Abstract

The “Stop the Steal” movement refers to a campaign that began in the aftermath of the 2020 US presidential election and culminated in the storming of the U.S. Capitol by a mob of Trump supporters. The "Stop the Steal" movement is a highly controversial and polarizing issue, with both sides defending their positions. U.S. Capitol by a mob. This thesis focuses on the sentiment analysis of the movement, with supporters expressing emotions, from strong positive to strongly negative, depending on the perspective of the individual expressing their sentiment. Sentiment analysis techniques can help to identify patterns and trends in the language used by supporters and critics of the movement, shedding light on the underlying emotions and attitudes that drive the discourse around this contentious issue.

Introduction

Twitter is a popular social media platform extensively used for networking ang microblogging where users put messages in the form of tweets. As of 2023, Twitter has more than 450 million active monthly users. As a result, the posts generated are on a variety of topics, with news being one of the major ones. Moreover, Twitter was the largest source of breaking news during the Presidential Elections in 2016. Twitter has a huge impact on the political sphere in the USA as the success of the leader always depends heavily on their ability to communicate with the masses.

This work presents the investigation of the "Stop the Steal" movement emerged in the aftermath of the 2020 United States presidential election. The “Stop the Steal” hashtags gained widespread attention on social media platforms, including Twitter, during the 2020 U.S. presidential election as supporters of former President Donald Trump claimed that the election was rigged, and that widespread voter fraud had taken place in key swing states. The movement gained traction on social media platforms and culminated in the storming of the U.S. Capitol by a mob of Trump supporters on January 6, 2021.

Performing sentiment analysis on the "Stop the Steal" movement is a complex and polarizing issue. Supporters of the movement express strong emotions, such as anger and frustration, and may make accusations against those they perceive as responsible for the alleged fraud. Critics of the movement, on the other hand, may express feelings of disbelief, disgust, or even fear at the actions of its supporters, and see the movement as a threat to democracy and the peaceful transfer of power.

Analyzing sentiment on the "Stop the Steal" movement requires a balanced and nuanced perspective, as there are diverse range of opinions on both sides of the debate. Sentiment analysis techniques can help to identify patterns and trends in the language used by supporters and critics of the movement, shedding light on the underlying emotions and attitudes that drive the discourse around this contentious issue.

Here, we look at the Twitter dataset regarding the “Stop the Steal” movement and build models for classifying positive, negative, and neutral sentiment. Every time someone tweets about something new, they include a "Hashtag" that is frequently the primary focus of analysis. Utilizing Twitter's streaming

API, the data is gathered. The fact that the tweets were gathered in a streaming method and so accurately reflect actual tweets in terms of language use and content gives this data one edge over manually used data sets. The Lexical method can be used for sentence-level sentiment analysis. Vader, a free tool that is a component of the NLTK Python library, performs sentence-level sentiment analysis. To determine the text's polarity and categorize it as positive, negative, or neutral, use the 'SentimentIntensityAnalyzer' method under Vader. Another important analysis can be done using LIWC analysis, which provides insights into the emotional tone of the movement's language by getting the measures of frequency of words in different categories, such as positive and negative emotions, cognitive processes, and social processes.

In this work, different analysis has been done focusing on sentiments regarding "Stop the Steal" movement.

Literature Review

Sentiment analysis is the computational analysis of a person's opinions, perceptions, behaviors, and feelings as they are expressed in written language. It is currently one of the most active study areas in natural language processing and machine learning. For the most part, there are two factors that account for its prominence. First off, it has a wide range of applications because emotions are fundamental to our behavior and a major factor in determining our patterns.

(Pang & Lee, 2008) Due in part to the vast rise in the number of tweets and messages expressing opinions that have become available over the past ten years, there has been a rapid growth in interest in sentiment and opinion mining in text. A number of fields, including the stock market, politics, and social movements have exploited sentiment in Twitter data for prediction or assessment (Bollen et al., 2011; Choy et al., 2011; Tumasjan et al., 2010; Zeitzoff, 2011). For instance, Choy et al. (2011) failed to predict via Twitter sentiment the ranking of the four candidates in Singapore's 2011 presidential election, although Tumasjan (2010) found tweet volume about the political parties to be a decent predictor for the outcome of the 2009 German election. Political sentiment on social networks has previously been studied post-hoc or using small, static samples.

Based on real-time Twitter data [26] that was obtained from Twitter via the Twitter-streaming application programming interface (API) [27], two politicians were compared. Positive and negative scores were discovered using SentiWordNet [28] and WordNet [29], two sentiment analyzers. Word sequence disambiguation (WSD) and negation handling were utilized to improve the model's accuracy [30, 31]. The authors of [32] also gathered data for the prediction of the Indonesian presidential elections using the Twitter streaming API. It was intended to leverage Twitter data to gauge public sentiment. After gathering data for this study, automatic buzzer recognition was used to remove meaningless tweets, and the tweets were then sentimentally examined by dividing each tweet into many sub-tweets. After gathering data for this study, automatic buzzer recognition was used to remove out meaningless tweets, and the tweets were then sentimentally examined by dividing each tweet into many sub-tweets. Then, it used mean absolute error (MAE) [33] to measure the performance of the prediction and claimed that this Twitter-based prediction was 0.61% better than the same type of surveys carried out traditionally. Next, it calculated sentiment polarity and used positive tweets associated with each candidate to predict election outcome.

Here, we created a special infrastructure and sentiment model to assess Twitter user sentiment in real-time in relation to the 2020 U.S. presidential election to address this problem related to “Stop the Steal” movement. Our approach combines real-time data processing and statistical sentiment modeling informed by and contributing to an understanding of the cultural and political practices at work by Twitter.

Methodology

We have presented our framework in which we have explained a process from the collection, sentiment analysis, and classification of Twitter opinions. We considered tweets that were posted by users in the form of hashtags to express their opinions about “Stop the Steal” movement. We then stored the retrieved tweets in the database and preprocessed the dataset. Here, the total of 330,958 tweets were extracted from 9/1/2020 till 1/21/2021.

1. Technology used:

Initially, Liwc-22 had been used to extract meaningful insights into psychological states, including their emotions, thinking styles, and social concerns. The twitter dataset was also preprocessed using NLTK python library. Furthermore, this framework uses other necessary libraries like Pandas, Numpy, Seaborn, Matplotlib, TextBlob and scikit learn for further processing.

2. Datapreprocessing

The following dataset was generated from up of user tweets, which typically contain information like mentions, hashtags, emoticons, and stop words that aren't very helpful for identifying user emotions. To prevent the model from being trained with an incorrect interpretation, cleaning the data is a crucial step. The following steps are used to achieve this and assure improved performance:

- Lowercasing all the text using the `lower()` method.
- Removing all the URLs using regular expressions `re.sub(r"https\S+|www\S+https\S+", '',text, flags=re.MULTILINE)`.
- Removing all the usernames and hashtags using regular expressions `re.sub(r'\@w+|\#','',text)`.
- Removing all the punctuation using regular expressions `re.sub(r'[^\w\s]','',text)`.
- Tokenizing the text using the `word_tokenize()` method.
- Removing all the stopwords using the `stopwords` package from NLTK.
- Stemming the text using the `PorterStemmer()` function from NLTK.

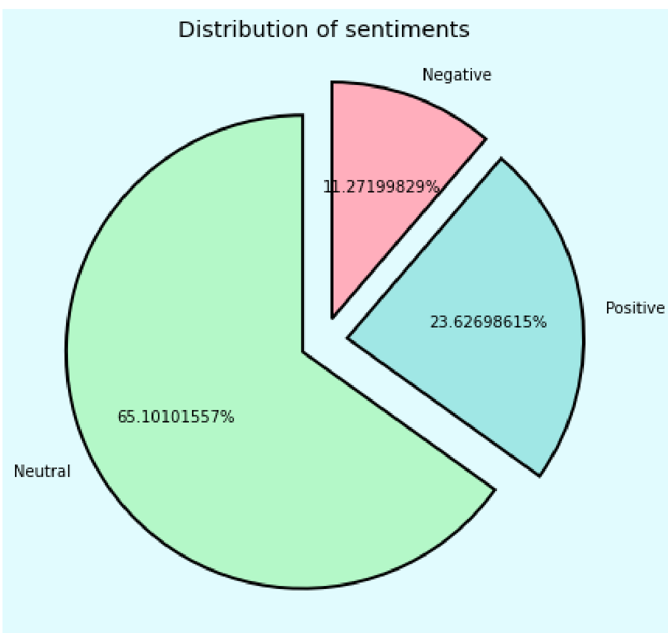
After the data preprocessing was done, the preprocessed data was stored in the ‘tweet’ column of the data-frame.

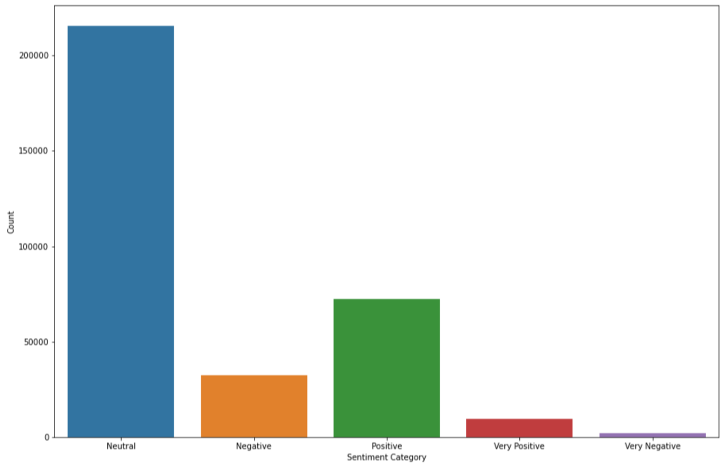
3. Liwc analysis

LIWC-22 application was used to analyze the twitter dataset to extract meaningful insights into the emotions, attitudes and linguistic characteristics associated with “stopthesteal” hashtag or related content. Here simply the raw csv file was uploaded and liwc analysis was conducted on tweets of the following dataset.

4. Sentiment Analysis

Sentiment analysis is further conducted to analyze sentiment analysis on user tweets from the liwc analysis result. The framework then defines a stemming function algorithm to convert each word to its root form and calculates the polarity of each tweet using TextBlob and classifies the sentiment based on the polarity. After that, it generates a countplot to show the distribution of sentiments and a pie chart to show the percentage of each sentiment. It also creates a word cloud for the most frequent words in positive and negative tweets. The code also sorts the tweets based on polarity to show the most positive or negative tweets.





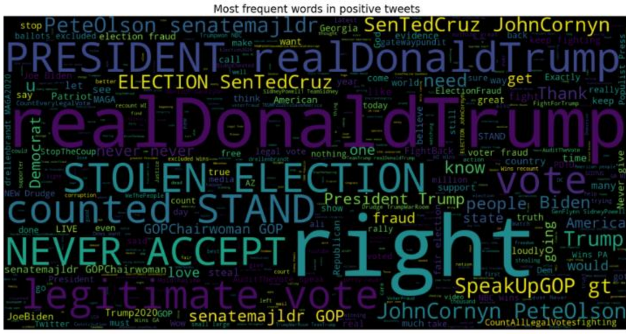
Also, various methods were analyzed to get the visualization of the precise information about this movement. With the period of 4-month in the dataset, it was analyzed on a weekly basis. The analysis such as Posting volume over time, word frequencies, sentiment score by week, and sentiment analysis over time had been analyzed.

5. Result

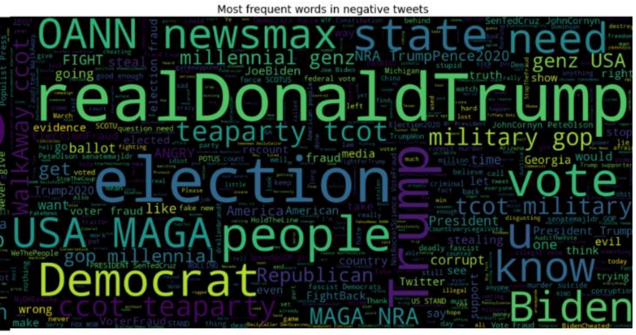
5.1 Word Frequencies

Various words such as “election”, “fraud” and “realDonaldTrump” were found to be one of the most common words after analyzing the frequency of words. This could suggest the primary theme of these discussions is the belief that the 2020 U.S. presidential election was fraudulent, and that former President Donald Trump was unfairly treated. By understanding the most used words and themes in “stopthesteal” discussions, it could furthermore give a better understanding of the conversation surrounding this topic and its potential impact on society.

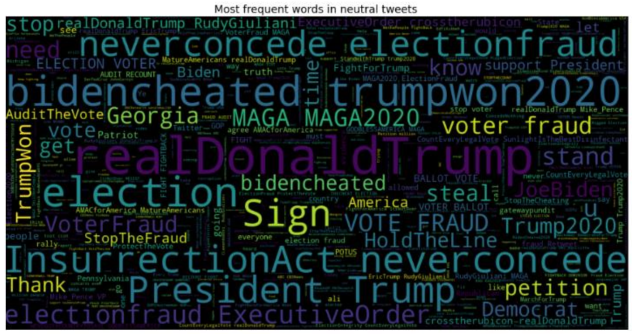
Most frequent words in positive tweets



Most frequent words in negative tweets



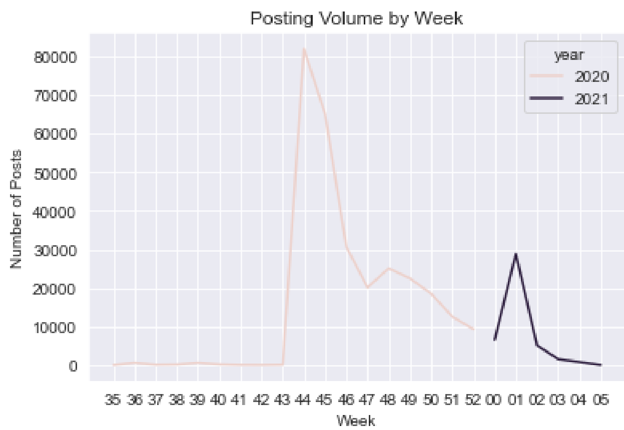
Most frequent words in neutral tweets



5.2 Posting Volume over Time (Weekly)

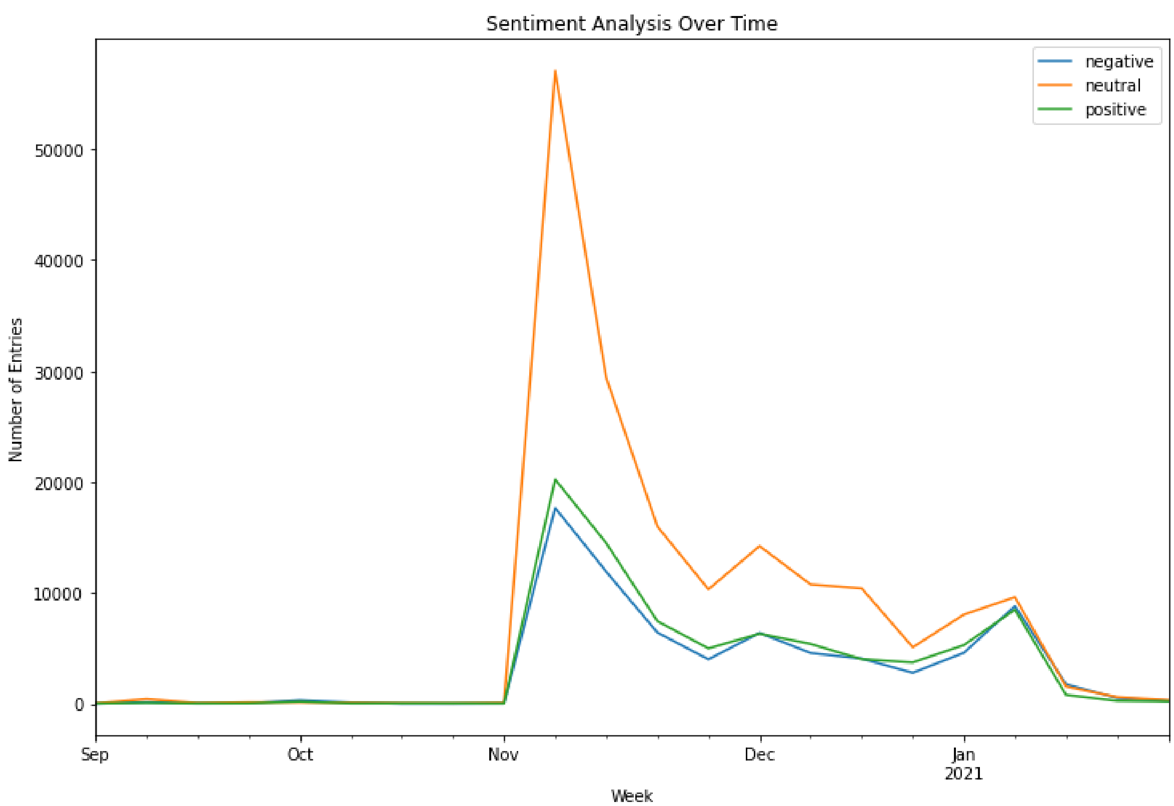
Posting volume is a method to track the frequency and volumes of posts related to a particular topic over time. By analyzing the posting volume related to "stopthesteal" on a weekly basis, we can get insights into the intensity of the conversation surrounding this topic. For example, a posting volume analysis might reveal that the volume of posts related to "stopthesteal" spiked in the weeks following the 2020 U.S. presidential election, particularly in the weeks leading up to the certification of the election results. This could suggest that the certification process and related legal challenges were driving the conversation around "stopthesteal" during that time. A posting volume analysis could also reveal shifts or changes in the conversation over time, such as a decline in the frequency of posts related to "stopthesteal" after the inauguration of President Biden, or a resurgence in the conversation after the January 6th Capitol riot.

Overall, a posting volume analysis can provide valuable insights into the ebb and flow of the conversation surrounding "stopthesteal" over time and can help to identify potential events or factors that are driving the conversation.



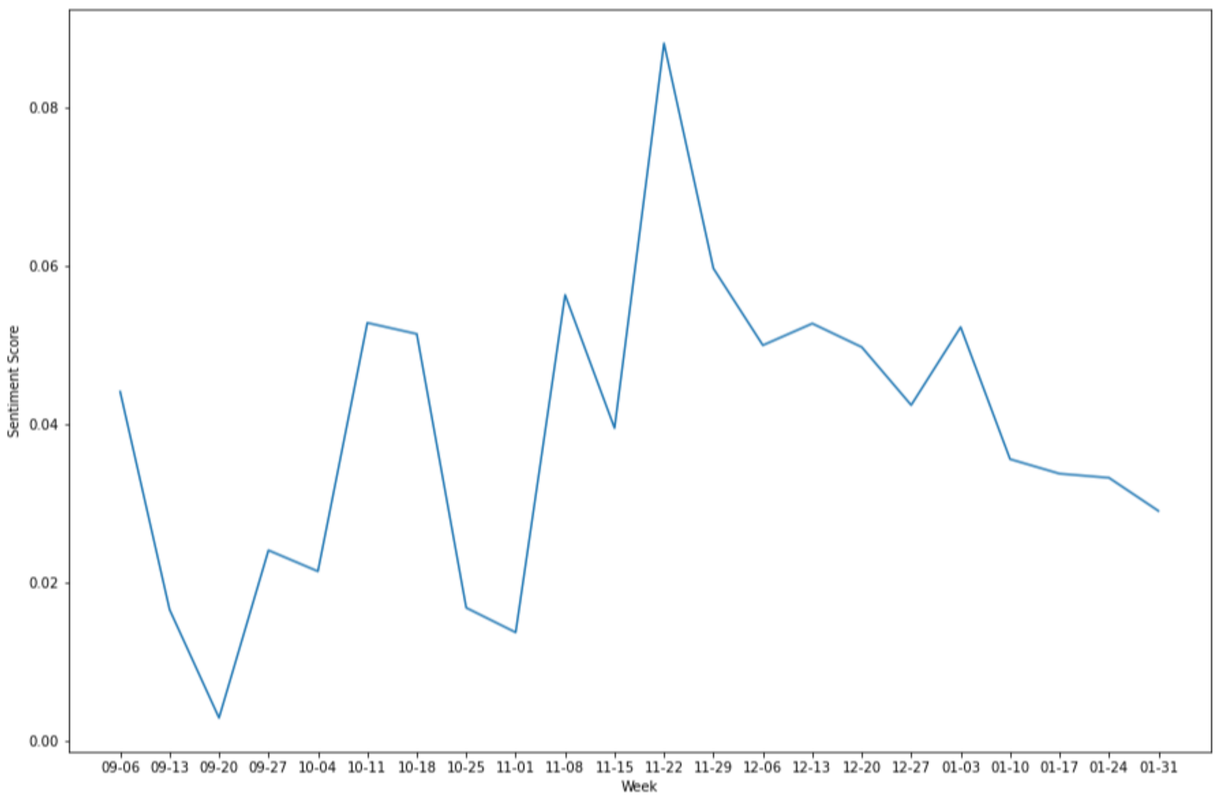
5.3 Sentiment Analysis over Time (Weekly)

Here, the sentiment analysis over time had been conducted to find the sentiments of tweets based on time.



5.4 Sentiment Score (Weekly)

Sentiment score is a numerical representation of the sentiment or emotional tone of a text or tweets. Here, sentiment score was analyzed based from the overall sentiments gained from sentiment analysis.



6. Limitations and Future Work
7. Conclusion