

## SOCIAL MEDIA AND STORMING OF THE UNITED STATES CAPITOL: AN ANALYSIS OF TWITTER SENTIMENT

Praveen Kumar P<sup>1</sup>, Dr. Sridhar Krishnaswami<sup>2</sup>

<sup>1</sup> Research Scholar, Department of Journalism and Mass Communication, College of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu.

Contact: [praveenp@srmist.edu.in](mailto:praveenp@srmist.edu.in)

<sup>2</sup> Professor (Retired), Department of Journalism and Mass Communication, College of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu.

Contact: [sridhar54k@gmail.com](mailto:sridhar54k@gmail.com)

### ABSTRACT:

Of all the Social Media platforms, Twitter has gained a lot of traction these days. It has become a space for people to explicitly share their critiques and perspectives about the world. Trending political events often explode in a number of tweets, offering a unique possibility to gauge the relationship between expressed public sentiment and political activity. Following his defeat on November 3, 2020 in the US Presidential election, President Donald Trump used his Twitter handle to endorse questionable reports, and in the process endorsed several conspiracies and theories regarding outcome of the Presidential election. An article in The Washington Post claimed that on an average Trump posts more than fifty misleading claims a day. To overturn his 2020 presidential election defeat, on January 6, 2021, a boisterous group of Trump supporters, invited and openly incited by the President, stormed Capitol Hill to prevent and intimidate members of Congress who had gathered there to certify and validate the victory of President-elect Joseph Biden. This paper has attempted to analyse Twitter users' sentiment in storming of the United States Capitol using R language and TM Package. Two thousand tweets were extracted from twitter users from USA which is then cleaned, processed, and categorized into ten sentiments finally summarizing the results as a whole. The nrc sentiment score method of the Syuzhet package has been used to classify sentiments and 'ggplot2' was used to visualize the sentiments obtained for this research.

**KEYWORDS:** Trump, US election, Twitter, Capitol Hill, Riots

### INTRODUCTION

Launched in March 2006, Twitter has hit the decade-and-a-half mark this year. With over 330 million month-to-month dynamic users and more than 500 million tweets posted every day, Twitter still keeps its status as a mainstream social media platform. What distinguishes Twitter from other social media platforms is its reachability among the big wigs including heads of the states, prime ministers, legislators, entertainment stars, and sportspersons. Using Twitter the users can tweet their up-to-date commentary on the ongoing issues as text, video, sound bites, and other arrangements, and the followers of the users can respond to those tweets. Because of this modus operandi the information and the reaction transport from bottom to top and vice versa at a much faster rate than other social media.

Apart from this, Twitter has an intriguing concept embedded in its tweets called hashtags. Alongside the tweets, users can utilize the hashtag image '#' before a pertinent watchword to sort their tweets and help other users to read more effectively during twitter search. Utilization of these hashtags makes text arrangement simpler since the hashtag itself can pass on a feeling or sentiment.

For example, #IndiaSupportsCAA became the official hashtag for Prime Minister's outreach campaign to garner backing for the Citizenship Amendment Act in India.

## **OVERTURE TO THE RIOTS**

Motivated by Trump, a large number of people belonging to his faction amassed in

Washington, D.C., a day before a joint meeting of Congress was scheduled to formalize Biden's win. The gathering was on the principle of the "stolen election conspiracy theory" tweeted by Trump. On the morning of the joint session, at a "Save America" rally, Trump said to his supporters, "if you don't fight like hell, you're not going to have a country anymore". Following this, many nonconformists strolled to the Capitol, where a joint meeting of Congress was taking place, and started rioting.

## **STORMING OF THE US CAPITOL**

On January 6, 2021, a mob of Trump Supporters stormed the US Capitol during a violent attack against the US Congress. Later named, 'The raging of the United States Capitol', the assault was orchestrated by a horde of Donald Trump's allies in an attempt to show their protest against Joe Biden's win in the presidential election. The rioters broke through police barricades, raged, started vandalizing, plundering the US capitol building, attacked the Capitol Police officials and news correspondents, raised scaffoldings on the Capitol grounds in an attempt to find and hold the legislators captive.

The mob recited the slogan "Hang Mike Pence" as they believed Pence has failed to dismiss the Electoral College votes, even though Mike Pence never had the power to do as such. The agitators focused on House Speaker Nancy Pelosi, vandalizing her workplaces, along with the rest of the congress members'. Following the storming of the gate, the Senate and House of Representatives, the chambers were emptied by the Capitol Police and all the personnel inside were secured. Agitators stripped the empty Senate chamber while police officers used handguns to guard the emptied House floor. A few pipe bombs were found later at the chambers occupied by the National Committee, and Molotov mixed drinks were also found later in a vehicle.

After hours of savagery and tumult, Trump called his supporters "very special" and advised them to "return home in peace" through a video delivered by the White House. Later in the day, Trump further went on to say that the attack as unavoidable. The Capitol was freed from rioters by midnight; the validation of the electoral votes resumed, and was finished in the early hours of morning. Pence announced President-elect Biden and Vice- President-elect Kamala Harris as winners of the presidential election 2020 and invited them to take the office on January 20, 2021. Constrained by his management, Trump later agreed to a change of command in a broadcast proclamation.

The storming of Capitol Hill was generally censured by political leaders and associations inside as well as outside of the United States.

### **#CapitolRiot**

As twitter has a close-knit relationship with several million citizens of the United States, it became one of the primary platforms through which they expressed their state of mind regarding the capitol riots. This was evident from the trending hastag “#CapitolRiot”. The hastag started trending in the micro blogging platform even before the mob took down barricades leading to the hill. Close to million tweets got embedded through Twitter before the mainstream media got the scent of what was happening. #CapitolRiot gained momentum as it suited both ends of the political spectrum to disseminate their perspective towards the riots.

### **LITERATURE REVIEW**

This section summarizes the understanding the researchers has gained by analyzing various approaches to the research question at hand.

### **SOCIALMEDIA AND 21ST CENTURY**

Deb Dutta Das et.al in the paper "Sentimental Analysis for Airline Twitter data" takes note of that social networking and its endless possibility has shocked the world at a praiseworthy speed. The sentiments of individuals all through the world are communicated to social media through their social media profiles, making them an excellent place for sentiment mining. Khanna, Pooja in the paper "sentiment analysis: an approach to opinion mining from Twitter data using r." expresses that different study devices and methods are accessible to gauge public opinion. With the unexpected expansion in text-based social media, the chunk of individuals simulcasts their perspective and thoughts on a huge scope of issues. We can examine this data accessible from the public to close populace attitudes to comprehend the latest things of the market

### **TWITTER AND SENTIMENT ANALYSIS**

Achin Jain and Vanita Jain in their paper note that ‘Sentiment analysis is utilized to analyze the opinions of the users for decision making with the assistance of natural language processing methods’. In their investigation to comprehend the feelings of individuals about alternate fuel sources, they have declared the exactness of sentiment characterization is better with including feature determination methods.

Madan A., et al, (2018) performed sentiment analysis using word reference-based procedure in Twitter information. The analysis was made to study the Goods and Service Tax (GST) introduction in India. The general assessment on GST was orchestrated into positive, nonpartisan, and impartial sentiments. 10,000 informational collections are used for assessment. Results show the public has a positive assessment of GST with 33 rate positive sentiments. Prabhsimran Singh., et al (2018) performed state-level sentiment analysis on demonetization. Sentiments of people in thirty states in India are arranged as hugely positive, positive, impartial, bleak, disastrous, and no information. Out of thirty states in India, just seven states gave indications of hugely positive.

### **RESEARCH METHODOLOGY**

This research paper attempts to do a sentiment analysis surrounding the hashtag #CapitolRiot in twitter using R programming language. The following tools were used.

### **TwitterR Package**

TwitterR is a commonly used R package. As the name suggests, this package allows the user to gain access to the Twitter API. This package was used to gather tweets garnered around #CapitolRiot

### **TM package**

TM package refers to Text Mining Package which helps in text record assortment. This is indicated as a corpus in linguistics. It represents an assortment of text documents and can be interpreted as a database for texts. Its components are Text Documents holding the actual text corpora and local metadata.

### **Procedure before directing Sentimental Analysis Data Set:**

The data collected from twitter have to be given as input to information analysis programming for sentiment analysis. Here R studio is used to dissect informational indexes from Twitter. A few other applications are additionally accessible to get to and examine twitter information.

### **Data editing:**

Editing the data is the next process after creating the data set. This stage is necessary to remove irrelevant details from the tweets. This is also called by the name “commotion reduction”. This includes removing HTML tags, blank areas, etc.,

### **Removing Emoji:**

An emoji is a pictorial representation of a facial expression. Due to the smaller number of the character count, several users tend to use emojis in their tweets to save up space. During the processing of tweets, all emojis will be removed and certain emojis were replaced by the words showcasing respective emotions in the tweets.

### **Stemming:**

Stemming refers to the process of removing prefixes. It is also done to cut down the complexity during analysis.

### **Stop Words:**

The words that provide no value to the tweets are called stop words. The lexicon used in the analysis has identified 90 such words. The stop words are removed to reduce the intricacy of the analysis.

### **Scoring of Sentiments**

The text provided is turned into a matrix and the score was then calculated by assigning value to each word and its association to a specific emotion. The score is then calculated using Breen’s approach. Breen’s approach states that  $\text{Sentiment Score} = (\text{Total number of positive words}) - (\text{Total number of negative words})$ .

The positive & negative attributes of a word are assessed by opinion lexicon, which for us luckily created by Hu and Liu. If the Sentiment score is higher than zero, then the text provided has a positive score and vice versa for the negative tweets

### **Sentiment Analysis Using Syuzhet**

Syuzhet is an R package that has four distinct sentiment dictionaries. All the dictionaries that are part of Syuzhet have proven to provide strong, but computationally expressive sentiment extraction.

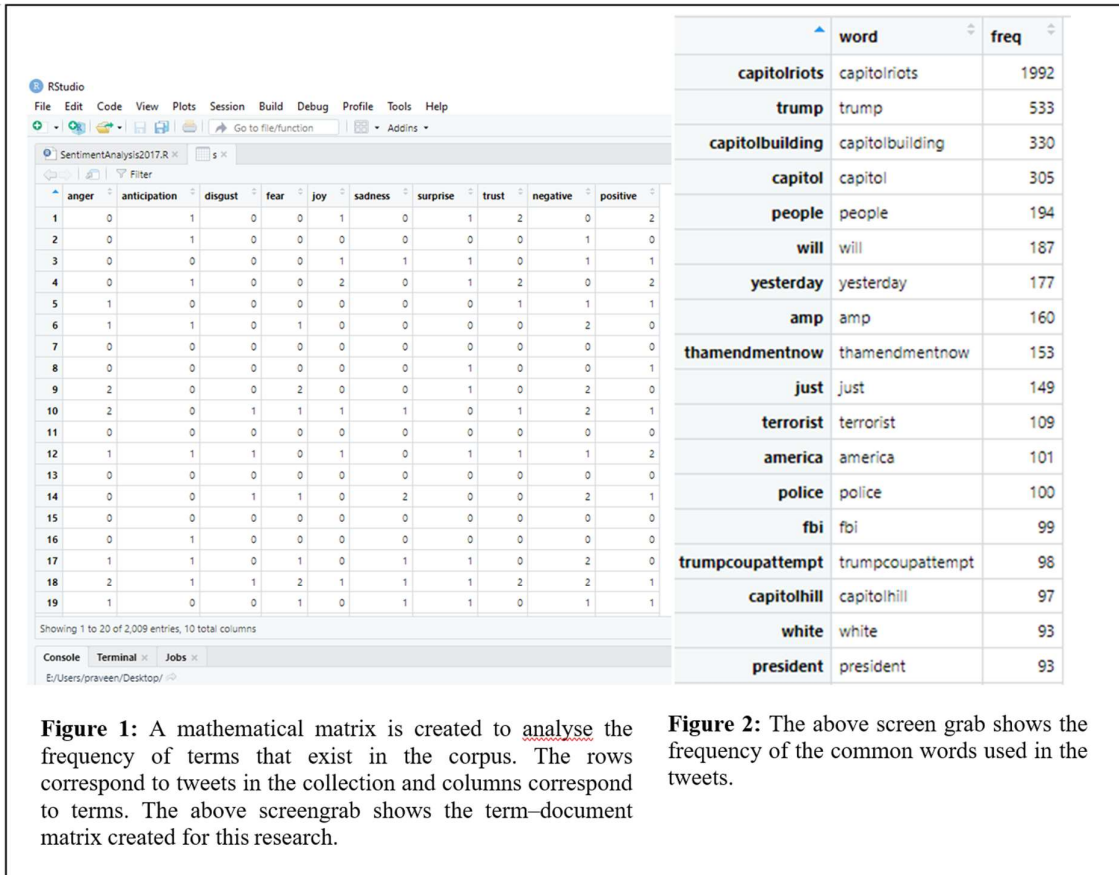
### **DATA COLLECTION AND INTERPRETATION**

Using the 'rtweet' package, a search query was requested to search for the tweets using the hashtag #CapitolRiots. Because of the limitation of the package, only 18,000 tweets were collected in the timeframe of fifteen minutes. All the 18,000 tweets were original tweets and not retweets made by different users.

The same query was repeated for the next five hours to accumulate 1,08,000 tweets. Because of the lack of definition provided in the coding, the search query collected 90 variables that are available with the tweets including favorites count, followers count, country, language, etc.,. For the process of compilation of the collected tweets, openxlsx package was installed in Rstudio. This package allows the user to export the collected data into .xls format. Because of the limitation of the hardware the .xls files were transferred to five different files. This allows for easier and quick access and compilation of the required dataset.

Because of the lack of proper definition, the corpus collected by the rtweet included original tweets from different languages. This was rectified by filtering the tweets which carried the language 'English' variable. Apart from this, the tweets which carried a geotag for the location were also filtered from the corpus minimizing the usable pool to 63,643 tweets. This data cleansing allowed us to eliminate more than 44,357 tweets which added no value to the research as they were either tweets using unrecognized language or the ones made by people living outside the United States of America.

With 63,643 tweets left in the corpus, the method of simple random sampling was used to select 2,000 tweets from the usable corpus. Sampling method was introduced to to compensate for the hardware constraint. The .xls file format was changed into comma-separated values (.CSV) format to make the corpus readable by the RStudio. The tweet variable was then distinguished into a separate corpus using TM package.



**Figure 1:** A mathematical matrix is created to analyse the frequency of terms that exist in the corpus. The rows correspond to tweets in the collection and columns correspond to terms. The above screenshot shows the term-document matrix created for this research.

**Figure 2:** The above screenshot shows the frequency of the common words used in the tweets.

After a separate corpus was created with the tweets, they were subjected to a series of data cleaning processes. The following command was used “tm\_map(corpus, to lower)” to change all the texts in the tweets to lower case. The command “tm\_map (corpus,removePunctuation)” was used to eliminate all the punctuation symbols used in the tweets as they might create hindrances during the process of sentiment analysis. As we are only interested in the sentiment of the user, the command

“tm\_map(corpus, removeNumbers)”

was used to remove any numbers that were part of the corpus.

The text mining package has a compilation of 90 ‘stopwords’ in its database that was used to remove words like the, he, have, etc.

that serve no purpose to the current analysis.

The command

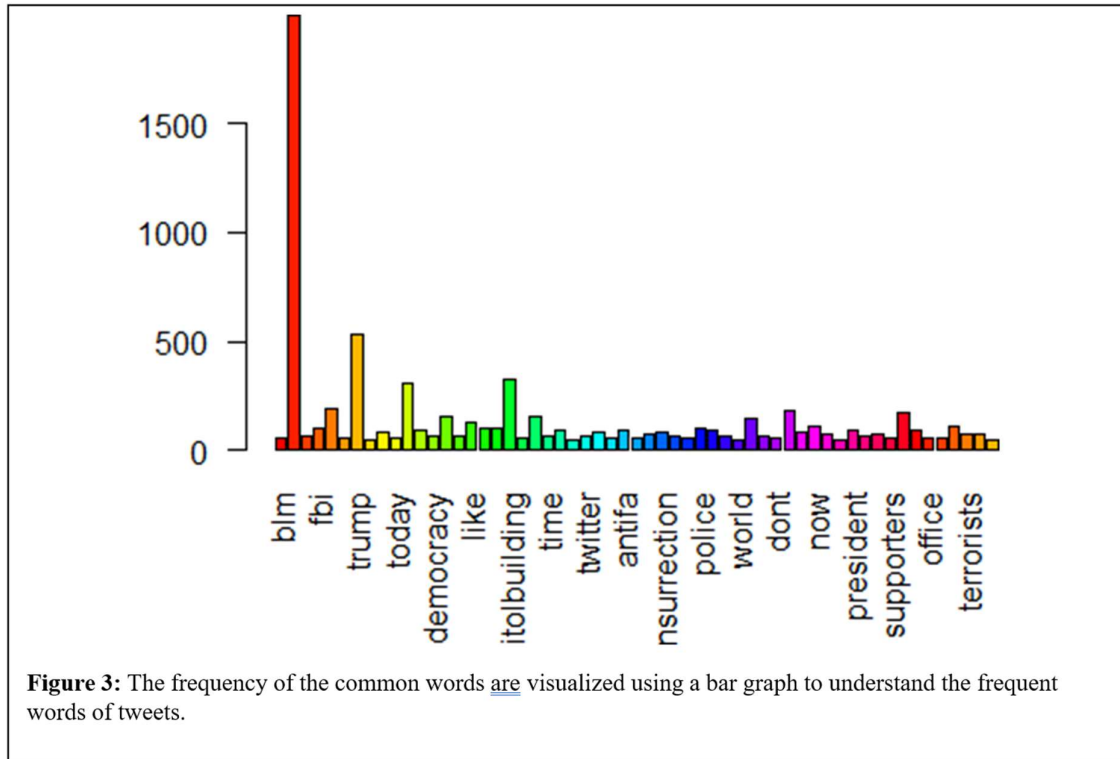
“tm\_map(corpus, removeWords, stopwords('english'))”

was used to eliminate the stop words that were part of the tweets. This process was followed by eliminating any links that were part of the collected tweets. The following command was used to accomplish the task “removeURL<- function(x)

gsub('http[[:alnum:]]\*', "", x) cleanset<- tm\_map(cleanset,content\_transformer(removeURL))”

The unnecessary white space that exists between different words was removed using the following command

.“cleanset<- tm\_map(cleanset, stripWhitespace)”.



A mathematical matrix is created to analyse the frequency of terms that exist in the corpus. The rows correspond to tweets in the collection and columns correspond to terms.

The frequency of the common words was then visualized using a bar graph to understand the frequent words which were part of tweets. The commonly used words in the tweets were found to be BLM, FBI, trump, democracy, Antifa, insurrection, terrorists. The most common one being 'BLM' referring to the Black Lives Matter social movement that highlights the racially motivated police brutality and violence against African Americans in the United States.

A qualitative dissection of the corpus showed us that several tweets were drawing parallels between the storming of Capitol Hill and the Black Lives Matter movement that reached its height before the corona outbreak. Several tweets also suggested the BLM as the precursor to the Capitol Hill riots.

### SENTIMENT SCORE

Syuzhet package was the primary package used in this analysis to deliver the sentiment score of the collected corpus. As discussed earlier, Syuzhet carries four sentiment dictionaries. The mathematical matrix corpus was then subjugated to the 'get\_sentiment' function of this package. This function analyses the sentiment and emotion exhibited in each sentence.

The method includes four available sentiment extraction methods to employ. The default method is called the "syuzhet". Apart from that, the other methods include "afinn", "nrc", and "stanford". For the purpose of this research, "nrc" method is used as this has been proved to be providing more accurate results.

The nrc sentiment score of the tweets is then displayed in a table with a score set for ten emotions. The emotions and sentiments include anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative and positive.

### **VISUALIZING THE SCORE**

Because of the colossal length of the table, understanding the sentiment score becomes a gruesome task. To simplify the scores, a visualization package called ‘ggplot2’ was installed in RStudio. The ‘barplot’ function of the package seems to be the right contender for the task as it allows for the variables to be plotted along a horizontal axis. The variables here are the emotions that are scored by the syuzhet package.

The frequency distribution of the sentiment scores can be visualized easily. Multiple emotion display allows for quick estimation of relativeness among different emotions

A quick bar chart was plotted using the data visualization package - ggplot2.

The code provided the visualization of the sentimental scores derived from the nrc sentiment analysis.

### **FINDINGS**

The visualization provided the much-needed clarity on the sentiments and emotions shared by Twitter users. A small note on the values scored by eight emotions and two sentiments are given below

#### **EMOTION ANALYSIS**

##### **Anger**

Anger seems to be the second-largest emotion distributed throughout the 2000 tweets. A little over 1500 words used in the tweets are associated with the emotion of anger in the corpus of nrc.

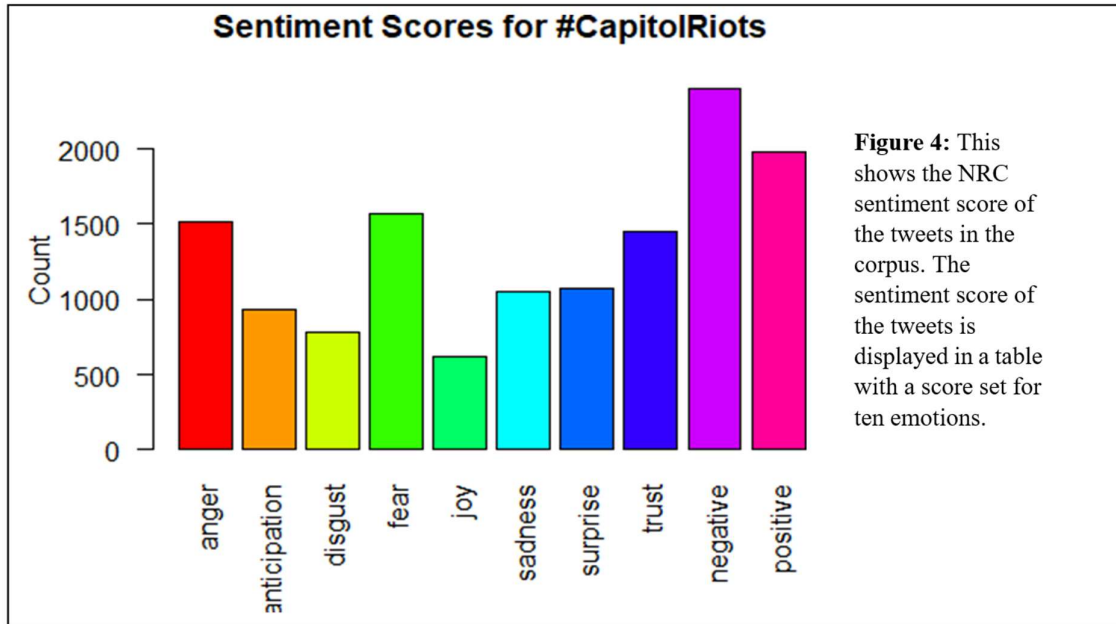
##### **Anticipation**

Anticipation scored the third last number. A little over 950 words associated with this emotion are seen throughout the tweets.

##### **Disgust**

With over 750 associated words, Disgust as an emotion is the second least emotion exhibited by Twitter users.





**Figure 4:** This shows the NRC sentiment score of the tweets in the corpus. The sentiment score of the tweets is displayed in a table with a score set for ten emotions.

**Fear**

Fear is the emotion caused by an apparent peril or danger, which causes physiological and eventually conduct changes. For example, mounting a forceful reaction or escaping the danger. This emotion recorded the largest response with over 1600 associated words out of the 2,000 tweets.

**Joy**

Joy scored the lowest in the analysis. A little over 500 words associated with joy are found in the 2,000 tweets taken for analysis.

**Sadness**

Sadness is exhibited throughout the corpus. A little over 1,100 words associated with this emotion are found inside the corpus.

**Surprise**

Because of the nature of the riots, this emotion scored big on the scorecard. Landing over 1,150 associated words, this is the fourth-largest number on the card.

**Trust**

Perhaps the most unanticipated outcome in this analysis is the score received by this emotion. Scoring over 1,400 points on the table, Trust seems to be the third-largest emotion exhibited by Twitter users on the day of capitol riots.

**SENTIMENT ANALYSIS**

With over 14,182 unigrams (words), the NRC is one of the largest crowd sourced databases to study the sentiment of the tweets. Thus, the words found in the corpus are articulated to the database to find out the sentiment that they carry with them. Understandably the negative sentiment seems to dominate the tweets. Over 2,500

words that align with negative sentiment are found to be inside the corpus whereas only 1,600 words were associated with positive sentiment.

This riot that was condemned by all the world leaders received a bizarre high number of positive scores in this analysis. However, on careful observation of the tweets, it was found that Twitter was used as a platform by several onlookers to spread positive messages to others during the riots. This explains the large number of positive sentiment scores calculated by the nrc database. However, the total negative score seems to be higher than the total number of tweets calculated. This proves that the Twitter users were largely distraught by the storming of Capitol Hill.

## CONCLUSION

Emotional and Sentimental Analysis is often seen as a path that many researchers take to understand the pulse of the public. The accessibility of the internet and yearning for social media has made it easier for researchers to collect a huge amount of data as they fall under the public domain. This research tried to analyze the sentiments and emotions exhibited by Twitter users during the storming of Capitol Hill.

The analysis established the assumption that the Twitter users largely showed a sign of negative sentiment towards the riot. The NRC lexicon used in the research displayed the variance displayed by the Twitter users regarding the topic in question. As several types of research have already established that Twitter users are more likely to align towards liberal thoughts, it is no doubt the riot rallied by an ultra-conservative group will receive negative emotions from Twitter. However, the approach used in this research showed us the huge number of words associated with the emotion of trust used in the tweets. This shows us that Twitter was also used by its users to reinforce their faith in the democratic process when the same was put to test during the storming of the Capitol Hill.

## LIMITATIONS AND SCOPE FOR FUTURE RESEARCH

It can be said without any doubt that there is still space for further research in this since the method of analyzing the text via machine learning is relatively new. Apart from that, there is the problem of hardware constraint, the researcher was able to analyze only 2,000 tweets. However, a report suggests that over a million tweets were tweeted within the first few hours of the riot. To understand the overall sentiment of the Twitter users a larger pool of tweets is suggested for future research. Even though NRC emotion lexicon is considerably a large database, a cross-sentiment analysis using other lexicons will provide richer findings.

## References:

1. Jockers, M. L. (2017). Syuzhet: Extract sentiment and plot arcs from text. Retrieved from <https://github.com/mjockers/syuzhet>
2. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexiconbased methods for sentiment analysis," *Comput. Linguist.*, vol. 37, pp. 267-307, 2011.
3. Singh P, Sawhney RS, Kahlon KS (2017) Sentiment analysis of demonetization of 500 and 1000 rupee banknotes by Indian government. *ICT Express*.
4. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

5. Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013.
6. Das, Deb & Sharma, Sharan & Natani, Shubham & Khare, Neelu & Singh, Brijendra. (2017). Sentimental Analysis for Airline Twitter data. IOP Conference Series: Materials Science and Engineering.
7. Gautam, Geetika & Yadav, Divakar. (2014). Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis. 2014 7th International Conference on Contemporary Computing.
8. Singh, Prabhsimran & Sawhney, Dr. Ravinder Singh & Kahlon, Karanjeet. (2017). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. ICT Express. 4. 10.1016/j.ict.2017.03.001.
9. Saif, Hassan, et al. "On stopwords, filtering and data sparsity for sentiment analysis of twitter." (2014): 810-817.
10. <https://www.washingtonpost.com/politics/2020/10/22/president-trump-is-averaging-more-than-50-false-or-misleading-claims-day/>
11. Agarwal, Apoorv, et al. (2011). "Sentiment analysis of twitter data." In Proceedings of the workshop on languages in social media, pp. 30-38. Association for Computational Linguistics.