

HW_04

Preet Dabhi

3/8/2022

```
#####  
# Company      : Stevens  
# Project      : HW_04  
# Purpose      : Perform Naïve Bayes Classification  
# First Name   : Preet  
# Last Name    : Dabhi  
# Id           : 10459151  
# Date         : 03/6/2022  
  
## Delete all the objects from your R- environment.  
#####  
  
rm(list=ls())  
  
#Import package 'e1071' for Naive Bayes Classifier and class package  
#install.packages("e1071")  
#install.packages(class)  
library(e1071)  
library(class)  
  
dataFrame <- read.csv("D:/SEM 3/CS 513/HW_02/breast-cancer-wisconsin.csv",header=TRUE, sep=",")  
summary(dataFrame)
```

```
##      Sample      F1      F2      F3  
## Min.   : 61634   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000  
## 1st Qu.: 870688   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000  
## Median : 1171710   Median : 4.000   Median : 1.000   Median : 1.000  
## Mean   : 1071704   Mean    : 4.418   Mean    : 3.134   Mean    : 3.207  
## 3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000  
## Max.   :13454352   Max.    :10.000   Max.    :10.000   Max.    :10.000  
##      F4      F5      F6      F7  
## Min.   : 1.000   Min.   : 1.000   Length:699   Min.   : 1.000  
## 1st Qu.: 1.000   1st Qu.: 2.000   Class :character   1st Qu.: 2.000  
## Median : 1.000   Median : 2.000   Mode  :character   Median : 3.000  
## Mean   : 2.807   Mean    : 3.216           Mean    : 3.438  
## 3rd Qu.: 4.000   3rd Qu.: 4.000           3rd Qu.: 5.000  
## Max.   :10.000   Max.    :10.000           Max.    :10.000  
##      F8      F9      Class  
## Min.   : 1.000   Min.   : 1.000   Min.    :2.00  
## 1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:2.00  
## Median : 1.000   Median : 1.000   Median :2.00
```

```
## Mean : 2.867 Mean : 1.589 Mean :2.69
## 3rd Qu.: 4.000 3rd Qu.: 1.000 3rd Qu.:4.00
## Max. :10.000 Max. :10.000 Max. :4.00
```

#Here we can see that by running summary on the dataframe F6 column there are some missing values in it
n <- as.numeric(as.character(dataFrame\$F6))

```
## Warning: NAs introduced by coercion
```

```
dataFrame$F6 <- n

#Remove the rows with missing values
dataFrame <- na.omit(dataFrame)
View(dataFrame)

#Replacing class column 2 and 4 with Benign and Malignant
dataFrame$Class<- factor(dataFrame$Class , levels = c("2","4") , labels = c("Benign","Malignant"))
is.factor(dataFrame$Class)
```

```
## [1] TRUE
```

```
newData<- dataFrame[2:11]
View(newData)

#70% of the sample size
sample_size <- floor(0.70 * nrow(newData))

#Set the seed to make your partition reproducible
set.seed(123)
trainData <- sample(seq_len(nrow(newData)), size = sample_size)

#Loading 70% Breast cancer record in training dataset
training <- newData[trainData, ]

#Loading 30% Breast cancer in test dataset
test <- newData[-trainData, ]

#Implementing NaiveBayes
naive_bayes_model<- naiveBayes(Class ~ ., data = training)

#Predicting target class for the Validation set
naive_bayes_predict <- predict(naive_bayes_model, test)

conf_matrix <- table(predict_nb=naive_bayes_predict,class=test$Class)
print(conf_matrix)
```

```
##           class
## predict_nb Benign Malignant
## Benign      132         1
## Malignant     7        65
```

```
#Output of Naive Bayes Classifier  
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}  
accuracy(conf_matrix)
```

```
## [1] 96.09756
```