**Heart Disease Prediction using SVM and Naïve bayes algorithm**

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Engineering

by

**Shashank Wankhede (49)**

**Sanket Wankhade (48)**

**Pritam Tawade (43)**

Under the guidance of **Prof. Ravita Mishra**



DEPARTMENT OF INFORMATION TECHNOLOGY

KOKAN GYANPEETH COLLEGE OF ENGINEERING

KARJAT

**2019-2020**

**DECLARATION**

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited references to the original sources. We also declare that we have adhered to all principals of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

**----------------------------------------------------**

**Shashank Wankhede (49)**

**----------------------------------------------------**

**Sanket Wankhade (48)**

**----------------------------------------------------**

**Pritam Tawade (43)**

**Date:**

**ACKNOWLEDGMENT**

It is an opportunity of immense pleasure for us to present the Mini project report on "Heart Disease Prediction" expressing our heart left gratitude to all those who have generously offered their valuable suggestions towards the completion of this report.

The credit goes to professor Ravita Mishra whose positive attitude, moral support and encouragement lead to the success of the report. Her generous help, excellent guidance, lucid suggestions and encouragement throughout the course of this work have greatly helped us in the successful completion of this work.

Shashank Wankhede (49)

Sanket Wankhade (48)

Pritam Tawade (43)

**ABSTRACT**

This project is used to predict potential Heart Diseases in people using Machine Learning algorithm. Machine Learning is used across many spheres in the world and it can be used in the field of medicine just as effectively. Attributes like age, sex, resting blood pressure are analyzed to detect the presence of heart disease in the patient. Such information, if predicted well in advance, can provide important insights to doctors for the diagnosis and treatment.

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Name of Chapter** | **Page No.** |
| 1 | Introduction | 1 |
| 2 | Heart Disease Prediction | 2 |
| 2.1 | Pre-processing | 2 |
| 2.2 | Requirements | 3 |
| 3 | Implementation | 3 |
| 3.1 | Algorithms | 3 |
| 3.2 | Sample Code | 4 |
| 4 | Analysing Data | 6 |
| 4.1 | Visualising data | 7 |
| 5 | Result | 9 |
| 6 | Conclusion | 10 |

**CHAPTER 1**

INTRODUCTION

Heart disease is a general term that means the heart is not working properly. Heart disease can also refers to bunch of different conditions. The most common type of “heart disease” is Coronary Artery Disease. It occurs when plague builds up in arteries that supply blood to the heart. Sometimes babies can be born with heart disease this is called as congential heart disease. The most common symptoms of heart disease is Pain in the chest, Trouble breathing, Swelling, Palpitations. According to the survey in 2016, the estimated prevalence of Heart disease in India was estimated to be

54.5 million. One in 4 deaths in India are beacause of Heart Disease and strokes.

We decided to use Data Mining algorithms to predict whether a person has any heart problem or not based on symptoms and data collected. We will use two algorithms such support vector machine and Naïve bayes

**CHAPTER 2**

**2.Heart Disease Prediction:**

We decided to use Data Mining algorithms to predict whether a person has any heart problem or not based on symptoms and data collected. We will use two algorithms such support vector machine and Naïve bayes

Data contains:-

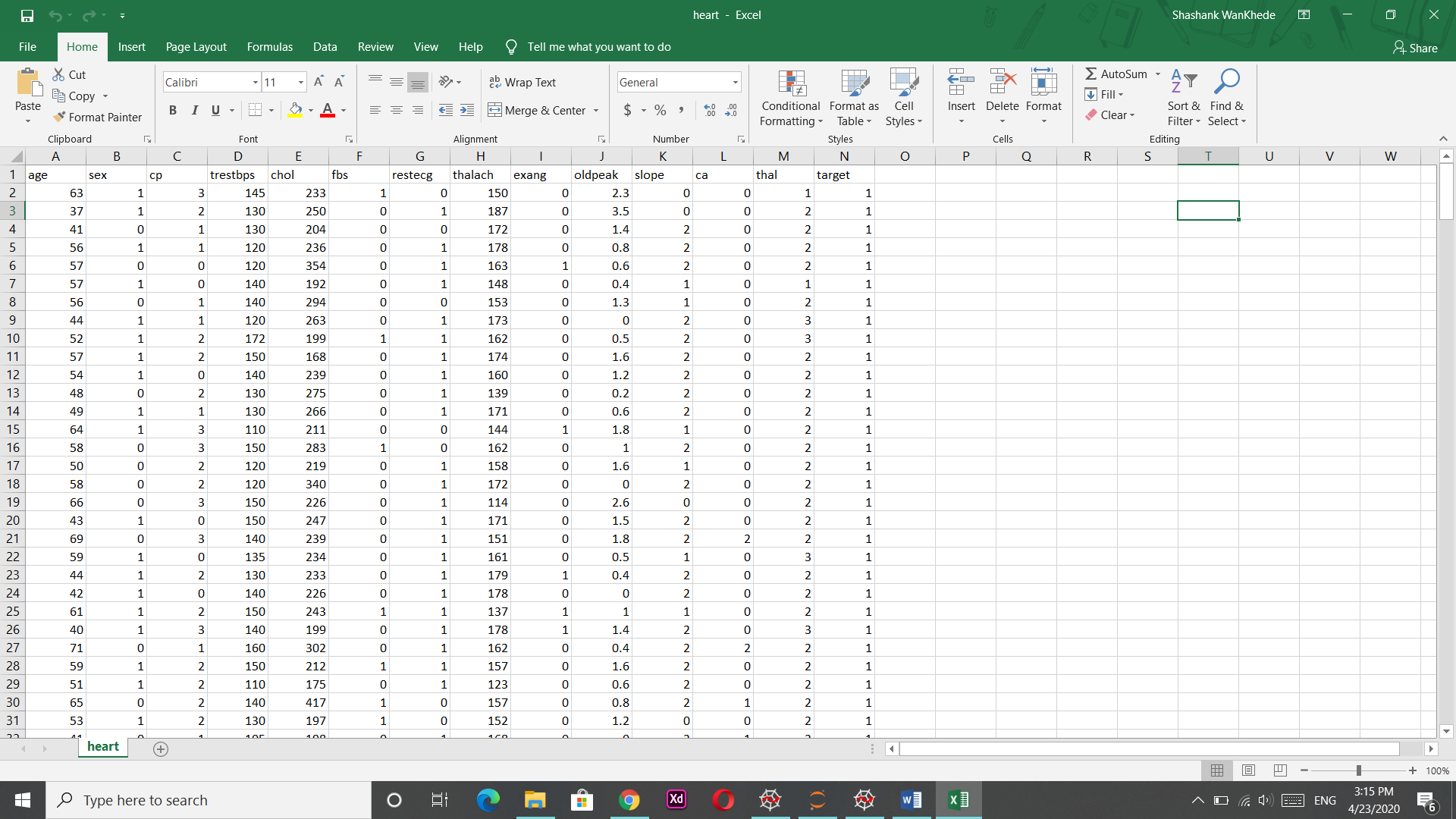
* age - age in years
* sex - (1 = male; 0 = female)
* cp - chest pain type
* trestbps - resting blood pressure (in mm Hg on admission to the hospital)
* chol - serum cholestoral in mg/dl
* fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
* restecg - resting electrocardiographic results
* thalach - maximum heart rate achieved
* exang - exercise induced angina (1 = yes; 0 = no)
* oldpeak - ST depression induced by exercise relative to rest
* slope - the slope of the peak exercise ST segment
* ca - number of major vessels (0-3) colored by flourosopy
* thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
* target - have disease or not (1=yes, 0=no)

**2.1 Preprocessing**

Checking for null values and removing them:

As our data is already preprocessed we do not need to clean it

**Dataset:-**



**2.2Requirements:**

* + **Sypder IDE**
  + **Jupyter notebook**
  + **Microsoft Excel**
  + **Python 3.8**
  + **Libraries: numpy , pandas, sklearn, seaborn, matplotlib**
  + **Dateset (downloaded from Kaggle.com)**
  + **Computer with internet**

**CHAPTER 3**

IMPLEMENTATION

**3.1 Algorithm:-**

**Support Vector Machine:-**

Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However,  it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well

**Naïve Bayes:-**

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

**3.2 Sample Code:**

#Importing Libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

#reading file

data = pd.read\_csv('heart.csv')

#dropping less corralted columns

data = data.drop(['sex','trestbps','chol','fbs'],axis=1)

x = data.iloc[:,:-1].values

y = data.iloc[:,[-1]]

x\_train,x\_test,y\_train,y\_test = train\_test\_split(x,y,test\_size=0.2,random\_state=0)

#using svm

from sklearn import svm

cls = svm.SVC(kernel='linear')

#Using Naïve Bayes algorithm

from sklearn.naive\_bayes import GaussianNB

nb = GaussianNB()

#fitting model

cls.fit(x\_train,y\_train.values.ravel())

nb.fit(x\_train,y\_train)

#making prediction

pred = cls.predict(x\_test)

nb\_pred=nb.predict(x\_test)

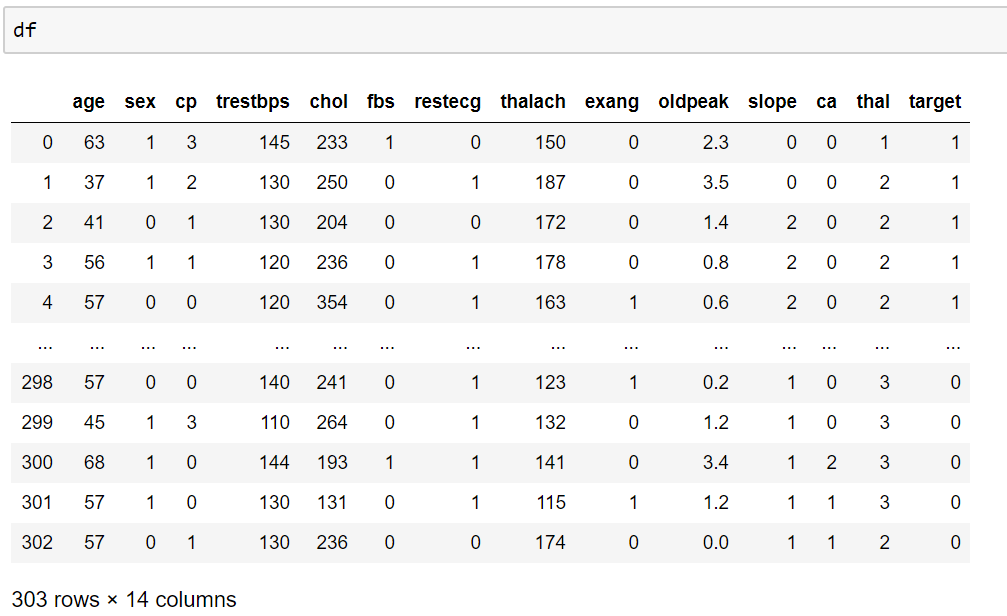
#calculating accuracy

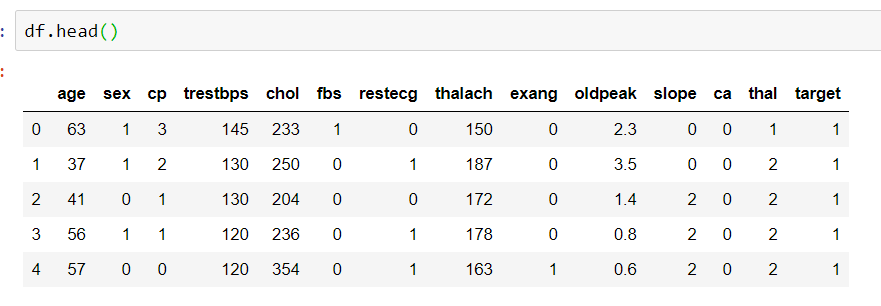
from sklearn.metrics import accuracy\_score

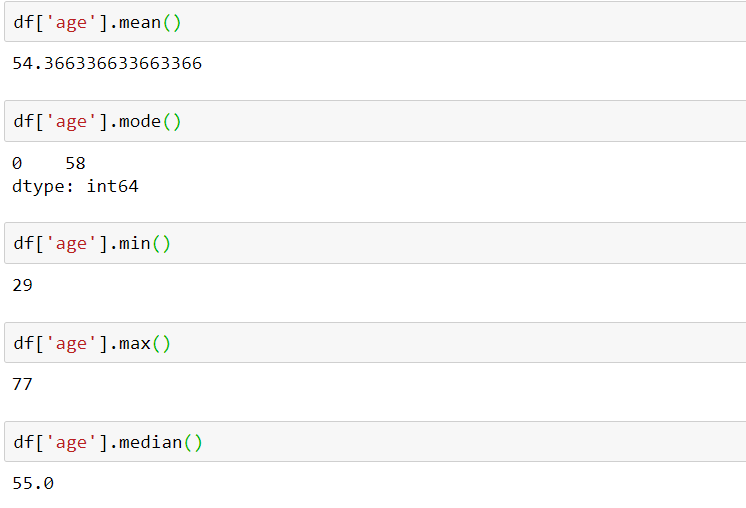
svm = accuracy\_score(y\_test, pred)

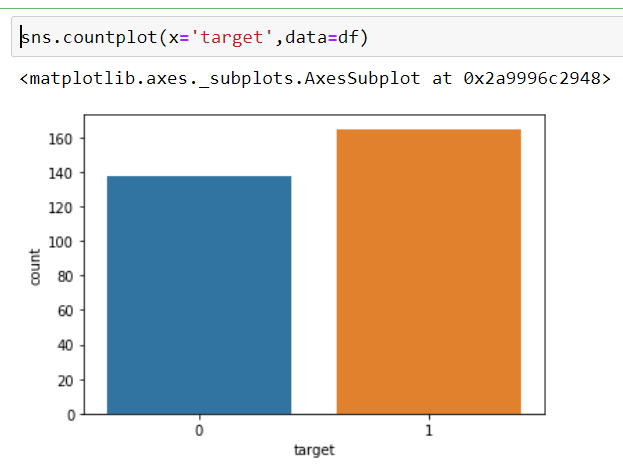
**CHAPTER 4**

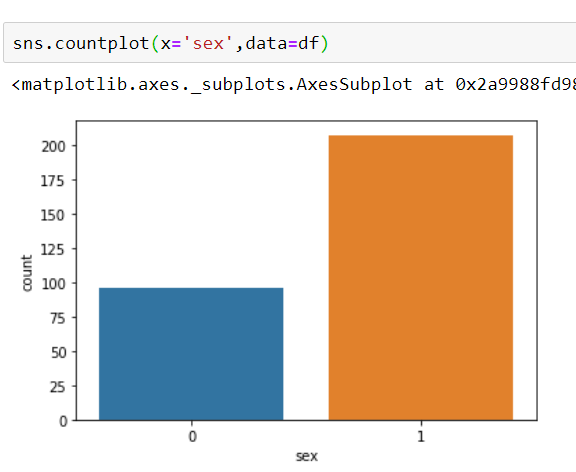
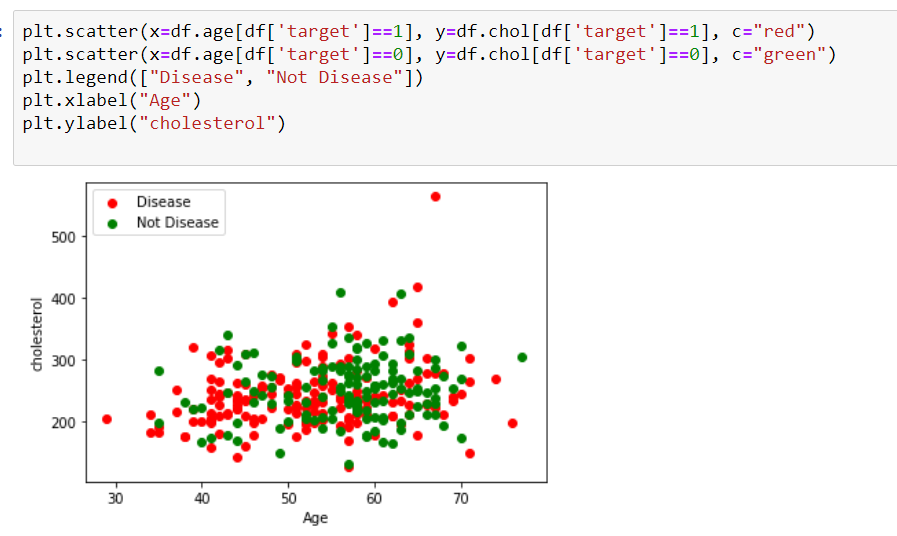
**4.1 Analysing data:**

****

****

**4.1 Visualizing Data:-**

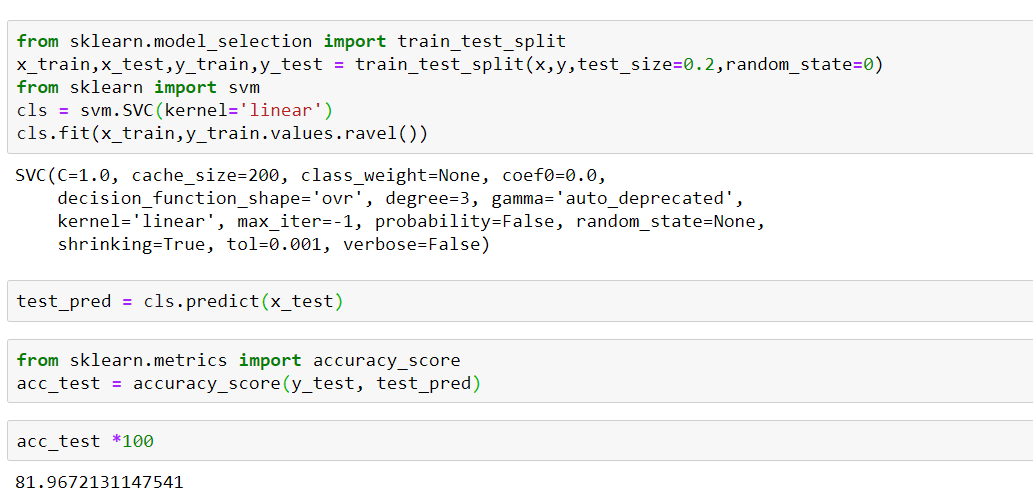
****counting target(1=yes,0=no)

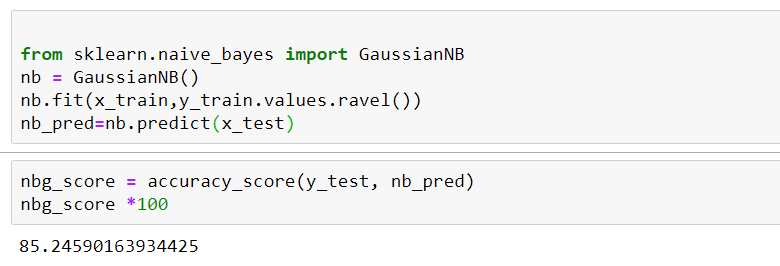
****

Counting male and female in dataset(1=male,0=female)

**Chapter 5**

**Result:**

**SVM algorithm:-**

**Naïve Bayes Algorithm:-**

Accuracy score of SVM is 81.9 %

Accuracy score of Naïve Bayes algorithm is 85.2 %

**CHAPTER 6**

CONCLUSION

* We came to know that Heart problem is beign a serious problem among all older persons specially a person who has high cholesterol ,chest pain are more likely to to have heart problem.
* We implemented two algorithms such as SVM and Naïve Bayes. Both perform well but Naïve bayes gives more accuracy than SVM i.e 85.2 % which simply means that Naïve bayes will predict or classify patients with heart disese

REFERENCES

* <https://www.kaggle.com/ronitf/heart-disease-uci>
* <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>
* <https://www.researchgate.net/publication/330981991_Analysis_of_Heart_Disease_Prediction_using_Various_Machine_Learning_Techniques>
* <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>
* <https://machinelearningmastery.com/naive-bayes-for-machine-learning/>