THE UNIVERSITY OF BURDWAN

Topic: Sentiment analysis on Self Derive Cars

Name : - Keya Mondal

Roll No : - 180312200063

Reg. No : - 201801015855 of 2018-19

Semester : - 6th (B.Sc.)

Subject : - Statistics

Course : - Project Work (DSE-4)

Session : - 2020-21

CONTENTS

Acknowledgement	3
Initial Proposal	4
Why twitter sentiment analysis	5
Abstract	5
Introduction	5
Data Acquisition	6
Objective	7
Implementing sentiment analysis application in R	7
Text pre-processing on tweets	8
Self –confidence label of the tweets	9
Tweets sentiment analysis	10
Word Cloud Visualization	12
R code	13
Conclusion	16
References	16

ACKNOWLEDGEMENT

Presentation, Inspiration and Motivation have always played a key role in the success any venture.

Primarily I would like to thank the supreme power the Almighty God who is obviously the one who has always guided me to work on the right path of life. I am greatly oblige in taking the opportunity to sincerely thanks our professors for helping and encouraging in this project even in this pandemic situation of covid19.

Lastly, I thank almighty, my parents, brother and friends for their constant encouragement without which this project not be possible.

I have no valuable words to express my thanks, but my heart is still full of favours receive from every persons.

Sentiment Analysis on Self Derive Cars

Initial Proposal:

Social Media has become an integral part of our life and thus, a mean to express our opinions about the things happening around us. If stats are to be believed on an average a person spends about 142 minutes daily on social media sites and there are about 3.77 billion active users! With the advancement in technology certain tools have been developed that can be used to understand people's emotions and classify them .I will explain Sentiment Analysis and also demonstrate how it can be used to analyze Social Media data using R. My primary goal is to classify twits as positive, negative, or neutral, and identify these words. I will implement learned material and implemented in the spirit of the class.

- Scrape twitter for data regarding the Self Drive Cars (message, date, Maybe geographical location)
- After cleaning the data, I will use to store information
- Analysis is going to be perform by querying and using ggplot2
- I will use R, Twitter, R packages (tidyr, dplyr, syuzhet, ggplot2)

Why twitter sentiment analysis?

Sentiment Analysis Dataset Twitter has a number of applications:

Business: Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customers' feelings towards products or brand, how people respond to their campaigns or product launches and also why consumers are not buying certain products.

Politics: In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analyzing election results.

Public Actions: Twitter Sentiment Analysis also is used for monitoring and analyzing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere.

Abstract

Mining social network data and developing user profile from unstructured and informal data are a challenging task. The proposed research builds user profile using Twitter data which is later helpful to provide the user with personalized recommendations. Publicly available tweets are fetched and classified and sentiments expressed in tweets are extracted and normalized. This research uses domain-specific seed list to classify tweets. Semantic and syntactic analysis on tweets is performed to minimize information loss during the process of tweets classification. After precise classification and sentiment analysis, the system builds user interest-based profile by analyzing user's post on Twitter to know about user interests.

Introduction

In the last decade, social networks have witnessed multifold advancements due to the rapid digitization of the service industry and other advancements in the field of information technology. A plethora of information sharing platforms and the increased connectivity with the Internet have also led to a change in the general perspective of networking, socialization, and personalization . For the month of December 2018, an average of 1.52 billion users were active on Facebook daily . This is besides auxiliary services offered by Facebook, such as WhatsApp, Messenger, and Instagram, each of which has over 1 billion active users, per month . Similarly, as identified from third-party reports, other platforms, such as YouTube owned by Google, iMessage by Apple, and WeChat by Tencent, are also a part of the, no longer elite, 1 billion-per-month-active-user-club. More significantly, three out of every four adult Internet users are now actively utilizing at least one social network platform . From a pure technological point of view, this enhanced connectivity has created unique challenges and opportunities by allowing the users to not only consume services but also to share their experiences, feelings, and thoughts. One of the most impactful and emerging social networks is Twitter, which allows its users to broadcast the latest (personal, communal, national, or international) events in the form of short messages, "tweets," which are typically comprised of text, audiovisual content, and/or links to external websites . Twitter is playing a key role in many fields such as social marketing , election campaigns , academia , and news.

However, such large use of social media has also introduced the problem of information overload. With an overwhelming amount of data on social media, users find it difficult to get personalized and concise information. Short and noisy text on social media also makes it hard to understand full context and classify data. In this paper, we propose a framework for providing personalized recommendations to the user by analyzing his health interest on social networks. While this work can be generalized in many domains, the research work presented henceforth is focused on processing self-driving data and information.

The proposed classification and sentiment analysis system uses a semantic structure, important keywords, and opinion words from tweets to monitor user interests and then generates personalized healthcare and wellness-related tweet recommendations. These personalized tweets consist of publicly available content which is precisely preclassified by our system. For tweet classification, the proposed system uses a domain-specific seed list which helps to decide which category a particular tweet belongs to. After classification, the proposed system also applies a lexicon-based sentiment analysis approach to extract topic level sentiments in tweets. To increase the accuracy of tweet analysis, the proposed system also uses synonyms with keywords. The proposed model performs more precise analyses of tweets enriching temporal patterns and semantics of keywords which optimize filtering result and help to extract more knowledge from tweets. For testing of profile generation, we collected 7015 tweets of users and generated user profile by extracting car-related keywords, entities, and sentiments.

Data Acquisition:

A simple Twitter sentiment analysis job where contributors read tweets and classified them as very positive, slightly positive, neutral, slightly negative, or very negative. They were also prompted asked to mark if the tweet was not relevant to self-driving cars. Added: June 8, 2015 by CrowdFlower | Data Rows: 7015.

Twitter-sentiment-self-drive-DFE.csv

Request more info

COLUMN NAME	TYPE
unit_id	integer
golden	boolean
unit_state	string
trusted_judgments	integer
last_judgment_at	datetime
sentiment	string
sentiment_confidence	decimal
our_id	integer
sentiment_gold	string
sentiment_gold_reason	string
text	string

Downloaded Now:

Source: https://www.crowdflower.com/data-for-everyone/

Data downloaded from "CrowdFlower" [Reference 1. Link]

Column Description:

unit_id: A unit id is the unique identifier user.

golden: Such as logical value (T|F)

unit_state: Such as string value relation with "golden" column (golden | finalized)

trusted_judgment_at: Such as numeric value

last Judgment At : This column define as date of Tweets (Y – M –D)

sentiment: This is users rating.

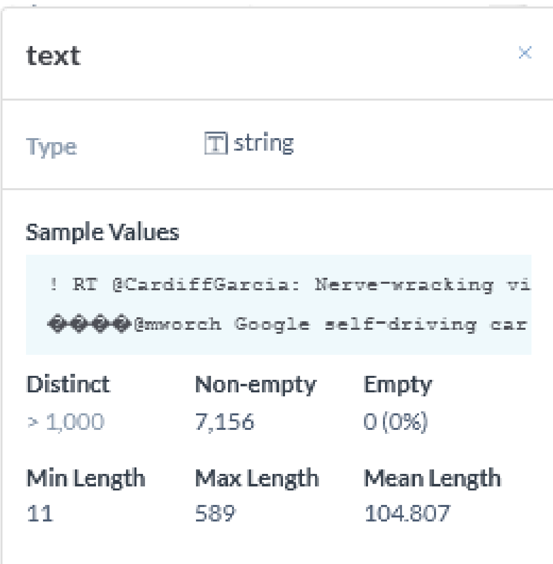
Sentiment Confidence: The polarity indicator takes on positive, neutral or negative as values and the polarity confidence is a number between 0-1

Our_id : a unique id creating on "unit_id" column.

Sentiment Gold: "Gold" here is short for "gold standard," so in this context it refers to sentiment tags that human reviewers have manually assigned to the data.

Sentiment_gold_resason: User's opinion based on self-car.

Text: *Users* share these *tweets*



An overview on text data using word cloud graph. A word cloud is "an image composed of words used in a particular text or subject, in which the size of each word indicates its frequency or importance."

Objective

1. Business impact of self-driveling cars.
2. Sentiment analysis on tweets.

Implementing sentiment analysis application in R

Now, we will try to analyze the sentiments of tweets made by a Twitter handle. We will develop the code in R step by step and see the practical implementation of sentiment analysis in R.

The code is divided into following parts:

- Extracting tweets using Twitter application
- Cleaning the tweets for further analysis
- Getting sentiment score for each tweet
- Segregating positive and negative tweet

Text pre-processing on tweets

Before we dive into analyzing text, we need to pre-process it. Text data contains white spaces, punctuations, stop words etc. These characters do not convey much information and are hard to process. For example, English stop words like “the”, “is” etc. do not tell you much information about the sentiment of the text, entities mentioned in the text, or relationships between those entities. Depending upon the task at hand, we deal with such characters differently. This will help isolate text mining in R on important words.

```
> df1$text[20:50]
 [1] "Audi Gets First Permit to Test Driverless Cars in California: Some Bay Area drag queens say they are... http://t.co/Bc5mlrrgEJ"
 [2] "Today state regs for autonomous cars go into effect: Audi of America gets 1st permit to test driverless cars on public roads, per @CA_DMV."
 [3] "Audi gets first permit to test self-driving cars in California: Think twice next time you tailgate that new Audi... http://t.co/JWTioM2nJT"
 [4] "Audi Gets First Permit to Test Driverless Cars in California: Some Bay Area drag queens say they are... http://t.co/YGoAqLEhUP"
 [5] "Audi Gets Permit to Test Self-Driving Cars in California http://t.co/V62MR5IaVN"
 [6] "I¢%âÂ\u0081ÒAudi becomes first automaker given permit to test self-driving cars. http://t.co/jNMKLEWrkT http://t.co/dqQ6NJfJlf\" @babydellz"
 [7] "Audi Snags First Permit To Test Self-Driving Cars In California: LOS ANGELES (CBS/AP) I¢%âÂ%Ò\u009d Audi is the first... http://t.co/PrtyPZYBSE"
 [8] "Here we go! @Audi gets first permit to test self-driving cars on California roads http://t.co/XDcPbgMAXh http://t.co/QOouYVOPLK #tech #auto"
 [9] "AGV (Automated Guided Vehicle)  @ TraPac Dock http://t.co/E4xiHwSb"
[10] "AGV (Automated Guided Vehicle)  @ TraPac Dock http://t.co/WCdgvwIj"
[11] "AGV (Automated Guided Vehicle) @ TraPac Dock http://t.co/WCdgvwIj"
[12] "Come join us at the SAE On-Road Automated Vehicle Standards Committee Open Meeting"
[13] "At the SAE On-Road Automated Vehicle Standards Committee Open Meeting"
[14] "Not possible for blind people to drive an automated vehicle that requires intervention during failure. Need L4 automation for that#AutoAuto"
[15] "Transportation becoming a service model? If OEMs are responsible for automated vehicle control, they might as well own it #AutoAuto"
[16] "Really good presentation from Jan Becker on Bosch's automated vehicle research. #AutoAuto check it out"
[17] "Automated vehicle vs connected vehicle #MITTRSummit http://t.co/5RqYp4eKb2"
[18] "WO2013070799A1 Automated Vehicle Conveyance Apparatus Transportation System #B61B #B61B13 http://t.co/HToCWDVEZM"
[19] "How is an automated vehicle supposed to behave in an emergency? #AutoAuto"
[20] "#ExperienceBosch My first ride on the automated vehicle is so surreal. Staring at the analysis, weI¢%âÂå_ http://t.co/tCYqtoHMgE"
[21] "Chris Urmson, director of Self-Driving Cars at Google, speaking at Automated Vehicle Symposium 2014. http://t.co/YFWlfzM39Z"
[22] "How is an automated vehicle supposed to behave in an emergency? #AutoAuto"
[23] "#AnnArbor is home to ongoing experiment w/ talking #cars? \"@freep: Automated vehicle test site groundbreaking today http://t.co/QwbtJb5Bwy\""
[24] "Thank the gods for automated vehicle condition evaluation kits. :D"
[25] "WO2012158248A1 Collaborative Vehicle Control Using Both Human Operator And Automated Controller ... #B60Q #B60Q1 http://t.co/Ku4mwYor"
[26] "Wow - \"@AUVSI: Google goes commercial: Company teams with Continental for automated vehicle systems... http://t.co/YHLH9Pq2Xa\""
[27] "Blogged: Automated vehicles and the Motor Vehicle Code: http://t.co/P5ccubJO"
[28] "California's new automated vehicle regulations (smartly) do not define the \"operator\" of automated vehicles. New term is \"designee.\""
[29] "Last year, Google warned it would have an automated vehicle  \"commercialized product in the not-too-distant future.\"  http://t.co/aAmWYEWiMf"
[30] "@ford video: sensors on Fusion Hybrid Automated Research Vehicle  https://t.co/ThUNZSkOqN"
[31] "Ford just revealed it's Automated Ford Fusion Hybrid Vehicle. Pretty amazing. #fordtrends @ Ford TestI¢%âÂå_ http://t.co/7axya8ogIW"
>
```

I’m done some text-mining pre-process on the user tweets text data.

Remove all special characters in the tweets.

Remove digits from the tweets.

Remove stop-words from the tweets.

Remove web link from the tweets.

Remove extra space/ tab delimiter from tweets.

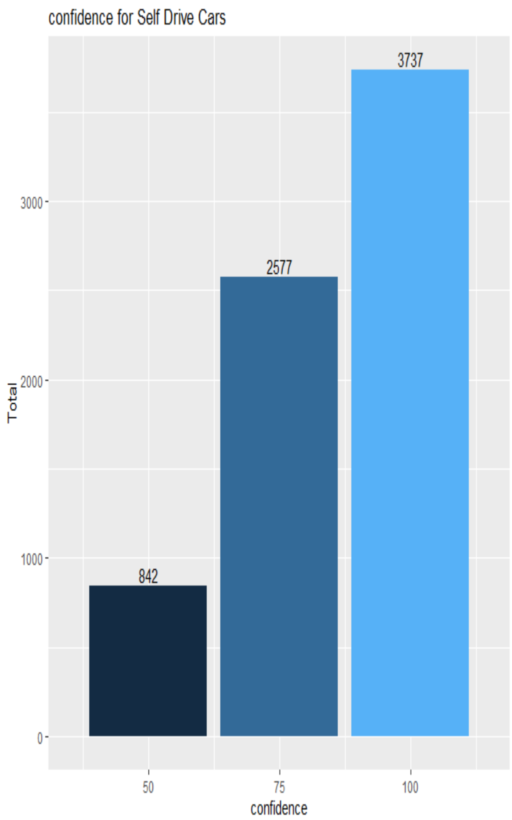
Those are my first challenging part in the text data. Then I'm use the text for sentiment analysis.

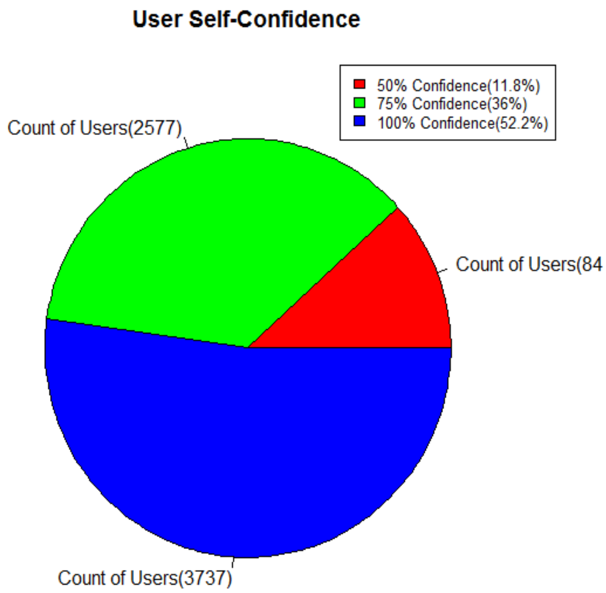
```
> df1$cleantext[20:50]
 [1] "Audi Gets First Permit to Test Driverless Cars in California Some Bay Area drag queens say they are"
 [2] "Today state regs for autonomous car go into effect Audi of America gets st permit to test driverless car on public roads per"
 [3] "Audi gets first permit to test selfdriving car in California Think twice next time you tailgate that new Audi"
 [4] "Audi Gets First Permit to Test Driverless Cars in California Some Bay Area drag queens say they are"
 [5] "Audi Gets Permit to Test SelfDriving Cars in California"
 [6] "I âA 0Audi becomes first automaker given permit to test selfdriving car"
 [7] "Audi Snags First Permit To Test SelfDriving Cars In California LOS ANGELES CBSAP I âA 0  Audi is the first"
 [8] "Here we go gets first permit to test selfdriving car on California roads tech auto"
 [9] "AGV Automated Guided Vehicle TraPac Dock"
[10] "AGV Automated Guided Vehicle TraPac Dock"
[11] "AGV Automated Guided Vehicle TraPac Dock"
[12] "Come join us at the SAE OnRoad Automated Vehicle Standards Committee Open Meeting"
[13] "At the SAE OnRoad Automated Vehicle Standards Committee Open Meeting"
[14] "Not possible for blind people to drive an automated vehicle that requires intervention during failure Need L automation for thatAutoAuto"
[15] "Transportation becoming a service model If OEMs are responsible for automated vehicle control they might as well own it AutoAuto"
[16] "Really good presentation from Jan Becker on Boschs automated vehicle research AutoAuto check it out"
[17] "Automated vehicle vs connected vehicle MITTRSummit"
[18] "WOA Automated Vehicle Conveyance Apparatus Transportation System BB BB"
[19] "How is an automated vehicle supposed to behave in an emergency AutoAuto"
[20] "ExperienceBosch My first ride on the automated vehicle is so surreal Staring at the analysis weI âÂå"
[21] "Chris Urmson director of SelfDriving Cars at Google speaking at Automated Vehicle Symposium"
[22] "How is an automated vehicle supposed to behave in an emergency AutoAuto"
[23] "AnnArbor is home to ongoing experiment w talking car Automated vehicle test site groundbreaking today"
[24] "Thank the gods for automated vehicle condition evaluation kits D"
[25] "WOA Collaborative Vehicle Control Using Both Human Operator And Automated Controller  BQ BQ"
[26] "Wow  Google goes commercial Company teams with Continental for automated vehicle systems"
[27] "Blogged Automated vehicles and the Motor Vehicle Code"
[28] "Californias new automated vehicle regulations smartly do not define the operator of automated vehicles New term is designee"
[29] "Last year Google warned it would have an automated vehicle commercialized product in the nottoodistant future"
[30] "video sensors on Fusion Hybrid Automated Research Vehicle"
[31] "Ford just revealed its Automated Ford Fusion Hybrid Vehicle Pretty amazing fordtrends Ford TestI âÂå"
>
```

Self –confidence label of the tweets

Confidence is a state of being clear-headed either that a hypothesis or prediction is correct or that a chosen course of action is the best or most effective. Confidence comes from a Latin word 'fidere' which means "to trust"; therefore, having self-confidence is having trust in one's self. Arrogance or hubris, in comparison, is the state of having unmerited confidence – believing something or someone is capable or correct when they are not.

The concept of self-confidence is commonly used as self-assurance in one's personal judgment, ability, power, etc. One's self confidence increases from experiences of having satisfactorily completed particular activities. It is a positive belief that in the future one can generally accomplish what one wishes to do. The reproduced approaches are also combined in an ensemble, averaging the individual classifiers' confidence scores for the three classes and deciding sentiment polarity based on these averages.

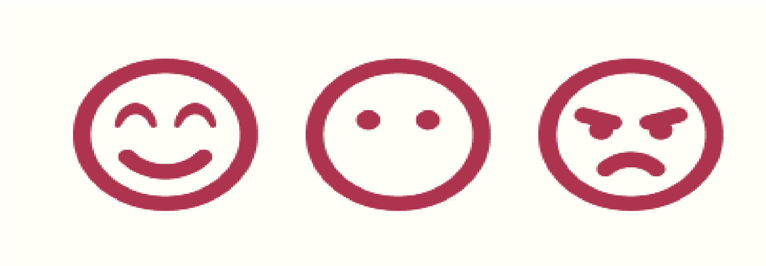





```
> summary(df1$confidence_label)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  50.00   75.00  100.00   85.11  100.00  100.00 
> 
```

This image represents the level of confidence of the users, their comments show confidence levels are over 70% so this is a valid data.

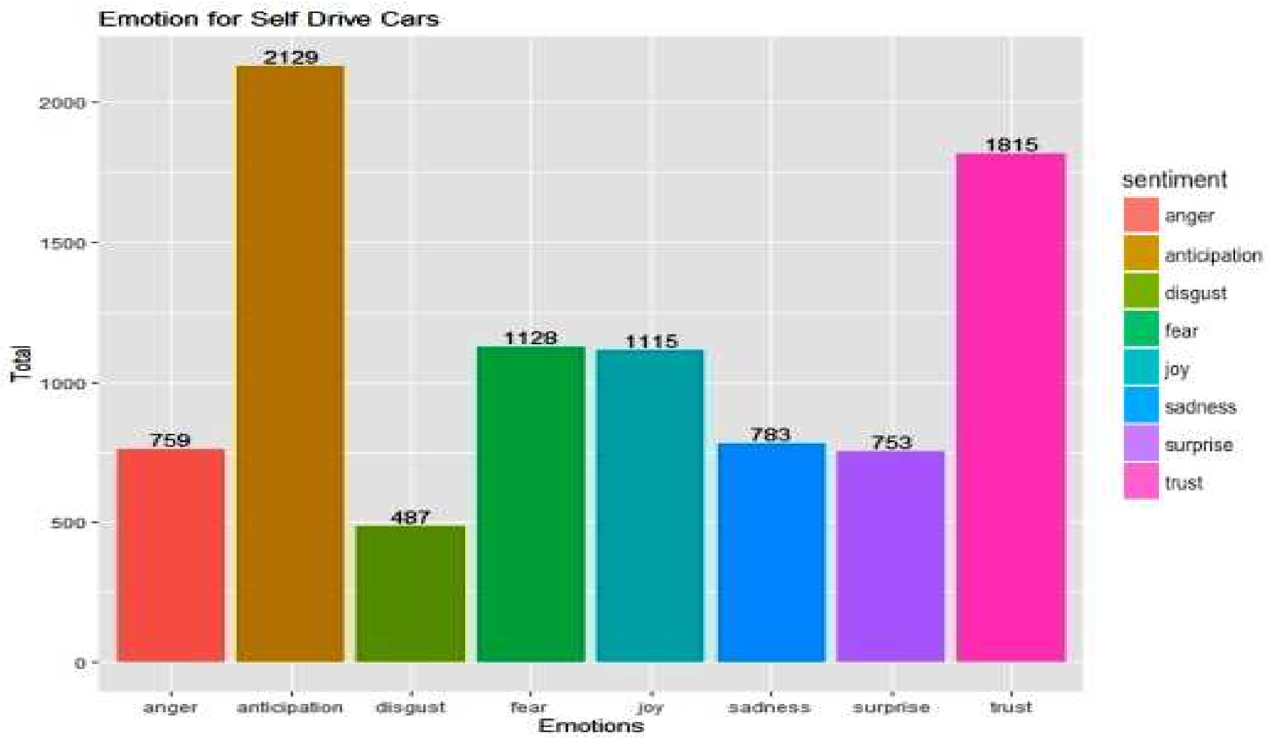
Tweets sentiment analysis



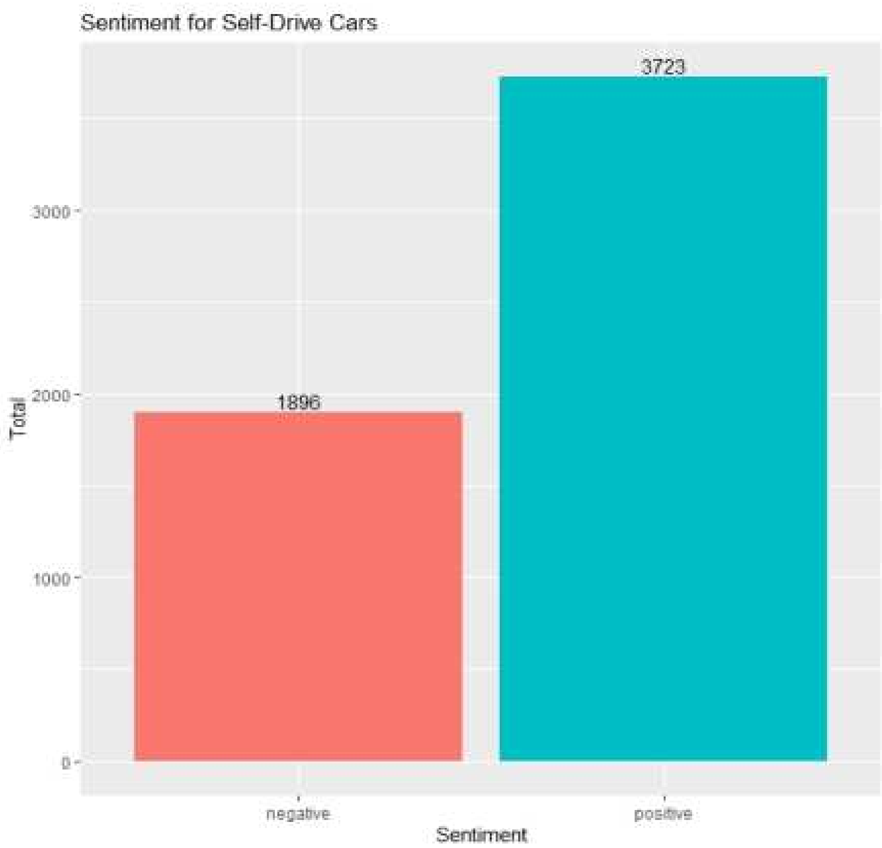
A common and intuitive approach to text is sentiment analysis. In a grand sense, we are interested in the emotional content of some text, e.g. posts on Facebook, tweets, or movie reviews. Most of the time, this is obvious when one reads it, but if you have hundreds of thousands or millions of strings to analyze, you'd like to be able to do so efficiently.

“syuzhet” uses NRC Emotion lexicon. The NRC emotion lexicon is a list of words and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

The **get_nrc_sentiment** function returns a data frame in which each row represents a sentence from the original file. The columns include one for each emotion type was well as the positive or negative sentiment valence. It allows us to take a body of text and return which emotions it represents — and also whether the emotion is positive or negative.



The gist is that we are dealing with a specific, pre-defined vocabulary. Of course, any analysis will only be as good as the lexicon. The goal is usually to assign a sentiment score to a text, possibly an overall score, or a generally positive or negative grade. Given that, other analyses may be implemented to predict sentiment via standard regression tools or machine learning approaches.

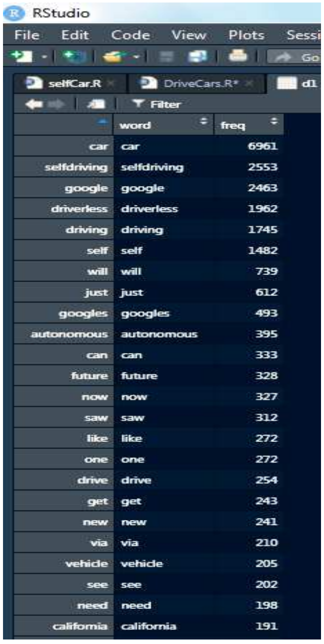


So, now we have analysed the twitter handle Self Driving Cars and got the sentiment around tweets. The break of total number of tweets by sentiment (negative – 1896 and positive-3723)

Word Cloud Visualization

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analyzing data from social network websites.

STEP 1: Retrieving the data and uploading the packages: To generate word clouds, you need to download the wordcloud package in R as well as the RcolorBrewer package for the colours. Note that there is also a wordcloud2 package, with a slightly different design and fun applications. I will show you how to use both packages.

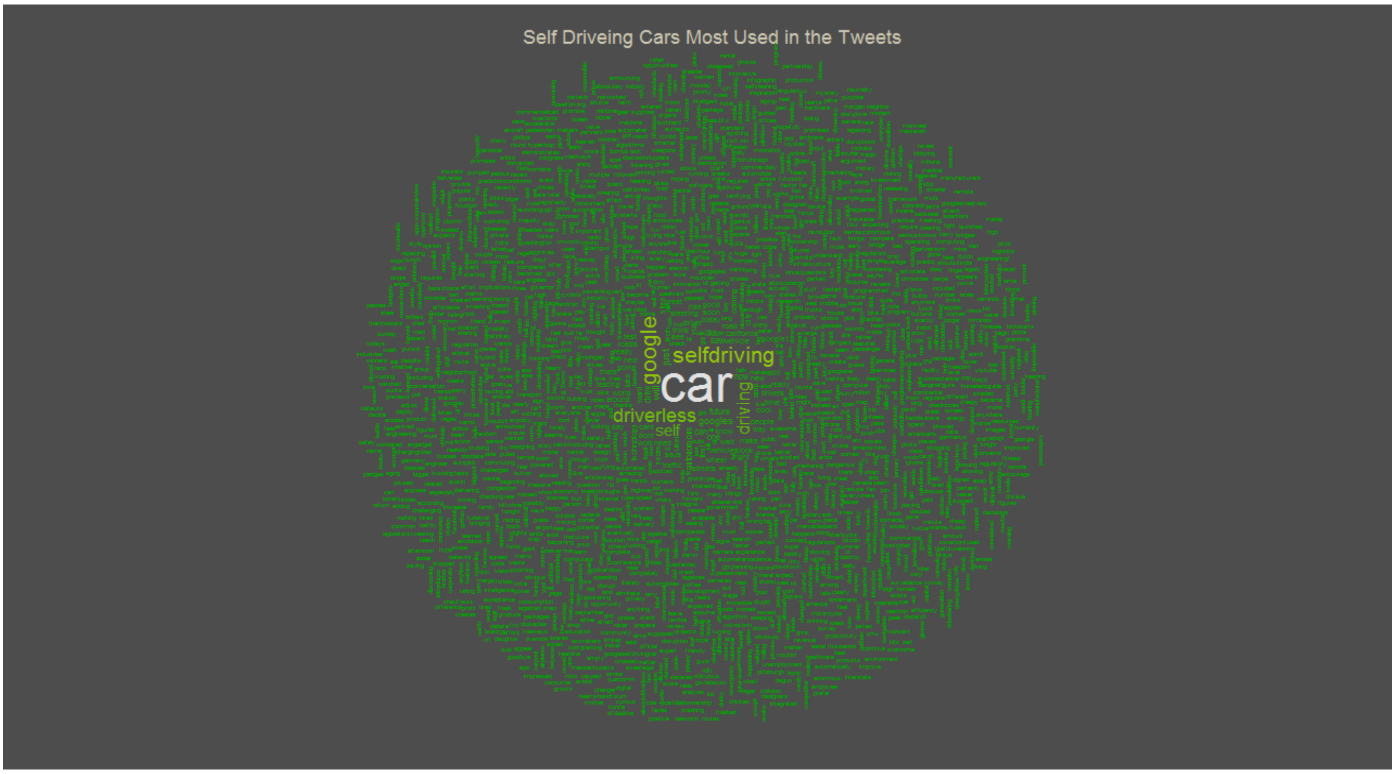


	word	freq
car	car	6961
selfdriving	selfdriving	2553
google	google	2463
driverless	driverless	1962
driving	driving	1745
self	self	1482
will	will	739
just	just	612
googles	googles	493
autonomous	autonomous	395
can	can	333
future	future	328
now	now	327
saw	saw	312
like	like	272
one	one	272
drive	drive	254
get	get	243
new	new	241
via	via	210
vehicle	vehicle	205
see	see	202
need	need	198
california	california	191

STEP 2: Clean the text data: Cleaning is an essential step to take before you generate your wordcloud. Indeed, for your analysis to bring useful insights, you may want to remove special characters, numbers or punctuation from your text. In addition, you should remove common stop words in order to produce meaningful results and avoid the most common frequent words such as “I” or “the” to appear in the word cloud.

STEP 3: Create a document-term-matrix: The next step is to have a dataframe containing each word in your first column and their frequency in the second column. This can be done by creating a document term matrix with the TermDocumentMatrix function from the tm package.

STEP 4: Generate the word cloud: The wordcloud package is the most classic way to generate a word cloud. The following line of code shows you how to properly set the arguments.



R code:

```
## sentiment analysis of twitter data..................
## @Keya Mondal /2021/06/19 .......................
options(repos = c(CRAN = "http://cran.rstudio.com")) #  repo-path setup
# install packages.............................
install.packages("syuzhet")
install.packages("tidyr")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("wordcloud")
install.packages("tm")
# libraries
library(RColorBrewer)
library(wordcloud)
library(ggplot2)
library(syuzhet)
library(tidyr)
library(dplyr)
library(tm)
## data input...................
cardata <- "D:/textAnalysis/Data/Twitter-sentiment-self-drive-DFE.csv" #datapath
lempath <- 'D:/textAnalysis/Data/lemmatization-en.txt'# Lemmatization  path
df<-read.csv(cardata,header=TRUE) # read data frame in csv format
df1 <- df[,-c(2:6,8,9,10),drop=FALSE] # drop data column
# Function..............................
# read lemmatization data
lemword <- scan(lempath,what='',sep='\n')
lemword<- strsplit(lemword,"[[:space:]]+")
fst_word<- sapply(lemword, `[[`, 1)# first element in a list
sec_word<- sapply(lemword, `[[`, 2)# Extract 2nd-vector element as list
# lemmatization function
lemword<-function(x1,fst_word,sec_word){
 linew<-unlist(strsplit(x1," "))
 word<- which(sec_word %in% linew)
 if(length((word)>0)){
   for(i in length(word)){
     linew[which(sec_word[word[i]] == linew)] <- fst_word[[word[i]]]
   }
   line<-paste(linew,collapse = " ")
 }
 return(x1)
}
```

```
# analysis part...................
# wordcloud on text
wordcloud(words = df$text,max.words=500, random.order=FALSE, rot.per=0.35,
        colors=brewer.pal(8, "Dark2"))
# clean text in the tweets raw data
df1['cleantext']<-""
for( i in 1:nrow(df1)){
 x2 <- df1$text[i]
 x2 <- tolower(x2) # all lower cases
 x2 <- lemword(x2,fst_word,sec_word) # using lemmatization
 line1 <-unlist(strsplit(x2," ")) #using split
 line1  <-line1[lapply(line1,function(x)  length(grep("@",x,value=FALSE)))  ==  0]  #  Remove  "@"  with
words
 line1 <- line1[lapply(line1,function(x) length(grep("http",x,value=FALSE))) == 0]# Remove weblink
 line1 <- gsub(", "," ",toString(line1))
 line1 <- gsub("(?<=[\\s])\\s*|^\\s+|\\s+$", "", line1, perl=TRUE)
 cleanTweet <- gsub("[[:punct:]]", "", line1) # remove punctuation
 cleanTweet <- gsub("[[:digit:]]", "", cleanTweet) # remove numbers/Digits
 cleanTweet = lemword(cleanTweet,fst_word,sec_word)# call lemmmatization
 cleanTweet <- gsub("googleì???ââ???ãs", " ", cleanTweet) # # remove a special word
 cleanTweet <- gsub("âââ", "", cleanTweet) # remove a special word
 cleanTweet <- gsub("cars", "car", cleanTweet) # remove tabs
 cleanTweet <- gsub("[^[:graph:]]"," ",cleanTweet)
 cleanTweet <- gsub(" $", "", cleanTweet) # remove blank spaces at the end
 df1$cleantext[i] <- cleanTweet
}
#  analysis on sentiment.confidence...............
# create frequency table
df1["confidence_label"]<- ifelse(df1$sentiment.confidence<=0.5,50,
                        ifelse(df1$sentiment.confidence<=0.75 & df1$sentiment.confidence>0.5,75,100))
x<-table(df1$confidence_label)
summary(df1$confidence_label)
# histogram plot on confidence_lables
plotData3 =gather(df1,"confidence","values",5)%>%
 group_by(df1$confidence_label) %>% summarise(Total=n())
names(plotData3)=c("confidence","Total")
# Plot
ggplot(data = plotData3, aes(x = plotData3$confidence, y = plotData3$Total),cex=1.5) +
 geom_bar(aes(fill = confidence), stat = "identity") +
 theme(legend.position = "none") +
 xlab("confidence") + ylab("Total") + ggtitle("confidence for Self Drive Cars")+
 geom_text(aes(label =plotData3$Total), position = position_dodge(width=30), vjust = -0.25,hjust=0.5)
```

```
# Pie chart on confidence_lables
pie_percent<- round(100*x/sum(x), 1)
labels <- c("Count of Users(842)", "Count of Users(2577)", "Count of Users(3737)")
pie(x,labels,main = "User Self-Confidence", col = rainbow(length(x)))
legend("topright",c("50% Confidence(11.8%)", "75% Confidence(36%)", "100% Confidence(52.2%)"),cex=0.8,fill = rainbow(length(x)))

# The sentiment analysis algorithm used here is based on the Word-Emotion Association...............
text_sentiment<- get_nrc_sentiment(df1$cleantext)
df2 <- cbind(df1,text_sentiment)# column join
plotData1 =gather(df2,"sentiment","values",6:13)%>%
 group_by( sentiment) %>% summarise(Total = sum(values))# sentiment/emotion
# Plot
ggplot(data = plotData1, aes(x = plotData1$sentiment, y = plotData1$Total)) +
 geom_bar(aes(fill = sentiment), stat = "identity") +
 theme(legend.position = "none") +
 xlab("Emotions") + ylab("Total") + ggtitle("Emotion for Self Drive Cars")+
 geom_text(aes(label =   plotData1$Total), position = position_dodge(width=0.75), vjust = -0.25)
# Positive vs negetive text
plotData2 =gather(df2,"Polarity","values",14:15)  %>%
 group_by( Polarity) %>%
 summarise(Total = sum(values))
# plot
ggplot(data = plotData2, aes(x = plotData2$Polarity, y = plotData2$Total)) +
 geom_bar(aes(fill = plotData2$Polarity), stat = "identity") +
 theme(legend.position = "none") +
 xlab("Sentiment") + ylab("Total") + ggtitle("Sentiment for Self-Drive Cars")+
 geom_text(aes(label =   plotData2$Total), position = position_dodge(width=0.75), vjust = -0.25)

# word cloud on modified text
x <- VectorSource(df1$cleantext)# text to word vector
x <- VCorpus(x)# create a corpus object
dtm <- TermDocumentMatrix(x, control = list(removePunctuation = TRUE,removeNumbers = TRUE, stopwords = TRUE))# docoment matrix on text
m <- as.matrix(dtm) # store a matrix
v <- sort(rowSums(m),decreasing=TRUE) # decreasing order sorting
d <- data.frame(word = names(v),freq=v)# create adata frame
d1 <- d[0:2000,]# consider ony first 2000 text
# plotting the word colud
par(bg='grey30')
png(file='WordCloud.png',width=1280,height=720, bg='grey30')
wordcloud(d1$word, d1$freq, col=terrain.colors(length(d$word), alpha=0.9), random.order=FALSE, rot.per=0.3 )
title(main = 'Self Driveing Cars Most Used in the Tweets', font.main = 1, col.main = 'cornsilk3', cex.main = 1.5)
dev.off()
```

Conclusion

In an early stage of developing emerging technologies, there is often great uncertainty regarding their future success. Companies can reduce this uncertainty by listening to the voice of customers as the customer eventually decides to accept an emerging technology or not. By analyzing 7,015 tweets, we could show classify social media automatically is a promising approach to analyze acceptance of emerging technologies such as self-driving cars. Even if data from Twitter is prone to certain biases, our results are in line with previous research. Our approach mitigates some of the methodological shortcomings regarding data collection, time-consuming manual coding, and biases of online surveys for emerging technologies.

Based on our results, we identified the need for developers and manufacturers to listen to the voice of future potential customers. Even the objectively best solution or superior new technology development can fail if it does not appeal to the customer or does not create public acceptance. Therefore, an active management of the public acceptance is mandatory to reduce the failure rate of new technologies.

In the instance we can say that Self driving car Manufacture Company become a big role in upcoming days.

Reference:

1. Liu, Bing. "Sentiment analysis and subjectivity." *Handbook of natural language processing* 2, no. 2010 (2010): 627-666.
2. Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." Proceedings of the workshop on language in social media (LSM 2011). 2011.
3. Bagheri, Hamid, and Md Johirul Islam. "Sentiment analysis of twitter data." arXiv preprint arXiv:1711.10377 (2017).
4. Arun, K., A. Srinagesh, and M. Ramesh. "Twitter sentiment analysis on demonetization tweets in India using R language." *International Journal of Computer Engineering and Information Technology* 9.6 (2017): 119.
5. Saini, Sonia, et al. "Sentiment analysis on twitter data using R." 2019 International Conference on Automation, Computational and Technology Management (ICACTM). IEEE, 2019.
6. Maas, Andrew, et al. "Learning word vectors for sentiment analysis." Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies. 2011.
7. Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
8. Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." *Journal of King Saud University-Engineering Sciences* 30.4 (2018): 330-338.
9. Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis on twitter." International Journal of Computer Science Issues (IJCSI) 9.4 (2012): 372.
10. Martínez-Cámara, Eugenio, et al. "Sentiment analysis in Twitter." Natural Language Engineering 20.1 (2014): 1-28.