University : Stevens Institute of Technology

Project : HW_06 C50

Purpose : Homework

First Name : Omkar

Last Name : Sinha

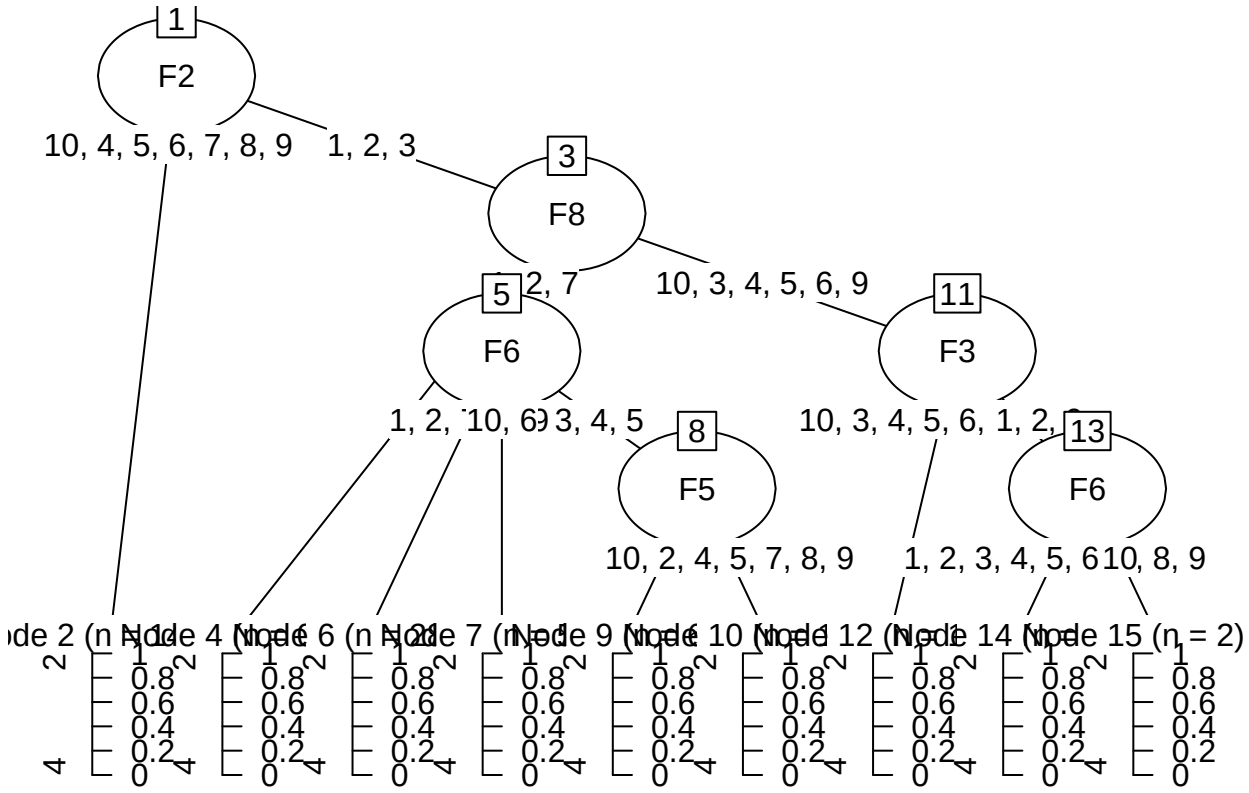
CWId : 10468312

Date : 11/10/2021

```
## Load the data

rm(list=ls())
file<-file.choose()
bc_C50 <-  read.csv(file,na.strings = "?",colClasses=c("Sample"="character",
                                                       "F1"="factor","F2"="factor","F3"="factor",
                                                       "F4"="factor","F5"="factor","F6"="factor",
                                                       "F7"="factor","F8"="factor","F9"="factor",
                                                       "Class"="factor"))


##installing required package
library(C50)
library(ggplot2)



index<-sort(sample(nrow(bc_C50),round(.30*nrow(bc_C50))))
training<-bc_C50[-index,]
test<-bc_C50[index,]


C50_class <- C5.0(Class~.,data=training[,-1])

summary(C50_class)
```

```
##
## Call:
## C5.0.formula(formula = Class ~ ., data = training[, -1])
##
##
## C5.0 [Release 2.07 GPL Edition]      Sun Nov 14 11:06:41 2021
## -------------------------------
```

```
##
## Class specified by attribute 'outcome'
##
## Read 489 cases (10 attributes) from undefined.data
##
## Decision tree:
##
## F2 in {10,4,5,6,7,8,9}: 4 (150/8)
## F2 in {1,2,3}:
## :...F8 = 8: 2 (0)
##     F8 in {1,2,7}:
##     :...F6 in {1,2,7,8,9}: 2 (285.5)
##     :   F6 in {10,6}: 4 (5.1/0.1)
##     :   F6 in {3,4,5}:
##     :   :...F5 in {1,3,6}: 4 (6.1/2.1)
##     :       F5 in {10,2,4,5,7,8,9}: 2 (16.4)
##     F8 in {10,3,4,5,6,9}:
##     :...F3 in {10,3,4,5,6,7,8,9}: 4 (16)
##         F3 in {1,2}:
##         :...F6 in {1,2,3,4,5,6,7,8,9}: 2 (8)
##             F6 = 10: 4 (2)
##
##
## Evaluation on training data (489 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##       8   10( 2.0%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     310    10    (a): class 2
##           169    (b): class 4
##
##
##  Attribute usage:
##
##  100.00% F2
##   69.33% F8
##   64.83% F6
##    5.73% F5
##    5.32% F3
##
##
## Time: 0.0 secs

plot(C50_class)
```




```
C50_predict<-predict(C50_class ,test , type="class")
table(actual=test[,"Class"],C50=C50_predict)
```

```
##       C50
## actual   2   4
##      2 126  12
##      4   0  72
```

```
wrong<- (test[,"Class"]!=C50_predict)
c50_rate<-sum(wrong)/length(test[,"Class"])
c50_rate
```

```
## [1] 0.05714286
```