Sentiment Analysis

and some stuff or whatever

Kursten Reznik
Steven Bate
Bryce Robinette
Jacelyn Villalobos

10/29/2020

Contents

1 Introduction

A major point of interest today is the outcome of the upcoming presidential election. Attempting to predict the presidential election is nothing new, however with the advent of social media, we have a new source of information from which to draw. In this report, we perform a *sentiment analysis* on data mined from twitter. Although we may not be able to extend our finding to the broad population, it still yields insight into how the people might vote. Indeed, it is nevertheless, one more tool that can be used in our endeavors to predict the outcome of the 2020 presidential election.

It will be written in a fancy and understandable way the totality of what we do once we have come closer to the end. As we get closer and know what will hold for the sentiment, votes, map, and random forest, I'll fill in the introduction.

Indeed, the outcome of a U.S. presidential election not only holds great ramifications for the U.S., but many other nations as well. As a heavy-hitter on the world stage, many governments, citizens, industries, economic experts, and policy makers are affected by the outcome of the U.S. presidential election. It is the intent of this report to contribute to the ongoing practice of statistics, and the methods used in predicting presidential elections.

More Yaba Yaba, if desired.

2 Data and Methods

We obtained data from several sources for this report. We created the main dataset from data mining twitter in the form of word text. These tweets were scraped by using relevant

hashtags, such as: #trump and #biden.

We also obtained electoral vote data from 1976 up to the 2016 election.

We created database in MySQL and performed our statistical analysis using the programming language R. Furthermore, we imported our data into Tableau in order to create relevant visualizations

We have other data as well. Like, citizen votes dating back to 1976, and I think Kursten has a set of data she has been working with as well for her wordcloud and map thing. Let me know what data you want included. Even if we don't perform any serious analysis on it, it can be used to give context.

It should be noted that this conclusions cannot be extended to the entire population because we are retrieving data from a social media platform in which not everyone uses. Even if the an individual has twitter, they may not be into politics and writing political hashtags, which our data is based on.

a bunch of fun and sightfulwords to be able to show:
44% of 18–24 year olds use Twitter. 31% of 25–30 year olds use Twitter.
26% of 30–49 year olds use Twitter. 17% of 50–64 year olds use Twitter.

2.1 Sentiment

2.2 Wordcloud

2.3 Data Regression

2.3.1 LDA

2.4 Random Forest

3 Results

4 Conclusion

5 References and Sources

6 Appendix

Hashtags used:

- #trump
- #republican
- #donaldtrump
- #maga
- #teamtrump
- #trump2020
- #biden
- #democrat
- #joebiden
- #teambiden
- #biden2020