

# Sentiment Analysis

and some stuff or whatever

Kursten Reznik  
Steven Bate

Bryce Robinette  
Jacelyn Villalobos

10/29/2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data and Methods</b>	<b>3</b>
2.1	Sentiment Analysis . . . . .	3
2.2	Electoral Vote Data . . . . .	4
2.2.1	Graphical Representation . . . . .	4
2.2.2	LDA . . . . .	4
2.3	Random Forest . . . . .	4
<b>3</b>	<b>Results (might be part of above)</b>	<b>4</b>
<b>4</b>	<b>Conclusion</b>	<b>4</b>
<b>5</b>	<b>References and Sources</b>	<b>5</b>
<b>6</b>	<b>Appendix</b>	<b>6</b>

## 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. In doing so, we were able to form a collection of the most common words used, as well as the text that was to be analyzed for our sentiment classification.

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

We created databases 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.

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

((What population does your data represent?)) *Indeed, a bunch of fun and insightful words 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 Analysis

Sentiment analysis refers to using natural language processing, text analysis, and computational linguistics to assign values to the words that we use, so that we may ascertain whether the writing is of a particular nature of interest. That is to say, does the text have a positive, negative, or neutral connotation to it. This is referred to as *polarity* classification. In this report, we perform *beyond polarity* sentiment classification, which also takes into account emotional states. These include anger, surprise, joy, disgust, and so on.

After scrubbing 10,000 tweets (per frickin day?) and assigning value to the text, we end up with the resulting sentiment figures:



## 5 References and Sources

## 6 Appendix

Hashtags used:

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