**Assignment 6**

**6.1: Use the C5.0 methodology to develop a classification model for the Diagnosis.**

# First Name: Utsav

# Last Name: Italiya

# Id : 10475248

rm(list=ls ())

library(C50)

#read data from csv file

df<-read.csv("F:/Sem1/CS513/lecture6/breast-cancer-wisconsin.csv",

na.strings = c("?") ,

colClasses=c("Sample"="character",

"F1"="factor","F2"="factor",

"F3"="factor","F4"="factor",

"F5"="factor","F6"="factor",

"F7"="factor","F8"="factor",

"F9"="factor","Class"="factor"))

df <-na.omit(df)

#70% training and 30% testing data

idx<-sort(sample(nrow(df),as.integer(.70\*nrow(df))))

training<-df[idx,]

testing<-df[-idx,]

#ploting c50 model

c50 <- C5.0(Class~. , training[,-1])

summary(c50)

plot(c50)

#prediction

prediction<-predict(c50,testing[,-1],type="class")

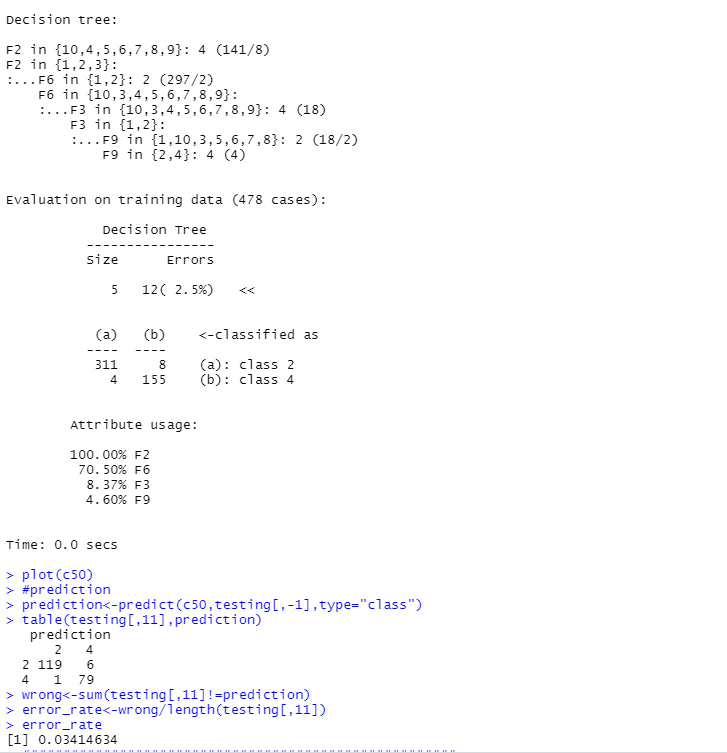
table(testing[,11],prediction)

wrong<-sum(testing[,11]!=prediction)

error\_rate<-wrong/length(testing[,11])

error\_rateDiagram, schematic

Description automatically generated



**6.2: Use the Random Forest methodology to develop a classification model for the Diagnosis and identify important features.**

######################################################

#random forest

######################################################

rm(list=ls())

library(C50)

#read data from csv file

df<-read.csv("F:/Sem1/CS513/lecture6/breast-cancer-wisconsin.csv",

na.strings = c("?") ,

colClasses=c("Sample"="character",

"F1"="factor","F2"="factor",

"F3"="factor","F4"="factor",

"F5"="factor","F6"="factor",

"F7"="factor","F8"="factor",

"F9"="factor","Class"="factor"))

df <-na.omit(df)

#70% training and 30% testing data

idx<-sort(sample(nrow(df),as.integer(.70\*nrow(df))))

training<-df[idx,]

testing<-df[-idx,]

#ploting random model

library(randomForest)

rm <- randomForest( Class~., data=training, importance=TRUE, ntree=1000)

importance(rm)

varImpPlot(rm)

#predictions

prediction<- predict(rm, testing)

table(actual=testing[,11],prediction)

wrong<-sum(testing$Class!=prediction)

error\_rate<-wrong/length(testing[,11])

error\_rate

#succection rate

successrate <- 1 - error\_rate

successrate

Table

Description automatically generated with medium confidence

