Cardio Vascular Disease Prediction System

Project report in partial fulfillment of the requirement for the award of the degree of

Bachelor of Technology

In

Computer Science Engineering

Submitted By

Anirban Das	University Roll No.12018009019385
Chandranil Bag	University Roll No.12018009019018
Pritam Kumar Roy	University Roll No.12018009019463
Richik Baidya	University Roll No.12018009019487
Samarth Thapa	University Roll No.12018009019629
Sourodeep Roy	University Roll No.12018009019168

Under the guidance of

Prof. Arunabha Tarafdar

&

Prof. Amartya Chakraborty

Department of Computer Science Engineering

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700160.

Cardio Vascular Disease Prediction System

Project report in partial fulfillment of the requirement for the award of the degree of

Bachelor of Technology

In

Computer Science Engineering

Submitted By

Anirban Das	University Roll No. 12018009019385
Chandranil Bag	University Roll No. 12018009019018
Pritam Kumar Roy	University Roll No. 12018009019463
Richik Baidya	University Roll No. 12018009019487
Samarth Thapa	University Roll No. 12018009019629
Sourodeep Roy	University Roll No. 12018009019168

Under the guidance of

Prof. Arunabha Tarafdar

&

Prof. Amartya Chakraborty

Department of Computer Science Engineering

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700160.

UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

UNIVERSITY
OF ENGINEERING & MANAGEMENT
'University Area', Plot No. III-B/5, Main Arterial Road, New Town, Action Area-III, Kolkata - 700 160, W.B., India
City Office : 'ASHRAM', GN-34/2, Salt Lake Electronics Complex, Kolkata - 700 091, W.B., India
(Established by Act XXV of 2014 of Govt. of West Bengal & recognised by UGC, Ministry of HRD, Govt. of India)

Ph. (Office) : 91 33 2357 7649
: 91 33 2357 2969
: 91 33 6888 8608
Admissions : 91 33 2357 2059
Fax : 91 33 2357 8302
E-mail : vc@uem.edu.in
Website : www.uem.edu.in

CERTIFICATE

This is to certify that the project titled Cardiovascular Disease Prediction System submitted by **Anirban Das (University Roll No. 12018009019385), Chandranil Bag (University Roll No. 12018009019018), Pritam Kumar Roy (University Roll No. 12018009019463), Richik Baidya (University Roll No. 12018009019487), Samarth Thapa (University Roll No. 12018009019629) and Sourodeep Roy (University Roll No. 12018009019168),** students of UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA, in partial fulfilment of requirement for the degree of Bachelor of Computer Science Engineering, is a bona fide work carried out by them under the supervision and guidance of Prof. Arunabha Tarafdar & Prof. Amartya Chakraborty during 8th Semester of academic session of 2021 - 2022. The content of this report has not been submitted to any other university or institute. I am glad to inform that the work is entirely original and its performance is found to be quite satisfactory.

_____

Prof. Arunabha Tarafdar

Assistant Professor

Department of Computer Science Engineering

UEM, Kolkata

_____

Prof. Amartya Chakraborty

Assistant Professor

Department of Computer Science Engineering

UEM, Kolkata

_____

Prof. Sukalyan Goswami

HOD, Department of Computer Science Engineering

UEM, Kolkata

ACKNOWLEDGEMENT

We would like to take this opportunity to thank everyone whose cooperation and encouragement throughout the ongoing course of this project remains invaluable to us.

We are sincerely grateful to our guide Prof. Arunabha Tarafdar and Prof. Amartya Chakraborty of the Department of Computer Science Engineering, UEM, Kolkata, for his wisdom, guidance and inspiration that helped us to go through with this project and take it to where it stands now.

We would also like to express our sincere gratitude to Prof. Sukalyan Goswami, HOD, Department of Computer Science Engineering, UEM, Kolkata and all other departmental faculties for their ever-present assistant and encouragement.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

Anirban Das

Chandranil Bag

Pritam Kumar Roy

Richik Baidya

Samarth Thapa

Sourodeep Roy

Cardio Vascular Disease Prediction System

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

Department of Computer Science Engineering

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

Abstract

Each year the number of deaths caused by Cardio Vascular diseases or CVDs are booming at a very rapid rate and hence it is very crucial to be able to predict any high-risk complications beforehand. The World Health Organization (WHO) estimated that around 17.9 million people died due to heart diseases in 2016, which then used to account for 31% of the deaths worldwide. Diagnosis is one of the most critical tasks since it requires a lot of precision and efficiency in order to determine the risk status of the patient i.e., to predict whether the individual is most likely to contract any kind of cardiovascular diseases. However, Machine Learning has expedited new prospects in increasing the effectiveness of analyzing, determining and predicting the possibilities based upon the enormous amount of data that has been gathered up and collected by the healthcare industry for research purposes.

The aim of this project is to establish a Machine Learning model that will allow us to accurately and efficiently predict the possibility of an individual who is most likely to have a Cardio Vascular Disease or heart diseases in the future with the help of five different Machine Learning classification algorithms. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using SVM and Naives Bayes which showed a good accuracy in comparison to the previously used classifier such as KNN, Logistic Regression etc. So, a quiet significant amount of pressure has been lifted off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The given heart disease prediction system enhances medical care and reduces the cost.

This project gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format.

Table Of Contents

Introduction

A class of disease that involves the heart or the blood vessels, are known as Cardio Vascular Diseases. Cardio Vascular Diseases are usually associated with the buildup of fatty deposits within the arteries (atherosclerosis) and increased risk of blood clots. CVDs are one of the most widespread diseases across the globe. It has been observed that deaths caused by CVDs have increased rapidly over the last few decades.

Identifying a heart disease is very difficult, since there are numerous factors that are needed to be taken into consideration, such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, etc. Most CVDs can be prevented by addressing and correcting behavioral habits such as smoking, unhealthy diet, physical inactivity, alcohol etc. According to the World Health Organization, it is important to detect CVDs as early as possible so that the figures of premature death amongst adults can be reduced, by initiating respective treatments for the patients.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future heart diseases by analysing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different 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 data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

The approach that our project takes, uses a technique in which multiple machine learning models are combined and the prediction outcome is on the basis of majority vote from all the models, which improves the overall accuracy and efficiency of the prediction system. Five classification algorithms namely **Logistic Regression**, **KNN**, **Naives Bayes**, **Support Vector Machine (SVM)** and **Random Forest** are used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the three algorithms are evaluated at numerous levels and types of evaluation strategies.

Problem Statement

The biggest challenge in heart disease is the detection of its presence. There are instruments available which can predict heart disease but either are very expensive or are inefficient to calculate chance of heart diseases in human. Early detection of cardiac vascular diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients daily at this scale accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

Literature Survey

1. Purushottam, et, al proposed a paper "Efficient Heart Disease Prediction System" using hill climbing and decision tree algorithms. They used Cleveland dataset and preprocessing of data is performed before using classification algorithms. The Knowledge Extraction is done based on Evolutionary Learning (KEEL), an open-source data mining tool that fills the missing values in the data set. A decision tree follows top-down order. For each actual node selected by hill-climbing algorithm a node is selected by a test at each level. The parameters and their values used are confidence. Its minimum confidence value is 0.25. The accuracy of the system is about 86.7%.

2. Diagnosis and prediction of heart disease and blood pressure along with other attributes using the aid of neural networks was introduced by R. Subramanian et al**.** A deep Neural Network was Built incorporating the given attributes related to the disease which were able to produce a output which was carried out by the output perceptron and almost included 120 hidden layers which is the basic and most relevant technique of ensuring a accurate result of having heart disease if we use the model for Test Dataset. The supervised network has been advised for diagnosis of heart diseases. When the testing of the model was done by a doctor using an unfamiliar data, the model used and trained from the previous learned data and predicted the result thereby calculating the accuracy of the given model.

3. On the other hand, Praveen Kumar Reddy, 2019, Try to reduce the occurrences of heart disease using decision tree algorithm. In this, Support Vector Machine algorithm classifies the data values by using hyper plane and decision tree is implemented by Gini index method in which highest gain of the attributes gives a better representation of decision tree algorithm.

4. Abhay Kishore, et, al proposed "Heart Attack Prediction Using Deep Learning" in which heart attack prediction system by using Deep learning techniques and to predict the probable aspects of heart related infections of the patient Recurrent Neural System is used. This model uses deep learning and data mining to give the best precise model and least blunders. This paper acts as strong reference model for another type of heart attack prediction models.

Methodology

• Existing System

Heart disease has been acknowledged as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its uncertain consequences, hence doctors and researchers around the world are focusing on creating a system which will be able to give accurate and efficient predictions. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different 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 data 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 to predict the possibility of heart disease.

• Proposed Solution

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre-processed 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 obtained by testing the system using the testing data.
This system is implemented using the following modules –

1. Dataset collection
2. Selection of attributes
3. Data pre-processing
4. Application of Algorithms

• Dataset Collection

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system. This dataset gives us the much-needed information i.e., the medical attributes such as age, resting blood pressure, fasting sugar level etc. of the patient that helps us in detecting the patient that is diagnosed with any heart disease or not.

- ## Dataset Attributes

S.No	Attribute Name	Description	Range of Values
1	Age	Age of the person in years	29 to 79
2	Sex	Gender of the person [1: Male, 0: Female]	0, 1
3	Cp	Chest pain type [1-Typical Type 1 Angina 2- Atypical Type Angina 3-Non-angina pain 4-Asymptomatic)	1, 2, 3, 4
4	Trestbps	Resting Blood Pressure in mm Hg	94 to 200
5	Chol	Serum cholesterol in mg/dl	126 to 564
6	Fbs	Fasting Blood Sugar in mg/dl	0, 1
7	Restecg	Resting Electrocardiographic Results	0, 1, 2
8	Thalach	Maximum Heart Rate Achieved	71 to 202
9	Exang	Exercise Induced Angina	0, 1
10	OldPeak	ST depression induced by exercise relative to rest	1 to 3
11	Slope	Slope of the Peak Exercise ST segment	1, 2, 3
12	Ca	Number of major vessels colored by fluoroscopy	0 to 3
13	Thal	3 – Normal, 6 – Fixed Defect, 7 – Reversible Defect	3, 6, 7
14	Num	Class Attribute	0 or 1

- ## Selection of Attributes

Attribute or Feature selection includes 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, exang, etc. are selected for the prediction. The Correlation matrix is used for attribute selection for this model.

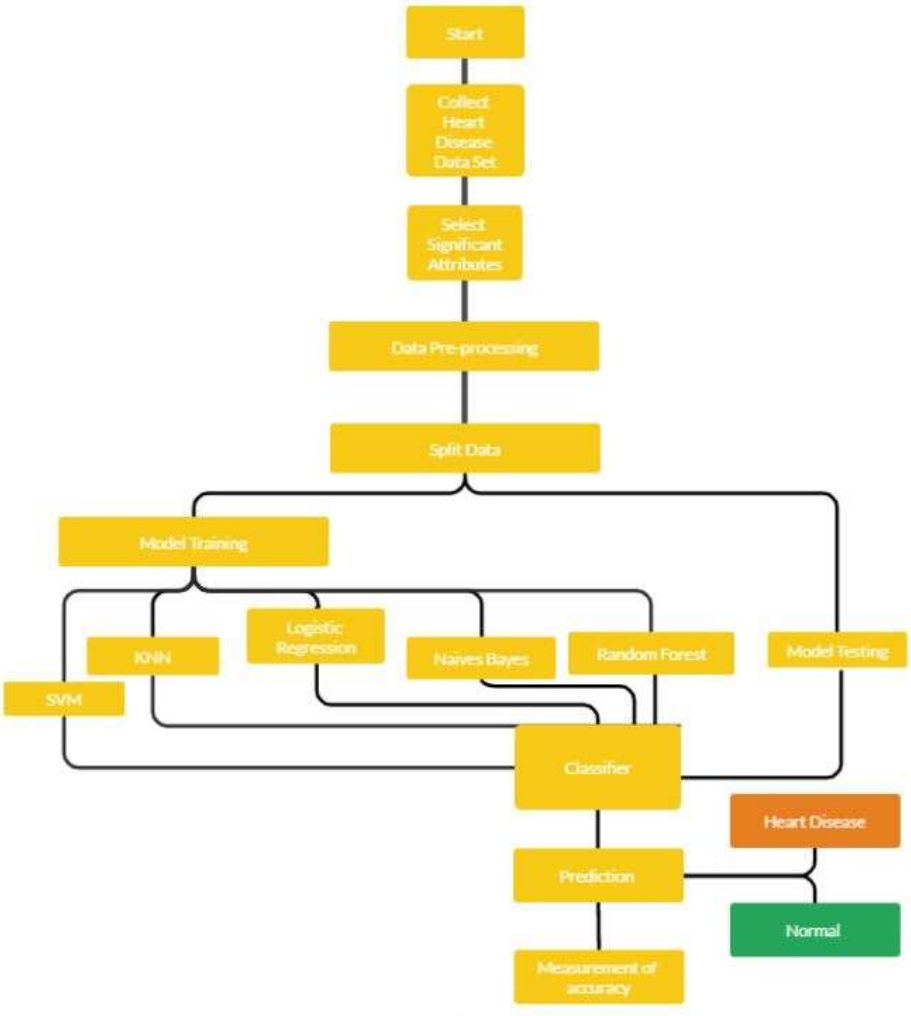
- ## Data Pre-processing

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

- ## Application of Algorithms

Various machine learning algorithms like **SVM, Naive Bayes, Random Forest, KNN, Logistic Regression,** are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction.

System Architecture



Machine Learning Algorithms

1. Logistic Regression

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. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

2. Naives Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

3. Support Vector Machine (Svm)

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine.

4. K-Nearest Neighbour

The second model that was built is the K-Nearest Neighbor classifier. The algorithm in this classifier involves finding the distances between the new instance and all of the training instances, then from a predefined K number it selects the nearest K data points to the new instance. Finally, classification occurs based on the majority class of the K data points selected. The K number in this project was chosen to be 7 since it produced the best results based on the GridsearchCV.

5. Random Forest Classifier

Random Forest is a supervised learning algorithm. It is an extension of machine learning classifiers which include the bagging to improve the performance of Decision Tree. It combines tree predictors, and trees are dependent on a random vector which is independently sampled. The distribution of all trees is the same. Random Forests splits nodes using the best among of a predictor subset that are randomly chosen from the node itself, instead of splitting nodes based on the variables. The time complexity of the worst case of learning with Random Forests is O(M(dnlogn)), where M is the number of growing trees, n is the number of instances, and d is the data dimension.

Experimental Setup

- Hardware Requirements

Processor	:	Minimum Intel i3 / AMD Ryzen 3
RAM	:	Min 4GB
Hard Disk	:	Min 100GB

- Software Requirements

Operating System:		Windows
Technology	:	Python
IDE	:	Jupyter Notebook

Code Base

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import os
print(os.listdir())

import warnings
warnings.filterwarnings('ignore')

data = pd.read_csv("C:/Users/user/Desktop/project/heart.csv")
data.head(6)
data.shape
data.info()
info = ["age","1: male, 0: female","chest pain type, 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic","resting blood pressure"," serum cholestoral in mg/dl","fasting blood sugar > 120 mg/dl","resting electrocardiographic results (values 0,1,2)"," maximum heart rate achieved","exercise induced angina","oldpeak = ST depression induced by exercise relative to rest","the slope of the peak exercise ST segment","number of major vessels (0-3) colored by flourosopy","thal: 3 = normal; 6 = fixed defect; 7 = reversable defect"]



for i in range(len(info)):
   print(data.columns[i]+":\t\t\t"+info[i])

plt.figure(figsize=(20,12))
sns.set_context('notebook',font_scale = 1.3)
sns.heatmap(data.corr(),annot=True,linewidth =2)
plt.tight_layout()

data["target"].describe()
data["target"].unique()
sns.set_context('notebook',font_scale = 1.3)
data.drop('target', axis=1).corrwith(data.target).plot(kind='bar', grid=True, figsize=(11, 7),
                                        title="Correlation with the target feature")
plt.tight_layout()

y = data["target"]

plt.figure(figsize=(10,5))
sns.set_context('notebook',font_scale = 1.5)
```

```
sns.countplot(data['target'])
plt.tight_layout()

target_temp = data.target.value_counts()

print(target_temp)



print("Percentage of patients without heart problems: "+str(round(target_temp[0]*100/303,2)))
print("Percentage of patients with heart problems: "+str(round(target_temp[1]*100/303,2)))

data["sex"].unique()

sns.barplot(data['sex'],y)

data["cp"].unique()

plt.figure(figsize=(15,7))
sns.set_context('notebook',font_scale = 1.3)
sns.barplot(data['cp'],y)
plt.tight_layout()

sns.barplot(data["fbs"],y)

sns.barplot(data["restecg"],y)

sns.barplot(data["exang"],y)

sns.barplot(data["slope"],y)

data["ca"].unique()

sns.countplot(data["ca"])

sns.barplot(data["thal"],y)

sns.distplot(data["thal"])

categorical_val = []
continous_val = []
for column in data.columns:
        print("--------------------")
        print(f"{column} : {data[column].unique()}")
        if len(data[column].unique()) <= 10:
                categorical_val.append(column)
        else:
                continous_val.append(column)
```

```
categorical_val.remove('target')
dfs = pd.get_dummies(data, columns = categorical_val)
dfs.head(6)

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
col_to_scale = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
dfs[col_to_scale] = sc.fit_transform(dfs[col_to_scale])
dfs.head(6)

from sklearn.model_selection import train_test_split
predictors = data.drop("target",axis=1)
target = data["target"]
X_train,X_test,Y_train,Y_test = train_test_split(predictors,target,test_size=0.20,random_state=0)



X_train.shape
X_test.shape
Y_train.shape
Y_test.shape

from sklearn.metrics import accuracy_score

from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(X_train,Y_train)
Y_pred_lr = lr.predict(X_test)
score_lr = round(accuracy_score(Y_pred_lr,Y_test)*100,2)
print("The accuracy score achieved using Logistic Regression is: "+str(score_lr)+" %")

from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(X_train,Y_train)
Y_pred_nb = nb.predict(X_test)
score_nb = round(accuracy_score(Y_pred_nb,Y_test)*100,2)
print("The accuracy score achieved using Naive Bayes is: "+str(score_nb)+" %")


from sklearn import svm
sv = svm.SVC(kernel='linear')
sv.fit(X_train, Y_train)
Y_pred_svm = sv.predict(X_test)
score_svm = round(accuracy_score(Y_pred_svm,Y_test)*100,2)
print("The accuracy score achieved using Linear SVM is: "+str(score_svm)+" %")

from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=7)
knn.fit(X_train,Y_train)
```

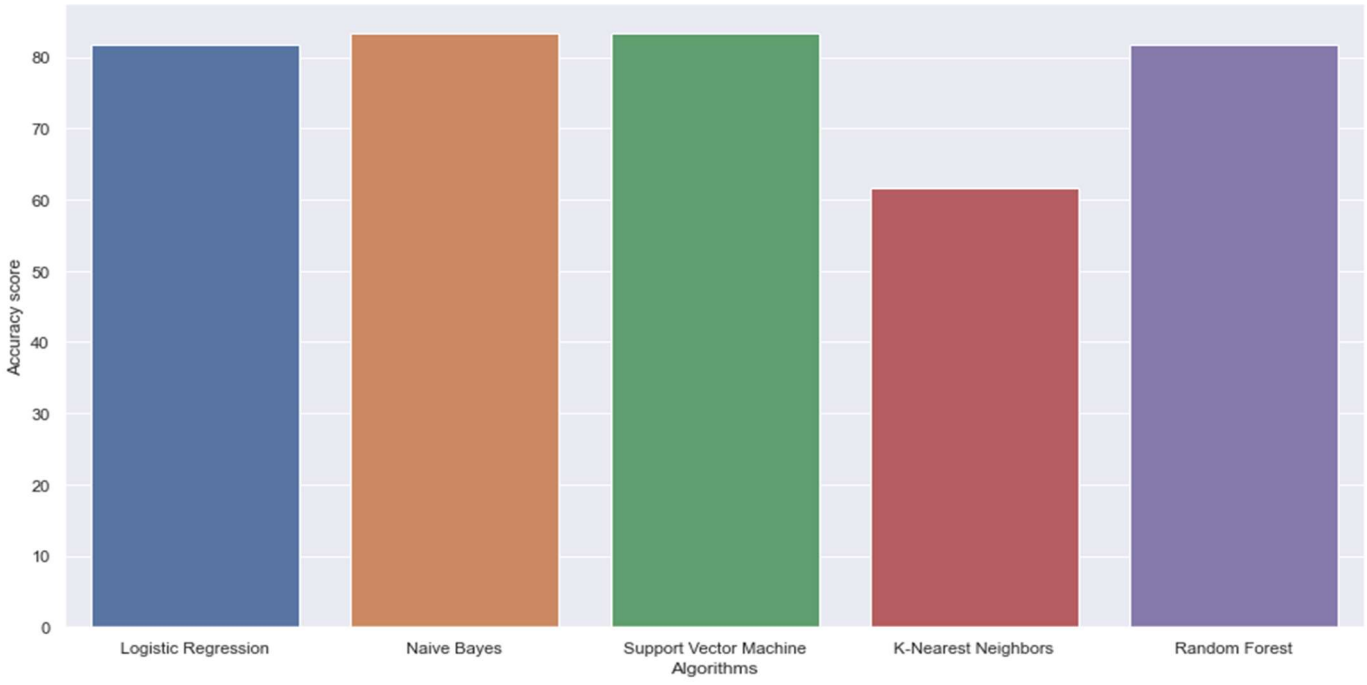
```
 Y_pred_knn=knn.predict(X_test)
score_knn = round(accuracy_score(Y_pred_knn,Y_test)*100,2)
print("The accuracy score achieved using KNN is: "+str(score_knn)+" %")

from sklearn.ensemble import RandomForestClassifier
max_accuracy = 0
for x in range(2000):
        rf = RandomForestClassifier(random_state=x)
        rf.fit(X_train,Y_train)
        Y_pred_rf = rf.predict(X_test)
        current_accuracy=round(accuracy_score(Y_pred_rf,Y_test)*100,2)
        if(current_accuracy>max_accuracy):
                max_accuracy = current_accuracy
                best_x = x

rf = RandomForestClassifier(random_state=best_x)
 rf.fit(X_train,Y_train)
 Y_pred_rf = rf.predict(X_test)

score_rf = round(accuracy_score(Y_pred_rf,Y_test)*100,2)
print("The accuracy score achieved using Random Forest is: "+str(score_rf)+" %")



scores = [score_lr,score_nb,score_svm,score_knn,score_rf]
algorithms = ["Logistic Regression","Naïve Bayes","Support Vector Machine","K-Nearest Neighbors","Random Forest”]
for i in range(len(algorithms)):
        print("The accuracy score achieved using "+algorithms[i]+" is: "+str(scores[i])+" %")

sns.set(rc={'figure.figsize':(15,8)})
plt.xlabel("Algorithms")
plt.ylabel("Accuracy score")
sns.barplot(algorithms,scores)
```

Result Analysis

The highest accuracy is given by SVM and Naives Bayes.

```
The accuracy score achieved using Logistic Regression is: 81.67 %
The accuracy score achieved using Naive Bayes is: 83.33 %
The accuracy score achieved using Support Vector Machine is: 83.33 %
The accuracy score achieved using K-Nearest Neighbors is: 61.67 %
The accuracy score achieved using Random Forest is: 81.67 %
```

After performing the machine learning approach for training and testing we find that accuracy of the SVM and Naives Bayes is better compared to other algorithms. Accuracy is calculated with the support of the confusion 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 Support Vector Machine and Naives Bayes is the best with 83.33% accuracy and the comparison is shown below.



Conclusion and Future Scope

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this project, the five different machine learning algorithms used to measure the performance are SVM, Random Forest, Naïve Bayes, Logistic Regression and K-Nearest Neighbors, which have been implemented on the dataset.

The expected attributes leading to heart disease in patients are available in the dataset which contains 76 features and 14 important features that are useful to evaluate the system are selected among them. If all the features taken into the consideration, then the efficiency of the system the author gets is less. To increase efficiency, attribute selection is done. In this n features have to be selected for evaluating the model which gives more accuracy. The correlation of some features in the dataset is almost equal and so they are removed. If all the attributes present in the dataset are taken into account, then the efficiency decreases considerably.

All the five machine learning methods accuracies are compared based on which one prediction model is generated. Hence, the aim is to use various evaluation metrics like correlation matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. Comparing all five the **Support Vector Machine** and **Naives Bayes** classifier gives the highest accuracy of **83.33**%.

Bibliography

[1] Soni J, Ansari U, Sharma D & Soni S (2011). *Predictive data mining for medical diagnosis: an overview of heart disease prediction.* International Journal of Computer Applications,***17*(8), 43-8**

[2] Dangare C S & Apte S S (2012). *Improved study of heart disease prediction system using data mining classification techniques*. *International Journal of Computer Applications*,***47*(10), 44-8**.

[3] Ordonez C (2006). *Association rule discovery with the train and test approach for heart disease prediction*. *IEEE Transactions on Information Technology in Biomedicine*, *10*(2), **334-43**.

[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). *An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm*. *International Journal of Computer Science and Information Technologies*, ***6*(1)**, 637-9.

[5] Bashir S, Qamar U & Javed M Y (2014, November). *An ensemble-based decision support framework for intelligent heart disease diagnosis*. In *International Conference on Information Society (i-Society 2014)* (pp. 259-64). **IEEE**.

[6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). *A coronary heart disease prediction model: the Korean Heart Study*. *BMJ open*, ***4*(5)**, e005025.

[7] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). *Multilocus genetic risk scores for coronary heart disease prediction*. *Arteriosclerosis, thrombosis, and vascular biology*, ***33*(9)**, 2267-72.

[8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). *Heart disease prediction using lazy associative classification*. In *2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s)* (pp. 40- 6). **IEEE**.

[9] Dangare Chaitrali S and Sulabha S Apte. *"Improved study of heart disease prediction System using data mining classification techniques." International Journal of Computer Applications* 47.10 (2012): 44-8.

[10] Soni Jyoti. "Predictive data mining for medical diagnosis: *An overview of heart disease prediction*." *International Journal of Computer Applications* 17.8 (2011): 43-8.

[11] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). *HDPS: Heart disease prediction system*. In *2011 Computing in Cardiology* (pp. 557-60). **IEEE**.

[12] Parthiban, Latha and R Subramanian. *"Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences* 3.3 (2008).

[13] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). *Wireless body area network for heart attack detection [Education Corner]*. *IEEE antennas and propagation magazine*, ***58*(5)**, 84-92.

[14] Patel S & Chauhan Y (2014). *Heart attack detection and medical attention using motion sensing device -kinect*. *International Journal of Scientific and Research Publications*, ***4*(1)**, 1-4.

[15] Zhang Y, Fogoros R, Thompson J, Kenknight B H, Pederson M J, Patangay A & Mazar S T (2011). *U.S. Patent No. 8,014,863*. Washington, DC: U.S. Patent and Trademark Office.

[16] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016,December). *Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design*. In *2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 299-303). **IEEE**.

[17] Buechler K F & McPherson P H (1999). *U.S. Patent No. 5,947,124*. Washington, DC: U.S. Patent and Trademark Office.

[18] Takci H (2018). *Improvement of heart attack prediction by the feature selection methods*. *Turkish Journal of Electrical Engineering & Computer Sciences*,***26*(1)**, 1-10.

[19] Worthen W J, Evans S M, Winter S C & Balding D (2002). *U.S. Patent No. 6,432, 124*. Washington, DC: U.S. Patent and Trademark Office.