DATA ANALYTICS – 4027

LAB-11

Name: Hari Krishna P

Reg No: 19BCE7675

DATE: 16/11/2021

Contents:

- **Classification Algorithm**

Submitted to:

Prof . Hari Seetha

Classification Algorithm implementation using R

Decision tree for iris dataset

```
ind <- sample(2, nrow(iris), replace=TRUE, prob=c(0.67, 0.33))

iris.training <- iris [ind==1, 1:4]

iris.trainLabels <- iris [ind==1, 5]

iris.test <- iris [ind==2, 1:4]

iris.testLabels <- iris [ind==2, 5]

install.packages("class")

library(class, lib.loc = "C:/Program Files/R/R-4.1.0/library")

a <- knn(train = iris.training, test = iris.test, cl = iris.trainLabels, k=3)

a

xtab=table(a, iris.testLabels)

install.packages("caret")

library(caret)

confusionMatrix(xtab)
```

Refer

https://www.tutorialspoint.com/r/r_decision_tree.htm for decion tree classification

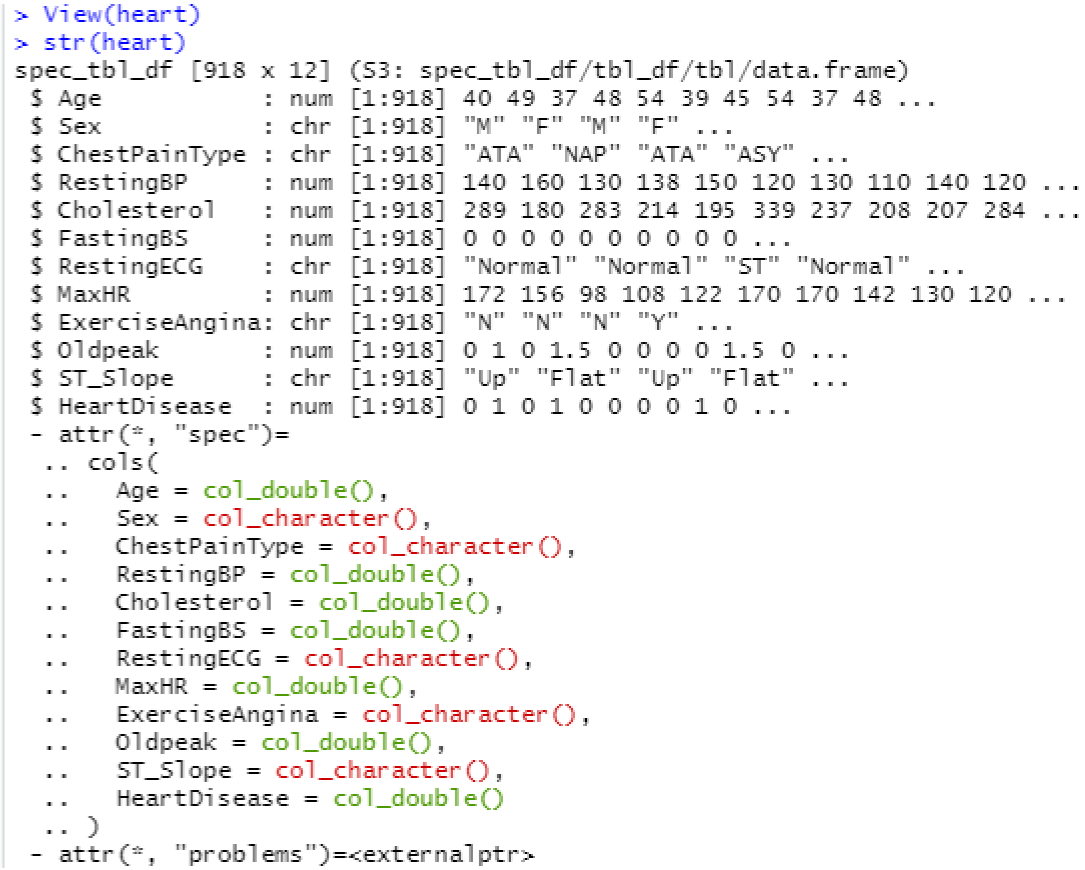
https://www.tutorialspoint.com/r/r_random_forest.htm for random forest classification

https://techvidvan.com/tutorials/svm-in-r/

Implement   classification algorithms on one of the classification datasets available in UCI ML repository and analyze the results

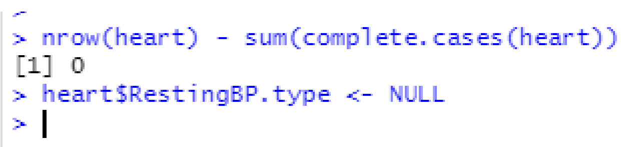
Dataset : Heart.csv

Data set downloaded from Kaggle:



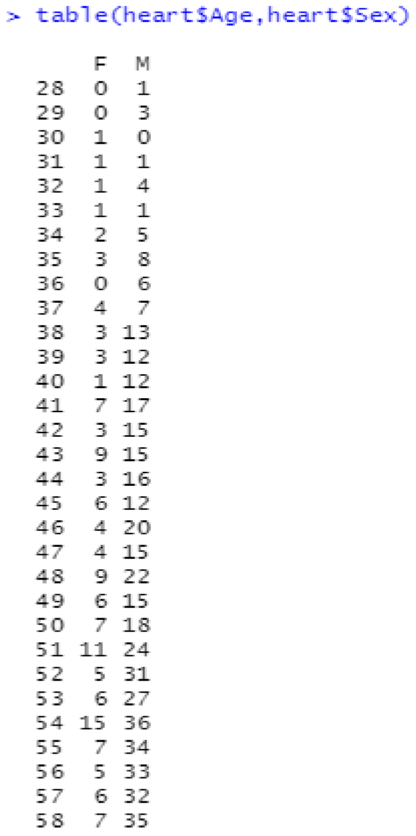
```
> View(heart)
> str(heart)
spec_tbl_df [918 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Age           : num [1:918] 40 49 37 48 54 39 45 54 37 48 ...
 $ Sex           : chr [1:918] "M" "F" "M" "F" ...
 $ ChestPainType : chr [1:918] "ATA" "NAP" "ATA" "ASY" ...
 $ RestingBP     : num [1:918] 140 160 130 138 150 120 130 110 140 120 ...
 $ Cholesterol   : num [1:918] 289 180 283 214 195 339 237 208 207 284 ...
 $ FastingBS     : num [1:918] 0 0 0 0 0 0 0 0 0 0 ...
 $ RestingECG    : chr [1:918] "Normal" "Normal" "ST" "Normal" ...
 $ MaxHR         : num [1:918] 172 156 98 108 122 170 170 142 130 120 ...
 $ ExerciseAngina: chr [1:918] "N" "N" "N" "Y" ...
 $ Oldpeak       : num [1:918] 0 1 0 1.5 0 0 0 0 1.5 0 ...
 $ ST_Slope      : chr [1:918] "Up" "Flat" "Up" "Flat" ...
 $ HeartDisease  : num [1:918] 0 1 0 1 0 0 0 0 1 0 ...
 - attr(*, "spec")=
  .. cols(
  ..   Age = col_double(),
  ..   Sex = col_character(),
  ..   ChestPainType = col_character(),
  ..   RestingBP = col_double(),
  ..   Cholesterol = col_double(),
  ..   FastingBS = col_double(),
  ..   RestingECG = col_character(),
  ..   MaxHR = col_double(),
  ..   ExerciseAngina = col_character(),
  ..   Oldpeak = col_double(),
  ..   ST_Slope = col_character(),
  ..   HeartDisease = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
```

DATA Cleaning:



```
> nrow(heart) - sum(complete.cases(heart))
[1] 0
> heart$RestingBP.type <- NULL
> |
```

Data Exploration and Analysis with plotting:



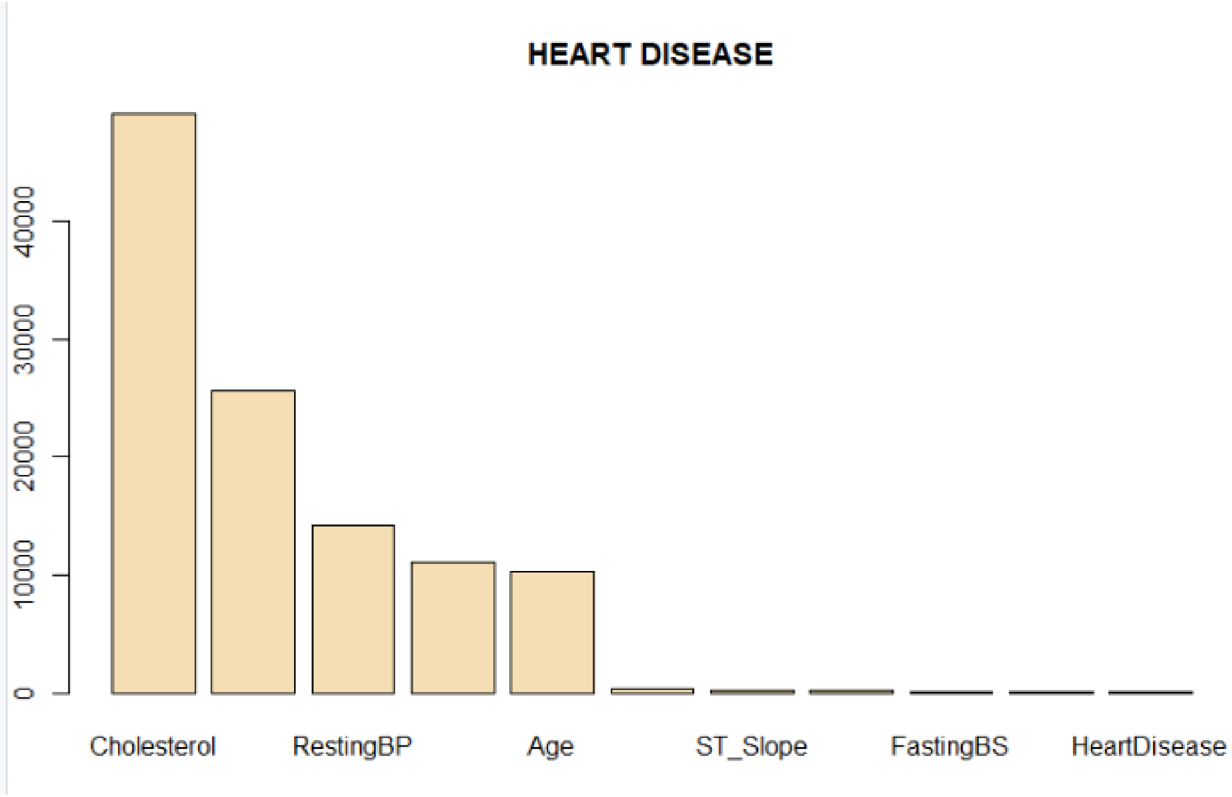
```
> table(heart$Age,heart$Sex)

      F  M
  28  0  1
  29  0  3
  30  1  0
  31  1  1
  32  1  4
  33  1  1
  34  2  5
  35  3  8
  36  0  6
  37  4  7
  38  3 13
  39  3 12
  40  1 12
  41  7 17
  42  3 15
  43  9 15
  44  3 16
  45  6 12
  46  4 20
  47  4 15
  48  9 22
  49  6 15
  50  7 18
  51 11 24
  52  5 31
  53  6 27
  54 15 36
  55  7 34
  56  5 33
  57  6 32
  58  7 35
```

```
number.perfect.splits <- apply(X=heart[-2],MARGIN = 2, FUN =
function(col){

  + a <- table(heart$Cholesterol,col)

  + sum(a == 0)

  + })

 order = order(number.perfect.splits,decreasing = TRUE)

number.perfect.splits <- number.perfect.splits[order]
```

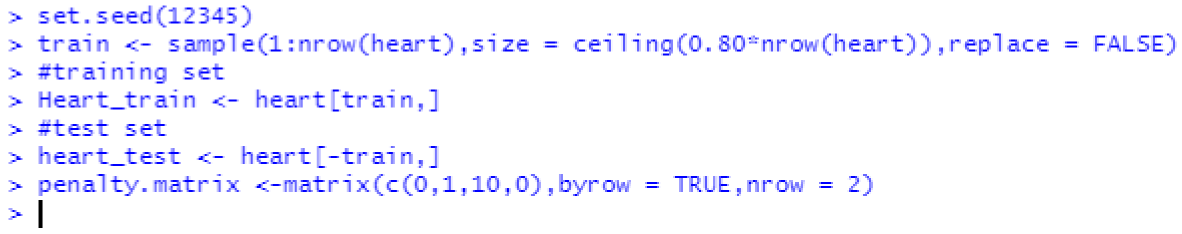
par(mar=c(10,2,4,2))

barplot(number.perfect.splits,main = "HEART DISEASE",xlab =)

barplot(number.perfect.splits,main = "HEART DISEASE",xlab = Heart,ylab = "Age",col = "colors")

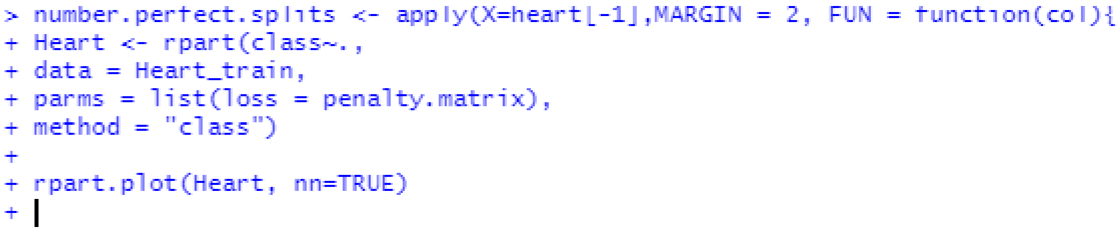


Data Splicing:



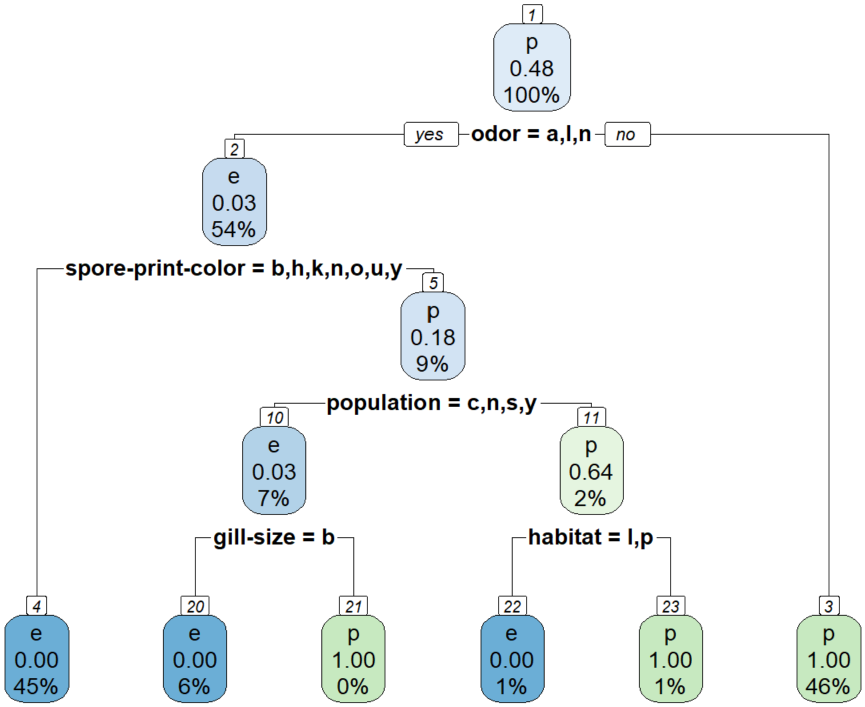
```
> set.seed(12345)
> train <- sample(1:nrow(heart),size = ceiling(0.80*nrow(heart)),replace = FALSE)
> #training set
> Heart_train <- heart[train,]
> #test set
> heart_test <- heart[-train,]
> penalty.matrix <-matrix(c(0,1,10,0),byrow = TRUE,nrow = 2)
> |
```

Building a model:



```
> number.perfect.splits <- apply(X=heart[-1],MARGIN = 2, FUN = function(col){
+ Heart <- rpart(class~.,
+ data = Heart_train,
+ parms = list(loss = penalty.matrix),
+ method = "class")
+
+ rpart.plot(Heart, nn=TRUE)
+ |
```

TREE:



1
p
0.48
100%
yes
odor = a,l,n
no
2
e
0.03
54%
spore-print-color = b,h,k,n,o,u,y
5
p
0.18
9%
population = c,n,s,y
10
e
0.03
7%
11
p
0.64
2%
gill-size = b
habitat = l,p
4
e
0.00
45%
20
e
0.00
6%
21
p
1.00
0%
22
e
0.00
1%
23
p
1.00
1%
3
p
1.00
46%