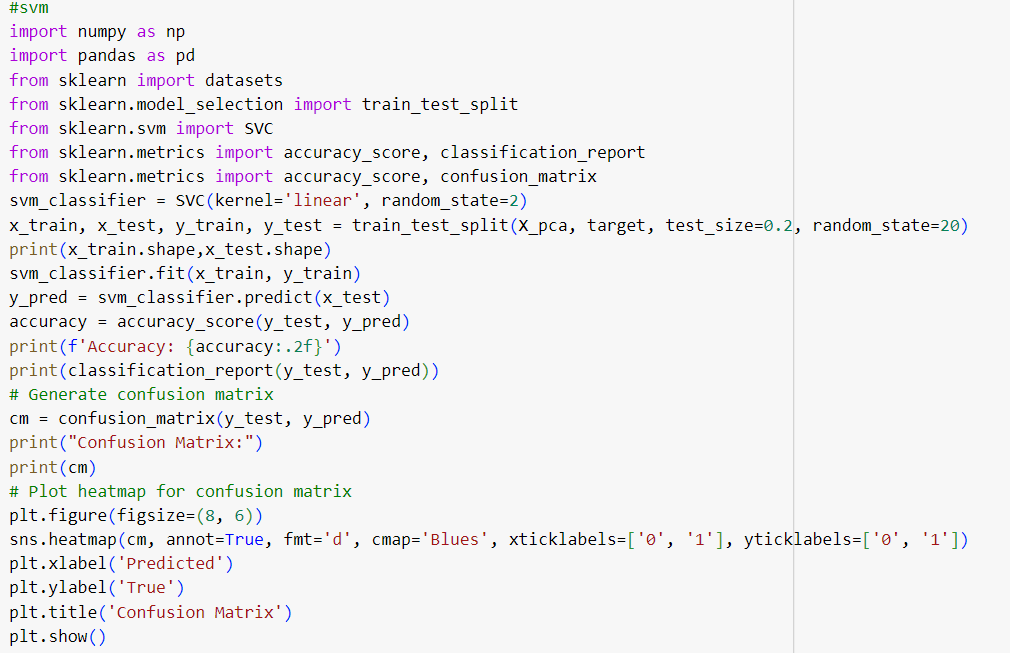


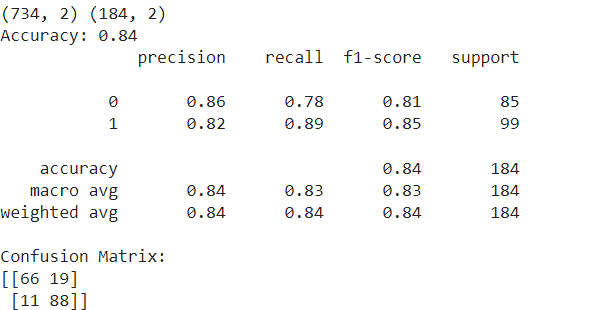


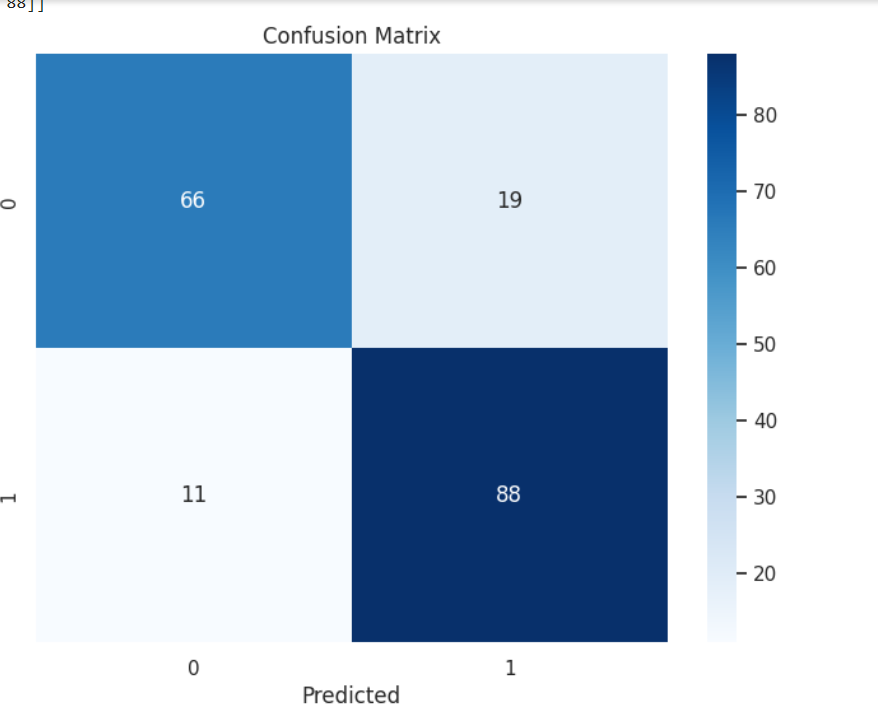


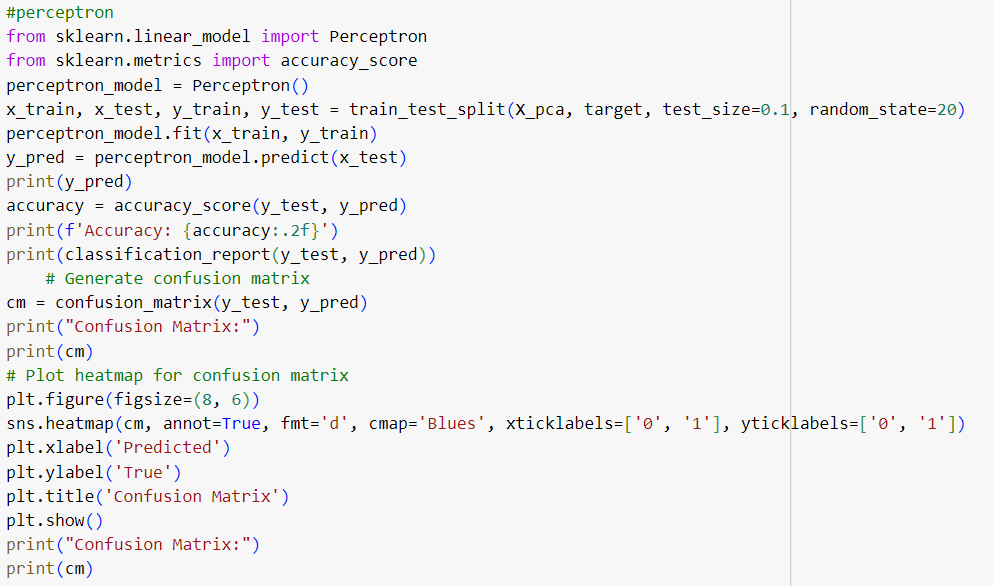


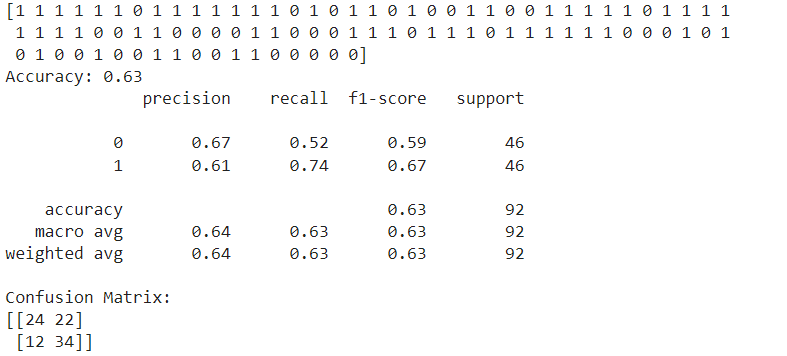


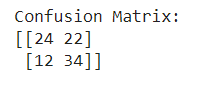
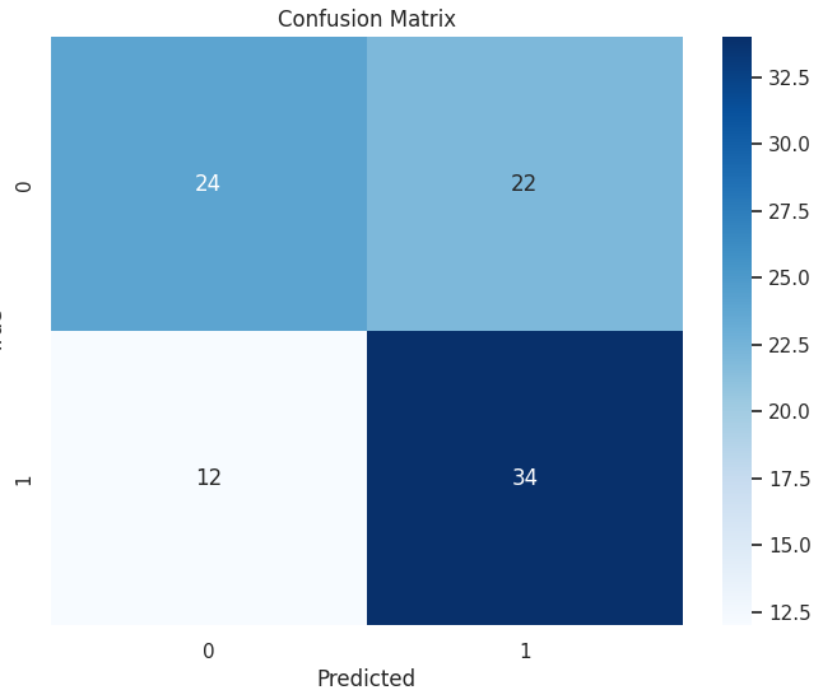


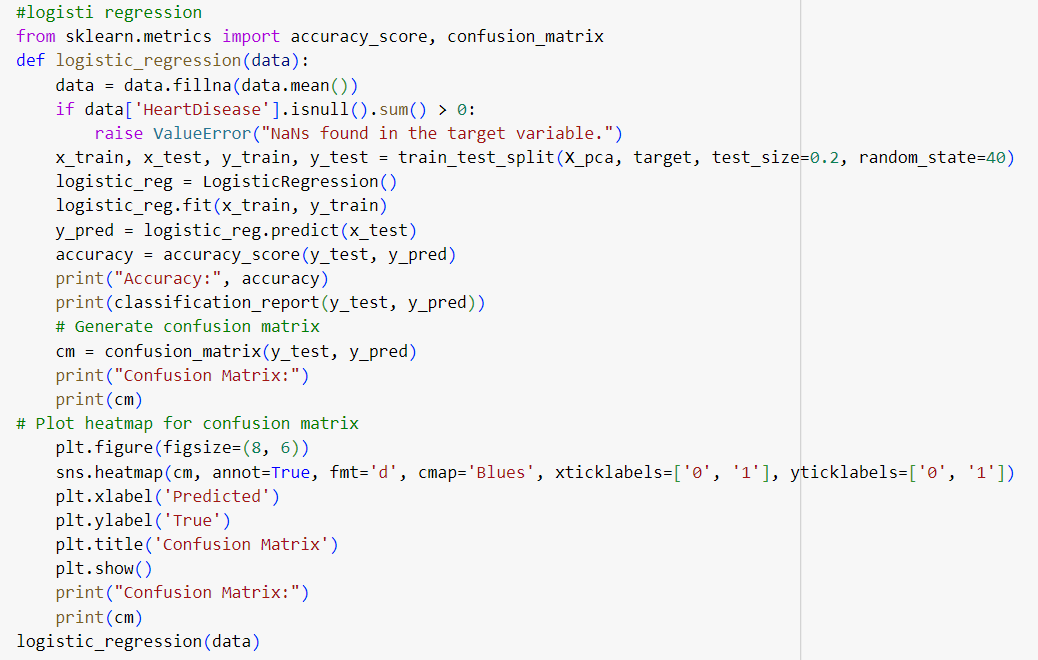


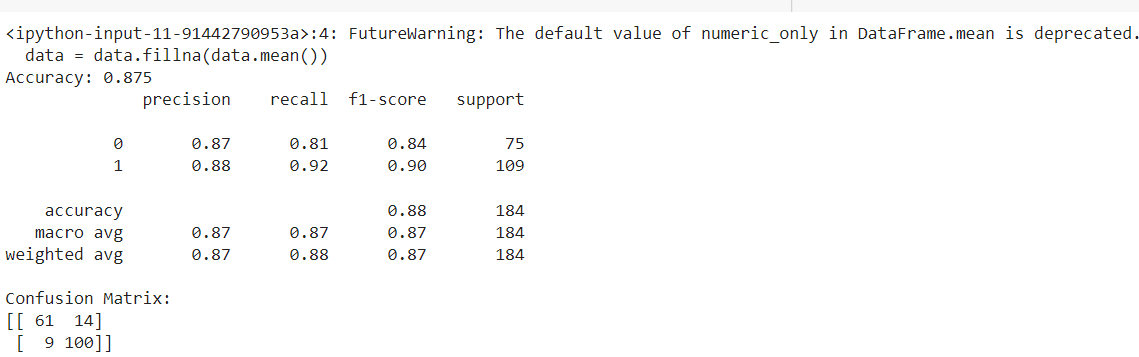


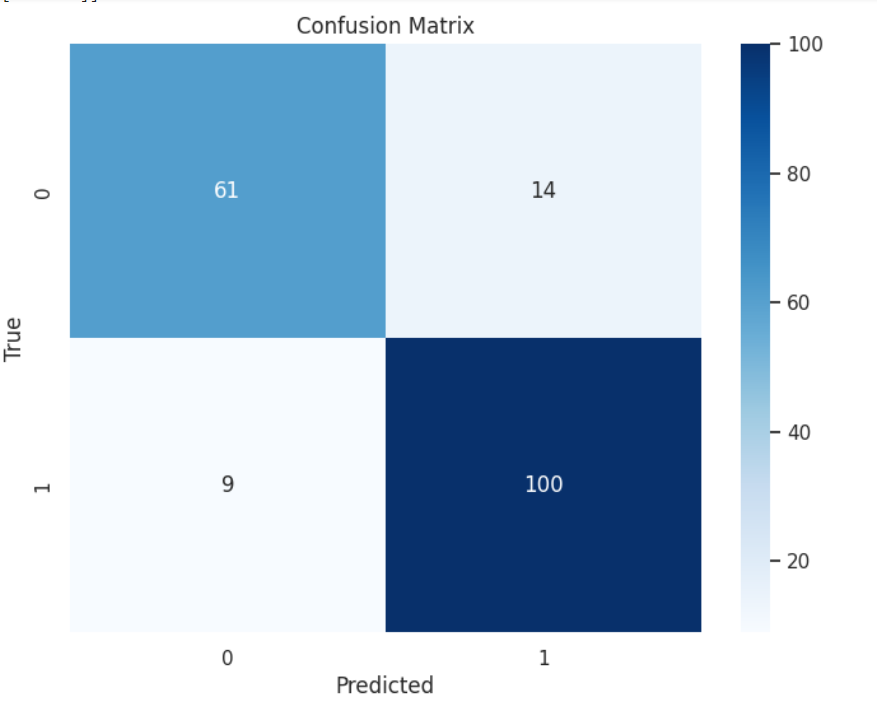


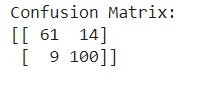




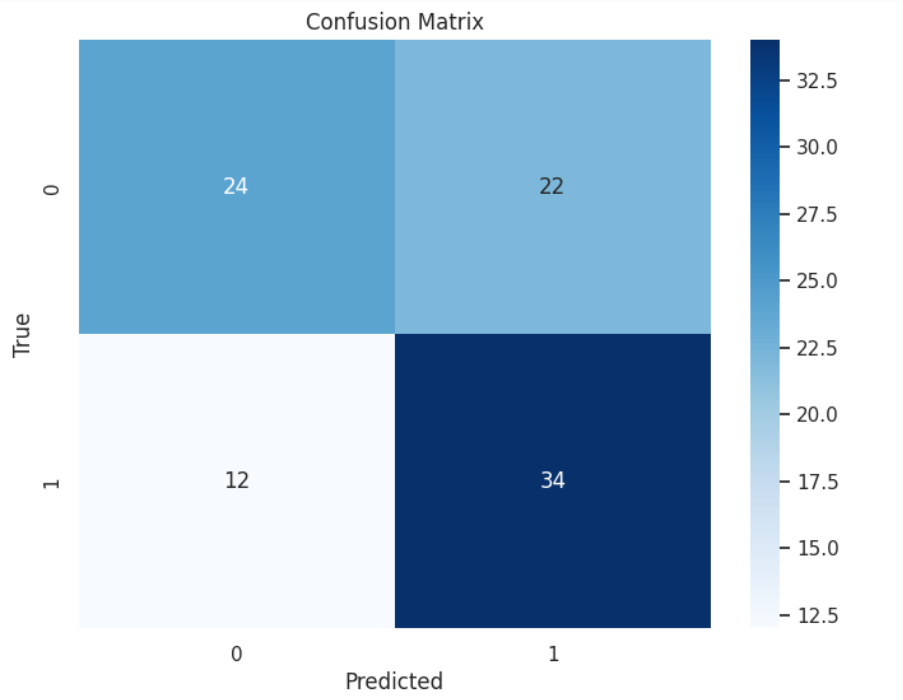
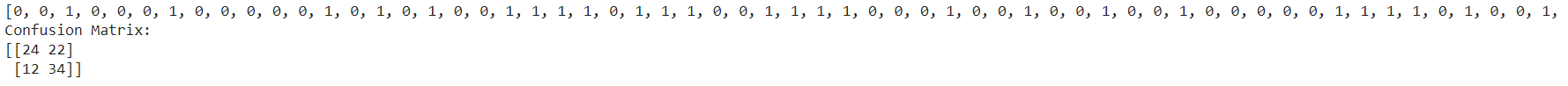


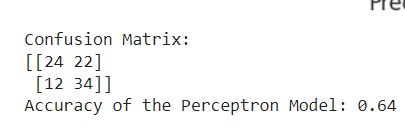




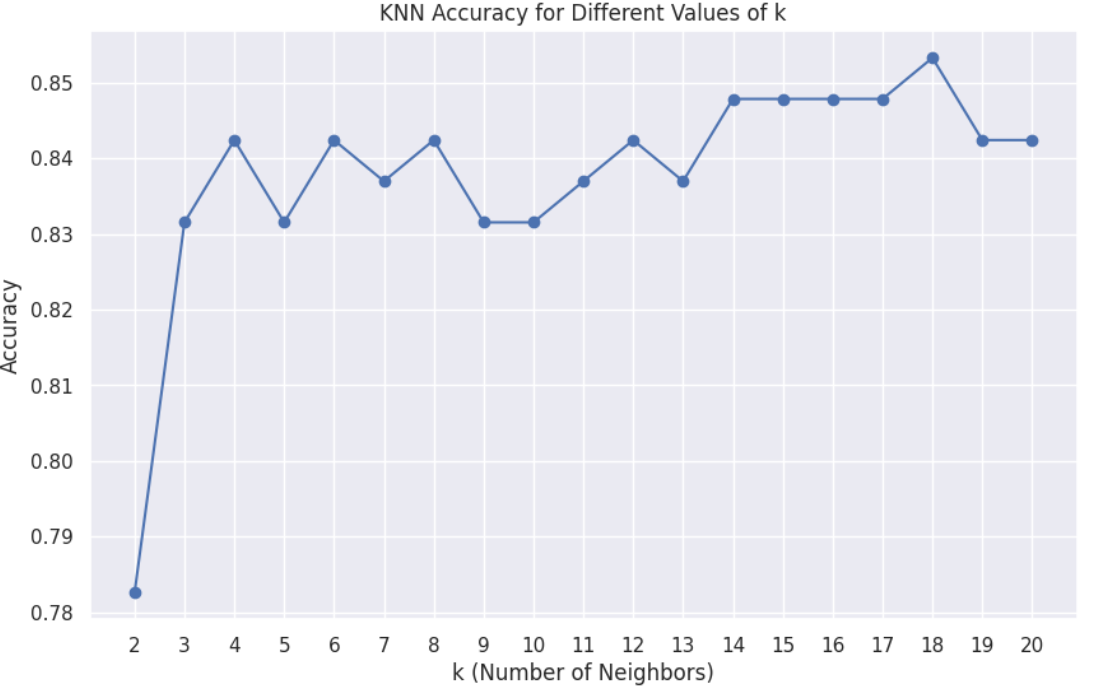






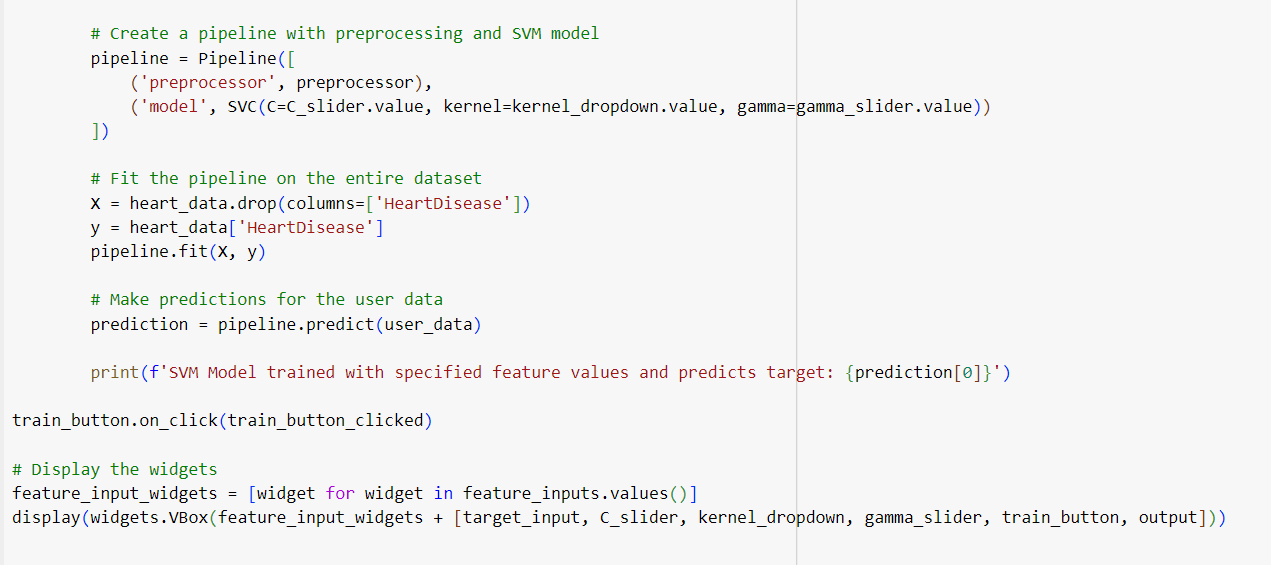


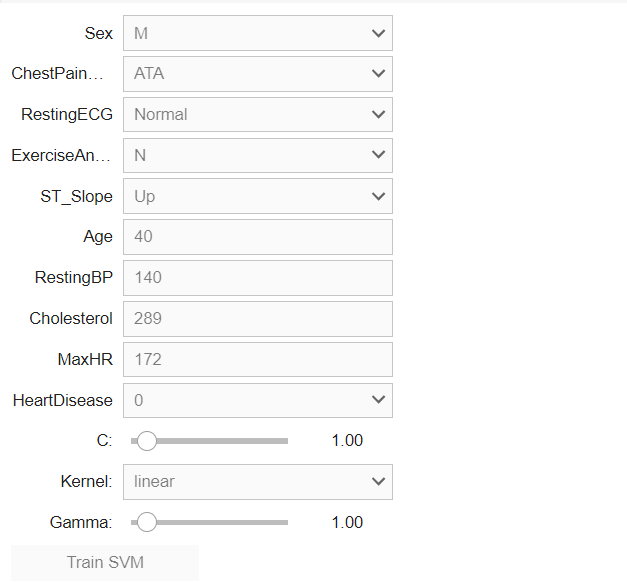












CONCLUSION:

**Machine learning algorithms** are computational models that allow computers to understand patterns and forecast or make judgments based on data without the need for explicit programming. These algorithms form the foundation of modern artificial intelligence and are used in a wide range of applications, including image and speech recognition, natural language processing, recommendation systems, fraud detection, autonomous cars etc.

Machine Learning can be a Supervised or Unsupervised. If you have lesser amount of data and clearly labelled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. If you have a huge data set easily available, go for deep learning techniques. You also have learned Reinforcement Learning and Deep Reinforcement Learning. You now know what Neural Networks are, their applications and limitations.

REFERENCES:

<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3759562>

<https://www.heart.org/en/health-topics/heart-failure>

<https://www.nhlbi.nih.gov/health-topics/heart-failure>

<https://www.hfsa.org/>

<https://www.sciencedirect.com/sc>

<https://www.jacc.org/doi/full/10.1016/j.jchf.2014.04.008>

<https://onlinelibrary.wiley.com/doi/pdf/10.1002/ehf2.12426>

<https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012072/meta>

<https://www.hindawi.com/journals/cin/2021/8387680/>