Social Media Insights to decode public sentiment during Political Elections

[1]Jovin Gnanaraj , [2]Yogavind Sarate, [3] Dr. Harshali Patil, [4]Dr. Jyotshna Dongardive

[1,2]STUDENT, [2,3]Professor
Department of Computer Science
University of Mumbai, Mumbai, India.

***Abstract*- Social media platforms have become a significant source of information and communication during political elections. This research paper aims to analyze social media data to gain insights into public sentiment and its impact on political campaigns. By leveraging data science techniques, sentiment analysis, and machine learning algorithms, we aim to understand the dynamics of public opinion, identify influential factors, and predict election outcomes based on social media sentiment.**

***Keywords:* Social media, Sentiment analysis, Political elections, Public opinion.**

I.INTRODUCTION:

1. What is an election?

One of the main important rights of democratic citizens is the establishment — the option to cast a ballot, which picks the delegates or competitors. The method followed to choose these representatives is referred to as an election. During the 5th century BC. Athens Greek shows the oldest democracy roots, in 1789 USA formed the first democratic country. In ancient times in Greek a ball or token was deposited in the ballot for a particular candidate, but in modern times electronic voting machines are used to vote. Election Campaigns are the means by which up-and-comers and ideological groups get ready and present their thoughts and positions on issues to the citizens social media is interactive technologies to form mass communication & sharing information. This research paper aims to analyze social media data to gain insights into public sentiment and its impact on political campaigns. By leveraging data science techniques, sentiment analysis, and machine learning algorithms, we aim to understand the dynamics of public opinion, identify influential factors, and predict election outcomes based on social media sentiment.

II.POLITICAL ELECTION:

A. Election Campaign

Political parties are the means the political parties prepare & show the ideas on societal issues
Types of election campaign:

- **Online**
- **Offline**

-Online

An online election campaign refers to the use of digital platforms and technologies to promote a political candidate or party during an election. It involves leveraging the power of the internet, social media, and other online tools to reach a wider audience and engage with voters.

-Offline

An offline election campaign involves traditional methods of campaigning outside of the online realm

B. Social Media

Social media's profound influence on society is examined in this abstract. It explores its impact on interpersonal relationships, mental health, and identity formation. Additionally, it discusses the role of social media in shaping public discourse, including the spread of misinformation. Social media touches on societal and cultural changes spurred by these platforms, from activism to polarization. Lastly, it acknowledges ethical concerns like data privacy and content moderation, emphasizing the need for responsible usage and ongoing research.

a. Why we choose twitter against all of others (introduction of twitter)

- ✓ **Twitter**

Choosing Twitter over other social media platforms depends on your specific goals and preferences. Here are some reasons why you might choose Twitter:

1. **Real-Time Updates: ** Twitter is known for its real-time nature, making it a great choice for staying up-to-date with breaking news, trends, and live events.
2. **Concise Communication: ** With its character limit (280 characters per tweet), Twitter encourages concise and to-the-point communication, making it ideal for sharing quick thoughts or information.
3. **Global Reach: ** Twitter has a vast user base, allowing you to connect with a diverse global audience, making it suitable for networking, promoting content, or engaging in conversations with people from different backgrounds.
4. **Hashtags and Trends: ** Twitter's use of hashtags and trending topics helps you discover and join discussions on specific subjects or current events easily.
5. **Engagement: ** The platform fosters engagement through features like likes, retweets, and replies, making it conducive for discussions, feedback, and interactions with your followers.
6. **Promotion and Branding: ** Twitter is popular among businesses and influencers for promoting products, services, and personal brands due to its potential for wide-reaching exposure.
7. **Direct Messaging: ** Twitter's direct messaging feature allows private conversations with other users, enabling networking and customer support.

However, the choice of social media platform should align with your goals, target audience, and content type. Different platforms cater to varying demographics and content styles, so it's essential to assess which one best suits your needs.

C. DATA COLLECTION AND PREPROCESSING

a. Data collection

Data collection or Data gathering is the most common way of social event and estimating data on designated factors in a laid out framework, which then empowers one to respond to applicable inquiries and assess results. Information assortment is an exploration part in all study fields, including physical and sociologies, humanities and business. While techniques change by discipline, the accentuation on guaranteeing precise and legit assortment continues as before. The objective for all information assortment is to catch proof that permits information investigation to prompt the detailing of dependable responses to the inquiries that have been presented.

b. Preprocessing

Data preprocessing can allude to control or dropping of data before it is utilized to guarantee or upgrade execution and is a significant stage in the data mining process. The expression "trash in, trash out" is especially material to data mining and AI projects. Data assortment strategies are frequently approximately controlled, coming about in out-of-range values, unthinkable data mixes, and missing values, among different issues.

Investigating information that has not been painstakingly evaluated for such issues can deliver deluding results. In this manner, portrayal and nature of information is vital prior to running any examination. Frequently, data preprocessing is the main period of an AI project, particularly in computational science. On the off chance that there is a high extent of unessential and repetitive data present or loud and temperamental information, then, at that point, information disclosure during the preparation stage might be more troublesome. Information readiness and sifting steps can take a lot of handling time. Instances of techniques utilized in data preprocessing incorporate cleaning, example determination, standardization, one-hot encoding, information change, highlight extraction and component choice.

The goal of this step is to collect a set P of social media posts from different sources (e.g., Twitter), related to the political event E under analysis. As a first step, the different factions, parties or candidates involved in the political event are identified, defined as the set F = {f1, f2, ..., fn}. In particular, in the case of the 2020 US presidential election, we focused on the two main candidates Joe Biden and Donald Trump.

1. Neutral keywords explanation

Neutral keywords typically refer to words or phrases that do not inherently convey a positive or negative sentiment. In the context of natural language processing, sentiment analysis, or search engine optimization, neutral keywords are those that do not evoke strong emotions or opinions.

For example, words like "table," "computer," or "book" are generally considered neutral because they don't carry a positive or negative sentiment on their own. These words are more factual and don't elicit strong emotions from people.

In sentiment analysis, where the goal is to determine the emotional tone of a piece of text, neutral keywords may not contribute significantly to the overall sentiment score. Instead, the focus might be on identifying positive or negative words that express opinions or emotions.

In the context of search engine optimization (SEO), neutral keywords are often used to attract a broader audience without targeting specific emotional responses. This can be useful when the content is intended to be informative or when the goal is to provide unbiased information.

In summary, neutral keywords are those that lack a distinct positive or negative sentiment and are often used in contexts where an objective or unbiased tone is desired.

2. Faction keywords explanation

The keywords selection process requires a small amount of domain knowledge, as these keywords can be manually selected among the trending hashtags that people commonly use to refer to E on social media. Moreover, the keyword selection process can be automatized by searching for specific patterns, like "#vote + candidate", often used for labeling politically polarized posts.

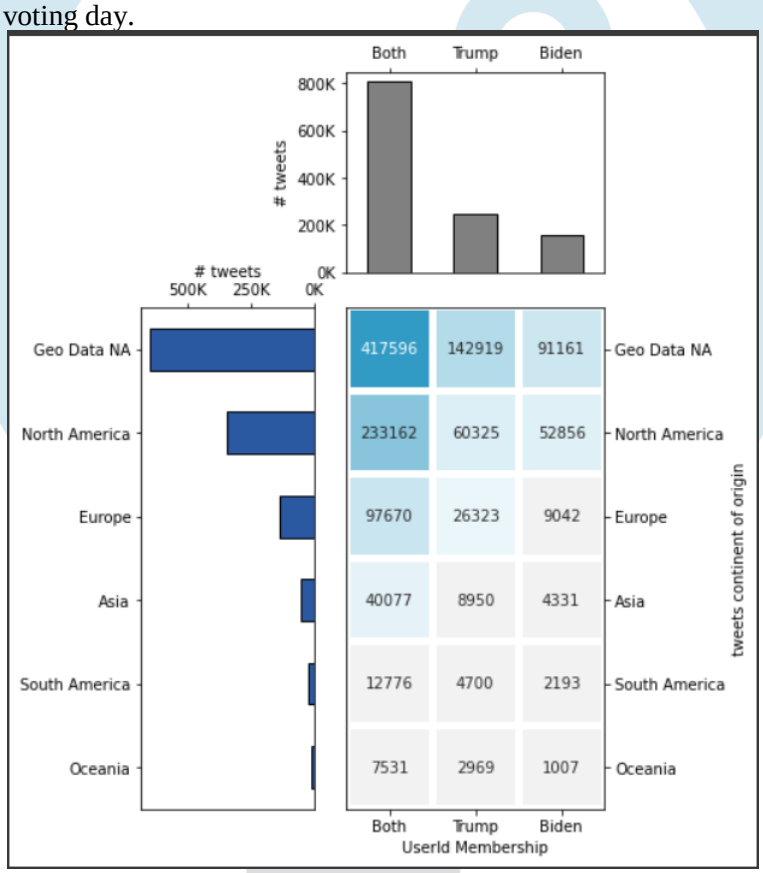
We surveyed the factual meaning of the gathered posts by concentrating on the age, orientation and topographical conveyance of Twitter users for understanding whether they can be viewed as electors of the political occasion. For this reason we utilized an extensive variety of information which can be straightforwardly removed from users

Twitter hashtag, keyword search, trends diagrams

Metadata (e.g., location and language), or examined starting from statistical reports about the usage of the social media platform in a given country (e.g., user distribution by age and gender). Furthermore, in order to improve the representativeness of the collected posts, user accounts are analyzed, filtering those that show anomalous publishing activity, such as social bots or news sites, or those that have inconsistent information in their profile, such as for example a location that is not defined or does not belong to any of the states considered.

D. Comparative Analysis

With assessments of public sentiment in this step we surveyed the adequacy of our methodology in deciding the extremity of social media clients fully intent on understanding which applicant or party popular assessment is most for. A first striking outcome was obtained through a constant examination, completed on Twitter data gathered during the fourteen days before Final voting day.



User Id's that post about both presidential candidates post the most, accounting for greater than 60% of the tweets. Furthermore as with the heatmaps above most tweets are from the America, at least the tweets that have geo data.

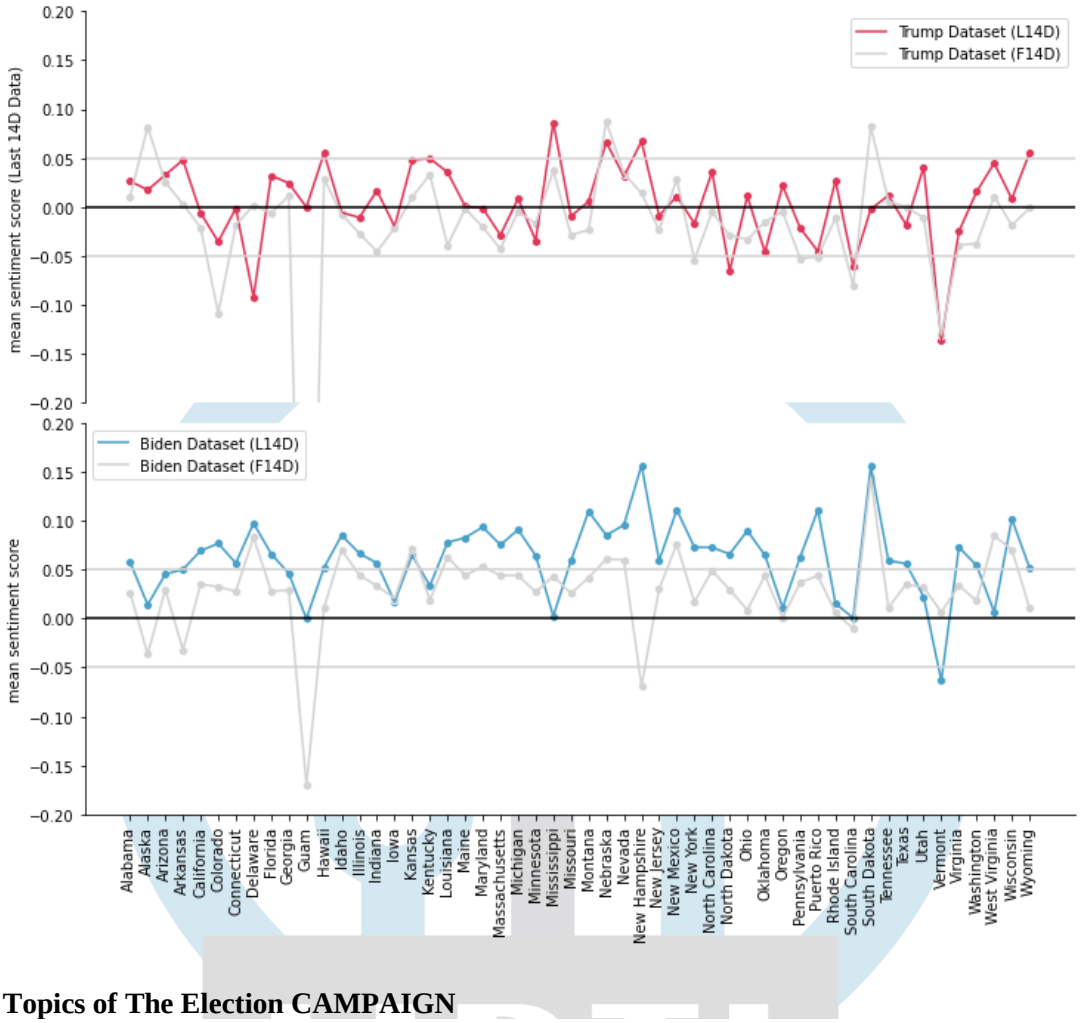
E. Emotion Analysis

This step breaks down the spellbound posts for distinguishing the clients' feeling underlying the online conversation about the official candidates. In particular, we consolidated the information about the political arrangement of social media with their feelings and close to home articulations.

The VADAR analysis produces a compound sentiment score, to create the visualization I took the mean compound score for the most recent 14 days (Labeled L14D on the charts) and the first 14 days (Labeled F14D and light gray on the charts) for each state.

It should be noted that any sentiment score between 0.05 and -0.05 is considered "Neutral".

The results seem to show a large number of states are trending to a “Positive” sentiment score for the democratic candidate from the previous more “Neutral” sentiment. Whereas most states are still largely "Neutral" for the republican candidate.



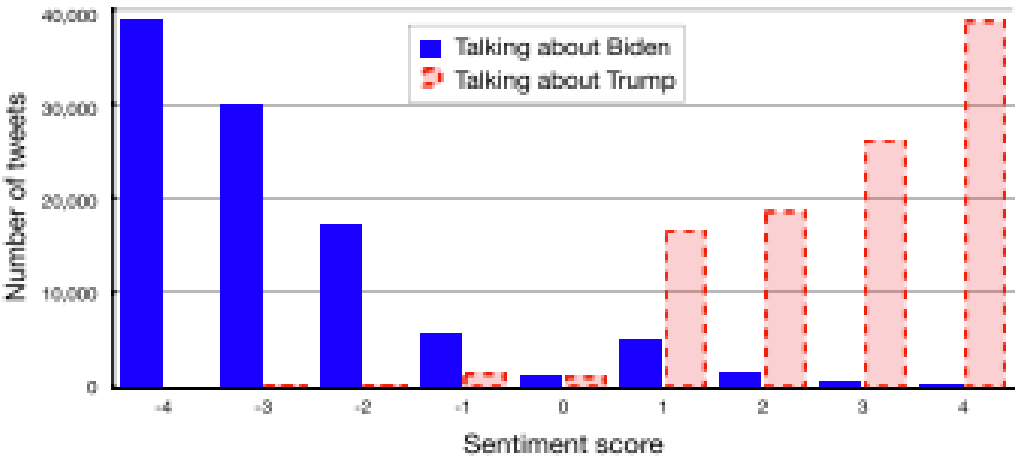
F. Trending Topics of The Election CAMPAIGN

In this step we identified the main politically related discussion topics characterizing the 2020 US election campaign. The above bar chart shows the most common Bi and Tri N-gram's in each of the respective datasets, filtered from words just from the "United States of America". The N-gram's show a clear relation to the upcoming election and each respective dataset seems to be related to each of the presidential candidates. Before the N-gram analysis we first must clean the tweets to remove stop words, strings with "http" etc. and then lemmatize the words.

Cluster ID	Topic	Top hashtags
#1	Bad management of Covid-19 emergency	#trumpknew, #trumpvirus, #covid, #trumpisaloser, #trumpisanationaldisgrace, #trumpliedpeopledied
#2	Town hall meetings; sub-topics: climate crisis, veterans, discrimination	#cnntownhall, #climatecrisis, #greennewdeal, #respectveterans, #hererightmatters, #stoptrumpsterror
#3	Encouraging peopleto vote	#election2020, #voteearly, #vote2020, #votebymail, #voteready, #electionday
#4	Accusations against Hunter Biden	#hunterbiden, #bidencrimefamily, #burisma, #ukraine, #hunterbiden-emails, #china
#5	The US Supreme Court;nomination of Amy Coney Barrett	#scotus, #amyconeybarrett, #filltheseat, #supremecourt, #riprbg, #scotushearings
#6	Support for Trump	#maga, #votetrump2020, #maga2020, #kag, #voteredtosaveamerica2020, #trumppence2020

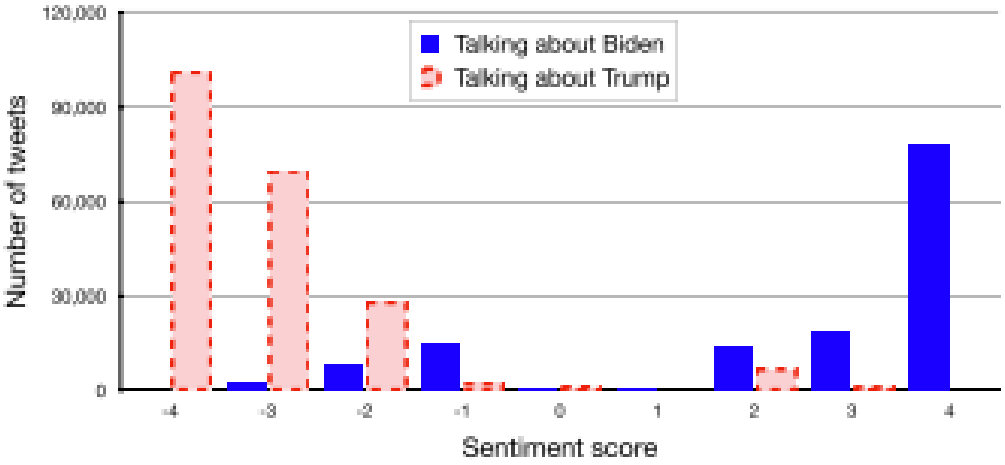
G. Sentiment Analysis

The opinion communicated in alluding to the two presidential candidates. Then, for each enraptured tweet we investigated the emotion the tweet conveys. Figures depict the feeling and the profound condition of the tweets with the general power of the tweets created by Trump and Biden supporters, respectively


Talking about Biden
Talking about Trump
Number of tweets
40,000
30,000
20,000
10,000
0
-4
-3
-2
-1
0
1
2
3
4
Sentiment score

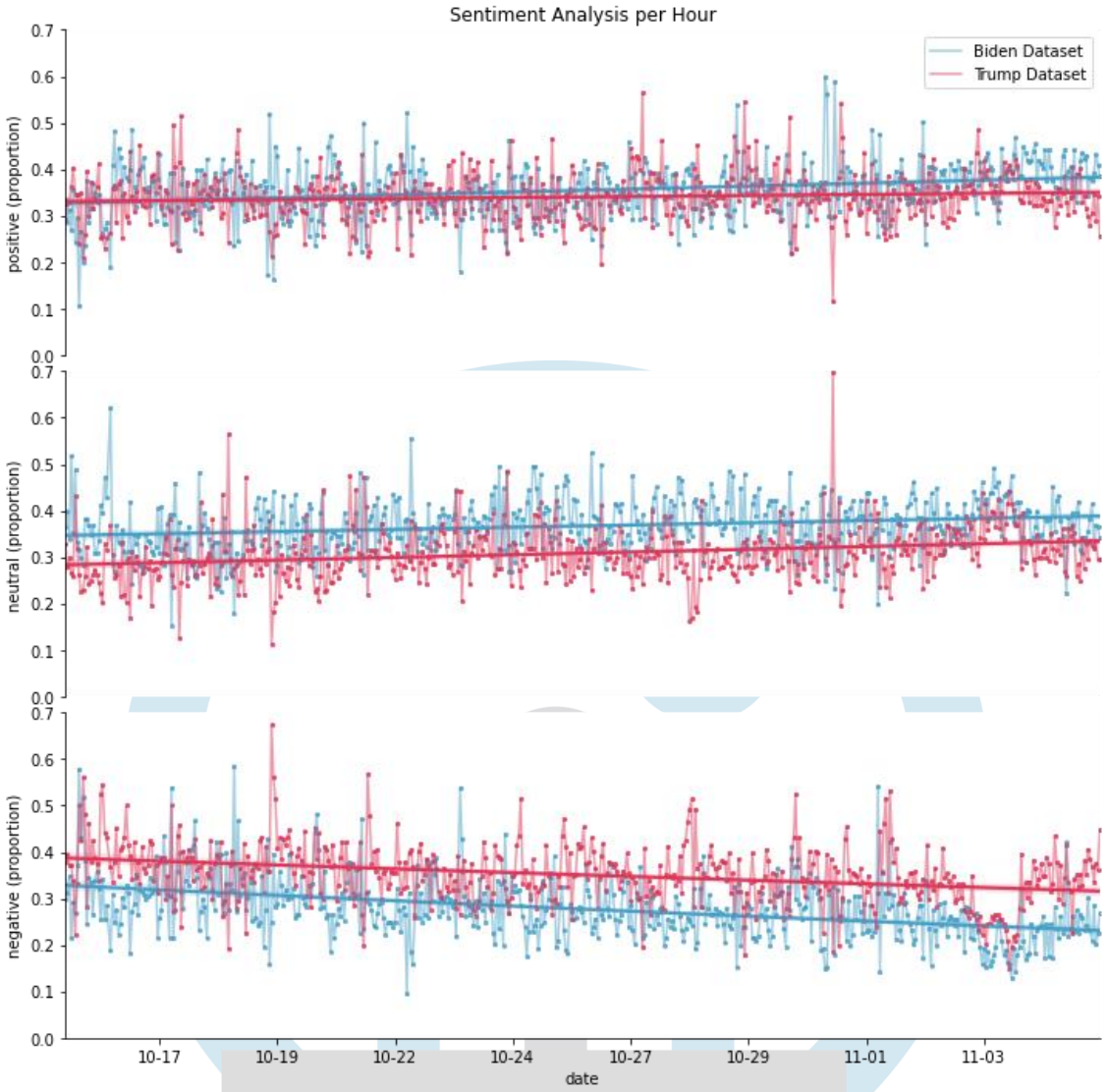
Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust

Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust


Talking about Biden
Talking about Trump
Number of tweets
120,000
90,000
60,000
30,000
0
-4
-3
-2
-1
0
1
2
3
4
Sentiment score

Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust

Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust



The above visualization was generated by first assigning each tweet a "Positive", "Neutral" or "Negative" sentiment then summing those for each day and calculating the proportions for each sentiment group. Then using logistic regression to find the best fit line to better show the sentiment trend overtime. When viewing the results:

The trend over the entire timeframe of the dataset for both presidential candidates is an increasing "Positive" and "Neutral" sentiment, with reducing negative sentiment.

However interesting developments occurred as we approached and subsequent to election day, where we see "Positive" sentiment increase quicker for the democratic candidate over the republican where we see a noticeable gap develop in the logistic regression lines. Moving onto the "Neutral" sentiment, the differences between two candidates remain largely steady until there is a noticeable blip on election day where the gap briefly disappears before then returning to the previous steady difference. Post-election day we see a shape increase in "Negative" sentiment for the republican candidate.

III. CONCLUSIONS:

There is interest in the US elections from many different countries in the world with tweets from 40 different languages, however with a large proportion of the tweets in English and originating from the US.

The sentiment analysis was performed only on data that had geo-data originating from the "United States of America" to try to ascertain the sentiment in each respective dataset and therefore each presidential candidate. When reviewing sentiment at the state level as we approached the election date a large number of states were trending to a “Positive” sentiment score for the democratic candidate from the previously more “Neutral” sentiment. Whereas most states are still largely “Neutral” for the republican candidate. This trend is correlatable when viewing the sentiment analysis from a date perspective.

REFERENCES:

1. Nakov P, Rosenthal S, Kiritchenko S, Mohammad SM, KozarevaZ, Ritter A, Stoyanov V, Zhu X (2016) Developing a successful semeval task in sentiment analysis of twitter and other social media texts. Lang Resour Eval 50(1):35–65
2. Plutchik R (2001) The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am Sci 89(4):344–350
3. Wong FMF, Tan CW, Sen S, Chiang M (2016) Quantifying political
4. leaning from tweets, retweets, and retweeters. IEEE Trans Knowl Data Eng 28(8):2158–2172
5. Manch hui (2020) us-presidential-election-sentiment-analysis
6. Kiritchenko S, Zhu X, Mohammad SM (2014) Sentiment analysis of
7. short informal texts. J Artif Intell Res 50:723–762
8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9288921/
9. https://ebin.pub/evolutionary-computing-and-mobile-sustainable-networks-proceedings-of-icecmsn-2020-1st-ed-9789811552571-9789811552588.html
10. http://www.ijctjournal.org/Volume2/Issue4/IJCT-V2I4P5.pdf
11. https://link.springer.com/article/10.1007/s13278-022-00913-9