# Twitter Sentiment Analysis for the 2020 U.S. and 2019 Australian Elections

Bharath Surianarayanan bs4224

Katerine Perdomo Moreno kpm8481

Nishanth Sanjeev ns5287

MD Shahedur Rahman mr6139

New York University

## Abstract

In this paper, we aim to use tweets from the social media platform Twitter to analyze public sentiment towards different parties in the U.S. 2020 Presidential Election and the Australian 2019 Election. The 2020 U.S. presidential election was a strongly contested election between the Democratic candidate Joe Biden and the Republican candidate Donald Trump. The 2019 Australian federal election was contested between Scott Morrison and Bill Shorten. We use our analytics to assign a sentiment score to various tweets and understand how the public sentiment towards the user affected their chances at the election. We find that this positive sentiment, on average corroborates the actual results of these elections.

## 1 Introduction

Sentiment refers to an idea or view based on an emotion about a particular topic. These sentiments are crucial in determining the attitude or emotional response about some person, event, or object. Prakruthi et al. [2018]

Twitter serves as one of the most widely used platforms that can be used to share opinions about a particular product or topic. If people find a topic interesting, they will share their views on it. An election is a formal group decision-making process by which a population chooses an individual or multiple individuals to hold public office. Understanding the public opinions about the election or the prospective candidates could help us understand/predict the outcome of the election. Our paper aims to analyze the tweets regarding various candidates and understand their sentiment scores (positive, negative, or neutral). By using multiple data sources, the analytic will extract the sentiment score and categorize the locations where a candidate is favored using the overall sentiment.

We test the goodness of our analysis by comparing our results against previously published election results and how the Twitter sentiment influenced the voting process. If our analytic results coincide with the expected output, it indicates that the results obtained are accurate.

## 2 Motivation

Typical users of this analysis include news outlets/media companies and political parties.

The information obtained will help the news outlets to predict the outcome of the results and give the general public an idea of the election outcome. Sentiment analysis can play a supporting role in any political campaign because political parties can get insights into the moods and opinions of voters, which might help them choose their future candidates or plan an appropriate strategy for forthcoming campaigns.

Understanding the public sentiment and mood towards a particular candidate can provide insights (and possibly predict) the

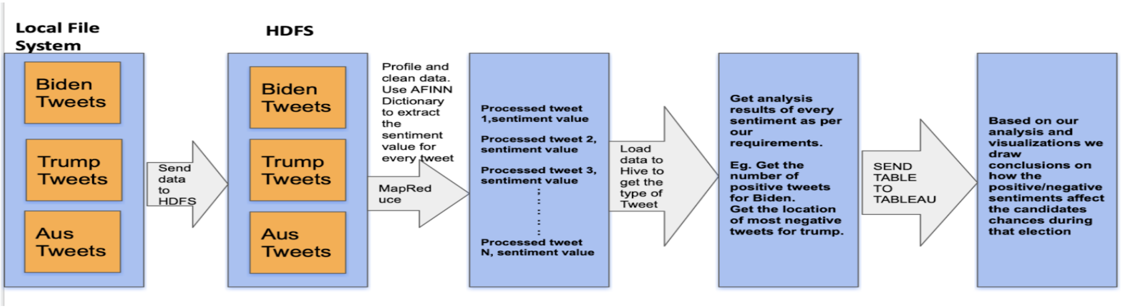


Figure 1: Flow diagram of the analytic process.

outcome of a nationwide election. For this reason, analyzing Twitter posts (a.k.a 'tweets') can be an inexpensive way to mine the collective sentiment of Twitter users towards particular electoral candidates in the time leading up to an election.

## 3 Related Work

Xia et al. [2021] conducted a tweet sentiment analysis of Donald Trump and Joe Biden's 2020 U.S. Presidential Election. They determined that the Multi-Layer Perceptron classifier performed the best on the Sanders Twitter benchmark dataset. They concluded from the results that the candidates had a very close negative to positive sentiment ratio, that negative sentiment is more common and prominent than positive sentiment within the social media domain, and that sentiment trends on social media can detect some key events. Compared to our project, they used a sample of data from a Sanders corpus to train the model and identified the Multi-Layer Perceptron model. In contrast, we only used a bank of words to calculate the sentiment values.

Furthermore, Chaudhry et al. [2021] analyzed and determined public views from Twitter sentiment related to the 2020 U.S. elections - before, during, and after these events. They also compared these views with the actual election results (similar to our study). In contrast to our study, however, they used TF-IDF to extract the relevant textual features and applied the Naive Bayes Classifier to obtain these public opinions. Upon analyzing swing states and cross-validating sentiment with the election results in reality, they found that election outcomes coincide with social media sentiment in most cases.

Additionally, Yaqub et al. [2020] considered two case studies - the U.S. elections of 2016 and the U.K. general elections of 2017. They performed a sentiment analysis of Twitter location data regarding these studies to evaluate the correlation between the real-life public opinion and the sentiment of location-based tweets (similar to our study, which considered tweet location). They discovered that there was indeed a high amount of corroboration between the two.

Dhawan et al. [2019] proposes an analytic for sentiment analysis of the Twitter dataset. The method aims to calculate the polarity of each tweet to distinguish whether the tweet is positive or negative, where sentiment polarity represents the emotions of users, such as anger, sadness, happiness, and joy. While the method offers meaningful insights, it is not scalable.

Kusrini and Mashuri [2019] proposes a technique for tokenization sentiment, elimination of stop words, and stemming. This study focused on developing sentiment analysis using lexicons and multiplication polarity. While the accuracy of the proposed analytic is lower than that of NLP based methods, the proposed tokenization sentiment is highly scalable and offers meaningful insights for our work.

Trupthi et al. [2017] provides an interactive automatic system to predict the sentiment of the tweets posted on social media using Hadoop, which can process a massive amount of data. The principal objective of this paper is to perform real-time sentimental analysis on the tweets extracted from Twitter and provide time-based analytics to the user. This paper addresses the challenges in sentiment analysis, particularly the issues of obtaining real-time Twitter data.

## 4 Dataset Description

In this project, we used the following datasets:

Trump Election Tweets, in this dataset,

includes #Trump or any other keyword; Mined between 15.10.2020 and 08.11.2020.

Biden Election Tweets, in this dataset, includes #biden or any other keyword; Mined between 15.10.2020 and 08.11.2020.

Australian 2019 Election Tweets, include tweets obtained during the 2019 Australian Election.

US 2020 Election Tweet Dataset Sabuncu [2020], tweets related to the US Presidential Election on November 3, 2020, collected between July 1, 2020, and November 11, 2020.

# 5 Design and Implementation

## 5.1 System Configuration

The code was compiled and tested in NYU'S Peel Hadoop Cluster.

- 18 compute nodes
- 48 Intel CPU cores each (864 total)
- 384 GB of RAM each (  7 TB total)
- Cloudera Hadoop distribution

Hadoop Filesystem (HDFS) Configuration on Peel:

- Total Storage: 1.0PB
- Block Size: 128MB
- Replication Factor: 3

Operating System

- Linux (RHEL 7.8)

Cloudera Hadoop distribution includes the following:

- Hadoop: 3.0.0
- Spark: 2.4.0

## 5.2 Design

- Figure 1 shows the design details of our project. The first stage in the pipeline represents the data transfer into the *Hadoop File System* for pre-processing to remove unwanted data such as; retweets, URLs, links, and tweets from paid users and robots to avoid the contradiction of the outcome.
- The next stage represents the cleaning and profiling of the datasets using *MapReduce* which leverages the use of AFINN dictionary Nielsen [2011] to assign the appropriate sentiment score.
- The data is then fed to Hive to get the type of "tweet", process it, and analyze the structured data. This is to determine the count and the effect of the "tweets". For example, we can find the number of positive tweets that mentioned Joe Biden or get the locations of the most negative tweets.
- The table is sent to Tableau for further analysis, which will provide clear visualizations of the data to facilitate straightforward interpretation of data.

# 6 Results

- The overall sentiment towards Biden is positive across the Eastern and Western states in the U.S.
- While the sentiment towards Trump across those states is largely negative.
- Figures 2 and 3 shows the average sentiment towards Biden and Trump in each U.S. state, and we can clearly see that Biden has the major support in the Eastern and Western states.
- The sentiment towards Scott Morrison in the Australian Election is higher than Bill Shorten's.
- Figures 4 and 5 shows the overall sentiment towards Scott Morrison and Bill Shorten in all locations, and we can clearly see that Scott Morrison has the better sentiment score.

Figure 2: State-wise sentiment analysis towards Biden in each US state

# 7 Obstacles

1. Parsing the tweets to remove unnecessary symbols is an important challenge,


Figure 3: State-wise sentiment analysis towards Trump in each US state

| location | sentiment_value_scott |
|---|---|
| WASHINGTON, DC | 307.0 |
| UNITED STATES | 182.0 |
| SYDNEY, AUSTRALIA | 176.0 |
| SYDNEY, NEW SOUTH WALES | 161.0 |
| NEW YORK, USA | 108.0 |
| LONDON, ENGLAND | 106.0 |
| MELBOURNE, VICTORIA | 85.0 |
| GLOBAL | 77.0 |
| MELBOURNE, AUSTRALIA | 75.0 |
| SINGAPORE | 73.0 |
| HONG KONG | 69.0 |
| CANBERRA | 69.0 |
| USA | 68.0 |
| FRANCE | 66.0 |
| PARLIAMENT HOUSE, CANBERRA | 66.0 |
| CANADA | 66.0 |
| INDIA | 62.0 |
| PAKISTAN | 60.0 |
| WORLDWIDE | 57.0 |
| NEW YORK, NY | 52.0 |
| LONDON | 52.0 |
| TORONTO, ONTARIO | 50.0 |
| SYDNEY | 50.0 |
| MELBOURNE | 49.0 |
| EARTH | 45.0 |
| CALIFORNIA, USA | 44.0 |
| NEW YORK | 40.0 |
| VICTORIA, AUSTRALIA | 40.0 |
| BURNABY, BRITISH COLUMBIA | 40.0 |
| UNITED KINGDOM | 39.0 |
| ADELAIDE, SOUTH AUSTRALIA | 38.0 |
| LOS ANGELES, CA | 36.0 |

Figure 4: State-wise sentiment analysis towards Scott Morrison in all locations

| location | sentiment_value_bill |
|---|---|
| MELBOURNE, AUSTRALIA | 109.0 |
| SYDNEY, AUSTRALIA | 87.0 |
| CANBERRA | 85.0 |
| SYDNEY | 70.0 |
| NASHCAPVILLE | 69.0 |
| MELBOURNE | 57.0 |
| PERTH, WESTERN AUSTRALIA | 55.0 |
| ADELAIDE, SOUTH AUSTRALIA | 52.0 |
| BRISBANE | 49.0 |
| SYDNEY, NEW SOUTH WALES | 48.0 |
| NEWCASTLE, AUSTRALIA | 46.0 |
| MELBOURNE, VICTORIA | 44.0 |
| SYDNEY AUSTRALIA | 43.0 |
| PERTH, AUSTRALIA | 40.0 |
| CAPITAL HILL, CANBERRA | 35.0 |
| CANBERRA, AUSTRALIAN CAPITAL TERRITORY | 34.0 |
| NEWCASTLE, NEW SOUTH WALES, AUSTRALIA | 32.0 |
| LIBERAL GOVERNMENT UTOPIA | 31.0 |
| WILSON, PERTH WA | 28.0 |
| WESTERN AUSTRALIA | 25.0 |
| AUSTRALIA MURDOCRACY HOSTAGE | 25.0 |
| BRISBANE, AUSTRALIA | 24.0 |
| MELBOURNE | 24.0 |
| UK | 23.0 |
| MELBOURNE, AUSTRALIA | 22.0 |
| VICTORIA, AUSTRALIA | 22.0 |
| RED EARTH , SASKATCHEWAN | 21.0 |
| CANBERRA, ACT | 21.0 |
| SYDNEY, AUSTRALIA | 19.0 |
| BRISBANE | 19.0 |
| INDONESIA, GLOBAL | 18.0 |
| HOBART, TASMANIA | 17.0 |
| WURUNDJERI, AUSTRALIA | 17.0 |
| MORNINGTON PENINSULA | 16.0 |
| SYDNEY AUS | 16.0 |
| AUSTRALIA UK | 16.0 |
| MELBOURNE, OZ | 15.0 |
| COOLANGATTA, GOLD COAST | 15.0 |
| WASHINGTON, DC | 14.0 |
| SUNSHINE COAST, AUSTRALIA | 14.0 |
| AUTHORISED P MALINAUSKAS PARLIAMENT HOUSE, NORTH TERRACE, ADELAIDE | 14.0 |
| SOUTH AUSTRALIA, AUSTRALIA | 14.0 |
| SINGAPORE | 14.0 |
| NSW, AUSTRALIA | 14.0 |
| VICTORIAQLD | 14.0 |
| PLANET MELMAC | 13.0 |
| LAUNCESTON, TASMANIA | 13.0 |

Figure 5: State-wise sentiment analysis towards Bill Shorten in all locations

and requires eliminating words, punctuation marks, and emojis that do not add value to sentiment analysis.

2. Performing context-based sentiment analysis is an important challenge that requires Natural Language Processing (NLP).

3. Tweets may not fully represent the mentality of all registered voters in a state. (e.g., not all Republicans who voted for Trump may have tweeted.)

4. Obtaining election tweets from more countries since we could only obtain tweet datasets from the U.S. and Australia.

# 8 Conclusion

- The Twitter sentiment towards Biden was more positive than Trump's. This result is accurate as Biden won the U.S. 2020 Election.
- The state-wise sentiment analysis towards Joe Biden and Donald Trump provided valuable insights that can be used for further studies.
- The Twitter sentiment towards Scott Morrison was more positive than Bill Shorten's. This is accurate as Morrison did win the Australian 2019 Election.

# 9 Acknowledgements


- Firstly, we would like to extend our gratitude to Prof. Yang Tang for his guidance and support throughout the course
- Secondly, we would like to thank IEEE-Dataport for making their dataset available free of cost, which proved invaluable for our analytics.
- We would also like to thank Tableau for making their license available to students.
- Last but not least, we would like to thank the NYU HPC Team for making the Peel Cluster available, where the entirety of this project was developed and tested.


# References


H. N. Chaudhry, Y. Javed, F. Kulsoom, Z. Mehmood, Z. I. Khan, U. Shoaib, and S. H. Janjua. Sentiment analysis of before and after elections: Twitter data of us election 2020. *Electronics*, 10(17):2082, 2021.

S. Dhawan, K. Singh, and P. Chauhan. Sentiment analysis of twitter data in online social network. In *2019 5th International Conference on Signal Processing, Computing and Control (ISPCC)*, pages 255–259, 2019. doi: 10.1109/ISPCC48220.2019.8988450.

Kusrini and M. Mashuri. Sentiment analysis in twitter using lexicon based and polarity multiplication. In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, pages 365–368, 2019. doi: 10.1109/ICAIIT.2019.8834477.

F. Å. Nielsen. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In M. Rowe, M. Stankovic, A.-S. Dadzie, and M. Hardey, editors, *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, volume 718 of *CEUR Workshop Proceedings*, pages 93–98, May 2011. URL http://ceur-ws.org/Vol-718/paper_16.pdf.

V. Prakruthi, D. Sindhu, and D. S. Anupama Kumar. Real time sentiment analysis of twitter posts. In *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pages 29–34, 2018. doi: 10.1109/CSITSS.2018.8768774.

İ. Sabuncu. Usa nov. 2020 election 20 mil. *Tweets (With Sentiment And Party Name Labels) Dataset*, 20, 2020.

M. Trupthi, S. Pabboju, and G. Narasimha. Sentiment analysis on twitter using streaming api. In *2017 IEEE 7th International Advance Computing Conference (IACC)*, pages 915–919, 2017. doi: 10.1109/IACC.2017.0186.

E. Xia, H. Yue, and H. Liu. Tweet sentiment analysis of the 2020 us presidential election. In *Companion Proceedings of the Web Conference 2021*, pages 367–371, 2021.

U. Yaqub, N. Sharma, R. Pabreja, S. A. Chun, V. Atluri, and J. Vaidya. Location-based sentiment analyses and visualization of twitter election data. *Digit. Gov.: Res. Pract.*, 1(2), apr 2020. ISSN 2691-199X. doi: 10.1145/3339909. URL https://doi.org/10.1145/3339909.