Detection of Cardiovascular disease using different machine learning algorithms followed by PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis)

Farhan Sharukh Hasan	**170204066**
Rahat Kader Khan	**170204074**
Shweta Bhattacharjee Porna	**170204111**

Project Report

Course ID: CSE 4214

Course Name: Pattern Recognition Lab

Semester: Spring 2021

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

14 March 2022

Detection of Cardiovascular disease using different machine learning algorithms followed by PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis)

Submitted by

Farhan Sharukh Hasan	**170204066**
Rahat Kader Khan	**170204074**
Shweta Bhattacharjee Porna	**170204111**

Submitted To

Faisal Muhammad Shah, Associate Professor

Md. Tanvir Rouf Shawon, Lecturer

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

14 March 2022

ABSTRACT

In this era of artificial intelligence and machine learning, diagnosis of health-related disorders based on data and its analysis is gaining a lot of momentum. Nowadays, machine learning data analysis play a pivotal role in predicting health-related disorders of various body parts of the human body. Over the past decades, heart disease is a common and dangerous disease caused by fat suppression. This disease occurs due to over pressure in the human body. We can predict cardiac disease using a variety of parameters in the dataset.By analyzing various parameters or features of the patient data, machine learning algorithms predict whether the patient is at a risk thereby saving time and reducing expenses We observed a dataset cover of 12 parameters and 70000 different data values to evaluate patient performance. The major objective of this project is to try to obtain improved accuracy for detecting heart disease using ML algorithms in which the target output calculates whether the person has heart disease or not. In our study, we used machine learning algorithms, which were Logistic Regression, Support Vector Machine, K-Nearest neighbour, Decision Tree, Random Forest Naive Bayes in combination with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). We also used Chi-Squared Statistical for features selection.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are many types of heart diseases, such as Coronary Artery Disease, Heart Failure(CAD), Cardiovascular disease etc. Among these diseases, cardiovascular disease is a fatal heart disease. In this report, the prediction of cardiovascular disease of men and women, varies in different ages in terms of height, weight, blood pressure, level of cholesterol and glucose, habit of smoking and alcohol. Our main goal is to apply a machine learning approach to figure out the best performance in predicting cardiovascular disease from the proposed modelsWe are classifying our disease into two categories, either the person has cardiovascular disease or not. The classification will be of binary classification between male and female of all ages

Machine Learning plays a decisive role in predicting outcomes of these diseases. The core topic is prediction using machine learning techniques. It becomes necessary to ensure that our cardiovascular system or any other system in the human body for that matter should remain healthy. Unfortunately, people around the world are suffering from cardiovascular diseases. Any technology that can help detect these diseases before much damage is done will shows that it is helpful in saving people's money and further significantly their life. Data mining techniques can be beneficial in predicting heart diseases. Analytical models can be made by finding earlier unknown patterns and trends in databases and using the obtained information. Machine learning is a technology that can help to accomplish a diagnosis of heart disease before much damage occurs to a person. Our study was based on some machine learning algorithms which were Logistic Regression, Support Vector Machine, K-Nearest neighbour, Decision Tree, Random Forest Naive Bayes in combination with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). We also used Chi-Squared Statistical for features selection.

Chapter 2

Literature Reviews

We searched paper that worked with same dataset and found three papers. Short review on those papers are given below:

Advait Shirvaikar, Advait Mandlik, Prof. Sangeeta Prasanna Ram, in their article [1] they had a comparative study was based on five distinct machine learning algorithms, which were Logistic Regression, Support Vector Machine, K nearest neighbour, Random Forest Naive Bayes in combination with Principal Component Analysis (PCA) for data selection, to predict whether a patient is prone to cardiovascular disease, based on analysis of the parameters or features of the patient data from the same Kaggle dataset comprising of 70000 values and 11 features. Based on their study , they found that the Random Forest algorithm is superior to the other algorithms, based on their 'sensitivity' in identifying the disease.that provides best results for Binary classification is the Random Forest Classifier which achieves a training Accuracy of about 75% and the Testing Accuracy of about 73%. The Recall or 'Sensitivity', an important metric in Medical Analysis, of the Random Forest Classifier model is also quite high, in comparison to the other classifier models, with a training Recall score of 74% and 73% on the Test Dataset. The conclusion can be finally drawn that machine learning is able to predict and restrict the implications done to a person's heart.

Anupama Yadav,Levish Gediya, Adnanuddin Kazi, in their article [2] They studied the prediction of heart disease using machine learning algorithms and python programming. Over the past decades, heart disease is a common and dangerous disease caused by fat suppression. This disease occurs due to overpressure in the human body. they predicted cardiac disease using a variety of parameters in the dataset. they observed the same dataset as us which covers of 12 parameters and 70000 different data values to evaluate patient performance. The major objective of the author's paper is to obtain improved accuracy for detecting heart disease using algorithms in which the target output calculates whether

the person has heart disease or not.By applying KNN in huge datasets takes a long time to process. The accuracy grown with this algorithm is 63.4% . By using the random forest classifier, the accuracy predicted result is approximately 71% but actually it is 71.4% The accuracy obtained with Decision tree clasifier algorithm is 68.4% SVM is also one of the classification algorithms in machine learning in which improved accuracy can be predicted. As compare to other algorithms, it is much better for expectedly predicting accuracy.

In our prediction, the predicted highest accuracy is 72.6% using linear SVM kernel.In our prediction, the predicted highest accuracy is 86.2% using Guassian SVM karnel.

Kumar, Digvijay and Bavithra,, in their article [3] They represented the various models based on such algorithms and techniques to analyze their performance. Such as Logistic Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, and ensemble models which are Supervised Learning algorithms. Using various important features that are necessary for the prediction of CVDs (like a person is having CVDs or not).In Logistic regression, they got an accuracy of 71.91% which is quite better than other algorithms.In Naïve Bayes algorithm they got an accuracy of 70.70% is not good concerning other algorithms.In SVM they got an accuracy of 71.83% for linear kernel, for kernel = 'rbf', gamma = .75 and c = 4 they got 70.94% and for kernel = 'poly', degree = 5, C=5 they got accuracy.In this Random Forest Classifier model, they got 68.46% which is worst among all models they have implemented.In Stacking model, They have implemented KNN, Random forest,naïve Bayes, and Logistic regression as a final estimator. Thus they got 72.16% accuracy which is very good and second-highest for this dataset. In KNeighbors Classifier model, they got 72.28% which is best among all models I have implemented.

Chapter 3

Data Collection & Processing

3.1 Source

Data is collected from the dataset published by Svetlana Ulianova as in the title of Cardiovascular Disease dataset. [4] The dataset consists of 70 000 records of patients data in 13 features, such as age, gender, systolic blood pressure, diastolic blood pressure, etc. The target class "cardio" equals 1, when the patient has cardiovascular disease, and it's 0 if the patient is healthy.

1. Age (Objective Feature): age in int (days)
2. Height (Objective Feature): height in int (cm)
3. Weight (Objective Feature): weight in float (kg)
4. Gender (Objective Feature): gender in categorical code
   (a) 1: female
   (b) 2: male
5. Systolic blood pressure (Examination Feature): ap.hi in int
6. Diastolic blood pressure (Examination Feature): aplo in int
7. Cholesterol (Examination Feature): cholesterol
   (a) 1: normal
   (b) 2: above normal
   (c) 3: well above normal
8. Glucose (Examination Feature): gluc

   (a) 1: normal

   (b) 2: above normal

   (c) 3: well above normal

9. Smoking | Subjective Feature | smoke | binary |

   (a) 0: non-smoker

   (b) 1: smoker

10. Alcohol intake | Subjective Feature | alco in binary |

   (a) 0: non-alchoholic

   (b) 1: alchoholic

11. Physical activity | Subjective Feature | active | binary |

   (a) 0: inactive

   (b) 1: active

12. Presence or absence of cardiovascular disease | Target Variable | cardio | binary | active | binary |

   (a) 0: absent

   (b) 1: present

3.2 Dataset Cleaning

1. **Removing unnecessary column:** We removed some unnecessary column here.And dropped the id.

2. **Checking duplicate rows and removing duplicate rows:** We checked for duplicate rows and removed if any.We removed 24 duplicate rows.

3. **Checking null values for each column:** We checked for null values,we didn't have any null values.

4. **Checking dataset for negative values and Taking only positive values:**We took only positive values.We found 8 negative value and removed it.

3.3 Dataset Visualization

3.3.1 Age in the dataset

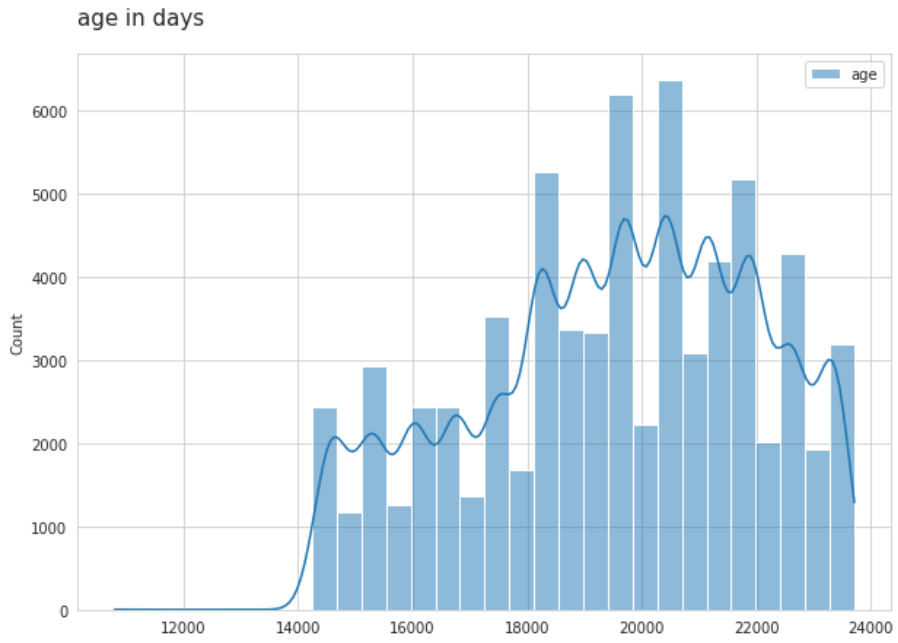

Figure 3.1: Age in the dataset

3.3.2 Gender

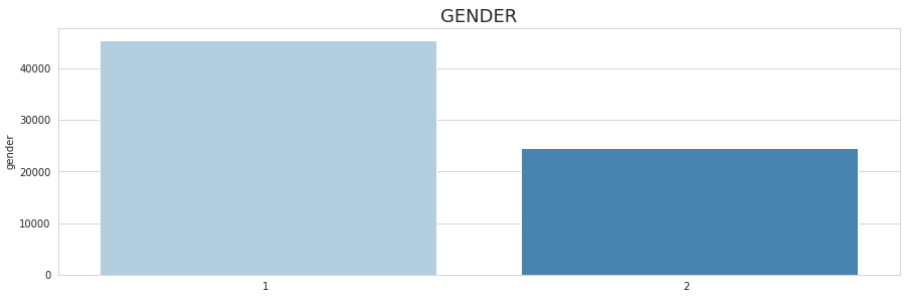

Figure 3.2: Gender

3.3.3 Weight in the dataset

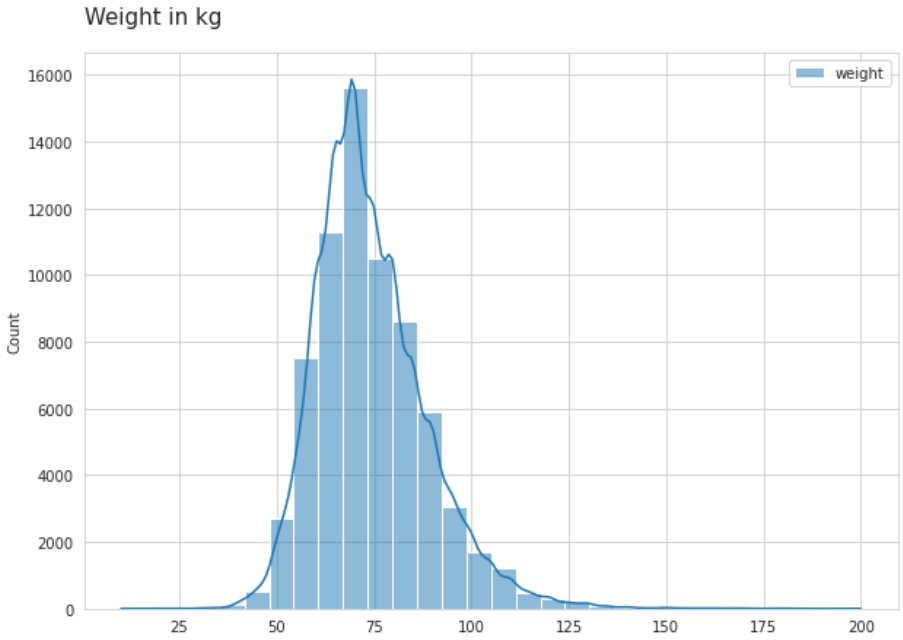

Figure 3.3: Weight in the dataset

3.3.4 Height in the dataset

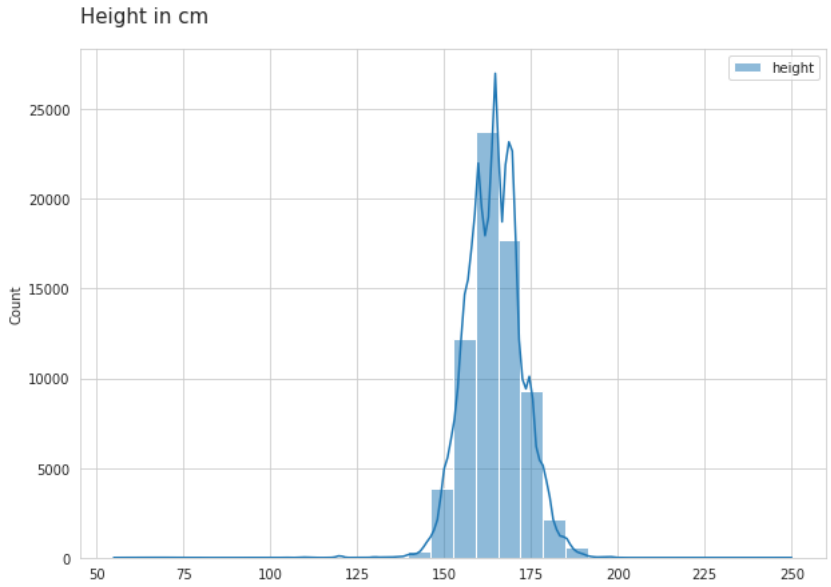

Figure 3.4: Height in the dataset

3.3.5 Cholesterol Level:

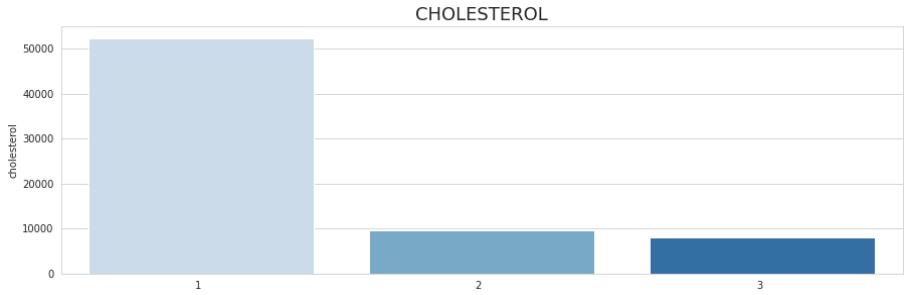

Figure 3.5: Cholesterol Level:

3.3.6 Glucose Level:

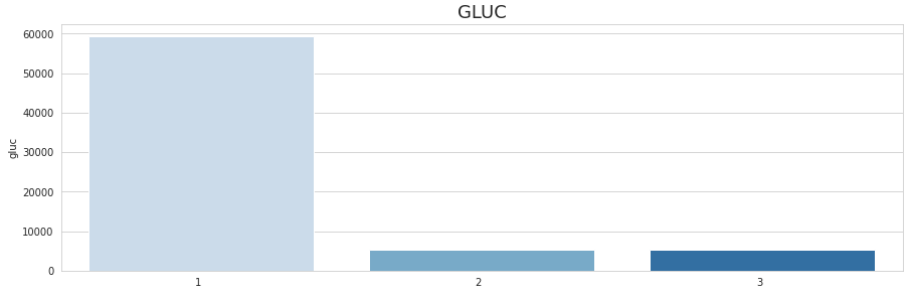

Figure 3.6: Glucose Level:

3.3.7 Smoker or Not:

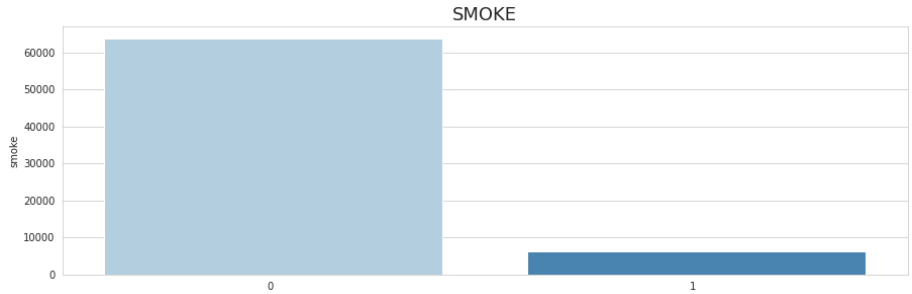

Figure 3.7: Smoker or Not:

3.3.8 Drink Alchohol or Not

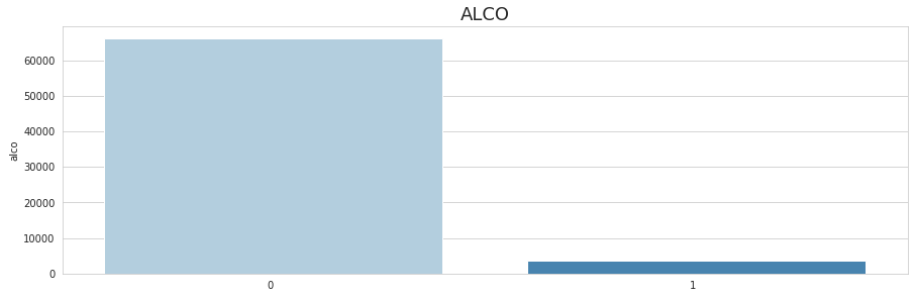

Figure 3.8: Drink Alchohol or Not

3.3.9 Phycically Active Or Not

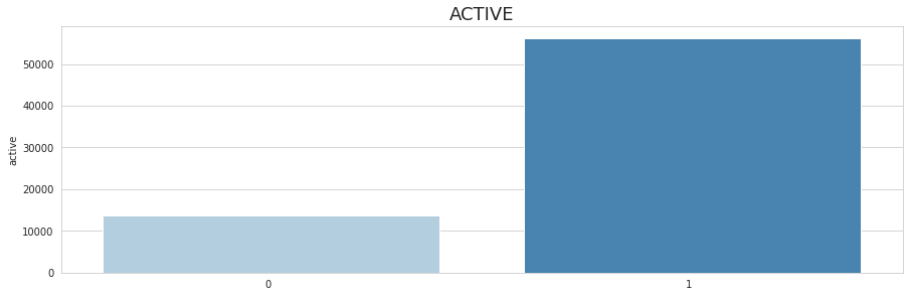

Figure 3.9: Phycically Active Or Not

3.3.10 Presence or absence of cardiovascular disease:

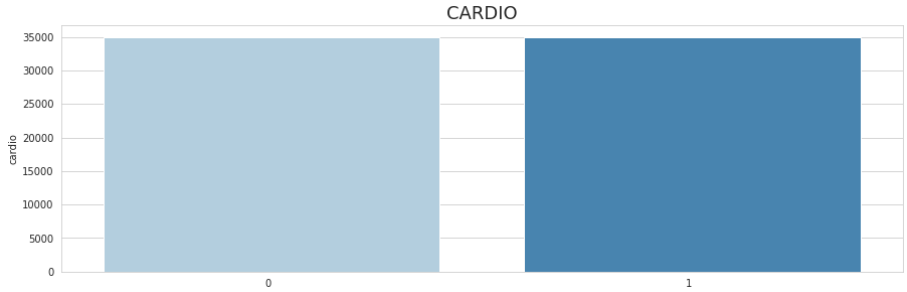

Figure 3.10: Presence or absence of cardiovascular disease:

3.4 Feature Engineering

3.4.1 Rounding of the age in the dataset

	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	age_year
0	18393	2	168	62.0	110	80	1	1	0	0	1	0	50.0
1	20228	1	156	85.0	140	90	3	1	0	0	1	1	55.0
2	18857	1	165	64.0	130	70	3	1	0	0	0	1	52.0
3	17623	2	169	82.0	150	100	1	1	0	0	1	1	48.0
4	17474	1	156	56.0	100	60	1	1	0	0	0	0	48.0

Figure 3.11: Rounding of the age in the dataset

3.4.2 Calculation of body mass index

	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	age_year	bmi
0	18393	2	168	62.0	110	80	1	1	0	0	1	0	50.0	21.967120
1	20228	1	156	85.0	140	90	3	1	0	0	1	1	55.0	34.927679
2	18857	1	165	64.0	130	70	3	1	0	0	0	1	52.0	23.507805
3	17623	2	169	82.0	150	100	1	1	0	0	1	1	48.0	28.710479
4	17474	1	156	56.0	100	60	1	1	0	0	0	0	48.0	23.011177

Figure 3.12: Calculation of body mass index

As we researched, Substracting Diastolic blood pressure from Systolic blood pressure gives Pulse Pressure that can't be negative. There are total 1227 observations where ap_hi < ap lwe got (68720, 14) shape,so we could reduce it.

3.5 Feature Selection

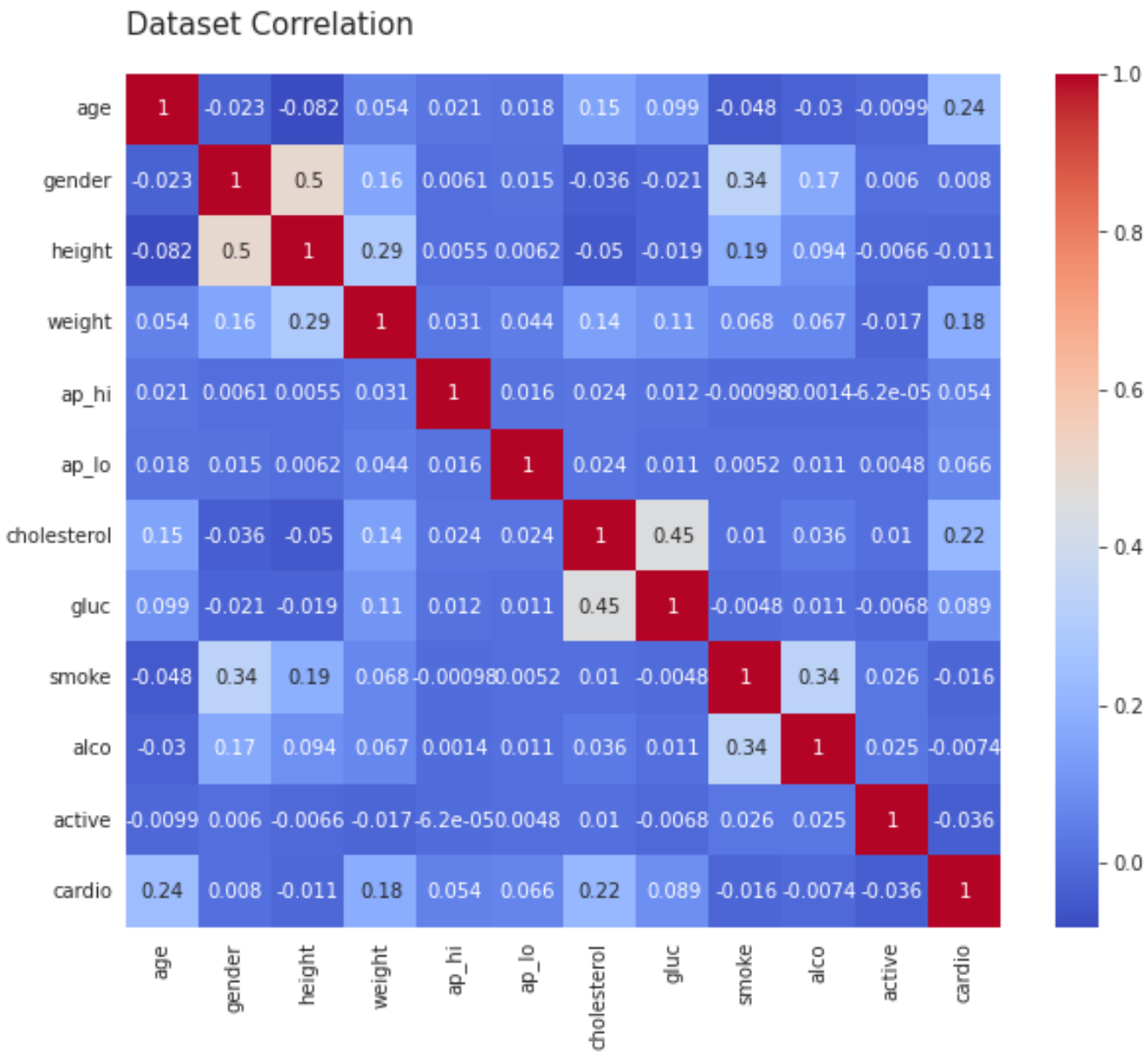

Figure 3.13: Dataset Correlation before feature engineering
Here,in this picture we can see the correlation value is very low,that's why the color is darker

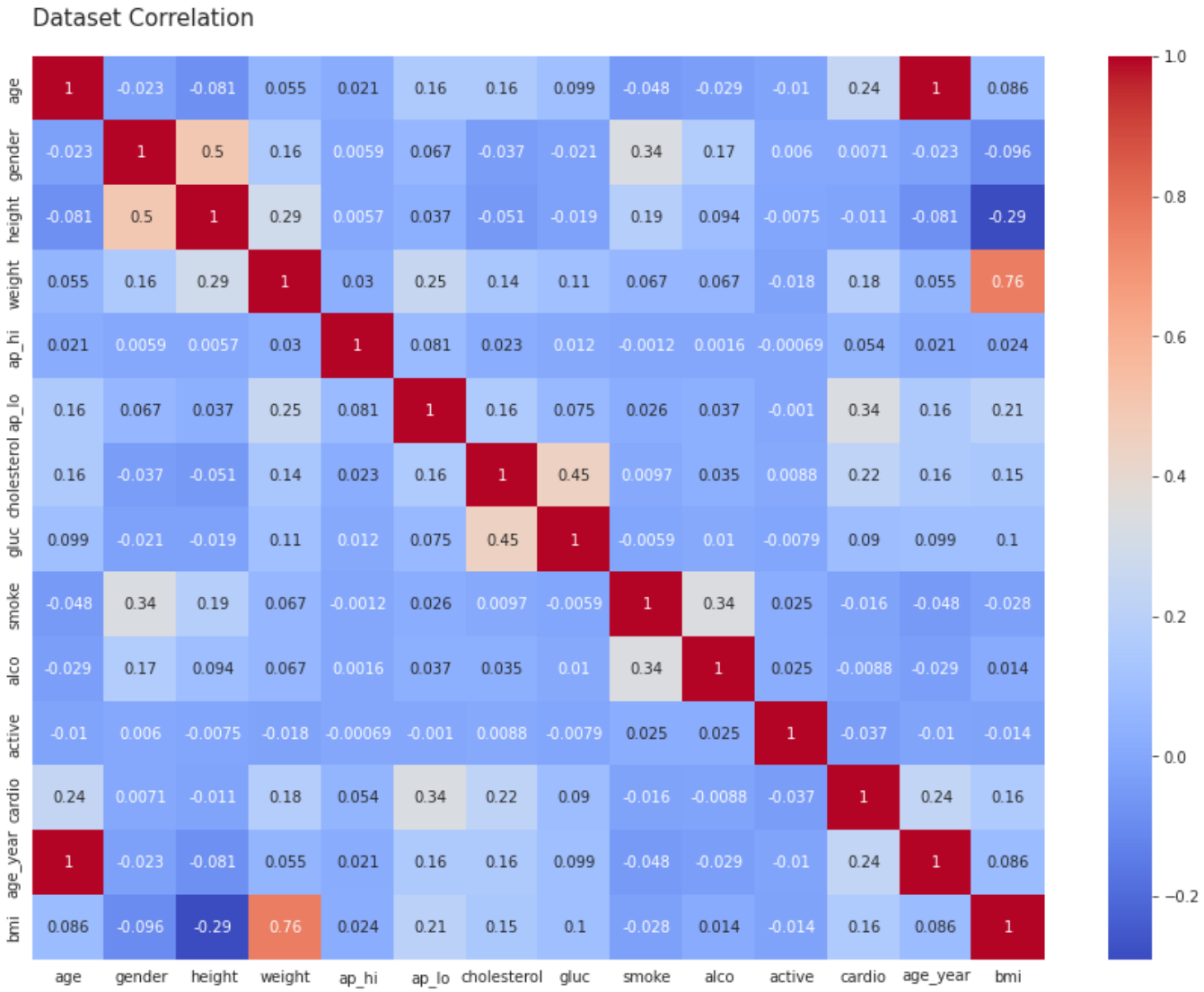

Figure 3.14: Dataset Correlation after feature engineering
Here,in this picture we can see after feature engineering the correlation value has increased and the colour became brighter.

3.6 Chi-Squared statistical

.A chi-squared test (also chi-square) is a statistical hypothesis test that is valid to perform when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test and variants thereof. Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table. In our experiment We did chi-squared for feature selection.We selected p-hi, ap-lo,cholesterol, gluc, smoke, alco, active, age, bmi.

Here,we can see age feature is more important and gender feature is less importantd .

```
[('gender', 0.5846240667063296),
 ('height', 3.098418214364506),
 ('alco', 5.010383623617884),
 ('smoke', 16.877352373644186),
 ('active', 18.84673384576003),
 ('gluc', 146.9869925540538),
 ('cholesterol', 1138.3240850070624),
 ('bmi', 2455.876482578835),
 ('age_year', 3366.719891049735),
 ('weight', 6139.336107669902),
 ('ap_lo', 8748.856824994302),
 ('ap_hi', 36785.87613042432),
 ('age', 1230912.1516819252)]
```

Figure 3.15: Chi-Squared statistical test for non-negative features to select 5 of the best features from the dataset

Chapter 4

Methodology

4.1 Model

4.1.1 Logistic Regression

the logistic model is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc.This technique is basically the relationship between features and the probability out of a particular event which deals with the sigmoid function which forms the basic concept of this algorithm. In logistic regression the sigmoid function fits the output of a linear equation between 0 and 1.

4.1.2 K-Neighbors Classifier

KNN is a non-parametric, lazy learning algorithm. Its aim is to use a database in which the given points are categorized into clusters to predict the classification of a new sample point. It is a supervised classifier that carry-out observations from within a test set to predict classification labels. KNN is one of the classification techniques used whenever there is a classification. By applying KNN in large datasets takes a long time to process.

4.1.3 Support Vector Machine

bSupport 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.. An SVM classifier creates a model

that allocates various new data points to a chosen category. Using the kernel trick, SVM is used for non-linear classification. It maps inputs into high dimensional feature spaces.

4.1.3.1 Linear SVM

Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

4.1.3.2 Guassian SVM

Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format; ||X1 — X2 || = Euclidean distance between X1 X2.

4.1.4 Random Forest Classifier

Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.A random forest classifier is a powerful tool in the machine learning library. With this classifier, we will be able to increase accuracy, and training time should be a smaller amount.Random forest is a very handy algorithm because the hyperparameters that it selects generally comes up with a good prediction result. It creates a tree for the data and makes predictions based on that. There are two steps in random forests, firstly design a random forest and make a prediction with the help of the classifier generated in the first stage.

4.1.5 Naive Bayes

Naive Byes is a popular classifier which assumes no feature has any relationship or dependence to each other. Naive Bayes classifiers are based on Bayes Theorem. The simplistic design of this model allows it to predict the data faster.

4.1.6 Decision Tree Classifier

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.

4.2 Dimension Reduction

4.2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

4.2.2 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function Analysis is a dimensionality reduction technique that is commonly used for supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space. For example, we have two classes and we need to separate them efficiently. Classes can have multiple features. Using only a single feature to classify them may result in some overlapping as shown in the below figure. So, we will keep on increasing the number of features for proper classification.

Chapter 5

Experiments and Results

For our experiment, first of all we selected the features via Chi-Squared and after that, we divided our dataset into training and testing samples, where 80% were used for training and the rest 20 percent% were used for testing. After the split, we fed our training samples to six individual models, and those results are shown below:

Here we used six model for our training the dataset. We applied simple model(without using dimension reduction techniques), one hot encoding (OHE), Principal Component Analysis (PCA), One Hot Encoding with Principal Component Analysis (OHE + PCA), Linear Discriminant Analysis (LDA), One Hot Encoding with Linear Discriminant Analysis (OHE + LDA), For each experiment accuracy and bar diagram is given bellow accordingly:

Model	Accuracy
Logistic Regression	0.719950524
KNN	0.705762515
Decision tree	0.633512806
Naive bayes	0.642316647
Linear SVM	0.727371944
Gaussian SVM	0.72548021
Random forest	0.70314319

Figure 5.1: Simple model

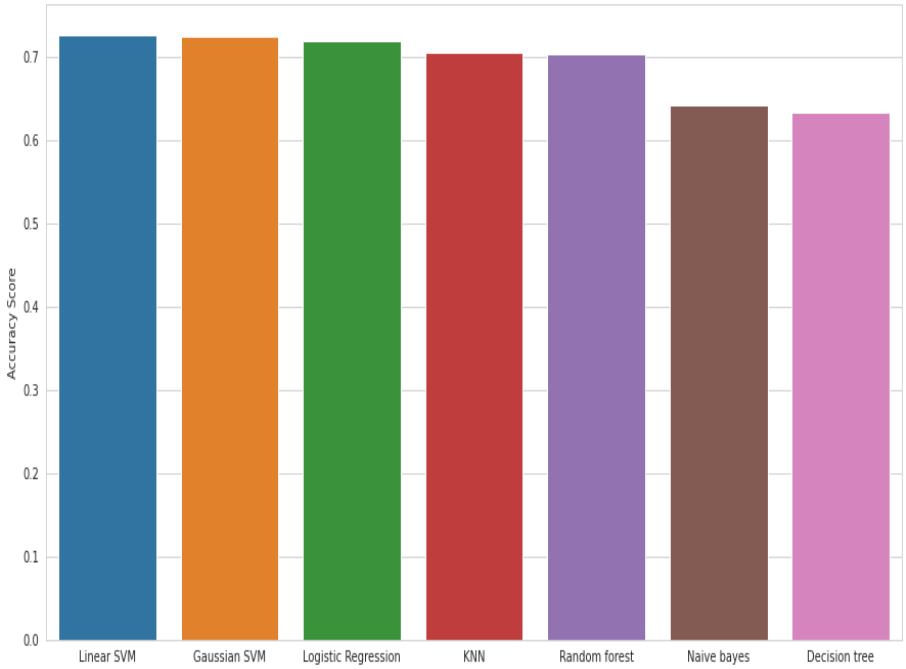

Figure 5.2: Simple model

Model	Accuracy
Logistic Regression	0.722351572
KNN	0.705544237
Decision tree	0.627328289
Naive bayes	0.612776484
Linear SVM	0.727590221
Gaussian SVM	0.726280559
Random forest	0.701396973

Figure 5.3: One Hot encoding

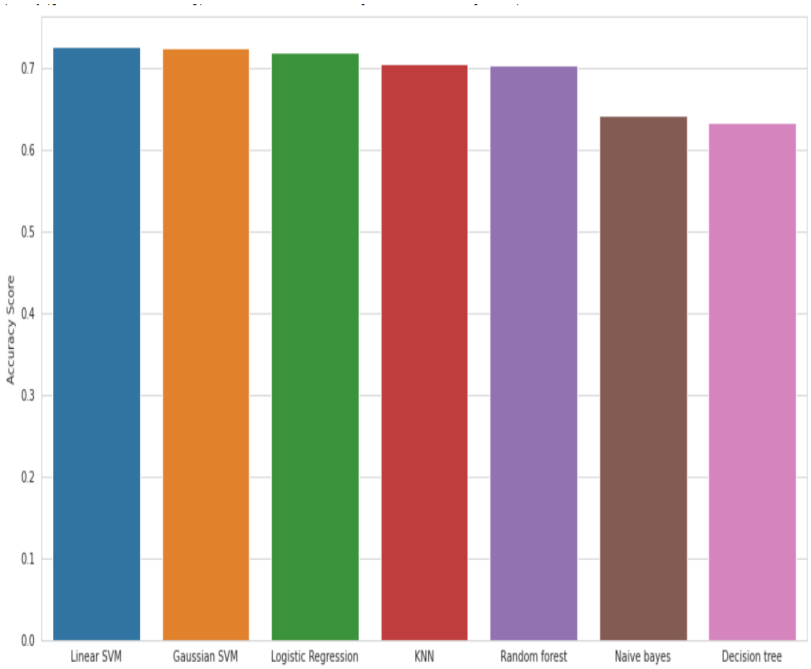

Figure 5.4: One Hot encoding

Dimension	Model	Accuracy
	Logistic Regression	0.659560536
	KNN	0.675203725
	Decision tree	0.579452852
2	Naive bayes	0.637878347
	Linear SVM	0.648355646
	Gaussian SVM	0.677459255
	Random forest	0.629729336
	Logistic Regression	0.689173458
	KNN	0.680151339
	Decision tree	0.589129802
3	Naive bayes	0.674476135
	Linear SVM	0.689610012
	Gaussian SVM	0.691428987
	Random forest	0.658760186
	Logistic Regression	0.68895518
	KNN	0.637223516
	Decision tree	0.551004075
4	Naive bayes	0.673675786
	Linear SVM	0.689391735
	Gaussian SVM	0.626600698
	Random forest	0.643408033
	Logistic Regression	0.649592549
	KNN	0.651411525
	Decision tree	0.552313737
5	Naive bayes	0.664217113
	Linear SVM	0.647118743
	Gaussian SVM	0.598297439
	Random forest	0.656359139

Figure 5.5: PCA

Dimension	Model	Accuracy
	Logistic Regression	0.643990105
	KNN	0.673166473
	Decision tree	0.569557625
6	Naive bayes	0.663198487
	Linear SVM	0.636568685
	Gaussian SVM	0.634749709
	Random forest	0.643771828
	Logistic Regression	0.593422584
	KNN	0.662107101
	Decision tree	0.544237485
7	Naive bayes	0.657668801
	Linear SVM	0.578215949
	Gaussian SVM	0.595096042
	Random forest	0.653012224
	Logistic Regression	0.591676368
	KNN	0.662616414
	Decision tree	0.549694412
8	Naive bayes	0.65774156
	Linear SVM	0.576251455
	Gaussian SVM	0.598006403
	Random forest	0.646100116

Figure 5.6: PCA

Dimension	Model	Accuracy
	Logistic Regression	0.600043655
	KNN	0.646100116
	Decision tree	0.563300349
2	Naive bayes	0.594732247
	Linear SVM	0.600189173
	Gaussian SVM	0.673457509
	Random forest	0.596332945
	Logistic Regression	0.599825378
	KNN	0.645227008
	Decision tree	0.547948196
3	Naive bayes	0.596842258
	Linear SVM	0.598806752
	Gaussian SVM	0.651338766
	Random forest	0.628346915
	Logistic Regression	0.663344005
	KNN	0.672075087
	Decision tree	0.478827125
4	Naive bayes	0.622307916
	Linear SVM	0.648137369
	Gaussian SVM	0.664289872
	Random forest	0.597933644
	Logistic Regression	0.68604482
	KNN	0.681388242
	Decision tree	0.486757858
5	Naive bayes	0.646900466
	Linear SVM	0.685826542
	Gaussian SVM	0.667564028
	Random forest	0.604554715

Figure 5.7: OHE + PCA

Dimension	Model	Accuracy
	Logistic Regression	0.685026193
	KNN	0.651338766
	Decision tree	0.536015716
6	Naive bayes	0.647191502
	Linear SVM	0.684807916
	Gaussian SVM	0.604845751
	Random forest	0.624781723
	Logistic Regression	0.647118743
	KNN	0.635331781
	Decision tree	0.54467404
7	Naive bayes	0.646463912
	Linear SVM	0.640570431
	Gaussian SVM	0.632857974
	Random forest	0.615395809
	Logistic Regression	0.637223516
	KNN	0.668728172
	Decision tree	0.521900466
8	Naive bayes	0.648355646
	Linear SVM	0.628419674
	Gaussian SVM	0.635986612
	Random forest	0.622235157

Figure 5.8: OHE + PCA

	Logistic Regression	0.590439464
	KNN	0.657959837
	Decision tree	0.534924331
9	Naive bayes	0.642898719
	Linear SVM	0.573268335
	Gaussian SVM	0.593931898
	Random forest	0.617942375
	Logistic Regression	0.58854773
	KNN	0.659123981
	Decision tree	0.547729919
10	Naive bayes	0.642971478
	Linear SVM	0.572249709
	Gaussian SVM	0.594222934
	Random forest	0.616487194
	Logistic Regression	0.581999418
	KNN	0.657959837
	Decision tree	0.549403376
11	Naive bayes	0.624854482
	Linear SVM	0.5669383
	Gaussian SVM	0.588838766
	Random forest	0.623908615

Figure 5.9: OHE + PCA

Model	Accuracy
Logistic Regression	0.697104191
KNN	0.689318976
Decision tree	0.595387078
Naive bayes	0.696158324
Linear SVM	0.698413853
Gaussian SVM	0.697395227
Random forest	0.595750873

Figure 5.10: LDA

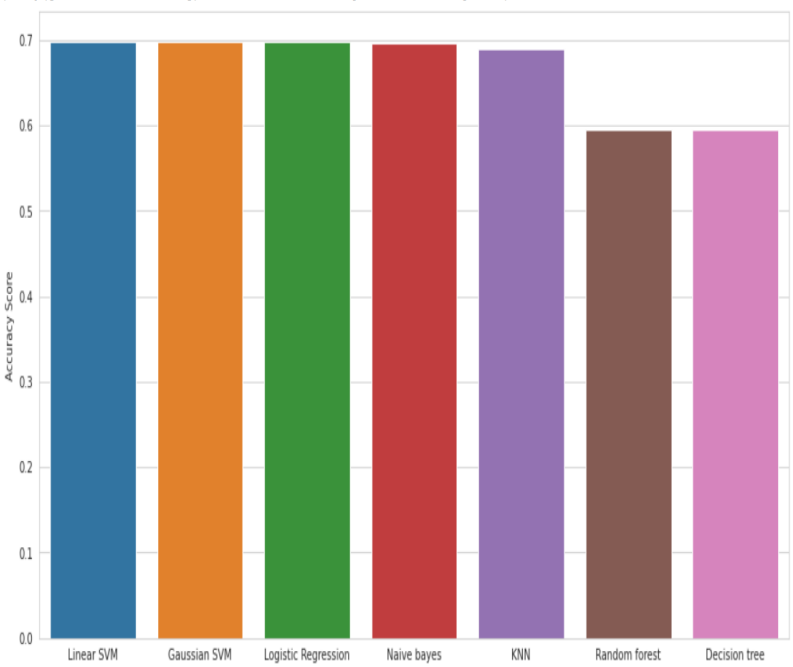

Figure 5.11: LDA

Model	Accuracy
Logistic Regression	0.697104191
KNN	0.689318976
Decision tree	0.595387078
Naive bayes	0.696158324
Linear SVM	0.698413853
Gaussian SVM	0.697395227
Random forest	0.595750873

Figure 5.12: OHE + LDA

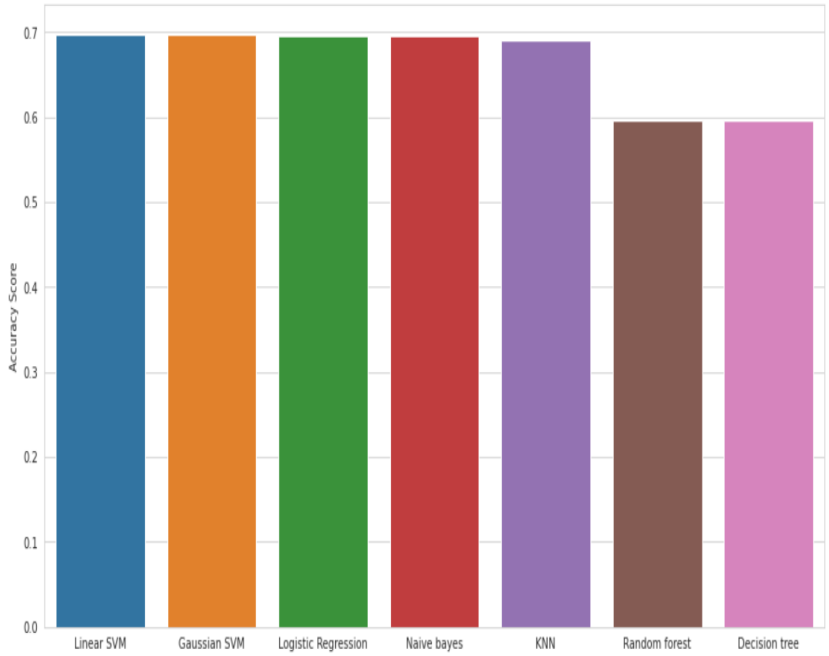

Figure 5.13: OHE + LDA

Chapter 6

Future Work and Conclusion

The motivation for the study was to find the most accurate Machine Learning classification algorithm for prediction of cardiovascular disease. This study compares the accuracy of 6 algorithms namely Logistic regression, K-Nearest Neighbour, Support Vector Machine, Decision Tree, Random Forests and Naive Bayes. After all 147 experiment that showed earlier we can say that SVM works better then any other algorithm. In the future we want to apply deep neural network for this dataset for better accuracy.

References

[1] P. S. P. R. Advait Shirvaikar, Advait Mandlik, "Prediction of cardiovascular disease by applying a combination of principal component analysis with machine learning techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 08, pp. 1937–1942, 09 2021.

[2] A. K. Anupama Yadav, Levish Gediya, "Heart disease prediction using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 08, pp. 1325–1329, 09 2021.

[3] D. Kumar and Bavithra, "Cardiovascular disease prediction using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 46–54, 09 2020.

[4] "ardiovascular-disease-dataset." https://www.kaggle.com/sulianova/cardiovascular-disease-dataset?fbclid=IwAR0TJxkG7VNF-M9KdZsDvFPXI0h4NVDfLJKdGON15oRIItUvs4TRYjaCun8.

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This project report was generated on Thursday 16th June, 2022 at 5:48pm.