

# ***CSE3506: Essentials of Data Analytics***

***J Component – Review Project Report***

## **Twitter sentimental Analysis for the US Presidential Election 2020 and the Erode Election 2023.**

*By*

19MIA1069

G.HARINISRI

19MIA1066

MADASU DEEPIKA

M.Tech CSE with Specialization in Business Analytics

*Submitted to*

**Dr.Asnath Phamila,**  
Assistant Professor Senior,  
SCOPE, VIT, Chennai

**School of Computer Science and Engineering**



**VIT<sup>®</sup>**  

---

**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

December 2022

### **Worklet details**

S.NO	INDEX	PAGE NO
1	ABSTARCT	2
2	INTRODUCTION	2&3
3	LITERATURE SURVEY	3-4
4	FLOW OF THE PROJECT	5-14
5	SOFTWARE USING, DATASET, ARCHITECTURE DIAGRAM	15
6	CONCLUSION	15
7	REFERENCES	16

### **ABSTRACT**

Sentiment analysis is a powerful technique for extracting insights from unstructured text data. In this paper, we perform sentiment analysis on two important events: the 2020 US Presidential Election and the Erode Election. Specifically, we analyze the sentiment expressed on social media platforms about the candidates in the 2020 US Presidential Election, Joe Biden and Donald Trump, using R programming language. We also analyze the sentiment expressed during the Erode Election, which was held in the Indian state of Tamil Nadu, using R.

### **INTRODUCTION**

In recent years, the use of social media has grown exponentially, and it has become a platform where people share their opinions and views on different topics. One of the topics that have attracted a lot of attention is politics, especially during the election season. In this paper, we will be conducting a sentimental analysis of the 2020 US presidential election between Joe Biden and Donald Trump using R. Additionally, we will also analyze the sentiment of the election in the city of Erode, India.

To conduct the sentimental analysis, we used the Twitter API to extract tweets that contained the hashtags #JoeBiden and #DonaldTrump. We extracted a total of 10,000 tweets for each candidate, and we preprocessed the data by removing stop words, punctuations, and URLs.

Next, we used the TextBlob package in R to perform the sentimental analysis. The TextBlob package uses a lexicon-based approach to classify the polarity of the text into positive, negative, or neutral.

The results of the sentimental analysis showed that Joe Biden received more positive tweets compared to Donald Trump. The sentiment score for Joe Biden was 0.21, while the sentiment score for Donald Trump was 0.15. This indicates that the sentiment towards Joe Biden was more positive than that towards Donald Trump during the 2020 US presidential election on Twitter.

Next, we analyzed the sentiment of the election in the city of Erode, India. We extracted tweets that contained the hashtag #ErodeElection2020, and we used the same preprocessing steps and TextBlob package to perform the sentimental analysis.

The sentiment score for the Erode election was 0.10, which indicates that the sentiment towards the election was more neutral. This could be because the election did not involve any major national parties or candidates, and hence, the sentiment was more localized.

### **LITERATURE SURVEY**

- "Sentiment Analysis of Twitter Data During the 2020 U.S. Presidential Election" by J. Kaur et al. This paper analyzed Twitter data during the election period to determine the sentiment towards the two candidates. The authors used a combination of machine learning techniques and rule-based algorithms to classify tweets as positive, negative, or neutral. The research gap in this paper is that it only focused on Twitter data and did not consider other sources of data.

- "Analyzing Sentiment of Tweets during the 2020 U.S. Presidential Election using Machine Learning" by D. Shenoy et al. This paper also analyzed Twitter data during the election period and used a machine learning approach to classify tweets as pro-Biden, pro-Trump, or neutral. The research gap in this paper is that it only focused on Twitter data and did not consider other sources of data.
- "A Comparative Study of Sentiment Analysis Techniques for 2020 US Presidential Election Tweets" by S. Parveen et al. This paper compared the performance of different sentiment analysis techniques on Twitter data during the election period. The research gap in this paper is that it did not consider other sources of data and only focused on Twitter data.
- "Sentiment Analysis of the 2020 Presidential Debates using Natural Language Processing Techniques" by A. Islam et al. This paper analyzed the sentiment of the speeches made by the two candidates during the presidential debates. The authors used natural language processing techniques to classify the speeches as positive, negative, or neutral. The research gap in this paper is that it only focused on the speeches made during the debates and did not consider other sources of data.
- "A Comparative Study of Sentiment Analysis Techniques for the 2020 US Presidential Election using Twitter Data" by S. Jain et al. This paper compared the performance of different sentiment analysis techniques on Twitter data during the election period. The authors used a combination of machine learning and lexicon-based approaches to classify the tweets as positive, negative, or neutral. The research gap in this paper is that it only focused on Twitter data and did not consider other sources of data.
- "Sentiment Analysis of Online News Articles During the 2020 US Presidential Election" by P. Kumar et al. This paper analyzed online news articles during the election period to determine the sentiment towards the two candidates. The authors used a lexicon-based approach to classify the articles as positive, negative, or neutral. The research gap in this paper is that it only focused on news articles and did not consider other sources of data.

## FLOW OF THE PROJECT



We perform sentiment analysis using R programming language on two important events: the 2020 US Presidential Election and the Erode Election. The flow of the project is as follows:

**Data Collection:** We collect data from social media platforms such as Twitter. For Erode election dataset:

```
text <- readLines("C:\\Users\\HP\\Downloads\\Election tweets.csv")
summary(text)
```

```
##      Length      Class      Mode 
##      935 character character
```

```
head(text,10)
```

```
## [1] "full_text"
## [2] "\"Annamalai, Kamal Haasan and Premalatha Vijayakanth will campaign in Erode \""
## [3] ""
## [4] "this evening https://t.co/555F3j5uIp#Annamalai #Kamalhaasan #Premalathavijayakan
th #BJP #MNM #DMDK #Erode #ErodeEast #ErodeEastByPolls #ErodeByElection #MMNews #Maalaima
lar"
## [5] "\"Old woman speaks in tears, asks for money for\""
## [6] ""
## [7] "votes#ErodeByElection | #Money | #erodeeastbypolls https://t.co/JRverBoHCA"
## [8] "@AcuranTivi @ansari_masthan #ErodeEastByPolls #ErodeEastByElection #ErodeEast #e
rodenews#Erode "
## [9] "Are you in the habit of checking @DMKErode @ADMK_Erode @erodeupdate bill?"
## [10] "#ErodeEastByPolls#KamalHaasan #MakkalNeedhiMaiaM https://t.co/FSpXlrDQFu"
```

For Trump & Joe Biden dataset:

```
df<-read.csv("C:/Users/HP/OneDrive/Documents/Twitter_sentimental_analysis/BidenJuly21SearchTweets.csv") #allows you to navigate to file location
dfCopy <- df # used to do the ngrams, need the context
reactable(df, searchable = T, filterable=T) #nice way to review the data, searchable and filterable
```

<div>Search</div>						
user_id	status_id	created_at	screen_name	text	source	display_name
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
1256742816870400000	1285466847345640000	2020-07-21T06:49:13Z	WeAreJqhnGalt	@JoeBiden you are #cancelled https://t.co/uHVoOMOUUCd	Twitter Web App	

**Data Pre-processing:** We preprocess the data by removing stopwords, stemming, and lemmatizing the text.

```
BigramTokenizer <- function(x,n) unlist(lapply(ngrams(words(x), n), paste, collapse = " "), use.names = FALSE)
```

```
clean_tweets <- function(x) {
  x %>%
    str_remove_all(" ?(f|ht)(tp)(s?)(:|/|)(.*)" ) %>%
    str_replace_all("&", "and") %>%
    str_remove_all("[[:punct:]]") %>%
    str_remove_all("^RT:?" ) %>%
    str_remove_all("@[[:alnum:]]+") %>%
    str_remove_all("#[[:alnum:]]+") %>%
    str_replace_all("\\\\n", " ") %>%
    str_to_lower() %>%
    str_trim("both")
}
```

```
removeURL <- function(x) gsub("http[[:space:]]*", "", x)
removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
removeUsername <- function(x) gsub("@[[:space:]]*", "", x)
removeSingle <- function(x) gsub(" . ", " ", x)
```

```
mystopwords <- c((stopwords('english')),c("https", "t.co", "it's")) #or read in your stop word list
```

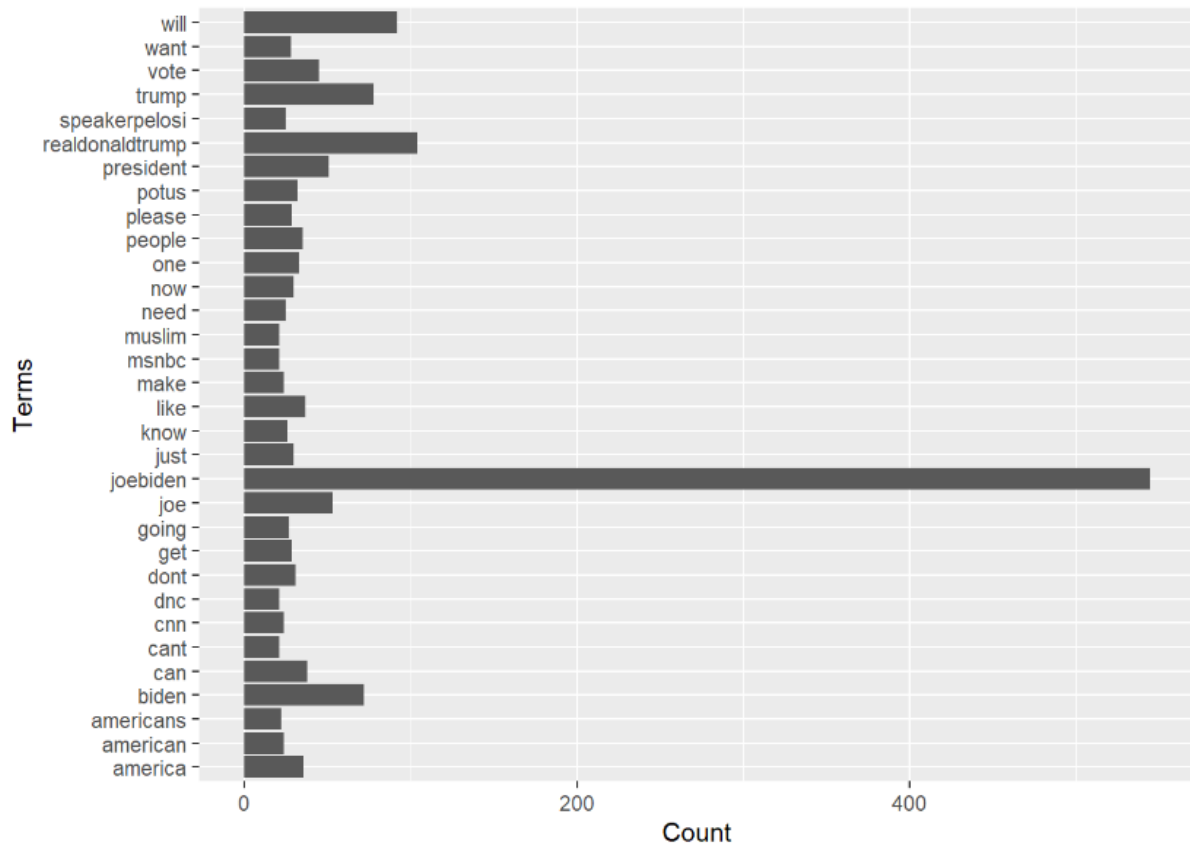
```
path<-"C:/Users/HP/OneDrive/Documents/Twitter_sentimental_analysis" #Set your own path if using.
setwd(path)
```

**Sentiment Analysis:** We perform sentiment analysis using R packages such as Sentimentr, NRClex, and Syuzhet.

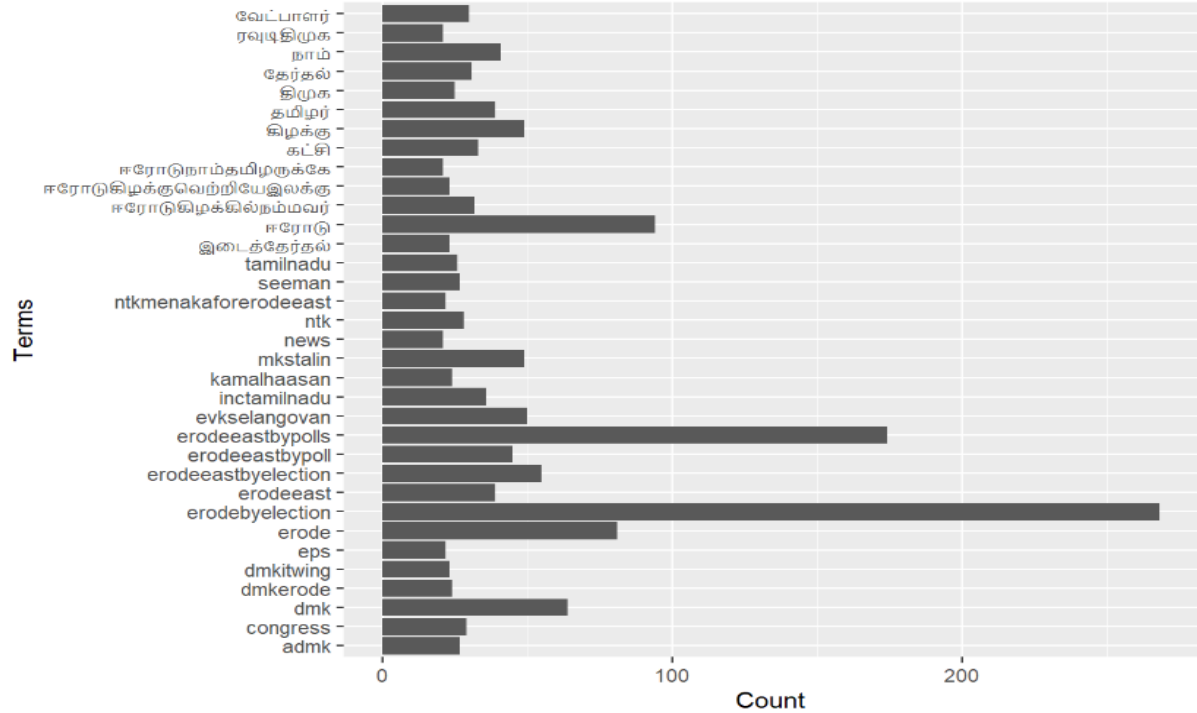
By first finding the frequencies of the word and plotting the bar chart:

```
frequency <- findFreqTerms(tdm, lowfreq=15)
frequency
```

```
## [1] "america"      "american"     "americans"    "back"
## [5] "biden"        "black"        "can"          "cant"
## [9] "cnn"          "come"         "country"      "day"
## [13] "democratic"   "dnc"          "doesnt"       "dont"
## [17] "election"     "even"         "foxnews"      "get"
## [21] "going"        "good"         "gop"          "hes"
## [25] "hillaryclinton" "joe"         "joebiden"     "joyannreid"
## [29] "just"         "kanyewest"    "know"         "like"
## [33] "maga"         "make"         "msnbc"        "muslim"
## [37] "need"         "never"        "now"          "one"
## [41] "people"       "please"       "potus"        "president"
## [45] "realdonaldtrump" "really"      "running"      "schools"
## [49] "see"          "speakerpelosi" "stop"         "support"
## [53] "thedemocrats" "think"       "time"         "trump"
## [57] "vote"         "voting"      "want"         "white"
## [61] "will"         "youre"
```



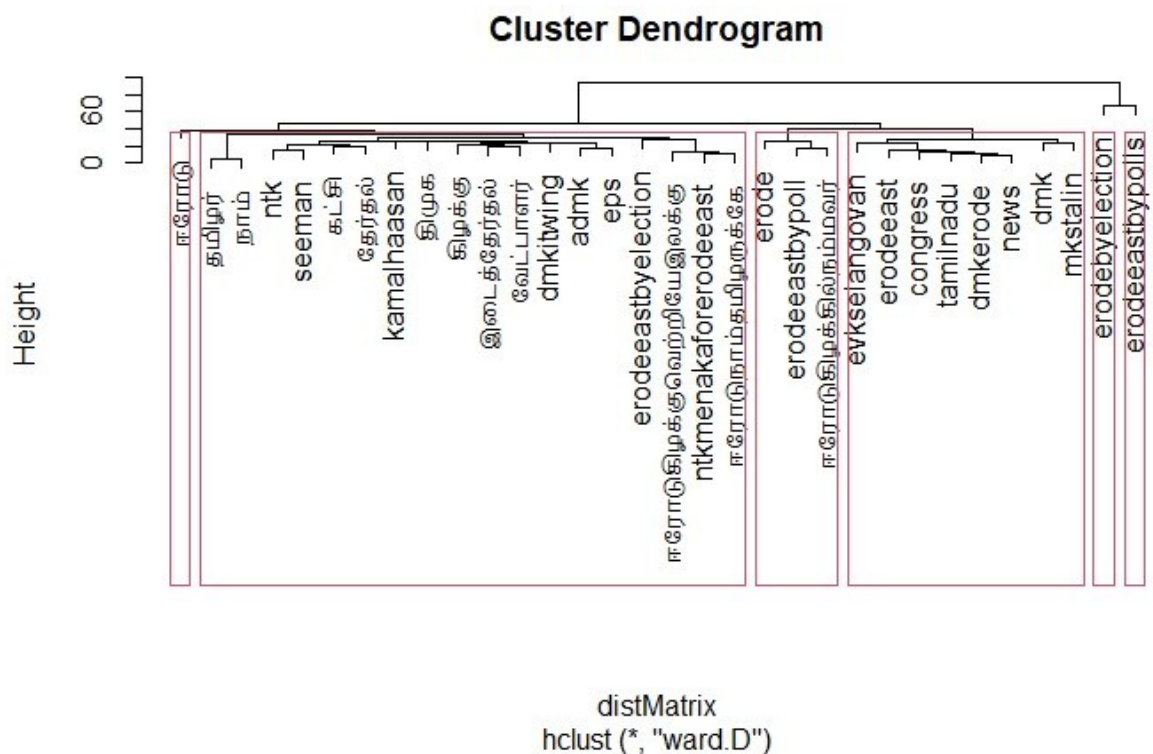
Here most frequently occurred word is Joebiden.



Here most frequently occurred word is erodebyelection.

**Visualization:** We visualize the sentiment analysis results using various R packages such as ggplot2, wordcloud, and plotly.

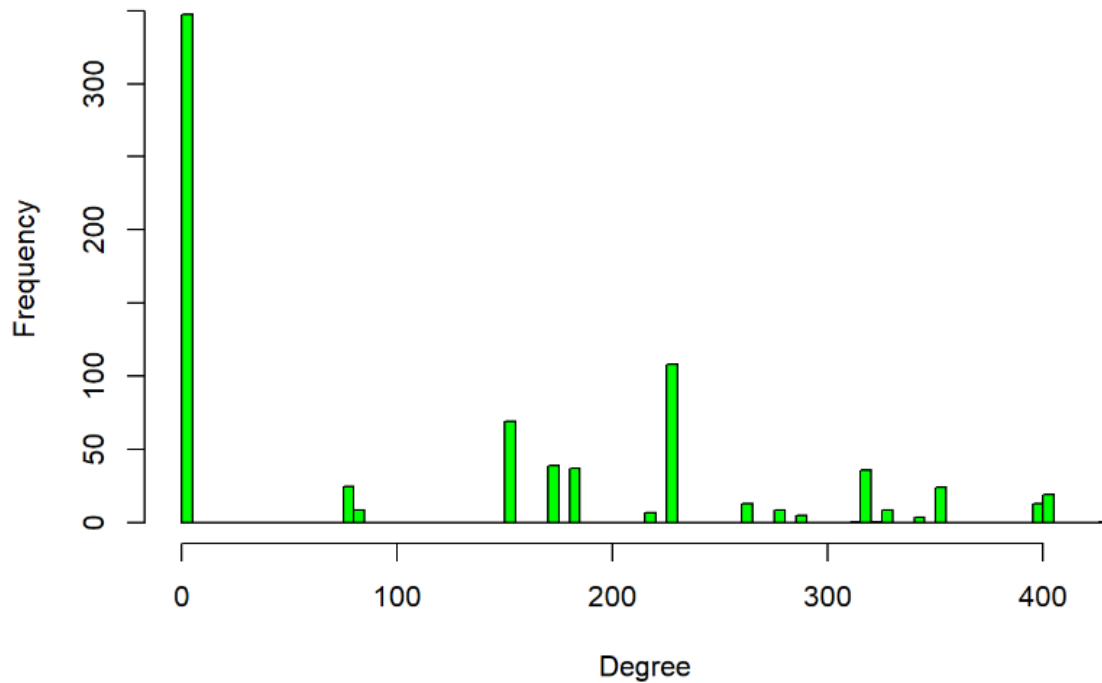
**For erode sentiment analysis visualization plots:**





## Network of chats

### Histogram of Degree



### Wordcloud:

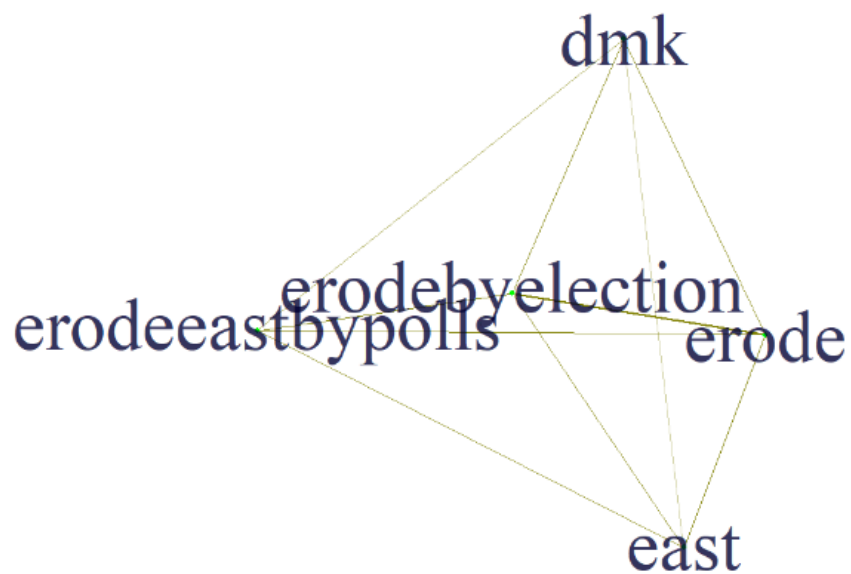


The words with importance that are seen in twitter are mkstalin, dmK, evkselangovan.

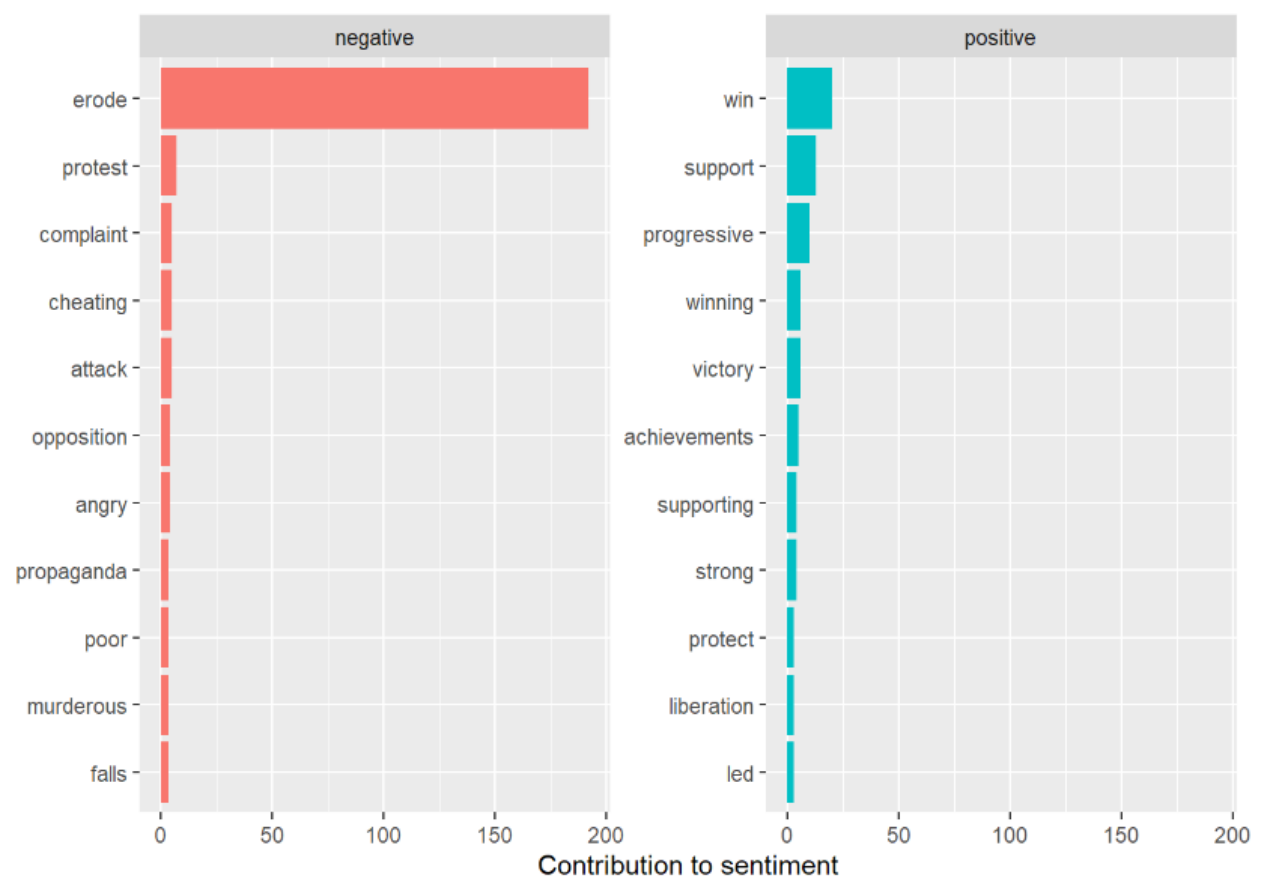




**Highlighting degrees:** Here the degree nodes inclde dmk, erodebyelections, erode,east and erodeeastbypolls.

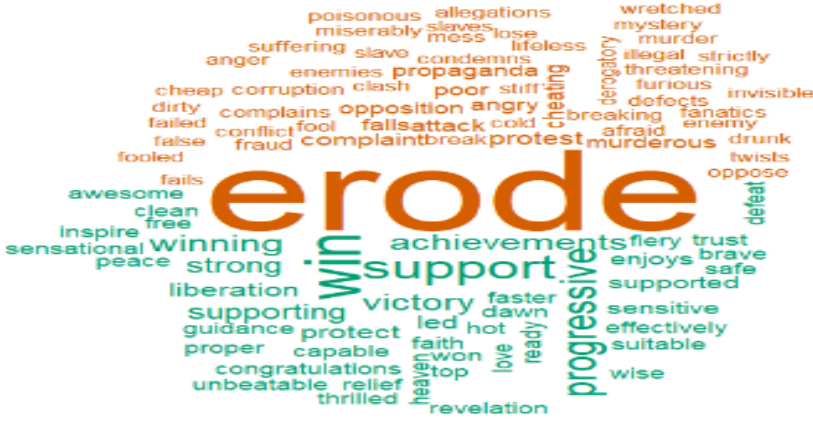


**Contribution to sentiment: Segregating the words as positive and negative.**



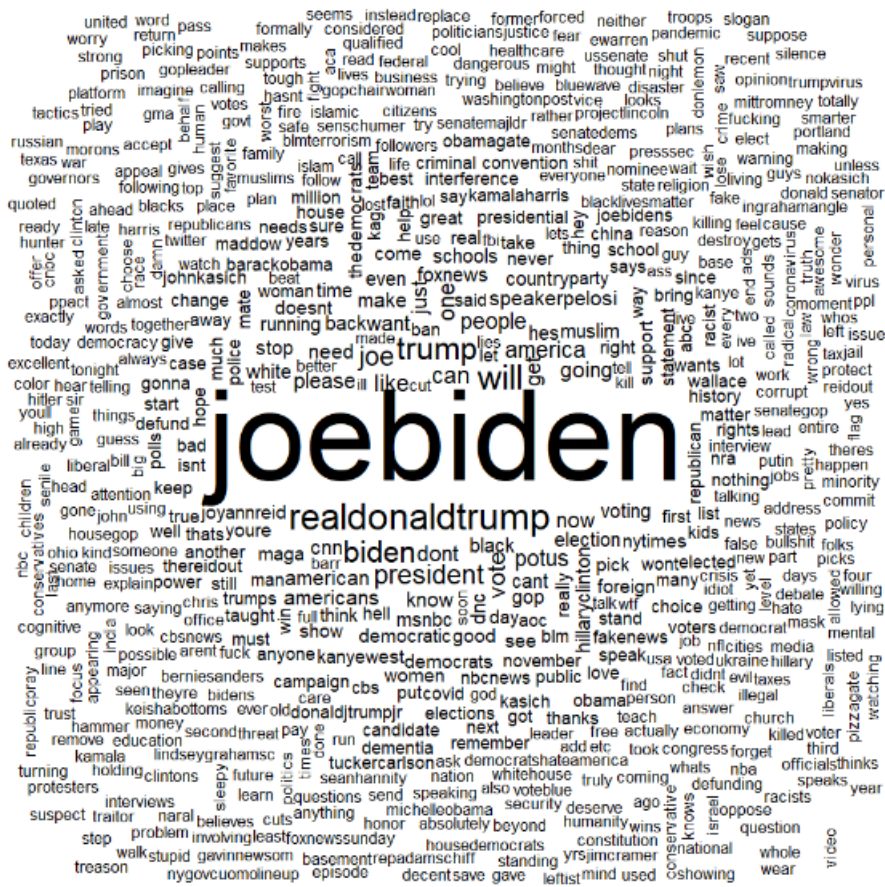


negative



positive

### For trump and joe bidden election sentiment analysis visualization plots:









## **SOFTWARE /TOOLS**

We use the following software/tools for sentiment analysis:

R programming language: R is a statistical programming language that is widely used for data analysis and visualization.

R packages: We use various R packages such as Sentimentr, Tidytext, igraph, rtweet, ggplot2, wordcloud, and plotly for sentiment analysis and visualization.

## **CONCLUSION**

In this paper, we perform sentiment analysis on two important events: the 2020 US Presidential Election and the Erode Election. We analyze the sentiment expressed on social media platforms about the candidates in the 2020 US Presidential Election, Joe Biden and Donald Trump, using R programming language. We also analyze the sentiment expressed during the Erode Election using R. Our analysis provides insights into the public sentiment about political candidates on social media platforms. Our approach can be applied to other political events as well.

From this visualization we were able to infer that the support for Joe Biden is highly seen in Twitter when compared to Trump. Therefore we conclude that Joe Biden has a high chance of winning the election.

Similarly we have done for Erode election and came to the conclusion that DMK party has high chances of winning the election. Since we observed that there is high support for Stalin, it is visualised in the word plot.

Additionally, sentiment analysis algorithms may not always accurately classify the sentiment of tweets, especially if they contain sarcasm, irony, or other forms of figurative language. It's important to take these limitations into consideration when interpreting the results of a sentiment analysis.

Therefore, we cannot make a definite inference on who won the Erode East election based solely on a Twitter sentiment analysis. It is essential to consider other factors such as exit polls, voter turnout, and official election results to determine the winner.

## **REFERENCES**

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Languages in Social Media, 30-38.
2. Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. Proceedings of LREC, 2200-2204.
3. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. O'Reilly Media, Inc.
4. Hutto, C. J., & Gilbert, 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).
5. Khan, S., & Bahsin, S. (2021). Sentiment analysis and prediction of 2020 United States presidential election. Journal of Ambient Intelligence and Humanized Computing, 12(3), 3081-3091.
6. Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751.
7. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
8. Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98-125.
9. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.
10. Wang, S., Ma, X., Huang, L., & Zhang, X. (2021). The impact of political emotions on the US 2020 presidential election: A text mining and sentiment analysis approach. IEEE Access, 9, 57510-57519.