



SRI GURU GOBIND SINGH COLLEGE OF COMMERCE
University Of Delhi



Data Mining Project

Heart Disease Prediction



Submitted by:

Anjali Rai (19078570005)

Garvit Bansal (19078570015)

ACKNOWLEDGEMENT

We would like to give our sincere thanks to our Data Mining teacher **Dr. Megha Ummat**, who guided us throughout our project in every possible way with her invaluable advice, useful suggestions and relevant ideas that facilitated the completion of this project. Our project would not have been possible without the proper and rigorous guidance of her and we feel honoured and privileged to work under her.

CERTIFICATE

This is to certify that the data mining project on **Heart Disease prediction** has been compiled by the students of the 'Sri Guru Gobind Singh College of Commerce' under the course 'BSc. (H) Computer Science' of 'Semester VI'.

The project compilation has taken place under my supervision. They have shown keen interest and utmost sincerity in the compilation of the project.

Dr. Megha Ummat

PROBLEM STATEMENT

In Today's time, the increase in heart disease is dramatically on the rise. The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heart disease in human. Early detection of heart diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases 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 data mining techniques to analyse the data for hidden information. This hidden information can be used for health diagnosis in medicinal data.

So, The Overall objective of this project is to predict the presence of heart disease in a patient on the basis of a labeled dataset.

DESCRIPTION OF DATASET

We are using dataset from one of the known sites that is **kaggle**.

The **dataset** : <https://www.kaggle.com/datasets/nareshbhat/health-care-data-set-on-heart-attack-possibility>

This database contains 14 attributes. The last "target" field refers to the presence of heart disease in the patient. It is integer valued 0 = no heart attack and 1 = more chance of heart attack.

The dataset consists of 303 individuals data. There are 14 columns in the dataset, which are described below:

1. **Age** – Displays the age of the individual
2. **Sex** – displays the gender of the individual using the following format :
1 = male
0 = female

3. **Chest-pain type** – displays the type of chest-pain experienced by the individual using the following format :

- 1 = typical angina
- 2 = atypical angina
- 3 = non — anginal pain
- 4 = asymptotic

4. **Resting Blood Pressure** – displays the resting blood pressure value of an individual in mmHg (unit)

5. **Serum Cholesterol** – displays the serum cholesterol in mg/dl (unit)

6. **Fasting Blood Sugar** – compares the fasting blood sugar value of an individual with 120mg/dl.

If fasting blood sugar > 120mg/dl then : 1 (true)
else : 0 (false)

7. **Resting ECG** – displays resting electrocardiographic results

- 0 = normal
- 1 = having ST-T wave abnormality
- 2 = left ventricular hypertrophy

8. **Max heart rate achieved** – displays the max heart rate achieved by an individual.

9. **Exercise induced angina** –

- 1 = yes
- 0 = no

10. **ST depression induced by exercise relative to rest –**
displays the value which is an integer or float.

11. **Peak exercise ST segment –**
1 = up sloping
2 = flat
3 = down sloping

12. **Number of major vessels (0–3) colored by flourosopy –**
displays the value as integer or float.

13. **Thal :** displays the thalassemia –
3 = normal
6 = fixed defect
7 = reversible defect

14. **Diagnosis of heart disease –** Displays whether the individual is suffering from heart disease or not :
0 = absence
1, 2, 3, 4 = present.

DATA PREPROCESSING

Data pre-processing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

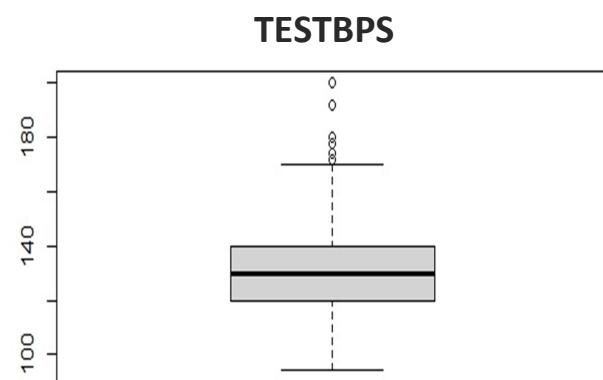
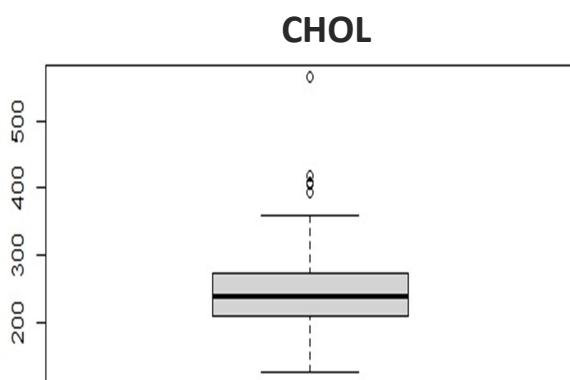
As, there are no **NA** or **missing** values in our dataset, so there is no need to filter the data or fill values using aggregate functions.

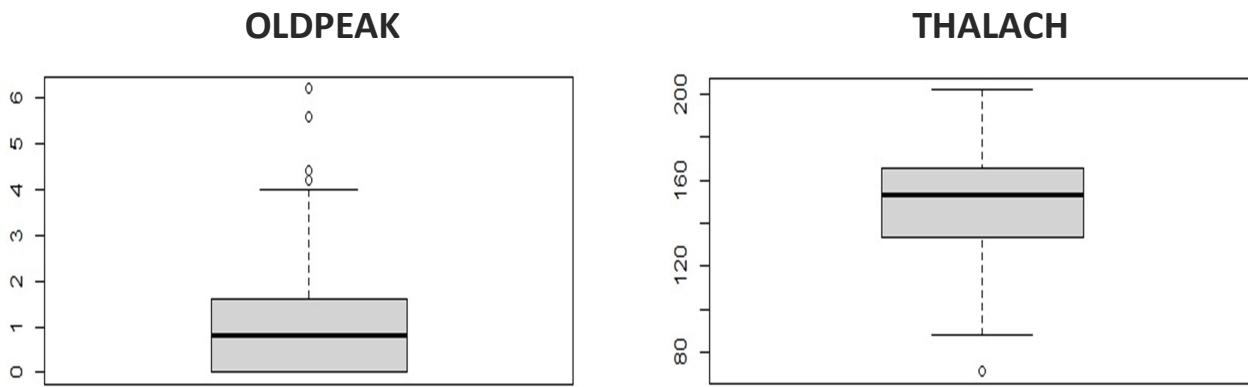
Outliers detection is done using Box plot:

Code:

```
for (i in c("trestbps", "chol", "thalach", "oldpeak"))
  boxplot(data1[, i])
```

Output:





The dataset also need normalization and standardization.

Summary of Dataset

Before normalization and standardization:

```
> summary(data1[c("trestbps","chol","thalach","oldpeak")])
   trestbps      chol      thalach     oldpeak
Min.   : 94.0  Min.   :126.0  Min.   : 71.0  Min.   :0.00
1st Qu.:120.0  1st Qu.:211.0  1st Qu.:133.5  1st Qu.:0.00
Median :130.0  Median :240.0  Median :153.0  Median :0.80
Mean   :131.6  Mean   :246.3  Mean   :149.6  Mean   :1.04
3rd Qu.:140.0  3rd Qu.:274.5  3rd Qu.:166.0  3rd Qu.:1.60
Max.   :200.0  Max.   :564.0  Max.   :202.0  Max.   :6.20
```

After normalization and standardization:

Code:

```
summary(data1[c("trestbps","chol","thalach","oldpeak")])

#After Normalizing and Standardizing
pp <- preProcess(data1[c("trestbps","chol","thalach","oldpeak")],method=c("scale","center"))
data2 <- predict(pp,data1[c("trestbps","chol","thalach","oldpeak")])
summary(data2)

> data2 <- predict(pp,data1[c("trestbps","chol","thalach","oldpeak")])
> summary(data2)
   trestbps      chol      thalach     oldpeak
Min.   :-2.14525  Min.   :-2.3203  Min.   :-3.4336  Min.   :-0.8954
1st Qu.:-0.66277  1st Qu.:-0.6804  1st Qu.:-0.7049  1st Qu.:-0.8954
Median :-0.09259  Median :-0.1209  Median : 0.1464  Median :-0.2064
Mean   : 0.00000  Mean   : 0.0000  Mean   : 0.0000  Mean   : 0.0000
3rd Qu.: 0.47760  3rd Qu.: 0.5448  3rd Qu.: 0.7139  3rd Qu.: 0.4827
Max.   : 3.89872  Max.   : 6.1303  Max.   : 2.2856  Max.   : 4.4445
```

PROCEDURE

Splitting dataset in training and testing data:

Code:

```
#classification
set.seed(1001)
trainIndex <- createDataPartition(data1$target,p=0.75,list=FALSE)
traindat <- data1[trainIndex,]
testdat <- data1[-trainIndex,]
```

Training Controls:

```
#Training Controls
control0 <- trainControl(method = "boot")
control1 <- trainControl(method = "LOOCV")
control2 <- trainControl(method = "cv" , number = 5)
control3 <- trainControl(method = "cv" , number = 10)
control4 <- trainControl(method = "repeatedcv" , number = 5)
control5 <- trainControl(method = "repeatedcv" , number = 10)
```

Creating KNN models using Trcontrols created before:

```
#Creating Model using KNN
modela0 <- train(target~.,traindat,"knn",trControl=control0)
modela1 <- train(target~.,traindat,"knn",trControl=control1)
modela2 <- train(target~.,traindat,"knn",trControl=control2)
modela3 <- train(target~.,traindat,"knn",trControl=control3)
modela4 <- train(target~.,traindat,"knn",trControl=control4)
modela5 <- train(target~.,traindat,"knn",trControl=control5)
```

Creating NB models using Trcontrols created before:

```
#Creating Model using Naive Bayes
modelb0 <- train(target~.,traindat,"nb",trControl=control0)
modelb1 <- train(target~.,traindat,"nb",trControl=control1)
modelb2 <- train(target~.,traindat,"nb",trControl=control2)
modelb3 <- train(target~.,traindat,"nb",trControl=control3)
modelb4 <- train(target~.,traindat,"nb",trControl=control4)
modelb5 <- train(target~.,traindat,"nb",trControl=control5)
```

Creating DT models using Trcontrols created before:

```
#Creating Model using Decision Tree
modelc0 <- train(target~.,traindat,"J48",trControl=control0)
modelc1 <- train(target~.,traindat,"J48",trControl=control1)
modelc2 <- train(target~.,traindat,"J48",trControl=control2)
modelc3 <- train(target~.,traindat,"J48",trControl=control3)
modelc4 <- train(target~.,traindat,"J48",trControl=control4)
modelc5 <- train(target~.,traindat,"J48",trControl=control5)
```

OBSERVATIONS

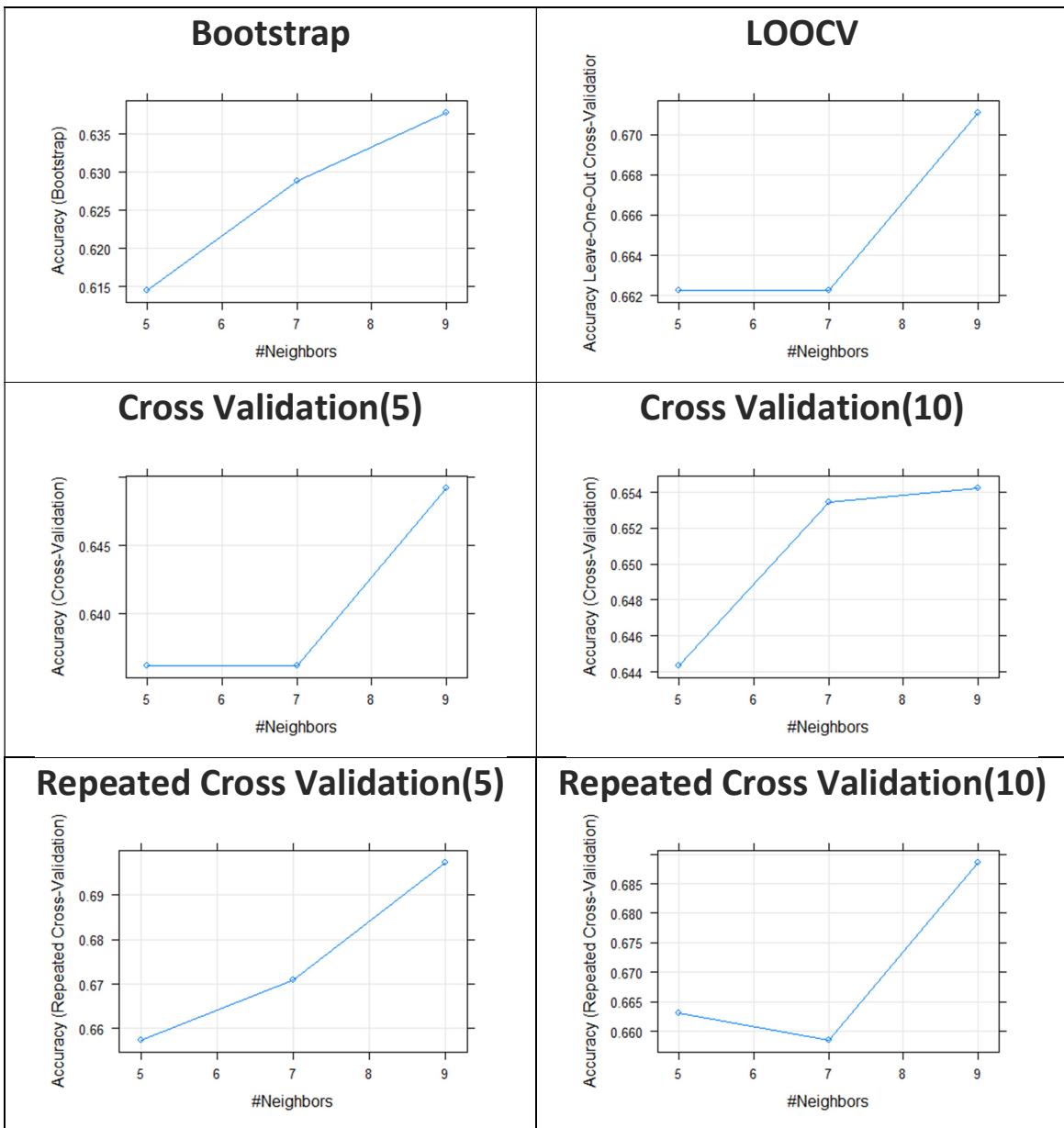
From graphs it is clear that decision tree and naïve bayes is more accurate than knn. But in this case study we can't rely on naïve bayes as attributes can have dependency as they are the symptoms and physical features of an individual. So Decision tree will be best suited.

Plotting Accuracy Models

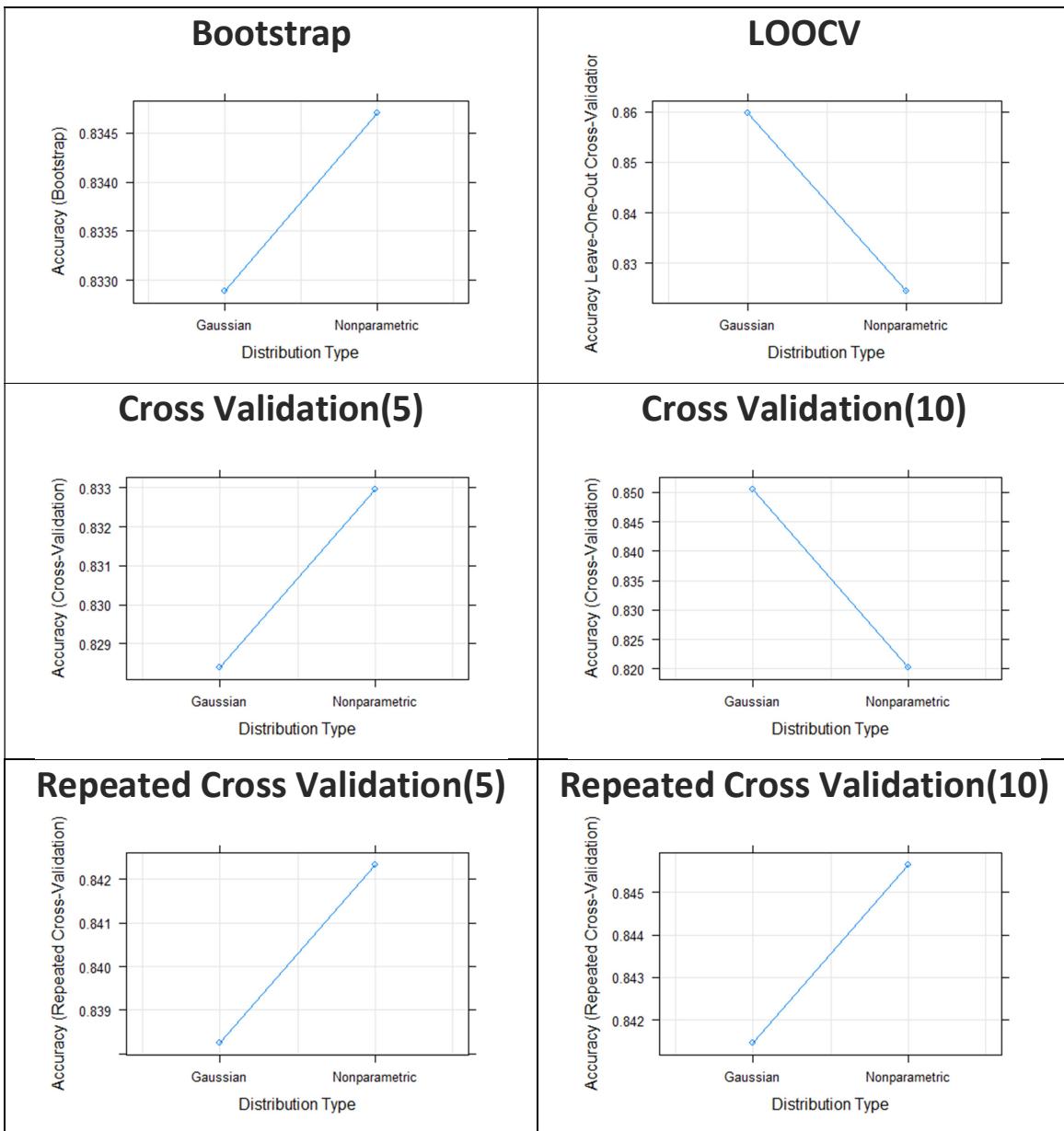
Code:

```
#Plotting Accuracy of models
plot(modela0)
plot(modela1)
plot(modela2)
plot(modela3)
plot(modela4)
plot(modela5)
plot(modelb0)
plot(modelb1)
plot(modelb2)
plot(modelb3)
plot(modelb4)
plot(modelb5)
plot(modelc0)
plot(modelc1)
plot(modelc2)
plot(modelc3)
plot(modelc4)
plot(modelc5)
```

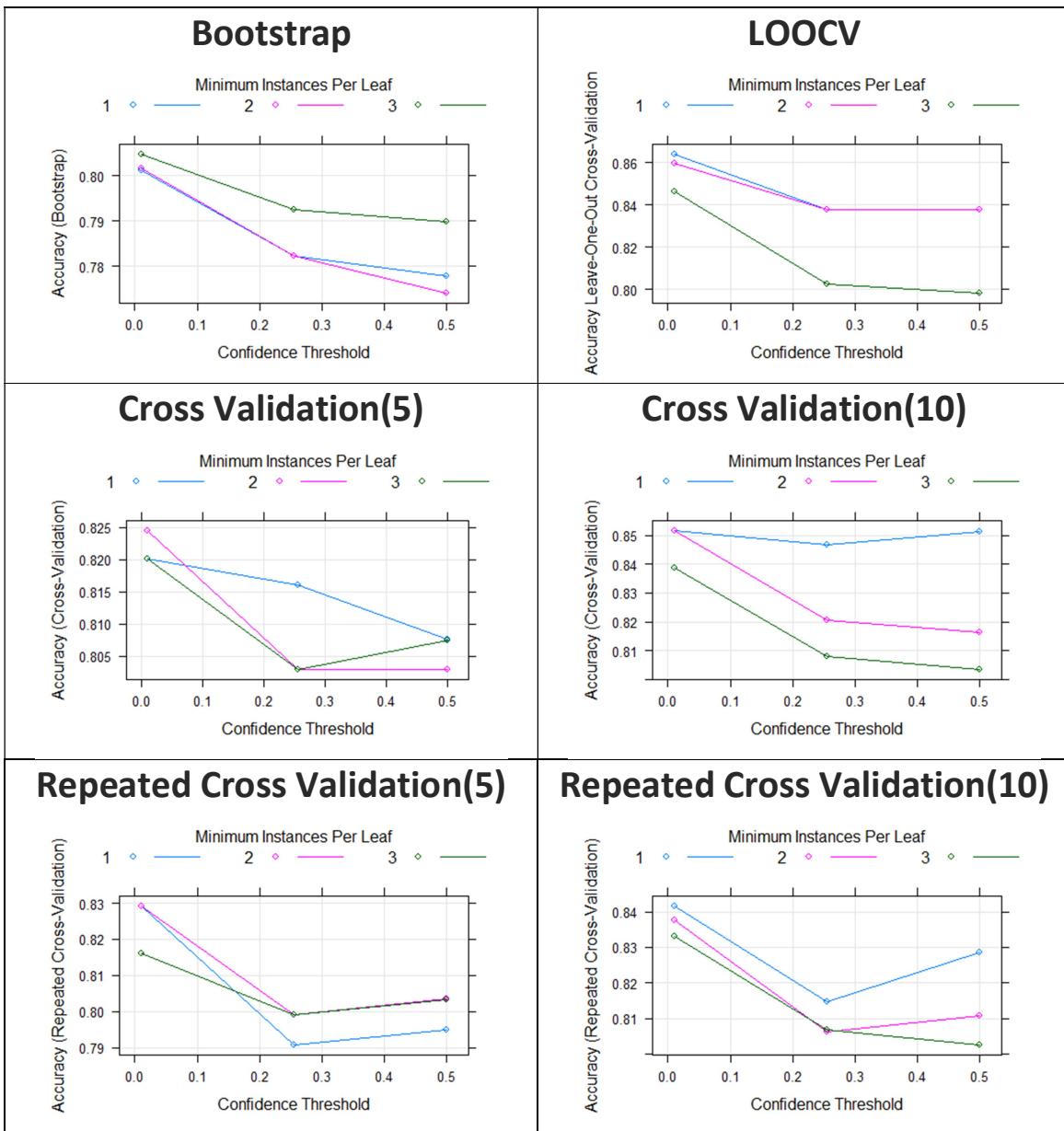
Accuracy graphs for KNN are as follows:-



Accuracy graphs for Naïve Bayes are as follows:-



Accuracy graphs for Decision Tree are as follows:-



Confusion Matrix for KNN:

<p>Bootstrap</p> <pre>> confusionMatrix(pa0,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-Value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037	<p>LOOCV</p> <pre>> confusionMatrix(pa1,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																
<p>Cross Validation(5)</p> <pre>> confusionMatrix(pa2,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-Value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037	<p>Cross Validation(10)</p> <pre>> confusionMatrix(pa3,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																
<p>Repeated Cross Validation(5)</p> <pre>> confusionMatrix(pa4,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-Value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037	<p>Repeated Cross Validation(10)</p> <pre>> confusionMatrix(pa5,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table border="1"> <thead> <tr> <th colspan="2">Reference</th> </tr> <tr> <th>Prediction</th> <th>0 1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>17 12</td> </tr> <tr> <td>1</td> <td>17 29</td> </tr> </tbody> </table> <p>Accuracy : 0.6133 95% CI : (0.4938, 0.7236) No Information Rate : 0.5467 P-Value [Acc > NIR] : 0.1482</p> <p>Kappa : 0.21</p> <p>McNemar's Test P-value : 0.4576</p> <table border="1"> <thead> <tr> <th>Sensitivity</th> <th>Specificity</th> </tr> </thead> <tbody> <tr> <td>0.5000</td> <td>0.7073</td> </tr> <tr> <td>Pos Pred Value</td> <td>0.5862</td> </tr> <tr> <td>Neg Pred Value</td> <td>0.6304</td> </tr> <tr> <td>Prevalence</td> <td>0.4533</td> </tr> <tr> <td>Detection Rate</td> <td>0.2267</td> </tr> <tr> <td>Detection Prevalence</td> <td>0.3867</td> </tr> <tr> <td>Balanced Accuracy</td> <td>0.6037</td> </tr> </tbody> </table> <p>'Positive' Class : 0</p>	Reference		Prediction	0 1	0	17 12	1	17 29	Sensitivity	Specificity	0.5000	0.7073	Pos Pred Value	0.5862	Neg Pred Value	0.6304	Prevalence	0.4533	Detection Rate	0.2267	Detection Prevalence	0.3867	Balanced Accuracy	0.6037
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																
Reference																																																	
Prediction	0 1																																																
0	17 12																																																
1	17 29																																																
Sensitivity	Specificity																																																
0.5000	0.7073																																																
Pos Pred Value	0.5862																																																
Neg Pred Value	0.6304																																																
Prevalence	0.4533																																																
Detection Rate	0.2267																																																
Detection Prevalence	0.3867																																																
Balanced Accuracy	0.6037																																																

Confusion Matrix for Decision Tree:

Bootstrap	LOOCV																				
<p style="text-align: center;">Bootstrap</p> <pre>> confusionMatrix(pa6,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-Value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37	<p style="text-align: center;">LOOCV</p> <pre>> confusionMatrix(pa7,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			
Cross Validation(5)	Cross Validation(10)																				
<p style="text-align: center;">Cross Validation(5)</p> <pre>> confusionMatrix(pa8,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37	<p style="text-align: center;">Cross Validation(10)</p> <pre>> confusionMatrix(pa9,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			
Repeated Cross Validation(5)	Repeated Cross Validation(10)																				
<p style="text-align: center;">Repeated Cross Validation(5)</p> <pre>> confusionMatrix(pa10,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-Value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37	<p style="text-align: center;">Repeated Cross Validation(10)</p> <pre>> confusionMatrix(pa11,as.factor(testdat\$target)) Confusion Matrix and Statistics</pre> <table style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">Reference</td></tr> <tr><td style="padding: 2px;">Prediction</td><td style="padding: 2px;">0</td><td style="padding: 2px;">1</td></tr> <tr><td style="padding: 2px;">0</td><td style="padding: 2px;">21</td><td style="padding: 2px;">4</td></tr> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">13</td><td style="padding: 2px;">37</td></tr> </table> <pre>Accuracy : 0.7733 95% CI : (0.6621, 0.8621) No Information Rate : 0.5467 P-Value [Acc > NIR] : 3.951e-05</pre> <p style="text-align: center;">Kappa : 0.5321</p> <p style="text-align: center;">McNemar's Test P-Value : 0.05235</p> <pre>Sensitivity : 0.6176 Specificity : 0.9024 Pos Pred Value : 0.8400 Neg Pred Value : 0.7400 Prevalence : 0.4533 Detection Rate : 0.2800 Detection Prevalence : 0.3333 Balanced Accuracy : 0.7600</pre> <p style="text-align: center;">'Positive' class : 0</p>	Reference	Prediction	0	1	0	21	4	1	13	37
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			
Reference																					
Prediction	0	1																			
0	21	4																			
1	13	37																			

As, mentioned before fields can have dependency so mapping rules using apriori algorithm at minimum support of 80% and minimum confidence of 70%

Rules are as follows:-

```
> inspect(rule)
lhs                                         rhs          support  confidence coverage lift   count
[1] {}                                     => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[2] {}                                     => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[3] {}                                     => {sex=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[4] {}                                     => {fbst=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[5] {}                                     => {exang=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[6] {slope=[1,2]}                           => {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[7] {thal=[2,3]}                           => {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[8] {slope=[1,2]}                           => {sex=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[9] {sex=[0,1]}                           => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[10] {slope=[1,2]}                          => {fbst=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[11] {fbst=[0,1]}                          => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[12] {slope=[1,2]}                          => {exang=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[13] {exang=[0,1]}                          => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[14] {thal=[2,3]}                           => {sex=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[15] {sex=[0,1]}                           => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[16] {thal=[2,3]}                           => {fbst=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[17] {fbst=[0,1]}                          => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[18] {thal=[2,3]}                           => {exang=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[19] {exang=[0,1]}                          => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[20] {sex=[0,1]}                           => {fbst=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[21] {fbst=[0,1]}                          => {sex=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[22] {sex=[0,1]}                           => {exang=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[23] {exang=[0,1]}                          => {sex=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[24] {fbst=[0,1]}                          => {exang=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[25] {exang=[0,1]}                          => {fbst=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[26] {slope=[1,2], thal=[2,3]}             => {sex=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[27] {sex=[0,1], slope=[1,2]}              => {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[28] {sex=[0,1], thal=[2,3]}              => {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[29] {slope=[1,2], thal=[2,3]}             => {fbst=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[30] {fbst=[0,1], slope=[1,2]}            => {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[31] {fbst=[0,1], thal=[2,3]}             => {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[32] {slope=[1,2], thal=[2,3]}             => {exang=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[33] {exang=[0,1], slope=[1,2]}           => {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[34] {exang=[0,1], thal=[2,3]}            => {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[35] {sex=[0,1], slope=[1,2]}              => {fbst=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[36] {fbst=[0,1], slope=[1,2]}            => {sex=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[37] {sex=[0,1], fbst=[0,1]}              => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[38] {sex=[0,1], slope=[1,2]}             => {exang=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[39] {exang=[0,1], slope=[1,2]}           => {sex=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[40] {sex=[0,1], exang=[0,1]}             => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[41] {fbst=[0,1], slope=[1,2]}            => {exang=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[42] {exang=[0,1], slope=[1,2]}           => {fbst=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[43] {fbst=[0,1], exang=[0,1]}            => {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[44] {sex=[0,1], thal=[2,3]}              => {exang=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[45] {fbst=[0,1], thal=[2,3]}             => {sex=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[46] {sex=[0,1], fbst=[0,1]}              => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[47] {sex=[0,1], thal=[2,3]}              => {exang=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[48] {exang=[0,1], thal=[2,3]}            => {sex=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[49] {sex=[0,1], exang=[0,1]}             => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[50] {fbst=[0,1], thal=[2,3]}             => {exang=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[51] {exang=[0,1], thal=[2,3]}            => {fbst=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[52] {fbst=[0,1], exang=[0,1]}            => {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[53] {sex=[0,1], fbst=[0,1]}              => {exang=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[54] {sex=[0,1], exang=[0,1]}             => {fbst=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[55] {fbst=[0,1], exang=[0,1]}            => {sex=[0,1]} 1.0000000 1.0000000 1.0000000 1.000000 303
[56] {sex=[0,1], slope=[1,2], thal=[2,3]} => {fbst=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[57] {fbst=[0,1], slope=[1,2], thal=[2,3]}=> {sex=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[58] {sex=[0,1], fbst=[0,1], slope=[1,2]}=> {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[59] {sex=[0,1], fbst=[0,1], thal=[2,3]}=> {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[60] {sex=[0,1], slope=[1,2], thal=[2,3]}=> {exang=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[61] {exang=[0,1], slope=[1,2], thal=[2,3]}=> {sex=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[62] {sex=[0,1], exang=[0,1], slope=[1,2]}=> {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[63] {sex=[0,1], exang=[0,1], thal=[2,3]}=> {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[64] {fbst=[0,1], slope=[1,2], thal=[2,3]}=> {exang=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[65] {exang=[0,1], slope=[1,2], thal=[2,3]}=> {fbst=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[66] {fbst=[0,1], exang=[0,1], slope=[1,2]}=> {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[67] {fbst=[0,1], exang=[0,1], thal=[2,3]}=> {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
[68] {sex=[0,1], fbst=[0,1], slope=[1,2]}=> {exang=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[69] {sex=[0,1], exang=[0,1], slope=[1,2]}=> {fbst=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[70] {fbst=[0,1], exang=[0,1], slope=[1,2]}=> {sex=[0,1]} 0.9306931 1.0000000 0.9306931 1.000000 282
[71] {sex=[0,1], fbst=[0,1], exang=[0,1]}=> {slope=[1,2]} 0.9306931 0.9306931 1.0000000 1.000000 282
[72] {sex=[0,1], fbst=[0,1], thal=[2,3]}=> {exang=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[73] {sex=[0,1], exang=[0,1], thal=[2,3]}=> {fbst=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[74] {fbst=[0,1], exang=[0,1], thal=[2,3]}=> {sex=[0,1]} 0.9339934 1.0000000 0.9339934 1.000000 283
[75] {sex=[0,1], fbst=[0,1], exang=[0,1]}=> {thal=[2,3]} 0.9339934 0.9339934 1.0000000 1.000000 283
[76] {sex=[0,1], fbst=[0,1], slope=[1,2], thal=[2,3]}=> {exang=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[77] {sex=[0,1], exang=[0,1], slope=[1,2], thal=[2,3]}=> {fbst=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[78] {fbst=[0,1], exang=[0,1], slope=[1,2], thal=[2,3]}=> {sex=[0,1]} 0.8745875 1.0000000 0.8745875 1.000000 265
[79] {sex=[0,1], fbst=[0,1], exang=[0,1], slope=[1,2]}=> {thal=[2,3]} 0.8745875 0.9397163 0.9306931 1.006127 265
[80] {sex=[0,1], fbst=[0,1], exang=[0,1], thal=[2,3]}=> {slope=[1,2]} 0.8745875 0.9363958 0.9339934 1.006127 265
```

RESULTS

Successful creation of 80-90% accurate model to predict whether an individual have heart problem or not.

CONCLUSION

Decision tree model will be best suited for these types of case studies in which some fields/attributes shows dependency or are related to each other.

REFERENCES

Data Reference : <https://www.kaggle.com/datasets/nareshbhat/health-care-data-set-on-heart-attack-possibility>