

MET Bhujbal Knowledge City, Nashik

Department of Computer Engineering

Academic Year : 2021-22

**Data Mining and Warehousing**

Project Report

On

**Heart Disease Risk Prediction**

By : Tanvi Nirmal (B-15)

Sakshi Thorat (B-17)

Guided by : Prof. Vishal Patil

**Contents**

1. Problem Statement …1

2. Introduction …2

3. Dataset …3

4. Classification Models …5

5. Result …6

6. Conclusion …7

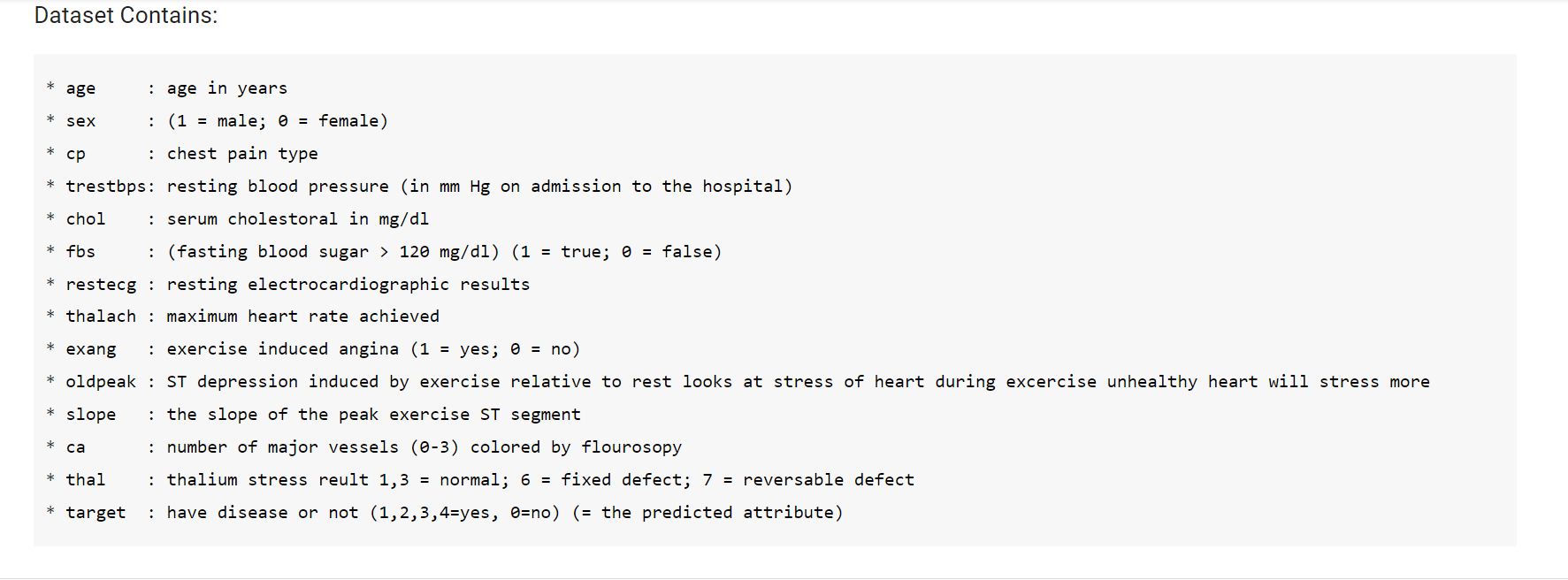
**Problem Statement**

Consider a labeled dataset belonging to an application domain. Apply suitable data preprocessing steps such as handling of null values, data reduction, discretization. For prediction of class labels of given data instances, build classifier models using different techniques (minimum 3), analyze the confusion matrix and compare these models. Also apply cross validation while preparing the training and testing datasets.

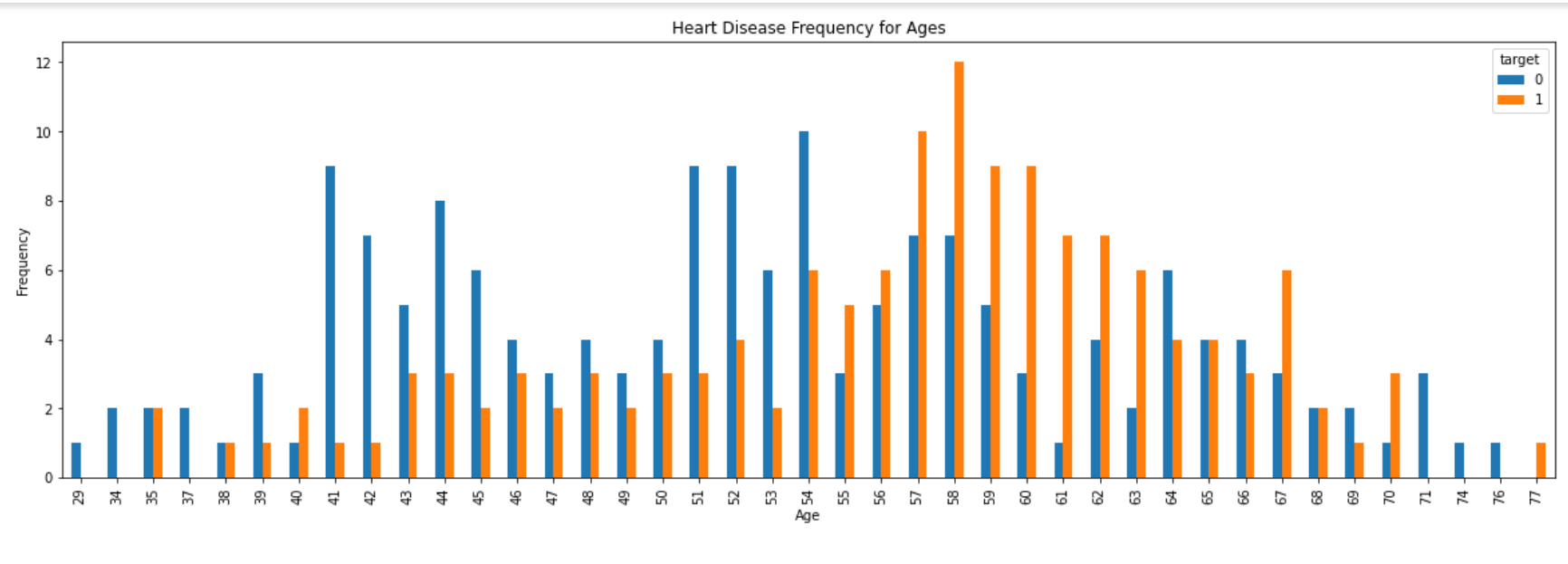
**Introduction**

It is a well known fact that Heart Diseases are currently the leading cause of death across the globe. The development of a computational system that can predict the presence of heart diseases in patients will significantly reduce the mortality rates and substantially reduce the costs of healthcare. Machine learning is used accross many spheres around the world. Especially it is gaining more popularity in the healthcare industry. Machine learning can play an essential role in predicting presense or absence of a critical disease, for an instance, Heart disease, etc. If such information is predicted well in advance, can provide important insights to the doctors, who can then carry out the treatments of the patients accordingly and efficiently. In addition to that, heart disease prediction is carried out using different approaches such as Support Vector Classifier, Logistic regression, Naive Bayes and Random Forest Classifier.

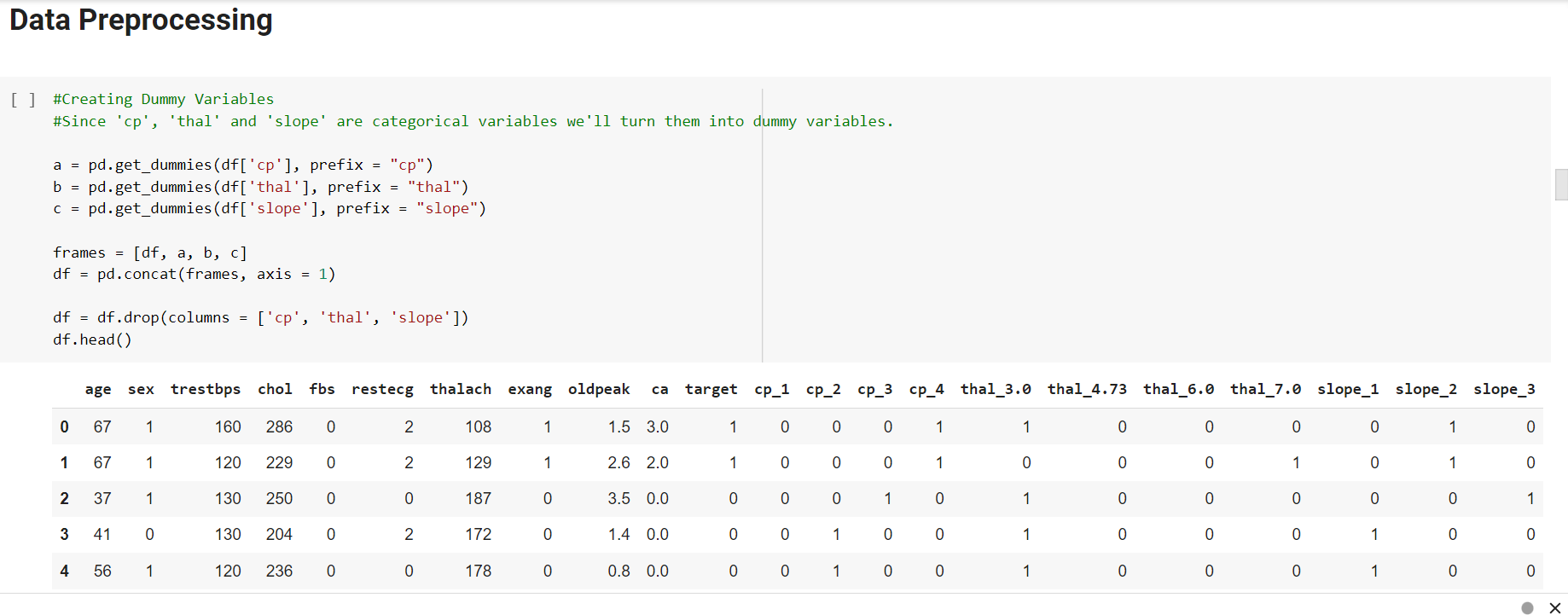
Data set :



Data Exploration :

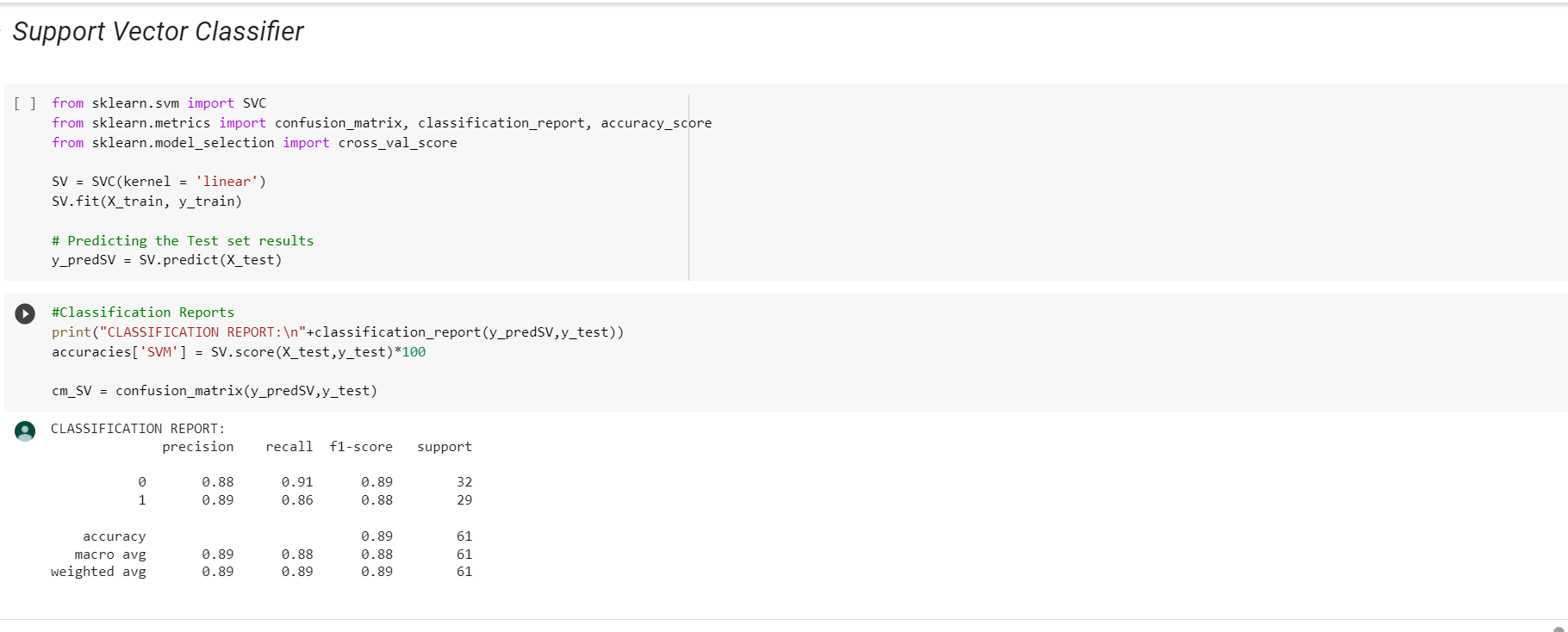


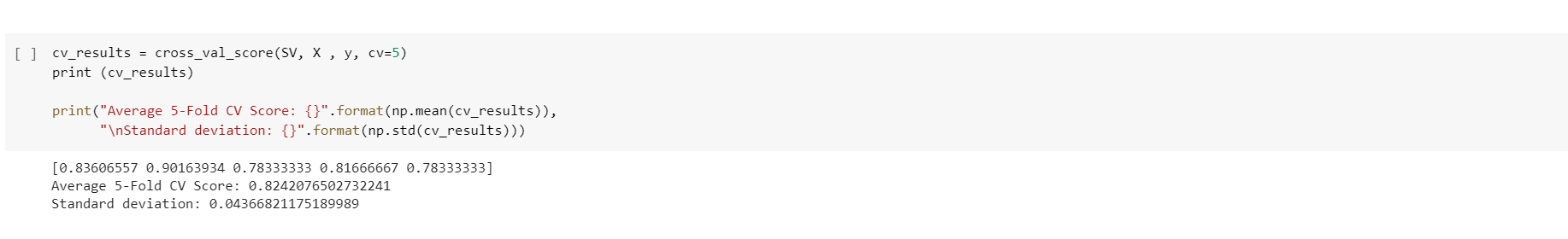
Data Preprocessing :



**Classification Models**

**Support** **Vector Classifier**

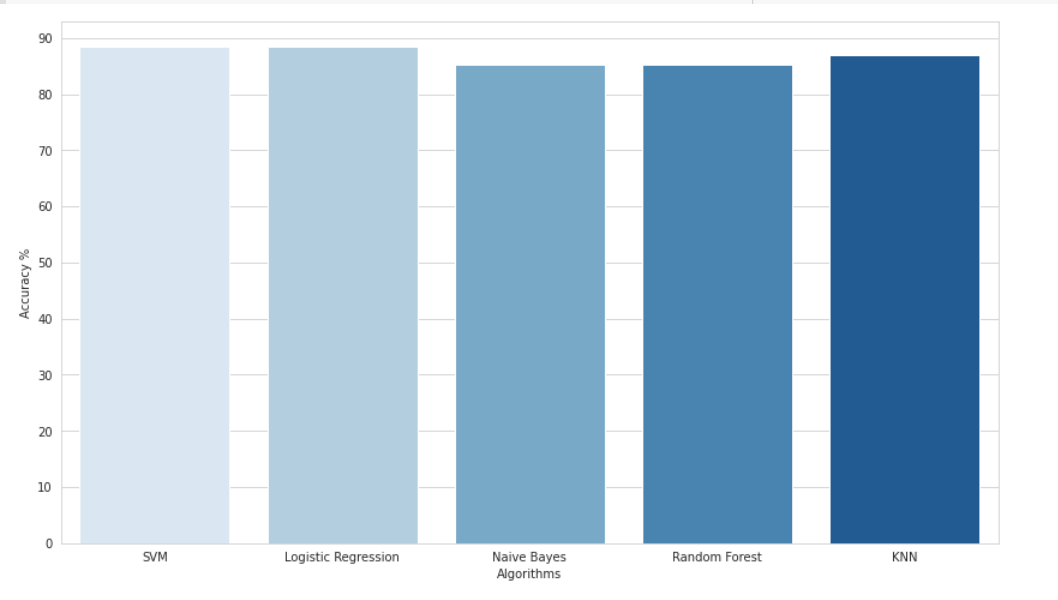
****

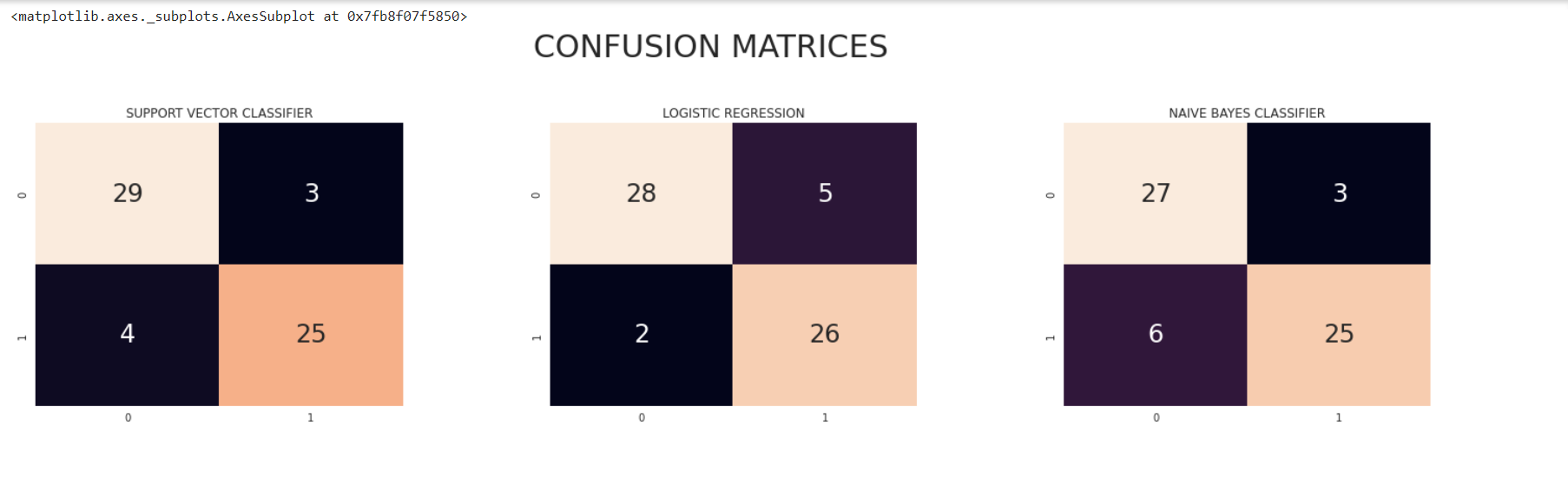


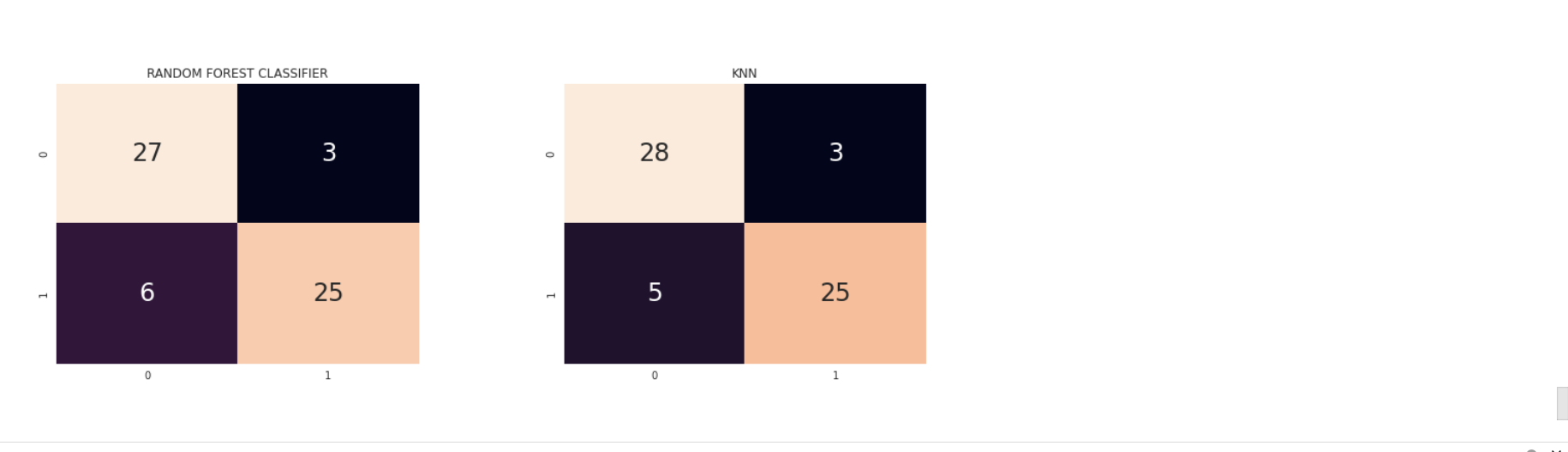
Similarly, we will use other classifiers Logistic regression, Naive Bayes and Random Forest Classifier on the dataset.

**Result :**

It was observed that Support Vector Classifier has more accuracy compared to other classifiers.







**Conclusion :**

Thus, we have successfully analysed the Heart Disease dataset and performed data preprocessing steps. We have also used different classifiers such as Support Vector Classifier, Logistic regression, Naive Bayes and Random Forest Classifier.