Poorvi_Raut_HW04_NaiveBayes.R

Owner

2023-03-01

```
#knowledge Discovery and Data Mining (CS 513) Homework 3: Naive Bayes classification
#Course : CS 513-A
# First Name : Poorvi
#Last Name : Raut
# ID : 20009560
# Purpose : HW_04_NaiveBayes

#clearing object environment
rm(list = ls())
#get working directory
getwd()
```

```
## [1] "C:/Users/Owner/Desktop/Spring 2023/CS 513 KDD"
```

```
#Import package e1071 for Naive Bayes Classifier and class , caret package to calculate confusio
n matrix metrics
library(e1071)
library(class)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```r
#Load the "breast-cancer-wisconsin.data.csv" from canvas into R and perform the Naive Bayes Meth
odology
dataSet<-read.csv("/Users/Owner/Desktop/Spring 2023/CS 513 KDD/breast-cancer-wisconsin.csv",na.s
tring = "?" )
#View Breast Cancer Dataset
View(dataSet)
#remove any rows with missing values
dataSet<-na.omit(dataSet)

View(dataSet)

#Converting the type of column F6 from character to numeric
dataSet$F6<-as.numeric(dataSet$F6)
View(dataSet)

#Converting the class column to factor class
dataSet$Class<-factor(dataSet$Class,levels = c("2","4"),labels = c("benign","malignant"))
is.factor(dataSet$Class)
```

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

```r
dataSet1<-dataSet[2:11]
View(dataSet1)

#partitioning 70% of size
sample_size<-floor(0.70*nrow(dataSet1))

#Set the seed to make your partition reproducible
set.seed(123)
traindata<-sample(seq_len(nrow(dataSet1)),size = sample_size)

# 70% of data in training set
train<-dataSet1[traindata,]

# 30% of data in testing set
test<-dataSet1[traindata,]

#Applying naive Bayes algorithm
Naive_Bayes<-naiveBayes(Class ~.,data=train)

#Predicting target class
predict_alg<-predict(Naive_Bayes,test)

#creating confusion matrix
conf_matrix<-table(predict_naive=predict_alg,class=test$Class)
print(conf_matrix)
```

```
##               class
## predict_naive benign malignant
##     benign        292         4
##     malignant      13       169
```

```
confusionMatrix(predict_alg,test$Class)
```

```
## Confusion Matrix and Statistics
##
##               Reference
## Prediction   benign malignant
##    benign        292         4
##    malignant      13       169
##
##                 Accuracy : 0.9644
##                   95% CI : (0.9437, 0.9791)
##      No Information Rate : 0.6381
##      P-Value [Acc > NIR] : < 2e-16
##
##                    Kappa : 0.9239
##
##   Mcnemar's Test P-Value : 0.05235
##
##              Sensitivity : 0.9574
##              Specificity : 0.9769
##           Pos Pred Value : 0.9865
##           Neg Pred Value : 0.9286
##               Prevalence : 0.6381
##           Detection Rate : 0.6109
##     Detection Prevalence : 0.6192
##        Balanced Accuracy : 0.9671
##
##          'Positive' Class : benign
##
```

```
#Calculating Accuracy of the algorithm
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix)
```

```
## [1] 96.44351
```

```
#Error rate
e<- 100- accuracy(conf_matrix)
print(e)
```

```
## [1] 3.556485
```