**Experiment 6**

**Student Name:** Sahil Kaundal **UID:** 21BCS8197

**Branch:** CSE (Lateral Entry)  **Section/Group:** 616/A

**Semester:** 6th **Date of Performance:** 27/04/2023

**Subject Name:** Data Mining Lab **Subject Code:** 20CSP-376

1. **Aim:**

To perform classification using Bayesian classification algorithm using R.

1. **Apparatus / Simulation Used:**

* Windows 7 or above
* R Studio

1. **Objective:**

* Demonstrate of the Bayesian classification algorithm using R.
* Performing classification using Bayesian classification algorithm.

1. **Theory and Output:**

Naive Bayes is a Supervised Non-linear classification algorithm in [R Programming](https://www.geeksforgeeks.org/introduction-to-r-programming-language/). Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Baye’s theorem with strong (Naive) independence assumptions between the features or variables. The Naive Bayes algorithm is called “Naive” because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

Naive Bayes algorithm is based on Bayes theorem. Bayes theorem gives the conditional probability of an event A given another event B has occurred.

* ***where,****P(A|B) = Conditional probability of A given B.   
  P(B|A) = Conditional probability of B given A.   
  P(A) = Probability of event A.   
  P(B) = Probability of event B.*

For many predictors, we can formulate the posterior probability as follows:

***P(A|B)****= P(B1|A) \* P(B2|A) \* P(B3|A) \* P(B4|A) …*

**Example** 

Consider a sample space:

          {HH, HT, TH, TT}

**where,**

H: Head

T: Tail

P(Second coin being head given  = P(A|B)

first coin is tail) = P(A|B)

= [P(B|A) \* P(A)] / P(B)

= [P(First coin is tail given second coin is head) \*

   P(Second coin being Head)] / P(first coin being tail)

= [(1/2) \* (1/2)] / (1/2)

= (1/2)

= 0.5

1. **Code:**

# Naive Bayes

# Importing the dataset

dataset = read.csv('D:/CU-College/Sem 6/Data Mining/Social\_Network\_Ads.csv')

dataset = dataset[3:5]

# Encoding the target feature as factor

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

# Splitting the dataset into the Training set and Test set

library(caTools)

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

training\_set = subset(dataset, split == TRUE)

test\_set = subset(dataset, split == FALSE)

# Feature Scaling

training\_set[-3] = scale(training\_set[-3])

test\_set[-3] = scale(test\_set[-3])

# Fitting Naive Baiyes Classifier to the Training set

library(e1071)

classifier = naiveBayes(x = training\_set[-3],

y = training\_set$Purchased)

print(classifier)

# Predicting train set results

y\_pred\_train = predict(classifier, newdata = training\_set[-3])

# Making the Confusion Matrix for training set

cm\_train = table(training\_set[, 3], y\_pred\_train)

print(cm\_train)

#Accuracy on training data

accuracy\_train <- sum(diag(cm\_train))/sum(cm\_train)

cat("\nAccuracy on training set: ", accuracy\_train)

# Predicting the Test set results

y\_pred\_test = predict(classifier, newdata = test\_set[-3])

# Making the Confusion Matrix for testing set

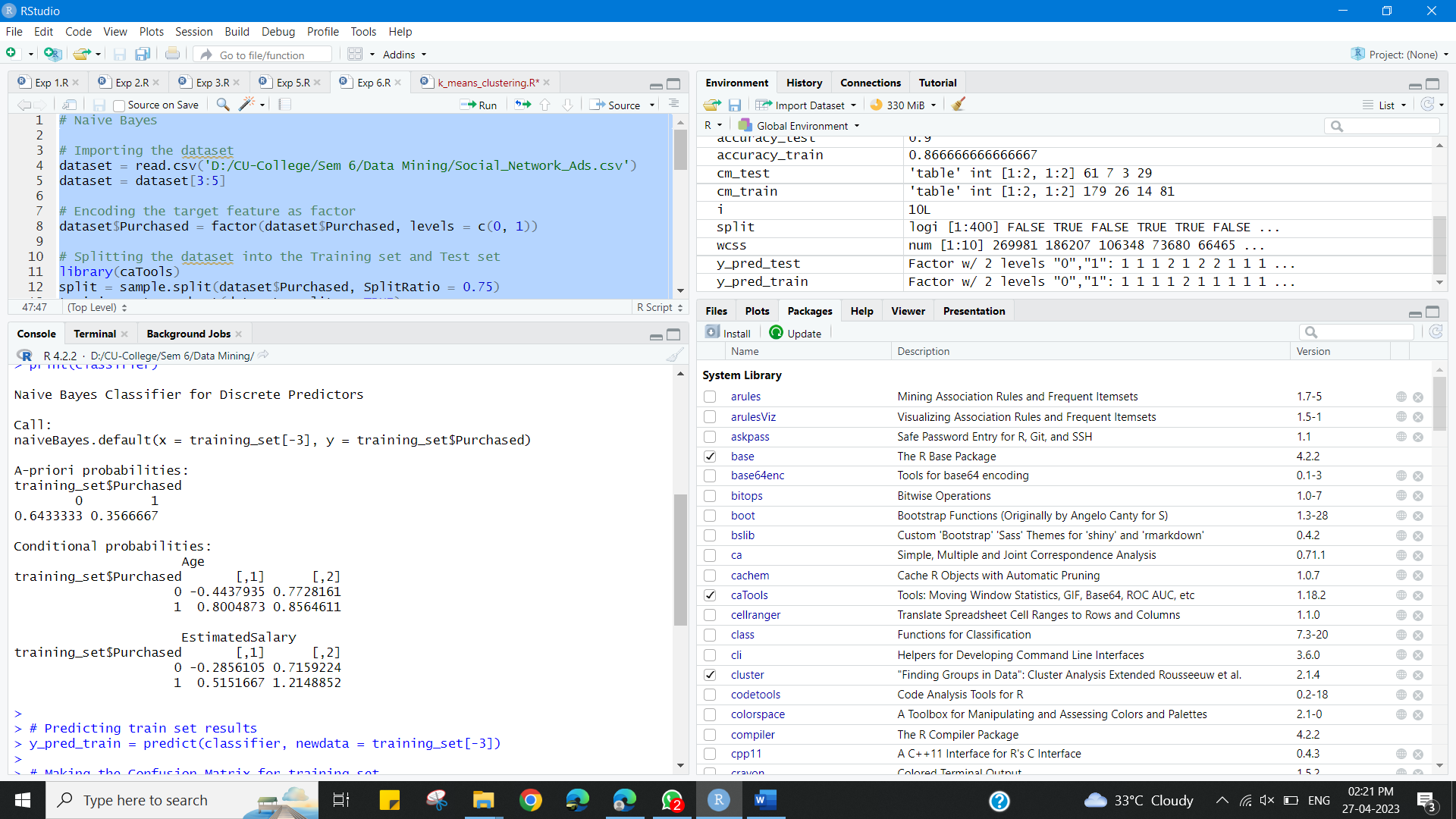
cm\_test = table(test\_set[, 3], y\_pred\_test)

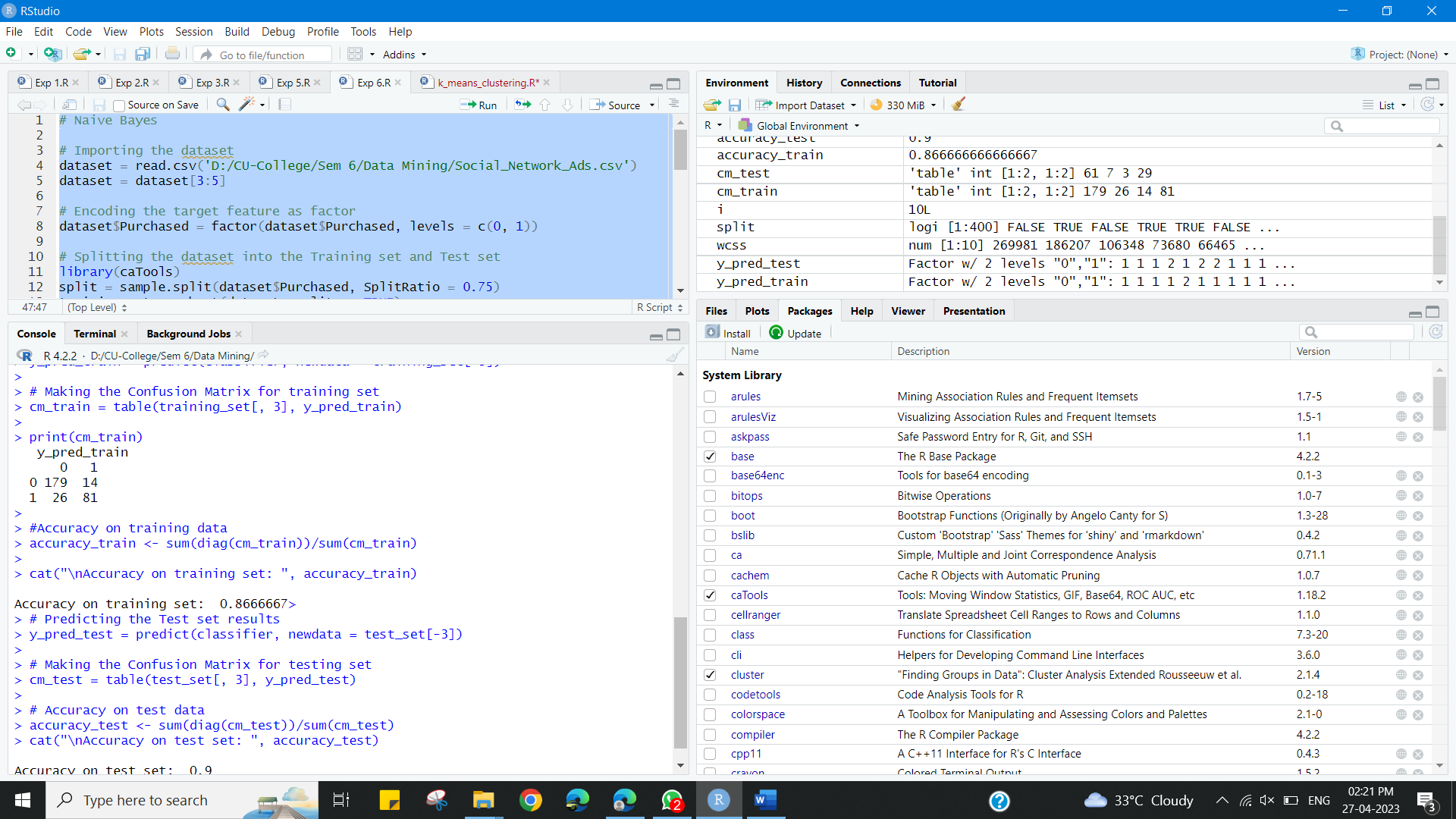
# Accuracy on test data

accuracy\_test <- sum(diag(cm\_test))/sum(cm\_test)

cat("\nAccuracy on test set: ", accuracy\_test)

1. **Output:**

****

****

**Learning outcomes (What I have learnt):**

* Demonstrate of the Bayesian classification algorithm using R.
* Performing classification using Bayesian classification algorithm.