**Exercise – 1**

**Implement Machine Learning Techniques for Predictive analytics**

**Program**:

install.packages(c("tm", "SnowballC", "caret", "e1071", "wordcloud"))

library(tm)

library(SnowballC)

library(caret)

library(e1071)

library(wordcloud)

url <- "https://archive.ics.uci.edu/ml/machine-learning-databases/00228/smsspamcollection.zip"

download.file(url, destfile = "smsspamcollection.zip")

unzip("smsspamcollection.zip")

data <- read.csv("SMSSpamCollection", sep = "\t", header = FALSE, col.names =

c("label", "text"))

data$label <- factor(data$label)

corpus <- VCorpus(VectorSource(data$text))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removeWords, stopwords("en"))

corpus <- tm\_map(corpus, stripWhitespace)

corpus <- tm\_map(corpus, stemDocument)

dtm <- DocumentTermMatrix(corpus)

dtm <- removeSparseTerms(dtm, 0.99)

set.seed(123)

trainIndex <- createDataPartition(data$label, p = 0.8, list = FALSE)

trainData <- dtm[trainIndex, ]

testData <- dtm[-trainIndex, ]

trainLabels <- data$label[trainIndex]

testLabels <- data$label[-trainIndex]

classifier <- naiveBayes(as.matrix(trainData), trainLabels)

predictions <- predict(classifier, as.matrix(testData))

confusionMatrix(predictions, testLabels)

wordcloud(words = colnames(as.matrix(dtm)),

freq = colSums(as.matrix(dtm)),

min.freq = 2,

colors = brewer.pal(8, "Dark2"))

**output**:



