

Predicting Elections: Can Social Media Provide Insight to Public Opinion?

Katherine Berwind 12338747 & Syed Irtaza Ali Shah 12453661

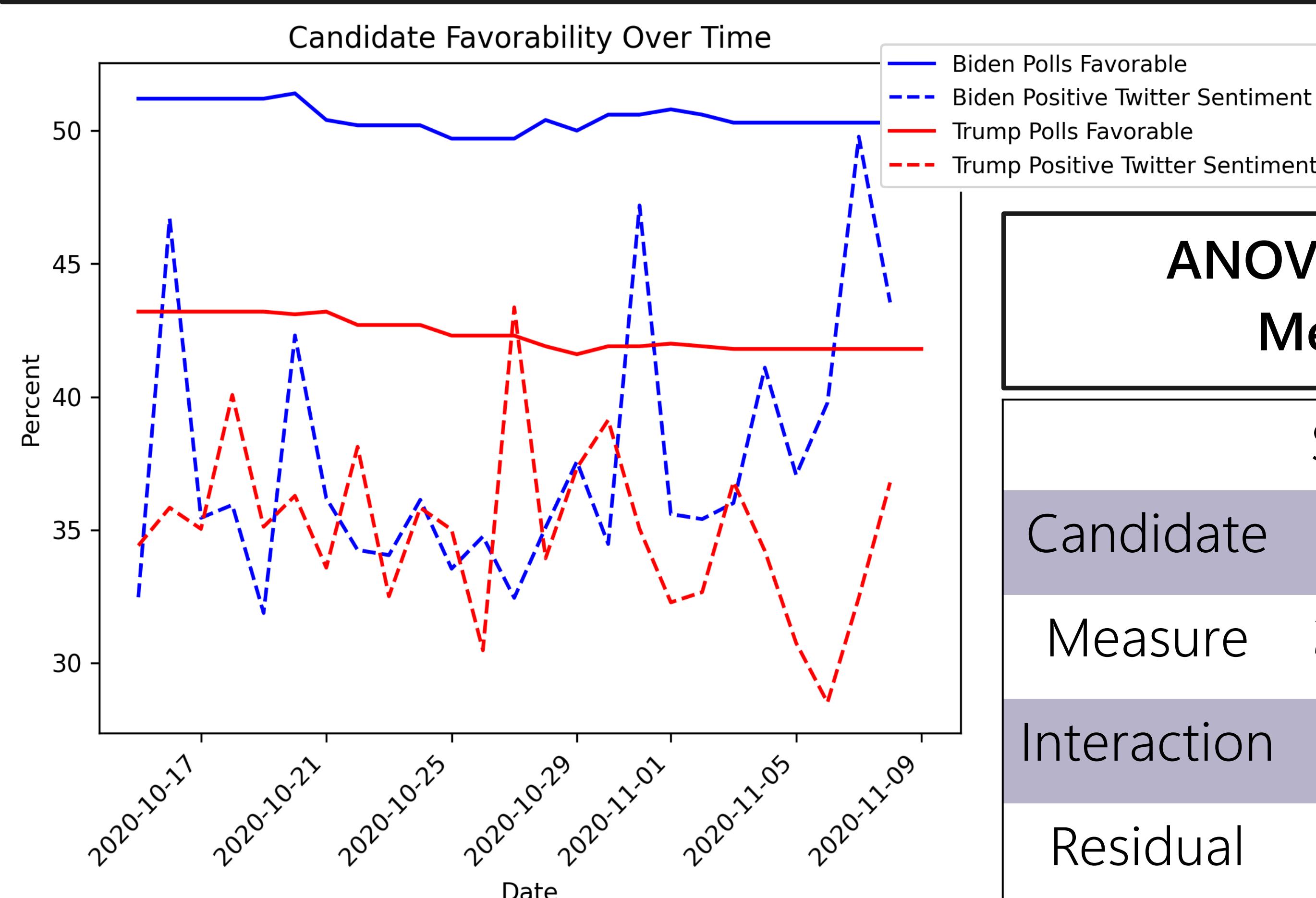
Introduction

- Political polling has limited reach, usually reaching an older demographic, and they are notoriously unreliable compared to actual voting habits.
- Social media is full of political opinions that can be mined to predict election outcomes.
- Does social media reflect polled public opinion?
- Are the opinion differences between candidates consistent?
- Does this hold geographically as compared to actual voting?

Methods

- Dataset: Historic tweets about the 2020 US Presidential election aggregated on Kaggle
- VADER Sentiment Analysis: (Valence Aware Dictionary and sEntiment Reasoner) Designed for informal and short text. Ideal for tweets
- 2-way ANOVA comparing opinion of candidates between measurement methods
- 2-way ANOVA comparing Twitter sentiment of candidates between California (CA) & Tennessee (TN)

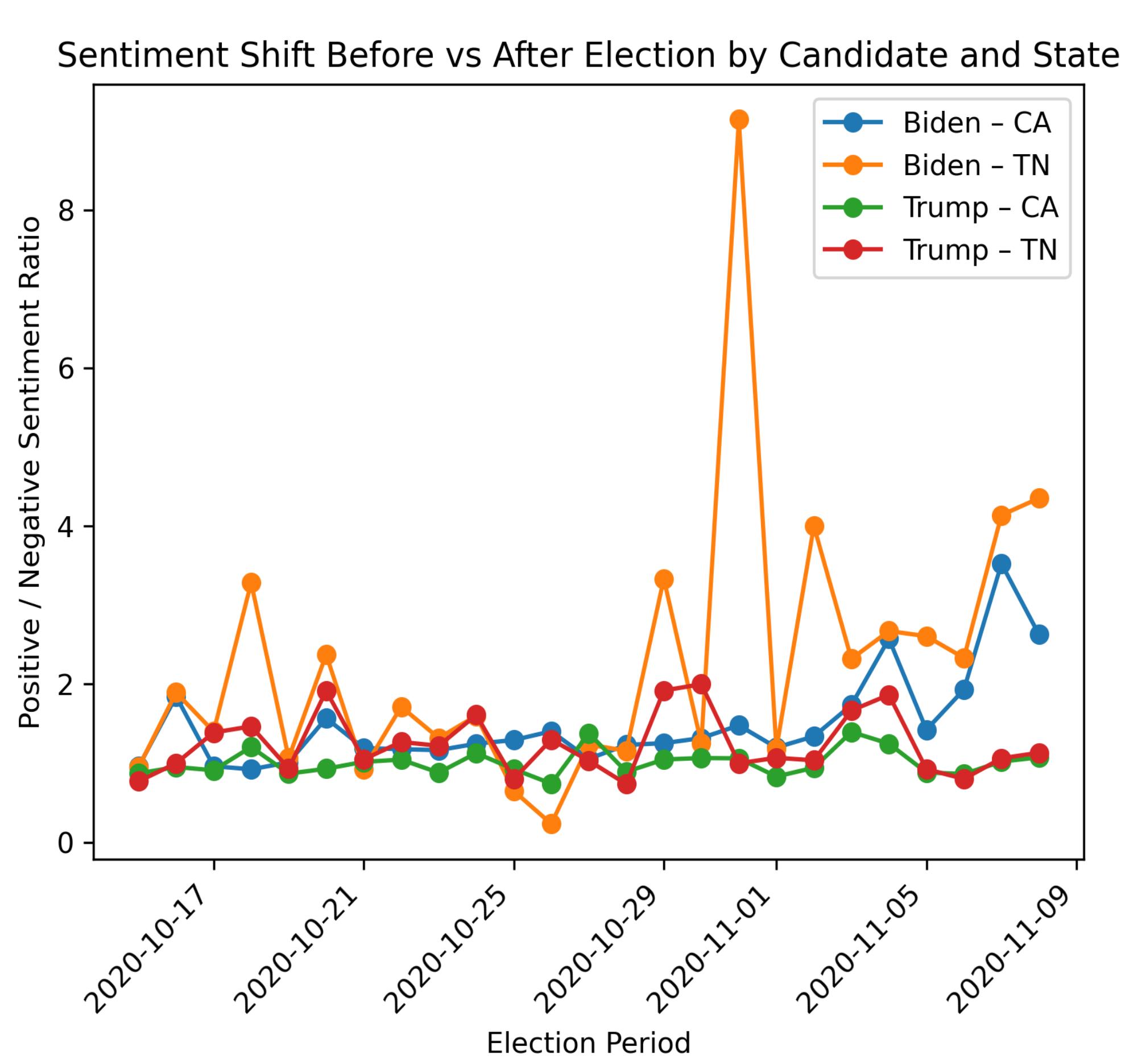
Results



ANOVA Score by Candidate & Measurement Method

	Sum Sq	df	F	p
Candidate	708.36	1	81.67	0.00
Measure	2573.44	1	296.71	0.00
Interaction	195.58	1	22.55	0.00
Residual	832.63	96	-	-

- Significant differences found between Twitter sentiment analysis and conventional polling when grouped by candidate and measure
- Differences in candidates similar across both methods of measurement
- Post-hoc analysis shows the candidate has an effect on the opinions in Twitter sentiment. Benefit from separate analysis?



ANOVA Twitter Sentiment by Candidate & State

	Sum Sq	df	F	p
Candidate	176.54	1	817.29	0.00
State	2.95	1	13.67	0.00
Interaction	1.34	1	6.20	0.01
Residual	13986.74	64750.00	-	-

- Significant differences found in sentiment between candidate and state of the poster
- Raw Twitter sentiment data contradicts actual election results (Biden won CA by the widest margin; Trump won TN by the widest margin)
- Post-hoc shows larger magnitude of difference in sentiment between candidates in CA as opposed to TN.

Discussion

- Twitter sentiment and polling data do not mirror each other
- Twitter sentiment by state does not hold to actual election outcome
- The variability of Twitter sentiment makes it unreliable for predicting election outcomes. Polling is the more reliable choice despite limits
- Actual use of twitter sentiment data requires a more in-depth methodology likely involving polarization analysis.

References

- Crespi, Irving. Pre-election polling: Sources of accuracy and error. Russell Sage Foundation, 1988.
- Federal Election Commission. (2022). FEDERAL ELECTIONS 2020 Election Results for the U.S. President, the U.S. Senate and the U.S. House of Representatives. Federal Election Commission. <https://www.fec.gov/resources/cms-content/documents/federalelections2020.pdf>
- RealClearPolitics. (n.d.). Trump: Favorable/unfavorable. RealClearPolling. <https://www.realclearpolitics.com/polls/favorability/donald-trump>
- RealClearPolitics. (n.d.). Biden: Favorable/unfavorable. RealClearPolling. <https://www.realclearpolitics.com/polls/favorability/donald-trump>
- Hui, M. (2020, November 9). US election 2020 tweets. Kaggle. <https://www.kaggle.com/datasets/manchunhui/us-election-2020-tweets>

