Team SPAMM

Sentiment analysis on tweets

University of Texas at Arlington

INSY 5377

Web & Social Analytics

***#MarchForOurLives***

**TABLE OF CONTENTS**

1. Abstract………………………………………………………………………………………………………………………….2
2. Introduction……………………………………………………………………………………………………………………3
3. Data Source…………………………………………………………………………………………………………………….4
4. About the Tweets……………………………………………………………………………………………………………5
5. Hypothesis………………………………………………………………………………………………………………………7
6. Sentiment Analysis………………………………………………………………………………………………………….13
7. Word Cloud…………………………………………………………………………………………………………………….15
8. Results.….……………………………………………………………………………………………………………………….16
9. Conclusion……………………………………………………………………………….…………………………………....17
10. Reference……………………………………………………………………………………………………………………...18

**ABSTRACT**

Twitter is the most popular online news and social networking service on which users post and interact with messages. As usage of social media is increasing these days, people use it to share knowledge, spread the word and express their opinions about the ongoing topic.

Sentiment Analysis is opinion mining using natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media.

Our project is about collecting tweets from twitter to find out the public sentiment. The hashtag we are using is *#**MarchForOurLives* which has gained tremendous pace in the social media because of the recent Florida school shooting. We are conducting the test to perform sentiment analysis on the given data. We are also looking to find and segregate the data based on the geographical area they belong to. This would allow us to understand the people’s response to the given case and understand what could be done. In addition, we are also going to do topic modeling to find out group of words (i.e. topic) that occur in the collection of all tweets. This will help us in discovering the hidden semantic structures in the text body (i.e. tweets).

**INTRODUCTION**

On February 14, 2018, the school shooting at Marjory Stoneman Douglas High School in Parkland, Florida, stirred the souls of thousands of people in the US and even the whole world. On March 24, 2018, a student led demonstration called **March For Our Lives** took place in support of tighter gun control in Washington D.C. which was followed by this event taking place in different cities in the USA.

The Mission Statement of **March For Our Lives –**

Not one more. We cannot allow one more child to be shot at school. We cannot allow one more teacher to make a choice to jump in front of an assault rifle to save the lives of students. We cannot allow one more family to wait for a call or text that never comes. Our children and teachers are dying. We must make it our top priority to save these lives.

This is not just schools, though. This is churches, nightclubs, concerts, movie theaters, airports, and more. A child should not fear a bullet on their walk home. We may be children, but we are not fighting for just children. All lives are precious, and our country must make the safety of its citizens a number one priority.

Many media outlets described this as a tipping point for gun control legislation[1]. The protesters demanded the universal background checks on all the gun sales and a ban on the sale of high-capacity magazines, among other demands. The estimated turnout for this march was 1.2 to 2 million people in the US in turn making it one of the largest protests in the history of America.

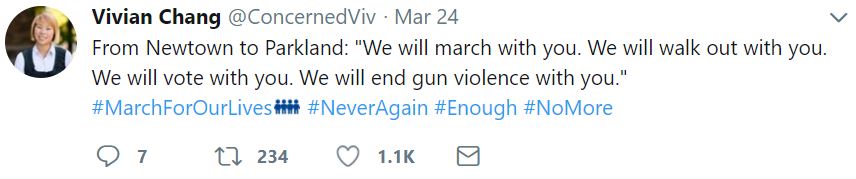
**DATA SOURCE**

The main data source we used is Twitter. We downloaded all the related tweets to the hashtag #MarchForOurLives using TwitterR package in R, then using data frame function we will change the dataset in data frame and save in the .csv file. This dataset was then analyzed for various topics, explaining various emotions in the world regarding the topic.

Few collected Tweets were as follows:







**ABOUT THE TWEETS**

We collected 30000 Tweets in a .csv file using R with the help of Twitter API’s and converted the data in data frame. After getting the data frame, we make a variable of tweets’s text. We then clean that text by removing all the user ID, HTML links, punctuation, numbers, ID name and stripping the text and saves it in variable name text6. The below mentioned is he code for the same:

consumerKey="xxxxxxxxxxxxxxxxxxxxxx"

consumerSecret="xxxxxxxxxxxxxxxxxxxxxxxx"

accesstoken="xxxxxxxxxxxxxxxxxxxxxxx"

accesstokensecret="xxxxxxxxxxxxxxxxx"

tweets<-searchTwitter("#MarchForOurLives", n=30000, lang = 'en', since = '2018-02-15')

tweets.df <- ldply(tweets, function(t) t$toDataFrame())

write.csv(tweets.df, "data.csv")

text = tweets.df$text

text1 = gsub("(RT|via)((?:\\b\\w\*@\\w+)+)","",text)

text2 = gsub("http[^[:blank:]]+", "", text1)

text3 = gsub("@\\w+","",text2)

text4 = gsub("[[:punct:]]"," ",text3)

text5 = gsub("[^[:alnum:]]"," ",text4)

text6 = gsub("RT", "", text5)

write.csv(text6, "text.csv")

After this we create a corpus of all the clean text with the help of tm package in R. We will apply some functions on this corpus to remove emoji’s and special characters. We will also lower the text size, remove all the stop words and some others also as per the dataset. Once doing all of this, we will stem our whole document. Stemming usually refers to removal of derivational affixes, for e.g. Dogs become Dog. The below mentioned is he code for the same:

text7 = Corpus(VectorSource(text6))

text7.removeEmoji = function(x) gsub("\\p{So}|\\p{Cn}", "", x, perl = TRUE)

text7.removeSpecialChar = function(x) gsub("[^[:alnum:]///' ]", "", x)

text7 = tm\_map(text7, content\_transformer(text7.removeEmoji))

text = tm\_map(text7, content\_transformer(text7.removeSpecialChar))

text7 = tm\_map(text7, removePunctuation, preserve\_intra\_word\_dashes = TRUE)

text7 = tm\_map(text7, content\_transformer(tolower))

text7 = tm\_map(text7, removeWords, c(stopwords("english"), "will", "can"))

text7 = tm\_map(text7, removeNumbers)

text7 = tm\_map(text7, stripWhitespace)

text7 = tm\_map(text7, stemDocument)

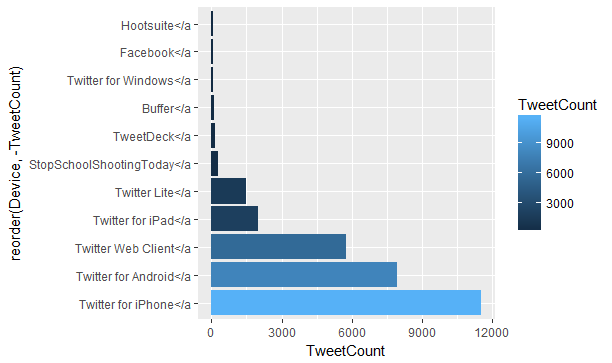
**HYPOTHESIS**

We are doing this sentiment analysis and many other types of data analysis and visualization on dataset to answer some of the following questions:

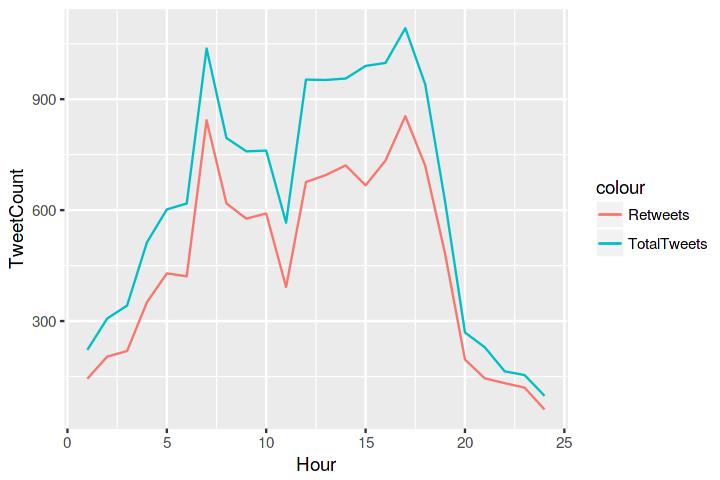
* Is this student led demonstration successful? What are there motives and how are they achieving that?
* Are people agreeing and supporting this student led group’s motive?
* What are the sentiments of people regarding this March For Our Lives movement and How are they supporting this movement?

The march was started by few school students and we wanted to focus on how these students managed to bring this revolutionary march into practice. We did the following research to understand how much this had impacted people.

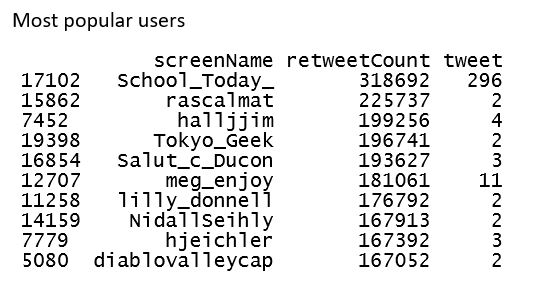
1. We started by analyzing how social media and technology helped to spread the word. Hence, we did a research on the various platforms and devices used and plotted the graph. The graph shows the tweet counts for various types of platforms. It was seen that the tweets and the hashtag were not only limited to Twitter, but it also spread to Facebook and other platforms as well. The maximum tweets with the hashtag were from the iPhone users.



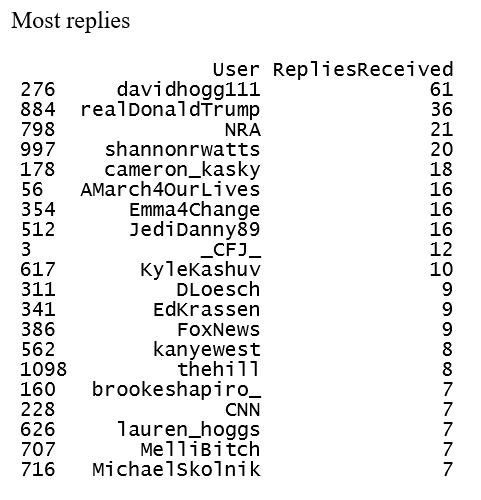
1. Next, we analyzed the total tweets and retweets posted per day which would help us to understand how people were considerate about this march taking place. The graph below helped us to analyze that maximum tweets and retweets occurred early in the morning and in the evening (approximately around 7am and 5pm). This implicates that the users were actively responding to march before and after the office hours, although there were tweets happening throughout the day. This shows that the people who couldn’t be a part of the march physically, were being part of it through social media.



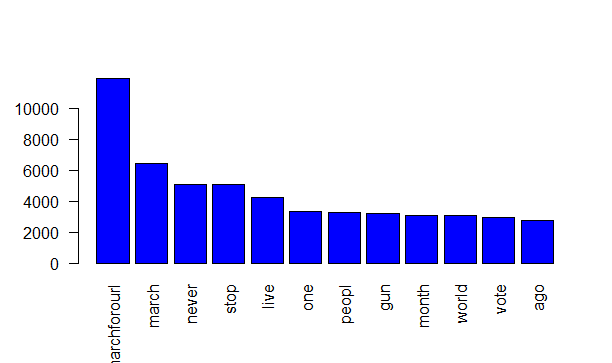
1. The third research we did was to see how twitter was used by public figures and celebrities to spread the word. Below is the table for most popular users and most replies received by a user. The twitter handle with the screenname School\_Today\_ posted the most tweets and retweets.



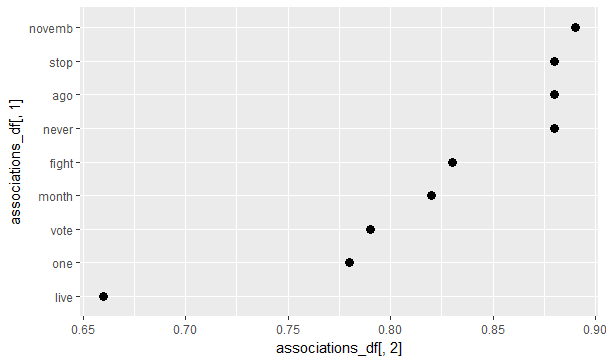
The famous #NeverAgain activist David Hogg, the President of the United States and the NRA (National Rifle Association of America) were some of the people who tweeted and got the most replies to their tweets.



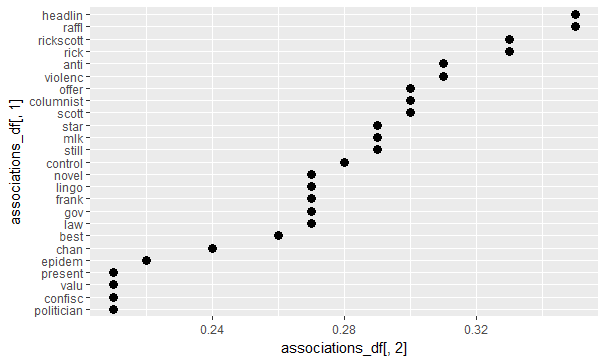
1. The fourth research that we have done is focused more on the words and topic that people have used in their tweets. Focusing on these words helps us in understand the topic that people are discussing on the twitter. It also gives us an idea of whether people are supporting this movement or not. We have conducted two different analysis on this to find out the people’s though process. One of these analyses is to find out the most common words used by people in tweets. This approach is quite useful to have a bird’s eye view on the dataset. It indicates that tweets of people are relevant to the topic of interest. The following generated plot is for the most frequent word that people have used in their tweets and the graph indicates that people have used words like march, stop, never, live and many others in their tweets for maximum time.



One final analysis is word relationships in dataset. This technique helps us in understanding the relationship between words in a corpus. Here it helps us in answering many different questions like which sequence of words are common across in our dataset (tweet’s text)? Given a word what different words are likely to follow, and to find out the strongest relationship between different words. The below is the graph for the word association of ‘march’ with other words. This word is highly corelated with words like stop, fight and never showing that people are supporting the motives of demonstration.



There is one more example for word association to get more clear idea of what are the sequence of words that people are write in their tweets. This association is for the word ‘gun’. The words here we have used for association are the most common words which many people have used just because to get the maximum idea of the correlation between the words and a person’s awareness through their tweets.



Based on the four researches that we carried out, we can now correctly say that social media like Twitter helped to spread the message that the students wanted the people to know. Moreover, it helped us to know that many people were in support of the movement through twitter, if not physically. The movement was also supported by big names like the activist, the various news channels like Fox and CNN and by the president of the United States. Additionally, it allowed us in understanding the thoughts of public for this student le demonstration and the measures they are taking to support this group’s motto.

**SENTIMENT ANALYSIS**

The sentiment analysis algorithm used here is based on the [NRC Word-Emotion Association Lexicon](http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm) of Saif Mohammad and Peter Turney. The idea here is that these researchers have built a dictionary/lexicon containing lots of words with associated scores for eight different emotions and two sentiments (positive/negative). Each individual word in the lexicon will have a “yes” (one) or “no” (zero) for the emotions and sentiments, and we can calculate the total sentiment of a sentence by adding up the individual sentiments for each word in the sentence. Not every English word is in the lexicon because many English words are neutral.

Below is the code for the sentiment Analysis:

#sentiment analysis

mysentiment <- get\_nrc\_sentiment(text6)

SentimentScores <- data.frame(colSums(mysentiment[,]))

names(SentimentScores) <- "Score"

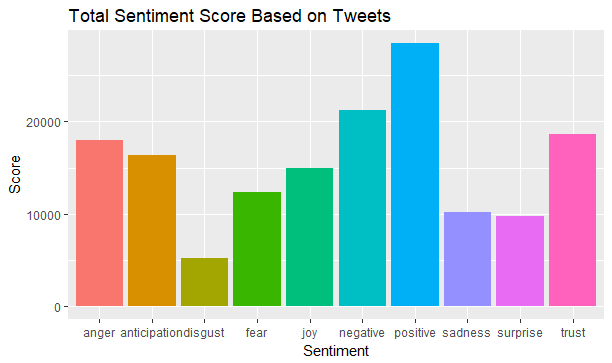
SentimentScores <- cbind("sentiment" = rownames(SentimentScores), SentimentScores)

rownames(SentimentScores) <- NULL

ggplot(data = SentimentScores, aes(x = sentiment, y = Score)) + geom\_bar(aes(fill = sentiment), stat = "identity") + theme(legend.position = "none") + xlab("Sentiment") + ylab("Score") + ggtitle("Total Sentiment Score Based on Tweets")

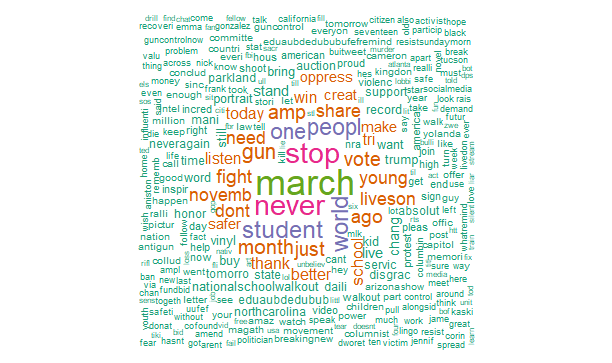
Using the above algorithm, we found the various sentiments of different people. The scores of sentiments like positive, trust, joy and anticipation show us that the people are positive of the demonstration’s purpose and are supporting by spreading the positivity through their tweets. They are showing the trust in the movement and are anticipating the changes that government should do by bringing the new laws of tighter gun control. The sentiments here like negative, anger, fear, sadness and surprise indicate that the people are shocked with the terrifying incident happened and are sharing their thoughts through tweets. Some people are feeling angry over what is happening in their society, while some in fear because of such a dangerous incident. There are people who feels negative about the society’s future by not trusting in government or movement which suggests that either they don’t want change or maybe they have lost their hope on this movement and government for tighter gun control laws.

Below is the chart for the sentiment scores which we obtained from the tweets.



**WORD CLOUD**

Word cloud is an image compromised of many different words in a text or subject. Here the size of each word is indicating the importance of that word in the dataset. As our text here is the collection of different tweets of people for #MarchForOurLives, it advocates that the words like march, never and many others hold the most importance in the tweets and then their maximum frequency in the tweets.



We obtained our word cloud using R and the library(wordcloud). The code for the same is as follows:

#wordcloud

pal <- brewer.pal(8, "Dark2")

wordcloud(text7, min.freq = 5, max.words = Inf, width = 1000, height = 1000, random.order = FALSE, color = pal)

**RESULTS**

All the different analysis here has helped us in answering the three question which we considered before in our hypothesis. The breakdown of different users tweeting via different platforms and tweet & Retweet count shows that users have actively participated in the movement and raised their and shared their thoughts on the violence going on in society. This help us in answering first question that social media like twitter successfully helped the student led demonstration in spreading their motive. They are achieving their motives by promoting their march on social media and bringing the awareness in the people.

Involvement of famous activist of #NeverAgain David Hogg, Big politician Figure Donald Trump and NRA in this movement answers our second question that people are supporting the movement and really looking for the change in law gun control. Analysis of most common words and word associations shows that people are still alive on the motto of March For Our Lives and they are continually talking about the topics like gun, stop, never and march.

People have shown various sentiment for this movement which we briefly explained earlier in sentiment topic. They have expressed their sentiments in way of trust, joy, anger, fear and in many other ways. This sentiment analysis gives us an idea of how people are supporting the movement. Many people are trusting in the movement and government to bring the changes while some are angry and sad because of the violence going on in society.

**CONCLUSION**

After all the analysis done on the dataset, we can conclude that the social media organizations have the more power than other resources to bring the change in the world regardless of the problem. Twitter is one of the most commonly used social media, particularly in the iPhone devices. Through the social media platform, even the students could make this march successful and could spread the message to the world.

These social media organizations can bring the awareness among the people about any issues or problems by conducting online surveys, march and interaction sessions. This can help to solve the problems easily.

**REFERENCES**

[1] <https://en.wikipedia.org/wiki/March_for_Our_Lives>

[2] <https://twitter.com/AMarch4OurLives>

[3] <https://en.wikipedia.org/wiki/Sentiment_analysis>

[4] <https://marchforourlives.com/mission-statement/>

[5] <https://earthdatascience.org/courses/earth-analytics/get-data-using-apis/text-mining-twitter-data-intro-r/>