Heart Disease Prediction Using Machine Learning Algorithms

A

Project Report

Submitted for the partial fulfilment

of B.Tech. Degree

in

COMPUTER SCIENCE & ENGINEERING

by

Gaurav Kumar (1805213020)

Arjun Singh (1805213014)

Sanjay Singh (1805213050)

Under the supervision of:

[Prof. M H Khan]

[Dr. Promila Bahadur]

Department of Computer Science and Engineering

Institute of Engineering and Technology

Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh.

Contents

Declaration

We hereby declare that this submission is our own work and that, to the best of our belief and knowledge, it contains no material previously published or written by another person or material which to a substantial error has been accepted for the award of any degree or diploma of university or other institute of higher learning, except where the acknowledgement has been made in the text. The project has not been submitted by us at any other institute for requirement of any other degree.

Submitted by: - Date: 23/05/2022

(1) Name: Gaurav Kumar

Roll No.: 1805213020

Branch: Information Technology

Signature:

(2) Name: Arjun Singh

Roll No.: 1805213014

Branch: Information Technology

Signature:

(3) Name: Sanjay Singh

Roll No.: 1805213050

Branch: Information Technology

Signature:

Certificate

This is to certify that the project report entitled "**Heart Disease Prediction Using Machine Learning Algorithms**" presented by **Gaurav Kumar, Arjun Singh and Sanjay Singh** in the partial fulfillment for the award of Bachelor of Technology in Computer Science and Engineering, is a record of work carried out by them under my supervision and guidance at the Department of Computer Science and Engineering at Institute of Engineering and Technology, Lucknow.

It is also certified that this project has not been submitted at any other Institute for the award of any other degrees to the best of my knowledge.

[Prof. M H Khan]

[Dr. Promila Bahadur]

Department of Computer Science and Engineering

Institute of Engineering and Technology, Lucknow

Acknowledgement

First of all, we all are indebted to the Almighty for giving us an opportunity to excel in our efforts to complete this project work on time.

We would like to thank our **Supervisor Prof. M H Khan**, and **Dr. Promila Bahadur** of Computer Science & Engineering Department, IET Lucknow, for their continued support and suggestions so far. They have been very constructive, supportive, kind and made it easier to write up this project work. Their continuous support gave us motivation to work on this write up and helped us to gain knowledge in various different fields. We would also like to thank for their critiques and comments on the initial project work. They allowed us to locate areas of improvement that helped us to write this Interim write up.

We would also like to express gratitude to all our friends who motivated us and helped us at each stepin completing this project work.

We will be failing in duty if we do not acknowledge with grateful thanks to the authors of the referencesand other literatures referred to in this project.

Last but not the least, we are very much thankful to our parents who guided us in every step which we took.

Thank You.

Gaurav Kumar [1805213020]
Sanjay Singh [1805213050]
Arjun Singh [1805213014]

Abstract

Machine Learning is used across many ranges globally. The medicare industry is no exclusion. It can play an important role in predicting presence or absence of locomotors disorders, heart diseases and more. Such information, if analysed well in advance, then it can provide important intuitions to doctors so that they can adapt their diagnosis and dealing per patient.

In this project the algorithms predict possible Heart Diseases in people. In this project we perform the comparative analysis of algorithms like Random Forest and we propose an ensemble algorithm which performs the hybrid classification by taking strong as well as weak classifiers since the algorithm can have multiple number of samples for the training data that is why we perform the analysis of existing classifier and proposed the classifier which can give the good accuracy and good analysis of the data.

List of Figures

List of Tables

Chapter 1 Introduction

According to World Health Organization(WHO), every year more than 14 million deaths occur world-wide due to the Heart Disease. Prediction of Heart Disease is referred as one of the most booming subject in data analysis. The load of Heart disease is increasing rapidly in whole world from past few years. Many researches have been conducted in the attempt to pin-point the most influenting factors of the heart disease as well as predicting accurately the overall risk. Heart Disease is even considered as a silent killer which leads to the sudden death of the person without obvious symptoms. The early diagnosis of disease plays a vital role in making decisions on lifestyle changes and in high -risk patients and which in turn reduces the complications.

Machine learning algorithms prove to be effective in assisting and in making the decisions and prediction of the data from the large quantity produced by the health care industries. This project aims to predict the future Heart Disease by analyzing dataset of patients which classifies whether they have or not the heart disease using machine-learning algorithm. Machine Learning techniques can be a boom in this regard. Even though it can occur in many forms, there is a common set of core risk factors that effect whether someone will ultimately be at risk for heart disease or not. By collecting the data through many sources, analysing them under suitable headings & finally analysing it to extract the desired data we can say that this technology can be adapted for the prediction of heart diseases.

Chapter 2
Literature Review

[1] Purushottam published the paper “Efficient Heart Disease Prediction System” using the Hill Climbing and the Decision Tree Algorithm. He used Cleveland dataset and performed preprocessing by using classification algorithms. The Knowledge Extraction is done on the basis of Evolutionary Learning (KEEL), an open-source data mining tool that fills the missing values in the data- set. A decision tree which follows top-down approach. For each actual node selected by hill-climbing algorithm anode is selected by a test at each level. The parameters and the values used are confidential. Its minimum value is 0.25.

The accuracy of the system is about 84%.

[2] Sonam Nikhar published paper “Prediction of Heart Disease Using Machine Learning Algorithms” their research gives point to point explanation of Naive Bayes and decision tree classifier that are used mainly in the prediction of Heart Disease.

Some analysis has been led to think about the execution of prescient data mining strategy on the same dataset, and the result decided that it has highest accuracy than Bayesian classifier.
Its accuracy is about 82.7%.

[3] Our team project is “Heart Disease Using Machine Learning Algorithm” using Random Forest Algorithm. It is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
We trained the algorithm with the most popular set of the data of the patients of the heart diseases so that it can predict the accurate values and can give satisfying results.

It’s accuracy is about 85-87%.

Chapter 3

Methodology

3.1 EXISTING SYSTEM

Heart disease is even highlighted as the silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is cause of anxiety about it & its results. Hence the efforts are being done continuously in predicting the possibility of this disease initially. So that the various tools and technology is continously being experimented with to suit the present-day health needs. Machine Learning techniques can be a boom to this regard. Even though heart disease can occur in many forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the information from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted in prediction of heart disease.

3.2 PROPOSED SYSTEM

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is attained by testing the system using the data. This system is implemented using the following modules.

1.) Collection of the Dataset
2.) Selection of the Attributes
3.) Data Pre-Processing
4.) Balancing of the Data
5.) Diseases Prediction

3.2.1 COLLECTION OF THE DATSET

Initially, we collect the dataset for the heart disease prediction system. After the collection, we split the dataset into the training and testing data. The training dataset used for prediction model learning and testing data used for evaluating the prediction model. For this whole project, 74% of training data is used and 34% of data is used for testing. The dataset used for this project is Heart Disease Uci. The dataset consists of 73 attributes.

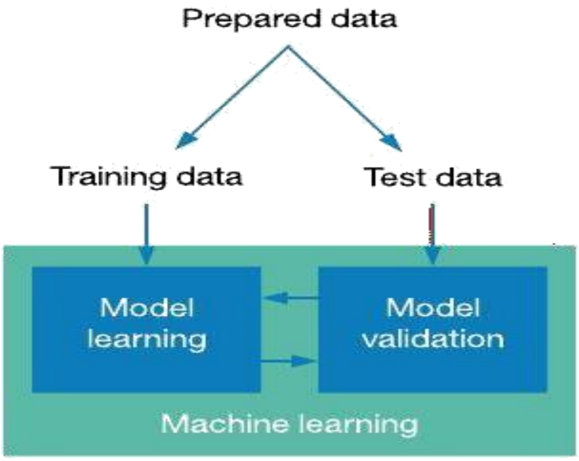
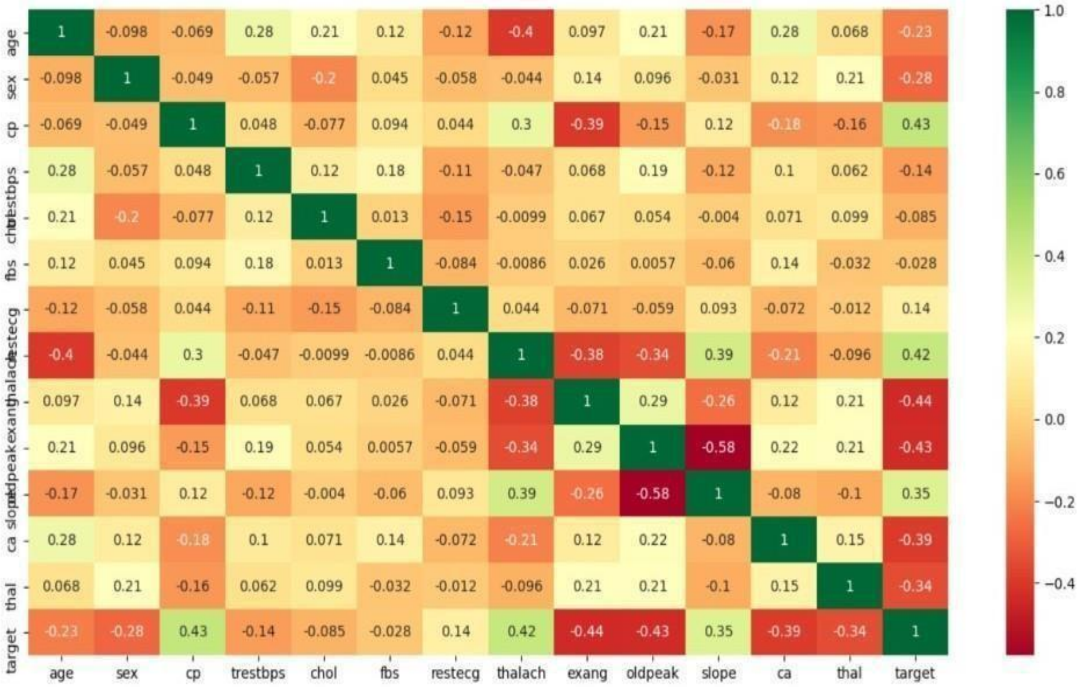


Figure: Collection of Data

3.2.2 Selection of attributes

Attribute selection includes in the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, etc are selected for the prediction. The Correlation matrix is being used for attribute selection for this model.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1	-0.098	-0.069	0.28	0.21	0.12	-0.12	-0.4	0.097	0.21	-0.17	0.28	0.068	-0.23
sex	-0.098	1	-0.049	-0.057	-0.2	0.045	-0.058	-0.044	0.14	0.096	-0.031	0.12	0.21	-0.28
cp	-0.069	-0.049	1	0.048	-0.077	0.094	0.044	0.3	-0.39	-0.15	0.12	-0.18	-0.16	0.43
trestbps	0.28	-0.057	0.048	1	0.12	0.18	-0.11	-0.047	0.068	0.19	-0.12	0.1	0.062	-0.14
chol	0.21	-0.2	-0.077	0.12	1	0.013	-0.15	-0.0099	0.067	0.054	-0.004	0.071	0.099	-0.085
fbs	0.12	0.045	0.094	0.18	0.013	1	-0.084	-0.0086	0.026	0.0057	-0.06	0.14	-0.032	-0.028
restecg	-0.12	-0.058	0.044	-0.11	-0.15	-0.084	1	0.044	-0.071	-0.059	0.093	-0.072	-0.012	0.14
thalach	-0.4	-0.044	0.3	-0.047	-0.0099	-0.0086	0.044	1	-0.38	-0.34	0.39	-0.21	-0.096	0.42
exang	0.097	0.14	-0.39	0.068	0.067	0.026	-0.071	-0.38	1	0.29	-0.26	0.12	0.21	-0.44
oldpeak	0.21	0.096	-0.15	0.19	0.054	0.0057	-0.059	-0.34	0.29	1	-0.58	0.22	0.21	-0.43
slope	-0.17	-0.031	0.12	-0.12	-0.004	-0.06	0.093	0.39	-0.26	-0.58	1	-0.08	-0.1	0.35
ca	0.28	0.12	-0.18	0.1	0.071	0.14	-0.072	-0.21	0.12	0.22	-0.08	1	0.15	-0.39
thal	0.068	0.21	-0.16	0.062	0.099	-0.032	-0.012	-0.096	0.21	0.21	-0.1	0.15	1	-0.34
target	-0.23	-0.28	0.43	-0.14	-0.085	-0.028	0.14	0.42	-0.44	-0.43	0.35	-0.39	-0.34	1

Figure: Correlation matrix

3.2.3 Pre-processing of Data

It is an important step in the creation of the machine learning model. Initially the data may not be clean or in required format for the model. In pre-processing of data, we convert the data into our required format. It is used to deal with the noises, duplicates, and missing values of the data-set. It has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

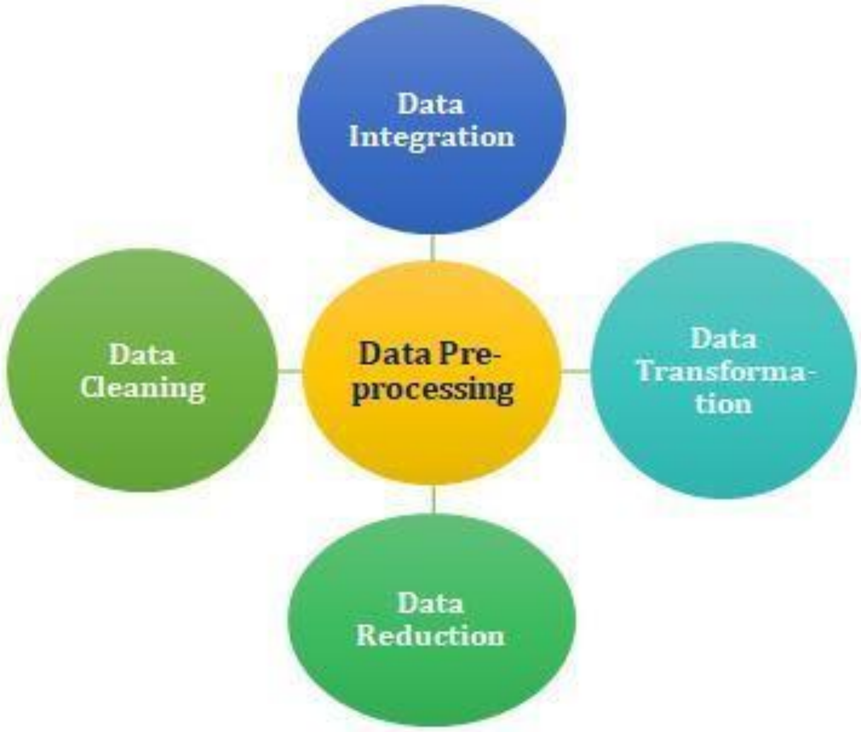


Figure: Data Pre-processing

3.2.3 Balancing of Data

Unbalanced datasets can be balanced in two ways. They are Under Sampling and Over Sampling:

(a) Under Sampling:

In Under Sampling, the dataset balance is done by the reducing the size of the ample class. This process is considered when the amount of data is adequate.

(b) Over Sampling:

In this, the dataset balance is done by increasing the size of the samples. It is considered when the amount of data is inadequate.

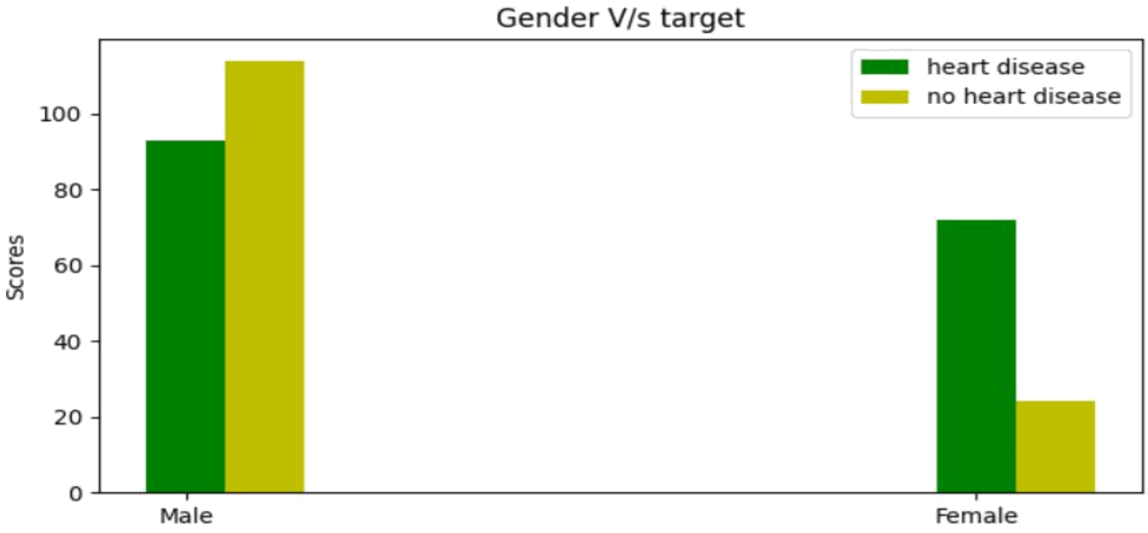


Figure: Data Balancing

3.2.4 Prediction of Disease

Various machine learning algorithms which are as follows SVM, Naive Bayes, Decision Tree, Random Forest, Logistic Regression and Ada-boost are used for prediction. Comparative analysis is performed among the algorithms and the one that gives the highest accuracy is used for heart disease prediction.

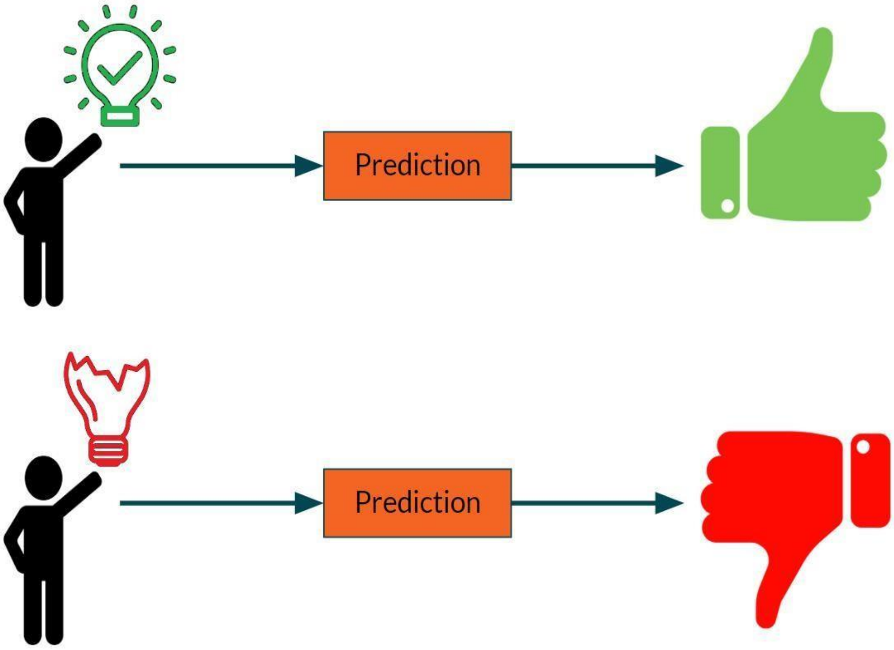


Figure: Prediction of Disease

WORKING OF SYSTEM

SYSTEM ARCHITECTURE

The system architecture gives an overview of the working system.

Dataset collection which contains patient details. Attributes selection process selects the useful attributes for the prediction of the disease. After identifying the available datasets, they are further selected, cleaned, made into the desired format. Different classification techniques as stated will be applied on the preprocessed data to predict the accuracy of disease. Accuracy measure compares the accuracy of different algorithms.

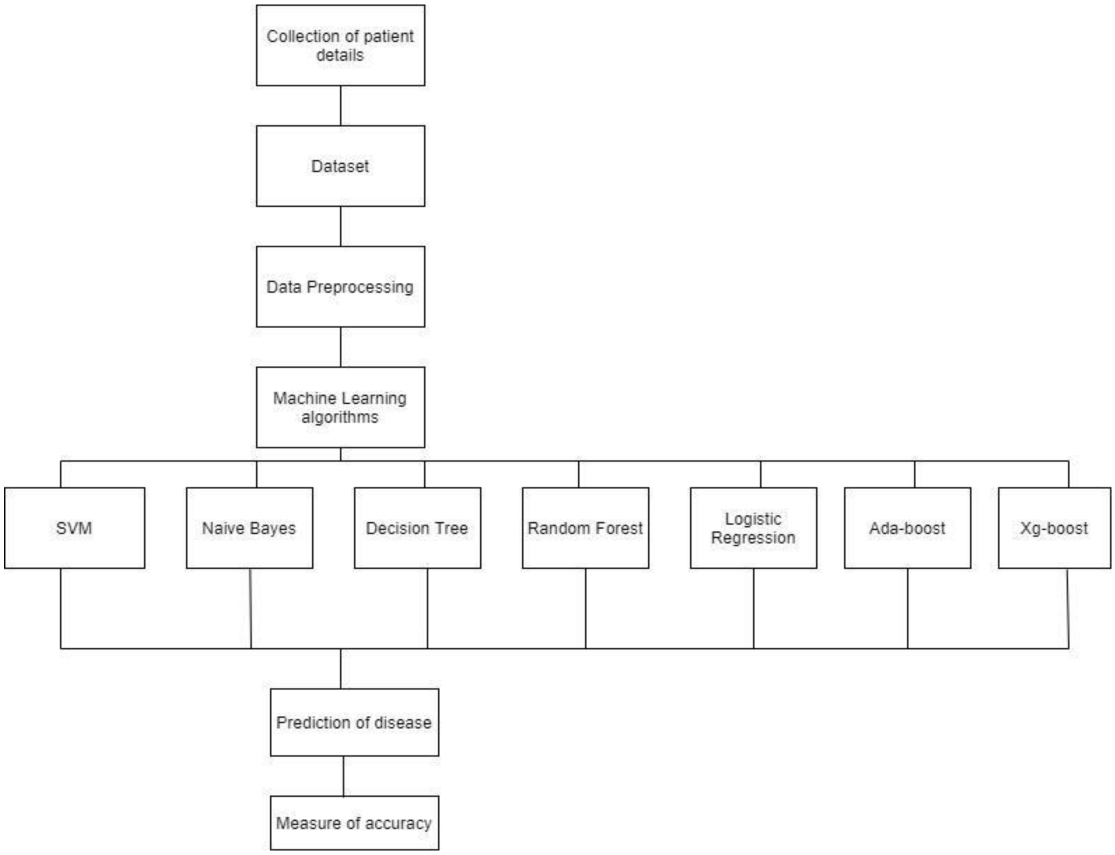


Figure: **System Architechture**

EXPERIMENTAL ANALYSIS

SYSTEM CONFIGURATION

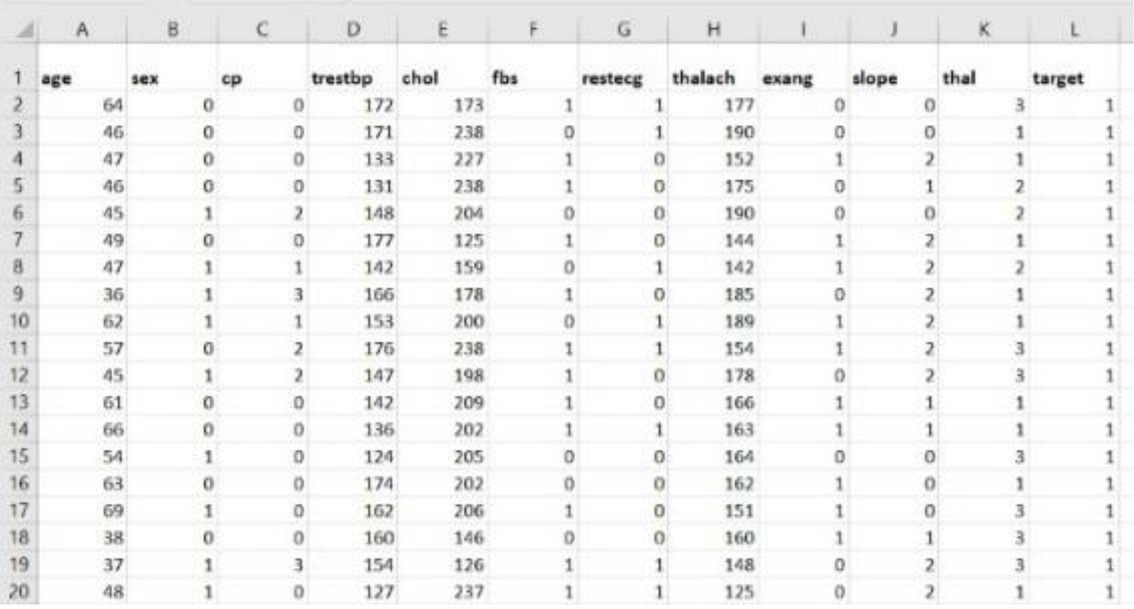
5.1.1 Hardware requirements:

Processer	:	Any Update Processer
Ram	:	Min 4GB
Hard Disk	:	Min 100GB

5.1.2 Software requirements:

Operating System	:	Windows family
Technology	:	Python3.7
IDE	:	Jupiter notebook

DATASET DETAILS



	A	B	C	D	E	F	G	H	I	J	K	L
1	age	sex	cp	trestbp	chol	fbs	restecg	thalach	exang	slope	thal	target
2	64	0	0	172	173	1	1	177	0	0	3	1
3	46	0	0	171	238	0	1	190	0	0	1	1
4	47	0	0	133	227	1	0	152	1	2	1	1
5	46	0	0	131	238	1	0	175	0	1	2	1
6	45	1	2	148	204	0	0	190	0	0	2	1
7	49	0	0	177	125	1	0	144	1	2	1	1
8	47	1	1	142	159	0	1	142	1	2	2	1
9	36	1	3	166	178	1	0	185	0	2	1	1
10	62	1	1	153	200	0	1	189	1	2	1	1
11	57	0	2	176	238	1	1	154	1	2	3	1
12	45	1	2	147	198	1	0	178	0	2	3	1
13	61	0	0	142	209	1	0	166	1	1	1	1
14	66	0	0	136	202	1	1	163	1	1	1	1
15	54	1	0	124	205	0	0	164	0	0	3	1
16	63	0	0	174	202	0	0	162	1	0	1	1
17	69	1	0	162	206	1	0	151	1	0	3	1
18	38	0	0	160	146	0	0	160	1	1	3	1
19	37	1	3	154	126	1	1	148	0	2	3	1
20	48	1	0	127	237	1	1	125	0	2	1	1

Input dataset attributes

● Gender (value 1: Male; value 0 : Female)

● Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value3: non-angina pain; value 4: asymptomatic)

● Fasting Blood Sugar (value 1: > 120 mg/dl; value 0:< 120 mg/dl)

● Exang – exercise induced angina (value 1: yes; value 0: no)

● CA – number of major vessels colored by fluoroscopy (value 0 – 3)

● Thal (value 3: normal; value 6: fixed defect; value 7:reversible defect)

● Trest Blood Pressure (mm Hg on admission to the hospital)

● Serum Cholesterol (mg/dl)

● Thalach – maximum heart rate achieved

● Age in Year

● Height in cms

● Weight in Kgs.

● Cholestrol

● Rest ecg

S. No.	Attribute	Description	Type
1	Age	Patient's age (29 to 77)	Numerical
2	Sex	Gender of patient (male-0 and female-1)	Nominal
3	Cp	Chest pain	Nominal
4	Trest bps	Resting blood pressure (in mm Hg on admission to hospital, values from 94 to 200)	Numerical
5	Chol	Serum cholesterol in mg/dl, value from 126 to 564)	Numerical
6	Fbs	Fasting blood sugar -120 mg/dl, true-1 false-0)	Nominal
7	Resting	Resting electro-cardio-graphics result (0 to 1)	Nominal
8	Thali	Maximum heart rate Achieved (71 to 202)	Numerical
9	Exang	Exercise angina	Nominal
10	Old-peak	ST depression introduced by exercise relative to rest (0 to .2)	Numerical
11	Slope	The slop of the peak exercise ST segment (0 to 1)	Nominal
12	Ca	Number of major vessels (0-3)	Numerical
13	Thal	3-normal	Nominal
14	Targets	1 or 0	Nominal

Table: Attributes of the dataset

PERFORMANCE ANALYSIS

In this project, various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Adaboost and XG-boost are used to predict heart disease. The dataset, has a total of 74 attributes, out of those only 14 attributes are considered for the prediction of heart disease. Various attributes of thepatient like gender, chest pain type, fasting blood pressure, serum cholesterol, etc are considered for this project. The accuracy for individual algorithms has to measure and whichever algorithm is giving the best accuracy, that is considered for the prediction. For evaluating the experiment, various evaluation metrics like accuracy, confusion matrix, precision, recall, and f1-score are considered.

Accuracy- It is the ratio of the number of correct predictions to the totalnumber of inputs
in the dataset.

Accuracy = (TP + TN) /(TP+FP+FN+TN)

Confusion Matrix- It gives us a matrix as output and gives the total performance of thesystem.

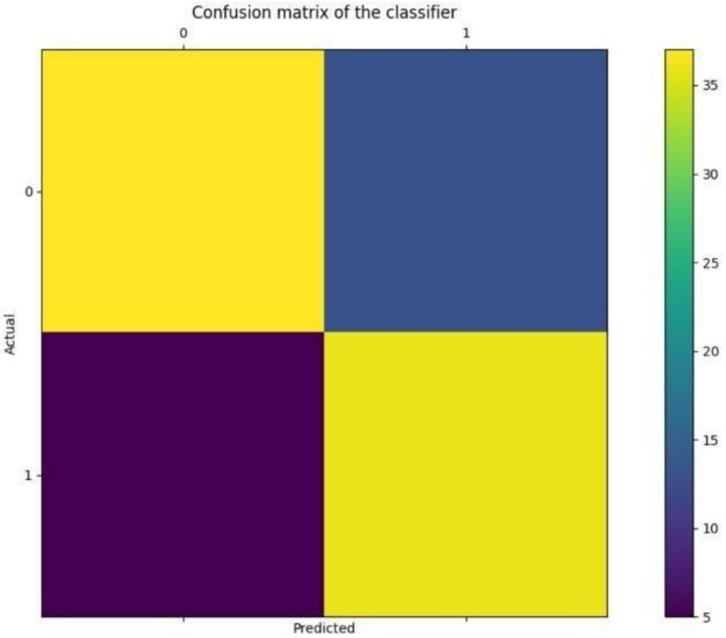


Figure: Confusion Matrix

Where :

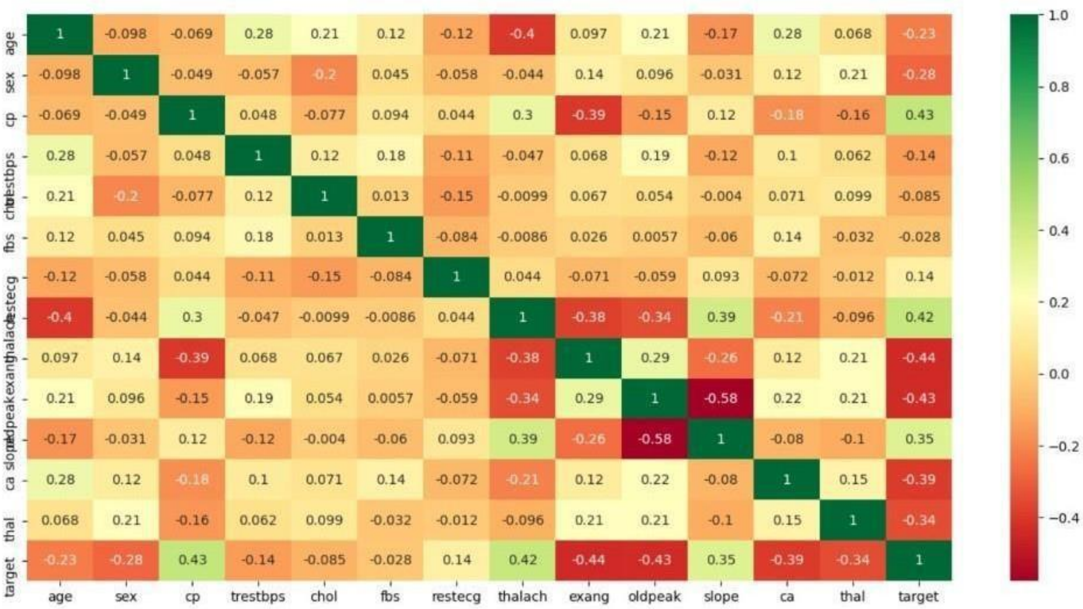
TP: True positive

FP: False Positive

FN: False Negative

TN: True Negative

Correlation Matrix: The correlation matrix in machine learning is used for feature selection. It represents dependency between various attributes.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1	-0.098	-0.069	0.28	0.21	0.12	-0.12	-0.4	0.097	0.21	-0.17	0.28	0.068	-0.23
sex	-0.098	1	-0.049	-0.057	-0.2	0.045	-0.058	-0.044	0.14	0.096	-0.031	0.12	0.21	-0.28
cp	-0.069	-0.049	1	0.048	-0.077	0.094	0.044	0.3	-0.39	-0.15	0.12	-0.18	-0.16	0.43
trestbps	0.28	-0.057	0.048	1	0.12	0.18	-0.11	-0.047	0.068	0.19	-0.12	0.1	0.062	-0.14
chol	0.21	-0.2	-0.077	0.12	1	0.013	-0.15	-0.0099	0.067	0.054	-0.004	0.071	0.099	-0.085
fbs	0.12	0.045	0.094	0.18	0.013	1	-0.084	-0.0086	0.026	0.0057	-0.06	0.14	-0.032	-0.028
restecg	-0.12	-0.058	0.044	-0.11	-0.15	-0.084	1	0.044	-0.071	-0.059	0.093	-0.072	-0.012	0.14
thalach	-0.4	-0.044	0.3	-0.047	-0.0099	-0.0086	0.044	1	-0.38	-0.34	0.39	-0.21	-0.096	0.42
exang	0.097	0.14	-0.39	0.068	0.067	0.026	-0.071	-0.38	1	0.29	-0.26	0.12	0.21	-0.44
oldpeak	0.21	0.096	-0.15	0.19	0.054	0.0057	-0.059	-0.34	0.29	1	-0.58	0.22	0.21	-0.43
slope	-0.17	-0.031	0.12	-0.12	-0.004	-0.06	0.093	0.39	-0.26	-0.58	1	-0.08	-0.1	0.35
ca	0.28	0.12	-0.18	0.1	0.071	0.14	-0.072	-0.21	0.12	0.22	-0.08	1	0.15	-0.39
thal	0.068	0.21	-0.16	0.062	0.099	-0.032	-0.012	-0.096	0.21	0.21	-0.1	0.15	1	-0.34
target	-0.23	-0.28	0.43	-0.14	-0.085	-0.028	0.14	0.42	-0.44	-0.43	0.35	-0.39	-0.34	1

Fig: Correlation matrix

Precision- It is the ratio of correct positive results to the total number of positiveresults predicted by the system.

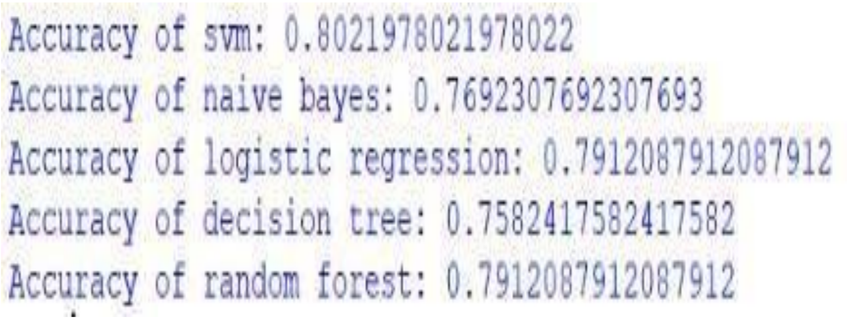
It is expressed as:

Recall- It is the ratio of correct positive results to the total number of positive resultspredicted by the system.

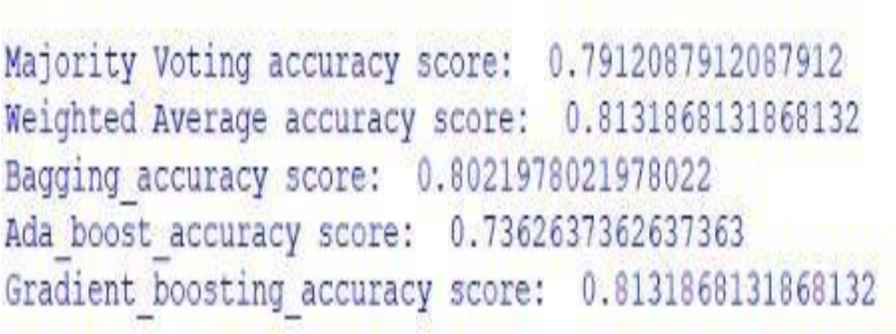
F1 Score- It is the harmonic mean of Precision and Recall. It measures the test accuracy. The range of this metric is 0 to 1.

PERFORMANCE MEASURES

The highest accuracy is given by Random Forest.



```
Accuracy of svm: 0.8021978021978022
Accuracy of naive bayes: 0.7692307692307693
Accuracy of logistic regression: 0.7912087912087912
Accuracy of decision tree: 0.7582417582417582
Accuracy of random forest: 0.7912087912087912
```



```
Majority Voting accuracy score:  0.7912087912087912
Weighted Average accuracy score:  0.8131868131868132
Bagging_accuracy score:  0.8021978021978022
Ada_boost_accuracy score:  0.7362637362637363
Gradient_boosting_accuracy score:  0.8131868131868132
```

Chapter 4

Experimental Results

After performing the best approach for training and testing data we find that accuracy of the Random forest is better compared to other algorithms. It is calculated with the support of the correlation matrix of each algorithm, here the number count of TP, TN, FP, FN is given and using the equation of accuracy, value has been calculated and it is concluded that extreme gradient boosting is best with 87.5% accuracy and the comparison is shown below.

Algorithm	**It’s Accuracy**
Random Forest	87.3%
SVM	80.2%
Logistic Regression	79.1%
XG-boost	79.1%
Naive Bayes	76.9%
Decision Tree	75.8%
Adaboost	73.6%

Table 2: **Accuracy Table**

Chapter 5
Conclusions

5.1 Conclusion

Cardiovascular diseases are a major killer in India as well as throughout the world, the application of promising technology like machine learning algorithms to the initial prediction of heart diseases will have a good impact on the society. The early diagnosis of the heart disease can help in making decisions on lifestyle changes in high-risk patients and in return to reduce the complications, which can be a great milestone in the medical field. The number of people who are facing this, are on a raise each year. This prompts for its early diagnosis and good treatment. The utilization of suitable technology support in this regard can be proved to be highly beneficial to the medical fraternity and for the patients. In this paper, the Random Forest algorithm of machine learning is used to measure the performance on the dataset.

5. 2 Future Works

The expected attributes leading to cardiovascular disease in patients are available in the dataset which contains more than 74 features and 14 most important features which are useful to evaluate the disease are selected. If all the features taken into consideration, then the efficiency of the algorithm the system gets is less. To increase efficiency, attribute selection is done. In this n features have to be selected for evaluating the pre-defined model which gives more accuracy. The correlation of some features in the dataset is almost equal and due to which they are removed. If all the attributes present in thedata-set are taken into account then the efficiency of the model decreases.

All the seven machine learning algorithm accuracies are compared on the basis of which, one prediction model is generated. Hence, the aim of the project is to use various evaluationmetrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. Comparing all seven the Random Forest gives the highest accuracy of 87.5%

References

Soni J, Ansari U, Sharma D & Soni S. Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of ComputerScience.

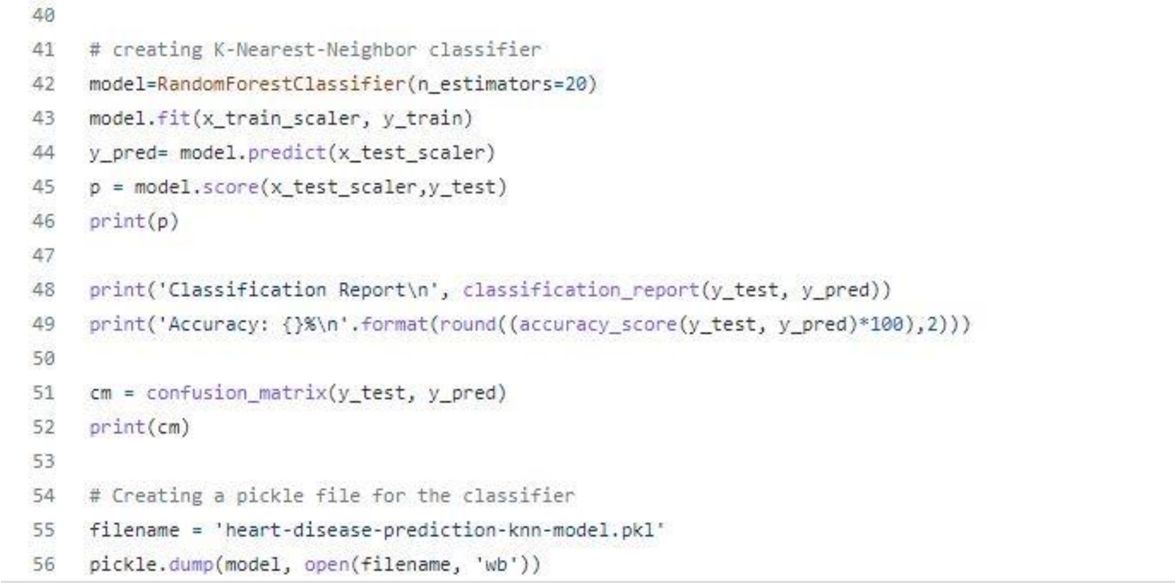
Dangare C S & Apte S S, Improved study of heart disease prediction system by using data mining classification techniques as mentioned in the International Journal of ComputerApplications.

Ordonez C in 2006. Association rule discovery with the train and test approachfor heart disease prediction as mentioned in the IEEE Transactions on Information Technology in Biomedicine.

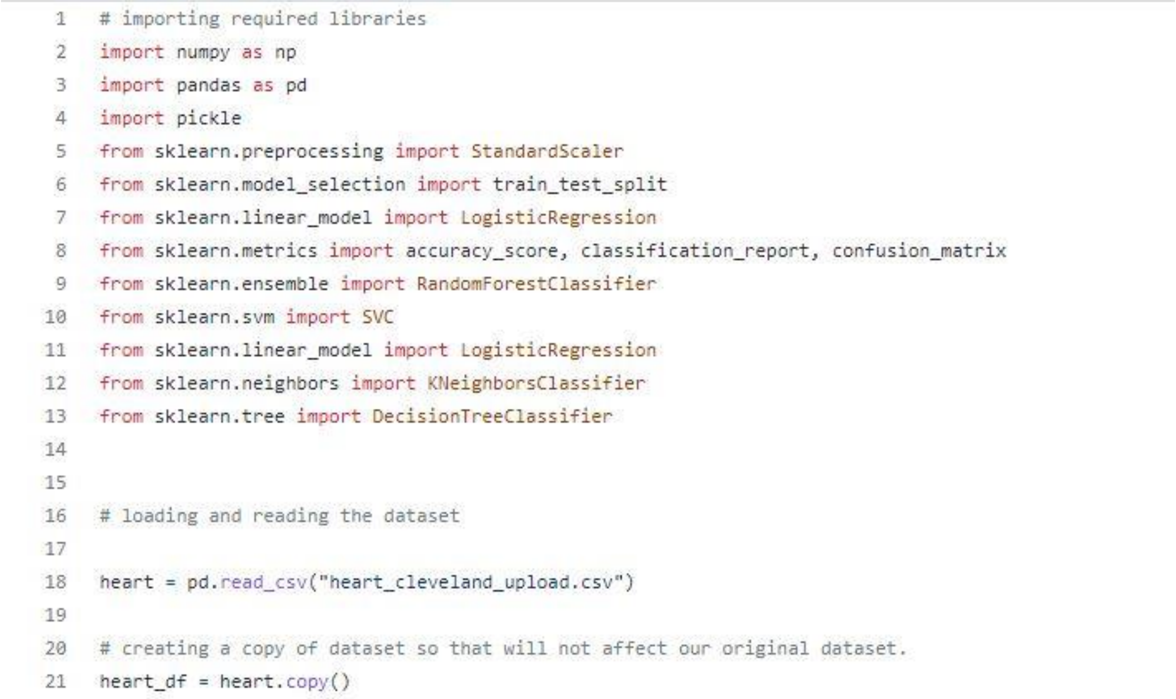
Shinde R, Arjun S, Patil P & Waghmare J in 2015. An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. InternationalJournal of Computer Science and Information Technology.

Bashir S, Qamar U & Javed M Y in November 2014. An ensemble-based decisionsupport framework for the intelligent heart disease diagnose as mentioned in the International Conferenceon Information Society.

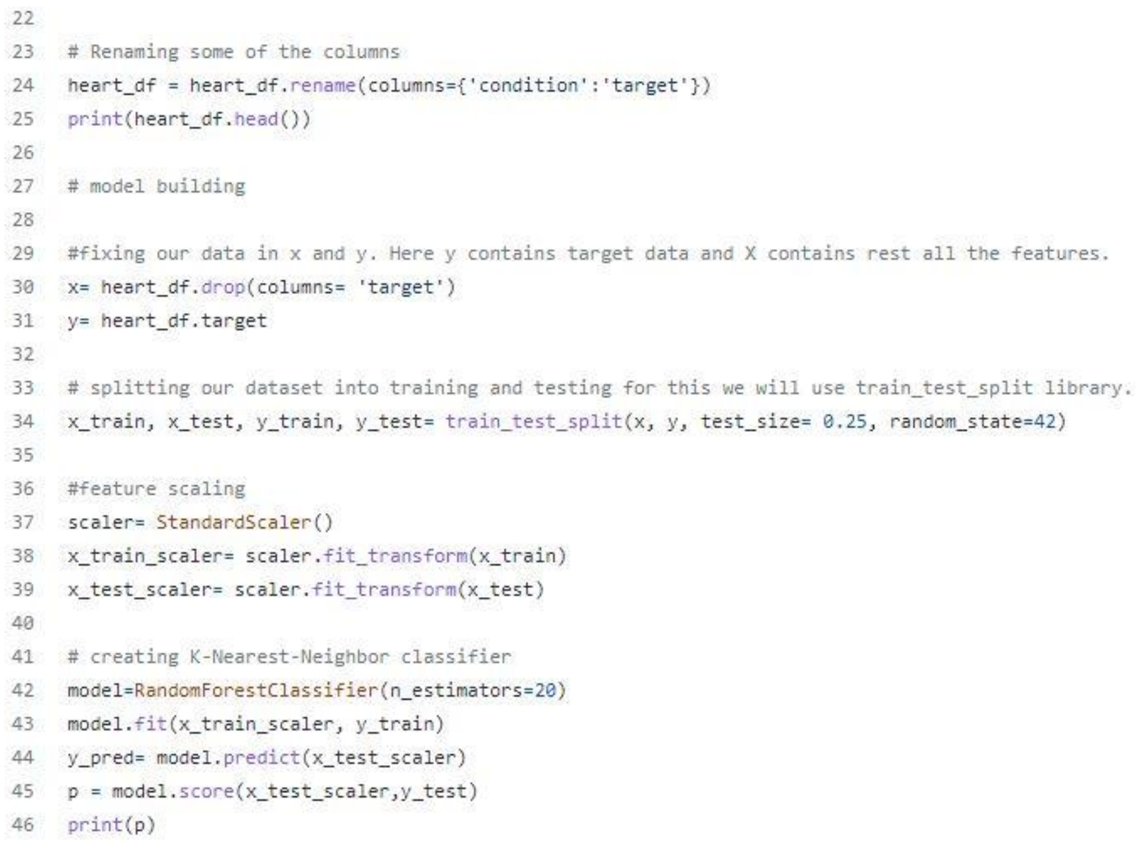
Annexure



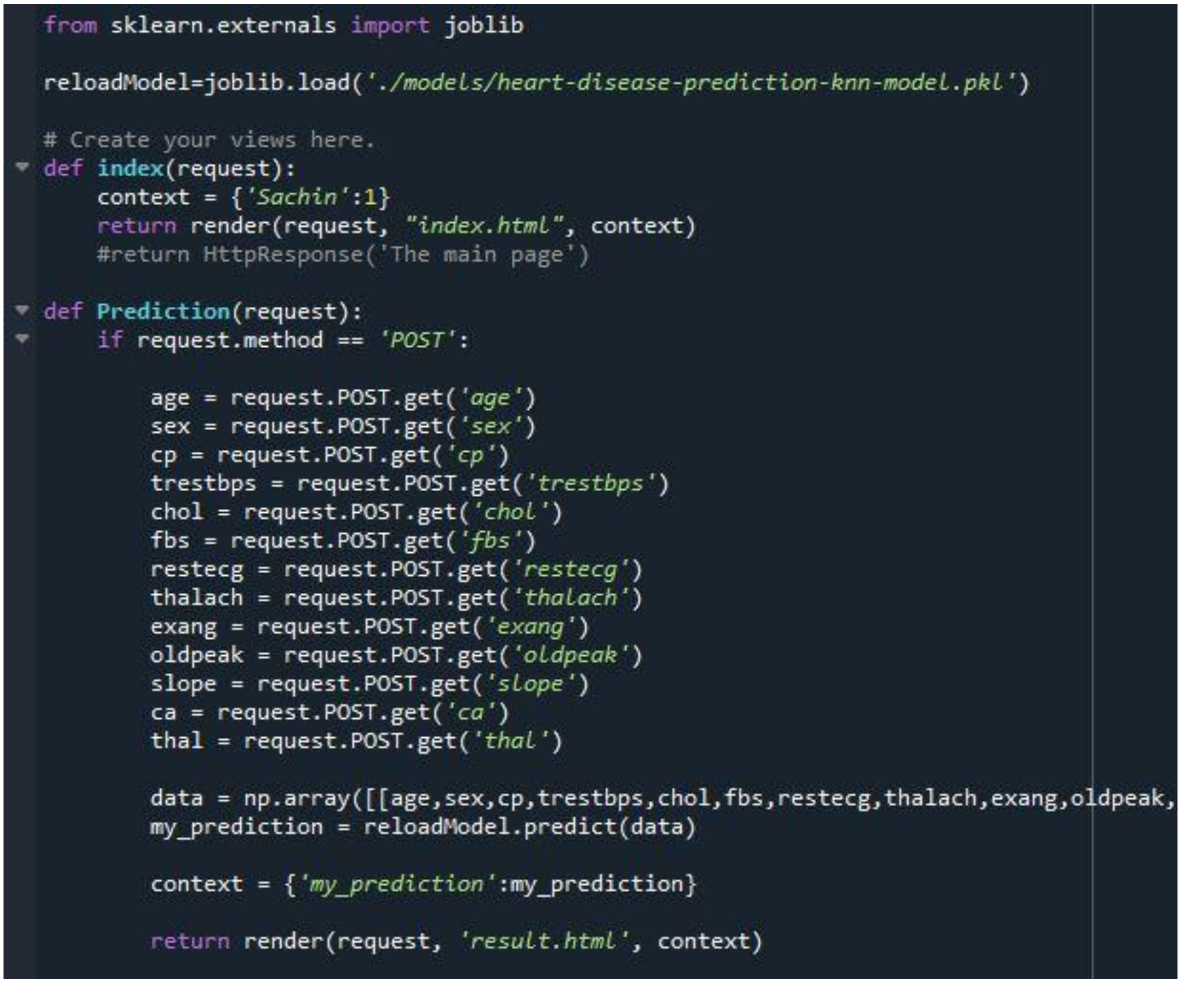
```

# creating K-Nearest-Neighbor classifier
model=RandomForestClassifier(n_estimators=20)
model.fit(x_train_scaler, y_train)
y_pred= model.predict(x_test_scaler)
p = model.score(x_test_scaler,y_test)
print(p)

print('Classification Report\n', classification_report(y_test, y_pred))
print('Accuracy: {}%\n'.format(round((accuracy_score(y_test, y_pred)*100),2)))

cm = confusion_matrix(y_test, y_pred)
print(cm)

# Creating a pickle file for the classifier
filename = 'heart-disease-prediction-knn-model.pkl'
pickle.dump(model, open(filename, 'wb'))
```



```
# importing required libraries
import numpy as np
import pandas as pd
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier


# loading and reading the dataset

heart = pd.read_csv("heart_cleveland_upload.csv")

# creating a copy of dataset so that will not affect our original dataset.
heart_df = heart.copy()
```



```

# Renaming some of the columns
heart_df = heart_df.rename(columns={'condition':'target'})
print(heart_df.head())

# model building

#fixing our data in x and y. Here y contains target data and X contains rest all the features.
x= heart_df.drop(columns= 'target')
y= heart_df.target

# splitting our dataset into training and testing for this we will use train_test_split library.
x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25, random_state=42)

#feature scaling
scaler= StandardScaler()
x_train_scaler= scaler.fit_transform(x_train)
x_test_scaler= scaler.fit_transform(x_test)

# creating K-Nearest-Neighbor classifier
model=RandomForestClassifier(n_estimators=20)
model.fit(x_train_scaler, y_train)
y_pred= model.predict(x_test_scaler)
p = model.score(x_test_scaler,y_test)
print(p)
```



```
from sklearn.externals import joblib

reloadModel=joblib.load('./models/heart-disease-prediction-knn-model.pkl')

# Create your views here.
def index(request):
    context = {'Sachin':1}
    return render(request, "index.html", context)
    #return HttpResponse('The main page')

def Prediction(request):
    if request.method == 'POST':

        age = request.POST.get('age')
        sex = request.POST.get('sex')
        cp = request.POST.get('cp')
        trestbps = request.POST.get('trestbps')
        chol = request.POST.get('chol')
        fbs = request.POST.get('fbs')
        restecg = request.POST.get('restecg')
        thalach = request.POST.get('thalach')
        exang = request.POST.get('exang')
        oldpeak = request.POST.get('oldpeak')
        slope = request.POST.get('slope')
        ca = request.POST.get('ca')
        thal = request.POST.get('thal')

        data = np.array([[age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,
        my_prediction = reloadModel.predict(data)

        context = {'my_prediction':my_prediction}

        return render(request, 'result.html', context)
```