Experiment-2.2

Student Name: Himanshu
UID: 20BCS7944
Branch: CSE
Section: 905/A
Semester: 6
Date of Performance: 12/04/2023
Subject Name: Data Mining Lab
Subject Code: 20CSP-376

1) Aim:

To perform the classification by Naive Bayesian classification algorithm.

2) Objective:

Classifying the data using Naive Bayesian on a pre dataset.

3) Code:

```
lilibrary(e1071)

library(caTools)

dataset = read.csv('/home/heefe/DMClassWork/Social_Network_Ads.csv')

dataset = dataset[3:5]

dataset$Purchased = factor(dataset$Purchased, levels = c(0, 1))

library(caTools)

split = sample.split(dataset$Purchased, SplitRatio = 0.75)

training_set = subset(dataset, split == TRUE)

test_set = subset(dataset, split == FALSE)
```

```
training_set[-3] = scale(training_set[-3])

test_set[-3] = scale(test_set[-3])

classifier = naiveBayes(x = training_set[-3], y = training_set$Purchased)

print(classifier)

y_pred_train = predict(classifier, newdata = training_set[-3])

cm_train = table(training_set[, 3], y_pred_train)

print(cm_train)

accuracy_train <- sum(diag(cm_train))/sum(cm_train)

cat("\nAccuracy on training set: ", accuracy_train)

y_pred_test = predict(classifier, newdata = test_set[-3])

cm_test = table(test_set[, 3], y_pred_test)

print(cm_test)

accuracy_test <- sum(diag(cm_test))/sum(cm_test)

cat("\nAccuracy on test set: ", accuracy_test)
```

4) Output:

Console | Terminal | Background Jobs

R 4.2.3 · ~/

```
> dataset = read.csv('/home/heete/Documents/DMClassWork/Social_Network_Ads.csv')
> dataset = dataset[3:5]
> summary(dataset)
      Age        EstimatedSalary    Purchased
 Min.   :18.00   Min.   : 15000   Min.   :0.0000
 1st Qu.:29.75   1st Qu.: 43000   1st Qu.:0.0000
 Median :37.00   Median : 70000   Median :0.0000
 Mean   :37.66   Mean   : 69742   Mean   :0.3575
 3rd Qu.:46.00   3rd Qu.: 88000   3rd Qu.:1.0000
 Max.   :60.00   Max.   :150000   Max.   :1.0000
> str(dataset)
'data.frame':   400 obs. of  3 variables:
 $ Age            : int  19 35 26 27 19 27 27 32 25 35 ...
 $ EstimatedSalary: int  19000 20000 43000 57000 76000 58000 84000 150000 33000 65000 ...
 $ Purchased      : int  0 0 0 0 0 0 0 1 0 0 ...
> dataset$Purchased = factor(dataset$Purchased, levels = c(0, 1))
> library(caTools)
> split = sample.split(dataset$Purchased, SplitRatio = 0.75)
> training_set = subset(dataset, split == TRUE)
> test_set = subset(dataset, split == FALSE)
> training_set[-3] = scale(training_set[-3])
> test_set[-3] = scale(test_set[-3])
> library(e1071)
> classifier = naiveBayes(x = training_set[-3],
+                         y = training_set$Purchased)
> print(classifier)
```

Console | Terminal | Background Jobs

R 4.2.3 · ~/

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = training_set[-3], y = training_set$Purchased)

A-priori probabilities:
training_set$Purchased
        0         1
0.6433333 0.3566667

Conditional probabilities:
                        Age
training_set$Purchased       [,1]      [,2]
                     0 -0.4553839 0.7760952
                     1  0.8213934 0.8192208

                        EstimatedSalary
training_set$Purchased       [,1]      [,2]
                     0 -0.2588014 0.7288695
                     1  0.4668099 1.2314915
```

```
> y_pred_train = predict(classifier, newdata = training_set[-3])
> cm_train = table(training_set[, 3], y_pred_train)
> print(cm_train)
   y_pred_train
      0   1
  0 179  14
  1  21  86
> accuracy_train <- sum(diag(cm_train))/sum(cm_train)
> cat("\nAccuracy on training set: ", accuracy_train)

Accuracy on training set:  0.8833333
> y_pred_test = predict(classifier, newdata = test_set[-3])
> cm_test = table(test_set[, 3], y_pred_test)
> print(cm_test)
   y_pred_test
     0  1
  0 60  4
  1  6 30
> accuracy_test <- sum(diag(cm_test))/sum(cm_test)
> cat("\nAccuracy on test set: ", accuracy_test)

Accuracy on test set:  0.9
>
```