Predicting Presidential Election Trends by Web Data

Deji Suolang*, Kaan Cem Ketenci
University of Michigan - Ann Arbor

Overview

This study is an experiment using web data to predict public support for 12 political candidates who have qualified for democratic presidential primary debate that was held on October 15, 2019.

Background

- Election data are the information can be used to produce predictions about people's likelihood of supporting candidates and changing their support with specific campaign interventions.
- Nationally representative surveys and opinion poll are costly and time-consuming.
- Big data may provide supplement and preliminary estimates, which allow us to monitor voters' support in real time.

Method

Data Collecting Period:
Oct 10, 2019 - Oct 20, 2019

Data Sources:
1.Twitter API: Stream Tweets using names and related terms of 12 candidates, compute volumetric information and sentiment score.
2.Google Trends API: Web scraping of daily and hourly number of Google searches of each of the 12 candidates' names.
3.Web Scraping from oddchecker.com: Oddschecker is a website that combines betting odds for various bets from all major betting companies. Convert betting odds data into implied probabilities.

A linear model is built using three parameters of estimates of change by twitter sentiments, scaled google searches, Oddschecker probabilities to predict the changes in the national-level benchmark Morning Consult Surveys .

Analysis

- Instead of focusing on the tweet counts, sentiment analysis is employed to interpret tweets that we collected for each candidate

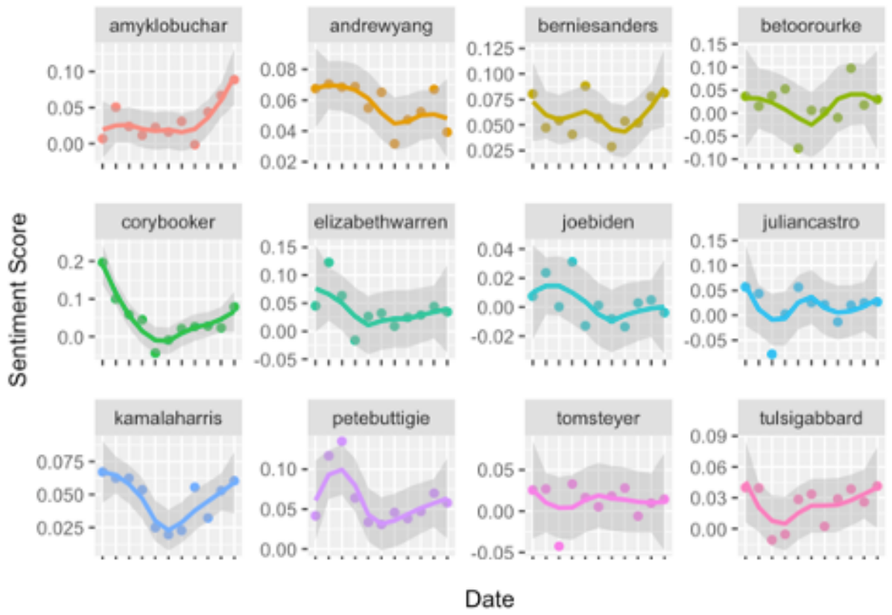

Figure 1: The Trends of Sentiment Scores

- Relative shares of each candidate's searches as a proportion of the total searches of all candidates on a given day are computed. Figure 2 reflects changes in scaled Google searches before and after the debate day for each of 12 candidates.

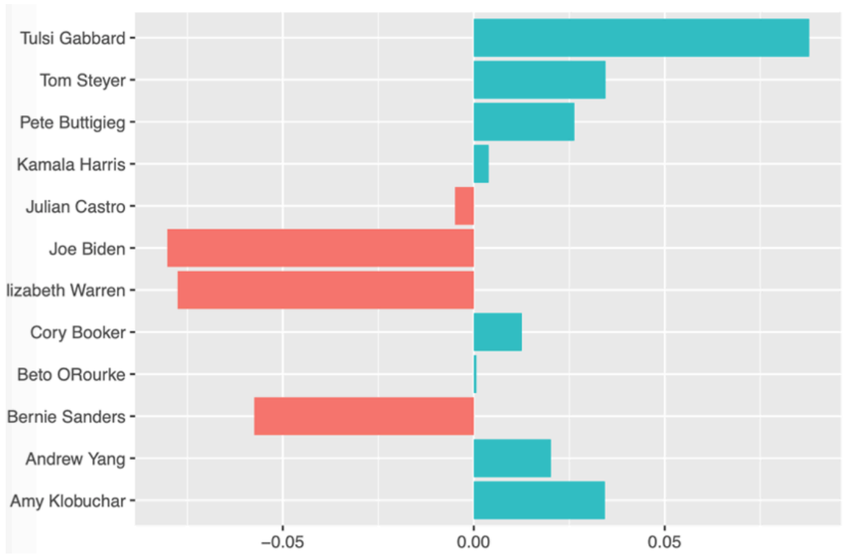

Figure 2: Changes in Scaled Google Search Index

Elizabeth Warren has the highest probability at around 50%, Joe Biden has the second rank with probabilities around 20%, and he is followed by Bernie Sanders and Pete Buttigieg both at around 10% probabilities of winning the nomination.

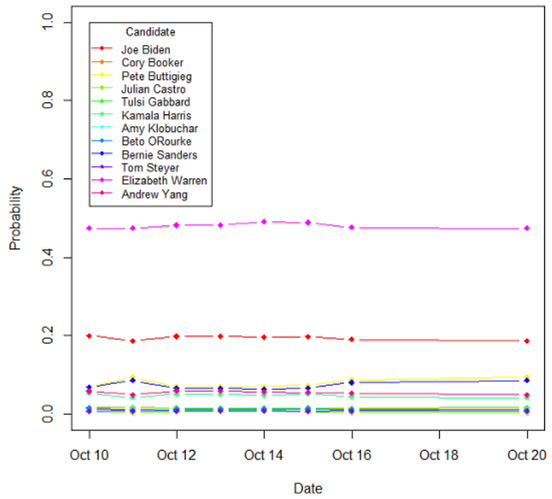

Figure 3: The Trends of Nomination Probabilities

Result

Coefficients	Estimate	Std.Error	t-value	p-value
Intercept	-0.0016	0.0024	-0.675	0.519
Twitter Sentiment	-0.1895	0.1740	-1.090	0.308
Google Searches	0.123	0.052	2.374	0.045*
OddscheckerProbabilities	0.157	0.157	0.157	0.157

Significant codes *p<0.05, **p<0.01, ***p<0.001

Discussion

- The model can be substantially improved if further data is collected from other debate periods and election cycles.
- Possible misjudged sentiment due to the auto-coding.
- Under coverage bias, large group of people do not use Twitter the target population does not coincide with the sampling frame
- Future research with state-level prediction could be very informative in terms of presidential elections.