setwd("C:/Users/Matrix/Desktop/New folder")

> filenameo <- "heart.csv"

> heartds <- read.csv(filenameo, header=TRUE)

> dim(heartds)

[1] 289 14

> heartds$output<- as.factor(heartds$output)

> sapply(heartds, class)

age sex cp trtbps chol fbs restecg

"integer" "integer" "integer" "integer" "integer" "integer" "integer"

thalachh exng oldpeak slp caa thall output

"integer" "integer" "numeric" "integer" "integer" "integer" "factor"

> head(heartds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2

4 55 1 1 120 236 0 1 178 0 0.8 2 0 2

5 56 0 0 120 354 0 1 163 1 0.6 2 0 2

6 55 1 0 140 192 0 1 148 0 0.4 1 0 1

output

1 yes

2 yes

3 yes

4 yes

5 yes

6 yes

> summary(heartds)

age sex cp trtbps

Min. :29.00 Min. :0.0000 Min. :0.000 Min. : 94.0

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:120.0

Median :54.00 Median :1.0000 Median :1.000 Median :130.0

Mean :54.01 Mean :0.6782 Mean :1.021 Mean :131.4

3rd Qu.:60.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:140.0

Max. :77.00 Max. :1.0000 Max. :3.000 Max. :200.0

chol fbs restecg thalachh

Min. :126 Min. :0.0000 Min. :0.0000 Min. : 71.0

1st Qu.:212 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:136.0

Median :243 Median :0.0000 Median :1.0000 Median :154.0

Mean :248 Mean :0.1453 Mean :0.5156 Mean :150.2

3rd Qu.:276 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:168.0

Max. :564 Max. :1.0000 Max. :2.0000 Max. :202.0

exng oldpeak slp caa

Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000

Median :0.0000 Median :0.600 Median :1.000 Median :0.0000

Mean :0.3183 Mean :1.008 Mean :1.419 Mean :0.7128

3rd Qu.:1.0000 3rd Qu.:1.600 3rd Qu.:2.000 3rd Qu.:1.0000

Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000

thall output

Min. :0.000 no :124

1st Qu.:2.000 yes:165

Median :2.000

Mean :2.315

3rd Qu.:3.000

Max. :3.000

> filenametr <- "train.csv"

> trainds <- read.csv(filenametr, header=TRUE)

> dim(trainds)

[1] 231 14

> trainds$output<- as.factor(trainds$output)

> sapply(trainds, class)

age sex cp trtbps chol fbs restecg

"integer" "integer" "integer" "integer" "integer" "integer" "integer"

thalachh exng oldpeak slp caa thall output

"integer" "integer" "numeric" "integer" "integer" "integer" "factor"

> head(trainds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2

4 56 0 0 120 354 0 1 163 1 0.6 2 0 2

5 55 1 0 140 192 0 1 148 0 0.4 1 0 1

6 56 0 1 140 294 0 0 153 0 1.3 1 0 2

output

1 yes

2 yes

3 yes

4 yes

5 yes

6 yes

> summary(trainds)

age sex cp trtbps

Min. :29.00 Min. :0.0000 Min. :0 Min. : 94.0

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0 1st Qu.:120.0

Median :54.00 Median :1.0000 Median :1 Median :130.0

Mean :54.28 Mean :0.6623 Mean :1 Mean :131.9

3rd Qu.:61.00 3rd Qu.:1.0000 3rd Qu.:2 3rd Qu.:140.0

Max. :77.00 Max. :1.0000 Max. :3 Max. :200.0

chol fbs restecg thalachh

Min. :126.0 Min. :0.0000 Min. :0.0000 Min. : 71.0

1st Qu.:212.5 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:133.0

Median :245.0 Median :0.0000 Median :1.0000 Median :155.0

Mean :249.7 Mean :0.1472 Mean :0.5368 Mean :150.1

3rd Qu.:279.5 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:168.0

Max. :564.0 Max. :1.0000 Max. :2.0000 Max. :202.0

exng oldpeak slp caa

Min. :0.000 Min. :0.0000 Min. :0.000 Min. :0.0000

1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000

Median :0.000 Median :0.6000 Median :1.000 Median :0.0000

Mean :0.316 Mean :0.9874 Mean :1.429 Mean :0.7056

3rd Qu.:1.000 3rd Qu.:1.6000 3rd Qu.:2.000 3rd Qu.:1.0000

Max. :1.000 Max. :5.6000 Max. :2.000 Max. :4.0000

thall output

Min. :0.000 no : 99

1st Qu.:2.000 yes:132

Median :2.000

Mean :2.338

3rd Qu.:3.000

Max. :3.000

> filenamete <- "test.csv"

> testds <- read.csv(filenamete, header=TRUE)

> dim(testds)

[1] 58 14

> testds$output<- as.factor(testds$output)

> sapply(testds, class)

age sex cp trtbps chol fbs restecg

"integer" "integer" "integer" "integer" "integer" "integer" "integer"

thalachh exng oldpeak slp caa thall output

"integer" "integer" "numeric" "integer" "integer" "integer" "factor"

> head(testds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall

1 55 1 1 120 236 0 1 178 0 0.8 2 0 2

2 44 1 1 120 263 0 1 173 0 0.0 2 0 3

3 48 0 2 130 275 0 1 139 0 0.2 2 0 2

4 59 1 2 150 212 1 1 157 0 1.6 2 0 2

5 51 1 2 110 175 0 1 123 0 0.6 2 0 2

6 44 1 1 130 219 0 0 188 0 0.0 2 0 2

output

1 yes

2 yes

3 yes

4 yes

5 yes

6 yes

> summary(testds)

age sex cp trtbps

Min. :34.00 Min. :0.0000 Min. :0.000 Min. :100.0

1st Qu.:45.75 1st Qu.:0.2500 1st Qu.:0.000 1st Qu.:113.5

Median :55.00 Median :1.0000 Median :1.000 Median :125.5

Mean :52.95 Mean :0.7414 Mean :1.103 Mean :129.1

3rd Qu.:59.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:138.0

Max. :69.00 Max. :1.0000 Max. :3.000 Max. :178.0

chol fbs restecg thalachh

Min. :157.0 Min. :0.0000 Min. :0.000 Min. :103.0

1st Qu.:209.0 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:138.2

Median :234.5 Median :0.0000 Median :0.000 Median :150.5

Mean :240.9 Mean :0.1379 Mean :0.431 Mean :150.6

3rd Qu.:264.5 3rd Qu.:0.0000 3rd Qu.:1.000 3rd Qu.:164.5

Max. :409.0 Max. :1.0000 Max. :1.000 Max. :188.0

exng oldpeak slp caa

Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000

Median :0.0000 Median :0.700 Median :1.000 Median :0.0000

Mean :0.3276 Mean :1.088 Mean :1.379 Mean :0.7414

3rd Qu.:1.0000 3rd Qu.:1.800 3rd Qu.:2.000 3rd Qu.:1.0000

Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000

thall output

Min. :0.000 no :25

1st Qu.:2.000 yes:33

Median :2.000

Mean :2.224

3rd Qu.:3.000

Max. :3.000

>

> library(caret)

Loading required package: ggplot2

Loading required package: lattice

> library(randomForest)

randomForest 4.7-1.1

Type rfNews() to see new features/changes/bug fixes.

Attaching package: ‘randomForest’

The following object is masked from ‘package:ggplot2’:

margin

> library(AUC)

AUC 0.3.2

Type AUCNews() to see the change log and ?AUC to get an overview.

Attaching package: ‘AUC’

The following objects are masked from ‘package:caret’:

sensitivity, specificity

> library(MASS)

> model.LDA <- lda(output~., data=trainds, na.action="na.omit")

> model.LDA

Call:

lda(output ~ ., data = trainds, na.action = "na.omit")

Prior probabilities of groups:

no yes

0.4285714 0.5714286

Group means:

age sex cp trtbps chol fbs restecg

no 56.10101 0.8383838 0.5353535 135.2323 255.0505 0.1313131 0.4242424

yes 52.90909 0.5303030 1.3484848 129.4773 245.7273 0.1590909 0.6212121

thalachh exng oldpeak slp caa thall

no 138.5253 0.5353535 1.5515152 1.222222 1.2020202 2.616162

yes 158.8561 0.1515152 0.5643939 1.583333 0.3333333 2.128788

Coefficients of linear discriminants:

LD1

age 1.153844e-02

sex -7.395488e-01

cp 3.214135e-01

trtbps -1.264734e-02

chol 1.628623e-05

fbs 3.937704e-01

restecg 3.726626e-01

thalachh 1.473600e-02

exng -4.846053e-01

oldpeak -3.111643e-01

slp 1.150970e-01

caa -5.407510e-01

thall -6.276274e-01

>

> pc <- predict(model.LDA, na.roughfix(trainds))

> summary(pc$class)

no yes

85 146

> pc <- predict(model.LDA, na.roughfix(testds))

> summary(pc$class)

no yes

23 35

> xtab <- table(pc$class, testds$output)

> caret::confusionMatrix(xtab, positive = "yes")

Confusion Matrix and Statistics

no yes

no 17 6

yes 8 27

Accuracy : 0.7586

95% CI : (0.6283, 0.8613)

No Information Rate : 0.569

P-Value [Acc > NIR] : 0.002123

Kappa : 0.5031

Mcnemar's Test P-Value : 0.789268

Sensitivity : 0.8182

Specificity : 0.6800

Pos Pred Value : 0.7714

Neg Pred Value : 0.7391

Prevalence : 0.5690

Detection Rate : 0.4655

Detection Prevalence : 0.6034

Balanced Accuracy : 0.7491

'Positive' Class : yes

>

> pb <- NULL

> pb <- pc$posterior

> pb <- as.data.frame(pb)

> pred.LDA <- data.frame(testds$output, pb$yes)

> colnames(pred.LDA) <- c("target","score")

> lift.LDA <- lift(target ~ score, data = pred.LDA, cuts=10, class="yes")

>

>

> xyplot(lift.LDA, main="LDA - Lift Chart", type=c("l","g"), lwd=2, scales=list(x=list(alternating=FALSE,tick.number = 10),y=list(alternating=FALSE,tick.number = 10)))

> labels <- as.factor(ifelse(pred.LDA$target=="yes", 1, 0))

> predictions <- pred.LDA$score

> auc(roc(predictions, labels), min = 0, max = 1)

[1] 0.8315152

> plot(roc(predictions, labels), min=0, max=1, type="l", main="LDA - ROC Chart")

>