Sentiment Analysis on Donald Trump and Barack Obama's Tweets

Maheen Asghar
School of Information, 2020

Dorsa Massihpour
Masters in Data Science, 2020

SI 671: Data Mining: Methods and Applications
Dr.Paramveer Dhillon

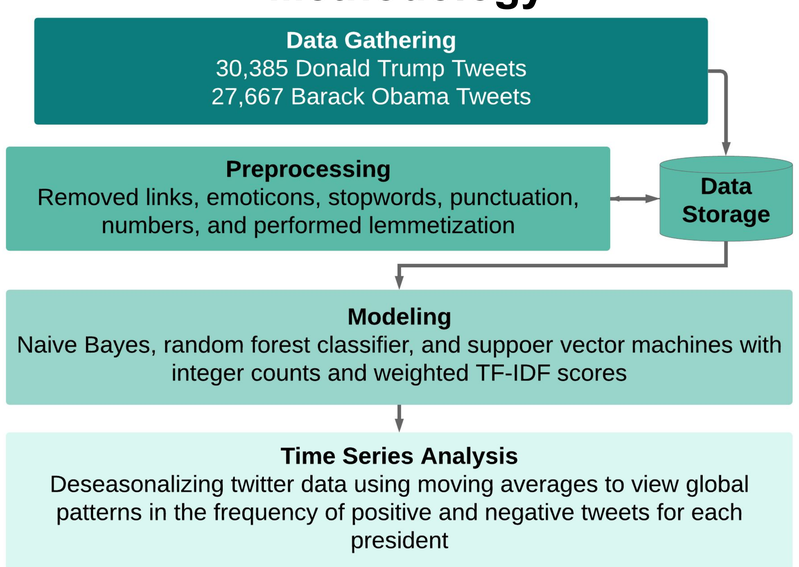
Aim

We examine sentiment analysis on Twitter data of the current and past presidents: Donald Trump and Barack Obama. Using machine learning, we create a sentiment analyzer that determines if tweets carry positive, neutral, or negative sentiment. In addition, we evaluate time series and identify trends for local variations to more accurately assess Trump's usual twitter sentiment vs Obama's.

Background

Sentiment analysis is the process of identifying and categorizing pieces of text according to its overall opinion. This type of analysis originated in the 20th century, but 99% of research papers in this field were written after 2004. Although there are many published papers on the twitter sentiment analysis of presidents, there are very few studies comparing Trump and Obama specifically, and simultaneously analyzing the frequency of their pos./neg./neutral tweets over time.

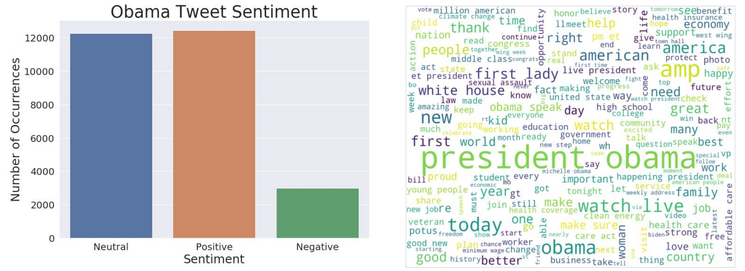
Methodology

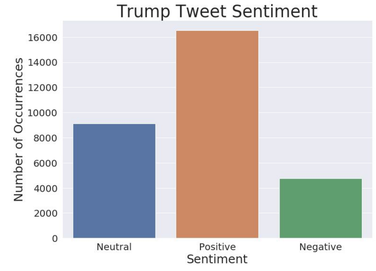


Findings

Exploratory Data Analysis

When looking at Trump versus Obama's tweets, we found that Obama had a more even distribution of positive and neutral sentiment of tweets, while having very little negative sentiment tweets. Trump had a disproportionately higher number of positive sentiment tweets over the other categories.





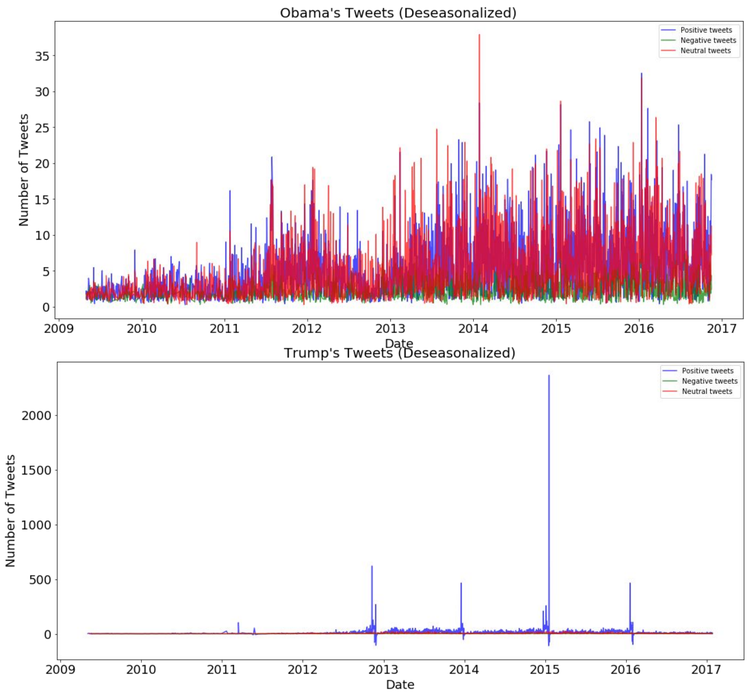
Model Selection and Time Series

Table 1: Obama Results

Method	Accuracy	F1 Score
Naïve Bayes (Baseline)	0.73	0.77
Naïve Bayes	0.76	0.69
Logistic Regression	**0.91**	**0.87**
Random Forest	0.70	0.65
SVM	0.44	0.20

Table 2: Trump Results

Method	Accuracy	F1 Score
Naïve Bayes (Baseline)	0.59	0.70
Naïve Bayes	0.72	0.74
Logistic Regression	**0.93**	**0.93**
Random Forest	0.79	0.80
SVM	0.54	0.71



Next Steps

1. Analyze a larger dataset of Donald Trump and Barack Obama tweets
2. Determine a range of sentiment: positive, slightly positive, neutral, slightly negative, and negative