Experiment 8

AIM: Write a program to perform Sentiment Analysis part 2

Code:

```
library("wordcloud")

positive <- subset(df,sentiment=="Positive")

head(positive)

wordcloud(positive$text, max.words = 100, scale = c(3,0.5))

negative <- subset(df,sentiment=="Negative")

head(negative)

wordcloud(negative$text, max.words = 100, scale = c(3,0.5))

neutral <- subset(df,sentiment=="Neutral")

head(neutral)

wordcloud(neutral$text, max.words = 100, scale = c(3,0.5))

convert_count <- function(x) {

  y <- ifelse(x > 0, 1,0)

  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))

  y

}

datasetNB <- apply(dtm, 2, convert_count)

dataset = as.data.frame(as.matrix(datasetNB))

dataset$Class = df$sentiment

dataset<-na.omit(dataset)

str(dataset$Class)

head(dataset)

dim(dataset)

set.seed(222)

split = sample(2,nrow(dataset),prob = c(0.75,0.25),replace = TRUE)
```
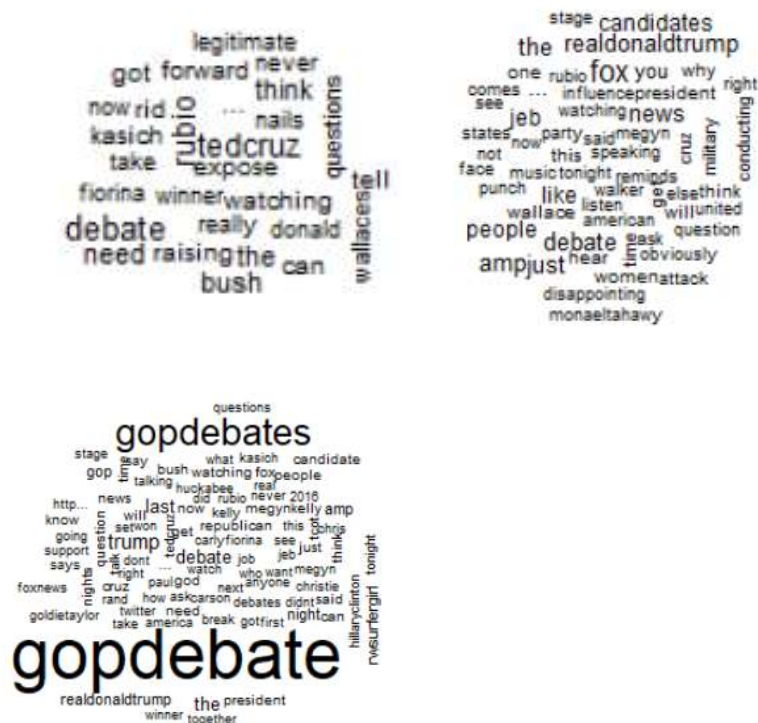
```
train_set = dataset[split == 1,]
test_set = dataset[split == 2,]


prop.table(table(train_set$Class))
prop.table(table(test_set$Class))
library(e1071)
svm_classifier <- svm(Class~., data=train_set,type = "C")
svm_classifier
svm_pred = predict(svm_classifier,test_set)
confusionMatrix(svm_pred,test_set$Class)
```

OUTPUT

```
> library("wordcloud")
> positive <- subset(df,sentiment=="Positive")
> head(positive)
                                                                        text
2  RT @ScottWalker: Didn't catch the full #GOPdebate last night. Here are some of Scott's best line
 in 90 seconds. #Walker16 http://t.co/ZSfFâ€¦
4  RT @RobGeorge: That Carly Fiorina is trending -- hours after HER debate -- above any of the men
n just-completed #GOPdebate says she's on â€¦
5  RT @DanScavino: #GOPDebate w/ @realDonaldTrump delivered the highest ratings in the history of p
esidential debates. #Trump2016 http://t.coâ€¦
6          RT @GregAbbott_TX: @TedCruz: "On my first day I will rescind every illegal executive acti
n taken by Barack Obama." #GOPDebate @FoxNews
12           RT @WayneDupreeShow: Just woke up to tweet this out #GOPDebate \n\nBest line of the nig
t via @GovMikeHuckabee http://t.co/6OV5hxHIcV
15                         RT @pattonoswalt: I loved Scott Walker as Mark Harmon's romantic
ival in SUMMER SCHOOL. Look it up. #GOPDebate
   sentiment
2   Positive
4   Positive
5   Positive
6   Positive
12  Positive
15  Positive
> wordcloud(positive$text, max.words = 100, scale = c(3,0.5))
Warning messages:
1: In tm_map.SimpleCorpus(corpus, tm::removePunctuation) :
  transformation drops documents
2: In tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x, tm::stopwords())) :
  transformation drops documents
> negative <- subset(df,sentiment=="Negative")
> head(negative)
                                                                        text
7  RT @warriorwoman91: I liked her and was happy when I heard she was going to be the moderator. No
 anymore. #GOPDebate @megynkelly  https://â€¦
9          Deer in the headlights RT @lizzwinstead: Ben Carson, may be the only brain surgeon who h
s performed a lobotomy on himself. #GOPDebate
10                     RT @NancyOsborne180: Last night's debate proved it! #GOPDebate #BATsAsk @B
dassTeachersA #TBATs  https://t.co/G2gGjY1bJD
11                                            @JGreenDC @realDonaldTrump In all fai
```

```
24   Neutral
33   Neutral
> wordcloud(neutral$text, max.words = 100, scale = c(3,0.5))
Warning messages:
1: In tm_map.SimpleCorpus(corpus, tm::removePunctuation) :
   transformation drops documents
2: In doTryCatch(return(expr), name, parentenv, handler) :
   invalid input 'ðÿ‡ºðÿ‡¸' in 'utf8towcs'
3: In doTryCatch(return(expr), name, parentenv, handler) :
   invalid input 'ðÿ‡ºðÿ‡¸' in 'utf8towcs'
4: In tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x, tm::stopwords())) :
   transformation drops documents
> convert_count <- function(x) {
+   y <- ifelse(x > 0, 1,0)
+   y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
+   y
+ }
> datasetNB <- apply(dtm, 2, convert_count)
Warning messages:
1: In doTryCatch(return(expr), name, parentenv, handler) :
   invalid input 'ðÿ‡ºðÿ‡¸' in 'utf8towcs'
2: In doTryCatch(return(expr), name, parentenv, handler) :
   invalid input 'ðÿ‡ºðÿ‡¸' in 'utf8towcs'
> dataset = as.data.frame(as.matrix(datasetNB))
> dataset$Class = df$sentiment
> dataset<-na.omit(dataset)
> str(dataset$Class)
 chr [1:13871] "Neutral" "Positive" "Neutral" "Positive" "Positive" "Positive" "Negative" ...
> head(dataset)
        â€¦ â€" â€œ aâ€¦ abc abl abort absolut abt accept accord across act actual add address
        admit adult advantag advic afford afraid agenda ago agre ahead air alien allenwestrepub
        allow almost alreadi also alway amaz amend america american among amp amymek analysi angri
        anim announc anoint anoth answer anymor anyon anyth anyway apolog appear appl applaus arent
        argument around articl ask ass asshol attack attempt attend attent audienc away awesom
        awkward babi back bad balanc band bankrupt bankruptci bar bare base bash basi basic batsask
        beat becam becom begin behind believ bell ben bencarson benefit berni berniesand best bet
        better bettyfckinwhit bias biff big bigger biggest bill billhemm billionair bimbo
        bipartisan birth bit bitch black blacklivesmatt blame bless blow bnrdebat bodi bomb boo
        book boom bor border bore boy brag brain break bretbaier brian brilliant bring brother
        brotherhood brought brutal budget build bulli bunch bush busi buy call calm cam came
        campaign can canada cancel candid cant canttrustabush car card care carlyfiorina carri
        carson case cat catch caus ccot certain chanc chang charact check cheer cherri child
        superpow support suppos sure surpris surreal system take takeaway taken talk talk teot
        teacher team teaparti ted tedcruz tell ten tenmen tequila term terribl terrifi terror
        terrorist thank that thatchrisgor theatr thebaxterbean thedailyshow thedemocrat thepatriot
        there theskimm theyr thing thingsmik think though thought threat three throw ticket tie
        tight time tlot today togeth told tomorrow tone tonight took toot top topic total touch
        tough toward tran transcript treat tri trip troll troop true trump trust truth tune turn
        tweet twice twitter twitterland two ugh unborn understand unfair unfathom union unit
        uniteblu unless unprofession until usa use usual vagina valu vet veteran via victim video
        view viewer vine violenc voic vote voter wage wait wakeupamerica walk walker wall wallac
        wanna want war washington washingtonpost wasnt wassermanschultz watch water way wed week
        well went weve what whererwomen whether whine white whole whos wilberforc will william win
        winner wisconsin wish without woman women womensright won wonder wont word work world worri
        wors worst worth wouldnt wow writeintrump wrong wtf yall yeah year yell yep yes yesterday
        yet young your youtub yrs zero zinger Class
[ reached 'max' / getOption("max.print") -- omitted 6 rows ]
> dim(dataset)
[1] 13871  1219
> set.seed(222)
> split = sample(2,nrow(dataset),prob = c(0.75,0.25),replace = TRUE)
> train_set = dataset[split == 1,]
> test_set = dataset[split == 2,]
>
> prop.table(table(train_set$Class))

  Negative    Neutral   Positive
 0.6146983  0.2282452  0.1570565
> prop.table(table(test_set$Class))

  Negative    Neutral   Positive
 0.6049884  0.2212877  0.1737239
> library(e1071)
> svm_classifier <- svm(Class~., data=train_set,type = "C")
```