

U.S. 2020 Election and Tweets

Alain Duplan
Robin Lovell
Harsul Shukla

Background

- The 2020 election has been one of the most controversial elections of our lifetimes
 - There is currently no precedent for how social media technology should be use or regulated (if at all) in the context of elections
- Twitter Controversy
 - Fake tweets
 - Liberal Biased?
 - Banned/Suspended accounts

Tasks/Guiding Questions

- Are democratic and republican users tweets representative of actual voting turnout by state?
- Are there a significant amount of tweets that are made by fake/spam accounts? If so, do they support one party more than the other?
- Do Twitter users display a biased sentiment towards one party or the other overall?

Data

- <https://www.kaggle.com/manchunhui/us-election-2020-tweets>
 - Limited to the last 2 months of presidential election period
 - Tweets separated as tweets about Biden and tweets about Trump
 - **Key Columns:** Tweet Text, Likes, Retweets, User Join Date, User Followers Count, State Code, Country, User ID
- <https://www.archives.gov/electoral-college/2020>
 - Electoral college results
- <https://cookpolitical.com/2020-national-popular-vote-tracker>
 - Popular vote
 - To be used in comparison

Data filtering

- Since our dataset was so large, filtering was needed to be able to run on our computers
 - Only selected for users that were in the U.S.
 - “Spam” was removed (only accounted for <5% of the dataset)
- Further filtered our data by only using tweets that were in English so our sentiment analysis could understand the text
 - Used Google Translate’s Python language detection feature
 - This removed about 10% of the Trump tweets and 15% of the Biden tweets

Users

- US Government Lawmakers/Regulators (Senators, sitting FCC members, etc.)
 - Goal: Draft and vote on legislation that curbs the power of social media and social media algorithms to influence elections
- Twitter Moderator
 - Goal: Understand Twitter's user base
- Journalism Student and Activist
 - Goal: Understand Twitter's user base
- Passionate Voter
 - Goal: Understand biases in the Twitter userbase that support or oppose their political viewpoint

Sentiment Analysis

- Considered using **ABSA** (Aspect Based Sentiment Analysis) in order to target Biden and Trump as the target of tweets
 - Tried the `aspect_based_sentiment_analysis` Python package
 - BERT based sentiment analysis model allows for tying sentiment to particular aspects in the text
 - Showed poor results overall, but could have been a valid approach if we had spent a lot of time/data tuning the model for social media text
- Instead, chose to use a simpler sentiment analysis tool, **VADER** (Valence-Aware Dictionary and Sentiment Reasoner)
 - Specifically tuned to social media text
 - Understands emoticons, slang, and acronyms that are often present in tweets
 - Returns a polarity score $[-1,1]$ that we map to a binary pro-biden or pro-trump
- Simplifying Assumption: Only considered Biden and Trump as candidates (no 3rd party)
 - If a Biden tweet had negative sentiment, we assumed it was pro Trump and vice versa

Demo

https://rmlovell.github.io/2020ElectionAnalysis/tweets_map.html

Findings AND Conclusions

- There were more overall #trump tweets than #biden tweets
- #biden tweets had more likes and retweets
- Sentiment analysis overwhelmingly leaned toward Biden
- Clusters present in tweet-vote and sentiment-vote scatter plots indicate there may be weak correlation between tweets and popular vote and tweet sentiments and popular vote

Questions?

Project in its entirety is posted on

<https://github.com/HarshulShuk/2020ElectionAnalysis>

https://rmlovell.github.io/2020ElectionAnalysis/tweets_map.html