MD2201: Data Science

Name of the student: Bhavin Patil **Roll No.** 78

Div: D **Batch:** B-3

Date of performance: 17/05/2022

Experiment No.6

Title: Classifier Performance

Aim: To measure the different performance parameters of a classifier.

Software used: Programming language R.

Data Set: Wisconsin Breast Cancer data set

Code Statement:

Apply KNN to the Wisconsin Breast Cancer data set. Split the data into training and testing samples. Scale the data and find the following

1. Accuracy
2. Sensitivity
3. Specificity
4. Precision

Code:

```
library(class)
dataset <- read.csv("wbc_csv.csv")
str(dataset)
dataset$diagnosis <- as.factor(dataset$diagnosis)
set.seed(123) #to generate the same random number
str(dataset)
r <- sample(nrow(dataset))
wbc <- dataset[r,]

#View(dataset)
#View(wbc)

wbc_mod <- wbc[,3:32]
```

```
#normalising the dataset
n2 <- function(b){
  (b-min(b))/(max(b)-min(b))
}

wbc_mod <- as.data.frame(lapply(wbc_mod, n2))

View(wbc_mod)

train <- wbc_mod[1:469,]
test <- wbc_mod[470:569,]
train_label <- wbc[1:469,2]
test_label <- wbc[470:569,2]


#KNN

p<-knn(train,test,train_label,k=7)
t<-table(Actual = test_label, Predicted = p)
print(t)

#Accuracy
accurary <- sum(diag(t)/sum(t))
cat("\n\nAccuracy is : ",accurary)


#Recall/Sensitivity
Re<-t[2,2]/sum(t[2,])
cat("\n\nRecall/Sensitivity is : ",Re)

#Specificity
spe<-t[1,1]/sum(t[1,])
cat("\n\nSpecificity is : ",spe)

#Precision
pr<-t[2,2]/sum(t[,2])
cat("\n\nPrecision is : ",pr)
```

Results:

Actual \ Predicted	B	M
B	50	1
M	4	45

Accuracy is : 0.95

Recall/Sensitivity is : 0.9183673

Specificity is : 0.9803922

Precision is : 0.9782609

Conclusion: We have successfully measured the difference performance of parameter of a classifier using knn and confusion matrix by splitting the given dataset into training and testing sample and also identified the accuracy, sensitivity, specificity and precision.