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 thalachh

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

exng oldpeak slp caa thall output

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

> head(heartds)

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

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

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

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

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

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

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

> summary(heartds)

age sex cp trtbps chol

Min. :29.00 Min. :0.0000 Min. :0.000 Min. : 94.0 Min. :126

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:120.0 1st Qu.:212

Median :54.00 Median :1.0000 Median :1.000 Median :130.0 Median :243

Mean :54.01 Mean :0.6782 Mean :1.021 Mean :131.4 Mean :248

3rd Qu.:60.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:140.0 3rd Qu.:276

Max. :77.00 Max. :1.0000 Max. :3.000 Max. :200.0 Max. :564

fbs restecg thalachh exng

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

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

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

Mean :0.1453 Mean :0.5156 Mean :150.2 Mean :0.3183

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

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

oldpeak slp caa thall output

Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.000 no :124

1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:165

Median :0.600 Median :1.000 Median :0.0000 Median :2.000

Mean :1.008 Mean :1.419 Mean :0.7128 Mean :2.315

3rd Qu.:1.600 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :6.200 Max. :2.000 Max. :4.0000 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 thalachh

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

exng oldpeak slp caa thall output

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

> head(trainds)

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

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

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

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

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

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

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

> summary(trainds)

age sex cp trtbps chol

Min. :29.00 Min. :0.0000 Min. :0 Min. : 94.0 Min. :126.0

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0 1st Qu.:120.0 1st Qu.:212.5

Median :54.00 Median :1.0000 Median :1 Median :130.0 Median :245.0

Mean :54.28 Mean :0.6623 Mean :1 Mean :131.9 Mean :249.7

3rd Qu.:61.00 3rd Qu.:1.0000 3rd Qu.:2 3rd Qu.:140.0 3rd Qu.:279.5

Max. :77.00 Max. :1.0000 Max. :3 Max. :200.0 Max. :564.0

fbs restecg thalachh exng

Min. :0.0000 Min. :0.0000 Min. : 71.0 Min. :0.000

1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:133.0 1st Qu.:0.000

Median :0.0000 Median :1.0000 Median :155.0 Median :0.000

Mean :0.1472 Mean :0.5368 Mean :150.1 Mean :0.316

3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:168.0 3rd Qu.:1.000

Max. :1.0000 Max. :2.0000 Max. :202.0 Max. :1.000

oldpeak slp caa thall output

Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.000 no : 99

1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:132

Median :0.6000 Median :1.000 Median :0.0000 Median :2.000

Mean :0.9874 Mean :1.429 Mean :0.7056 Mean :2.338

3rd Qu.:1.6000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :5.6000 Max. :2.000 Max. :4.0000 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 thalachh

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

exng oldpeak slp caa thall output

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

> head(testds)

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

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

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

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

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

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

6 44 1 1 130 219 0 0 188 0 0.0 2 0 2 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(rpart)

> fit = rpart(output~., data = heartds, control = rpart.control(cp = 0.0001))

> printcp(fit)

Classification tree:

rpart(formula = output ~ ., data = heartds, control = rpart.control(cp = 1e-04))

Variables actually used in tree construction:

[1] caa chol cp slp thalachh thall

Root node error: 124/289 = 0.42907

n= 289

CP nsplit rel error xerror xstd

1 0.427419 0 1.00000 1.00000 0.067855

2 0.064516 1 0.57258 0.73387 0.063677

3 0.040323 3 0.44355 0.56452 0.058735

4 0.024194 5 0.36290 0.49194 0.055945

5 0.012097 6 0.33871 0.48387 0.055606

6 0.000100 8 0.31452 0.54839 0.058154

> fit.pruned = prune(fit, cp = 0.0002)

> plot(fit.pruned)

> fit.pruned = prune(fit, cp = 0.0002)text(fit.pruned, cex = 0.9, xpd = TRUE)

Error: unexpected symbol in "fit.pruned = prune(fit, cp = 0.0002)text"

> text(fit.pruned, cex = 0.9, xpd = TRUE)

> pred<- predict(fit.pruned, testds, type = "class")

> data.frame(testds ,pred)

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

7 51 1 3 125 213 0 0 125 1 1.4 2 1 2

8 65 0 2 160 360 0 0 151 0 0.8 2 0 2

9 45 1 0 104 208 0 0 148 1 3.0 1 0 2

10 44 1 2 140 235 0 0 180 0 0.0 2 0 2

11 53 0 0 138 234 0 0 160 0 0.0 2 0 2

12 63 0 2 135 252 0 0 172 0 0.0 2 0 2

13 48 1 0 122 222 0 0 186 0 0.0 2 0 2

14 34 1 3 118 182 0 0 174 0 0.0 2 0 2

15 55 0 1 135 250 0 0 161 0 1.4 1 0 2

16 41 1 2 112 250 0 1 179 0 0.0 2 0 2

17 60 0 2 102 318 0 1 160 0 0.0 2 1 2

18 42 0 0 102 265 0 0 122 0 0.6 1 0 2

19 45 0 1 112 160 0 1 138 0 0.0 1 0 2

20 59 1 3 178 270 0 0 145 0 4.2 0 0 3

21 69 1 3 160 234 1 0 131 0 0.1 1 1 2

22 50 0 0 110 254 0 0 159 0 0.0 2 0 2

23 56 1 3 120 193 0 0 162 0 1.9 1 0 3

24 41 1 1 110 235 0 1 153 0 0.0 2 0 2

25 57 1 0 110 201 0 1 126 1 1.5 1 0 1

26 44 0 2 118 242 0 1 149 0 0.3 1 1 2

27 66 1 0 160 228 0 0 138 0 2.3 2 0 1

28 64 1 3 170 227 0 0 155 0 0.6 1 0 3

29 35 1 1 122 192 0 1 174 0 0.0 2 0 2

30 58 1 1 125 220 0 1 144 0 0.4 1 4 3

31 41 1 1 120 157 0 1 182 0 0.0 2 0 2

32 38 1 2 138 175 0 1 173 0 0.0 2 4 2

33 38 1 2 138 175 0 1 173 0 0.0 2 4 2

34 56 1 2 130 256 1 0 142 1 0.6 1 1 1

35 58 1 1 120 284 0 0 160 0 1.8 1 0 2

36 60 1 0 130 206 0 0 132 1 2.4 1 2 3

37 57 1 0 150 276 0 0 112 1 0.6 1 1 1

38 61 0 0 130 330 0 0 169 0 0.0 2 0 2

39 58 1 2 112 230 0 0 165 0 2.5 1 1 3

40 44 1 0 112 290 0 0 153 0 0.0 2 1 2

41 50 1 2 140 233 0 1 163 0 0.6 1 1 3

42 62 0 0 160 164 0 0 145 0 6.2 0 3 3

43 59 1 0 110 239 0 0 142 1 1.2 1 1 3

44 49 1 2 120 188 0 1 139 0 2.0 1 3 3

45 61 0 0 145 307 0 0 146 1 1.0 1 0 3

46 56 1 0 125 249 1 0 144 1 1.2 1 1 2

47 50 1 0 110 239 0 1 126 1 2.8 1 1 3

48 64 1 2 125 309 0 1 131 1 1.8 1 0 3

49 55 1 0 160 289 0 0 145 1 0.8 1 1 3

50 56 1 1 132 184 0 0 105 1 2.1 1 1 1

51 56 0 0 134 409 0 0 150 1 1.9 1 2 3

52 58 1 1 128 259 0 0 130 1 3.0 1 2 3

53 66 0 0 178 228 1 1 165 1 1.0 1 2 3

54 52 1 0 130 283 1 0 103 1 1.6 0 0 3

55 58 1 0 100 234 0 1 156 0 0.1 2 1 3

56 50 1 0 128 204 1 1 156 1 1.0 1 0 0

57 59 1 2 126 218 1 1 134 0 2.2 1 1 1

58 53 1 0 110 335 0 1 143 1 3.0 1 1 3

output pred

1 yes yes

2 yes yes

3 yes yes

4 yes yes

5 yes yes

6 yes yes

7 yes yes

8 yes yes

9 yes yes

10 yes yes

11 yes yes

12 yes yes

13 yes yes

14 yes yes

15 yes yes

16 yes yes

17 yes yes

18 yes yes

19 yes yes

20 yes yes

21 yes no

22 yes yes

23 yes yes

24 yes yes

25 yes yes

26 yes no

27 yes yes

28 yes yes

29 yes yes

30 yes yes

31 yes yes

32 yes yes

33 yes yes

34 no no

35 no yes

36 no no

37 no no

38 no yes

39 no no

40 no no

41 no no

42 no no

43 no no

44 no no

45 no no

46 no no

47 no no

48 no no

49 no no

50 no no

51 no no

52 no no

53 no no

54 no no

55 no no

56 no yes

57 no no

58 no no

> library(caret)

Loading required package: ggplot2

Loading required package: lattice

> confusionMatrix(testds$output,pred)

Confusion Matrix and Statistics

Reference

Prediction no yes

no 22 3

yes 2 31

Accuracy : 0.9138

95% CI : (0.8102, 0.9714)

No Information Rate : 0.5862

P-Value [Acc > NIR] : 3.242e-08

Kappa : 0.8234

Mcnemar's Test P-Value : 1

Sensitivity : 0.9167

Specificity : 0.9118

Pos Pred Value : 0.8800

Neg Pred Value : 0.9394

Prevalence : 0.4138

Detection Rate : 0.3793

Detection Prevalence : 0.4310

Balanced Accuracy : 0.9142

'Positive' Class : no