

Sentiment Analysis

A Modern Approach to Political Polling

Suleyman Qayum

INTRODUCTION

- ❖ Sentiment Analysis is a viable approach to public opinion polling
- ❖ This will be demonstrated by:
 - *predicting Presidential approval rates (national + by state)*
 - *comparing predictions against actual polling data*

OUTLINE



BUSINESS
CONTEXT



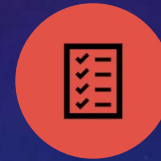
DATA



MODEL



RESULTS



EVALUATION



LIMITATIONS &
NEXT STEPS

BUSINESS CONTEXT

- ❖ Current polling techniques are inefficient
- ❖ Make them efficient by utilizing **sentiment data**
 - *less time consuming*
 - *more data available*
 - *makes polling more robust*

DATA

❖ Training Data

- *reputable dataset comprised of ~1,600,000 tweets*

❖ Test Data

- *collection of ~100,000 tweets*
- *recently extracted from Twitter*
- *mentioned Joe Biden*

❖ Sentiment labels used were

- *POSITIVE*
- *NEUTRAL*
- *NEGATIVE*

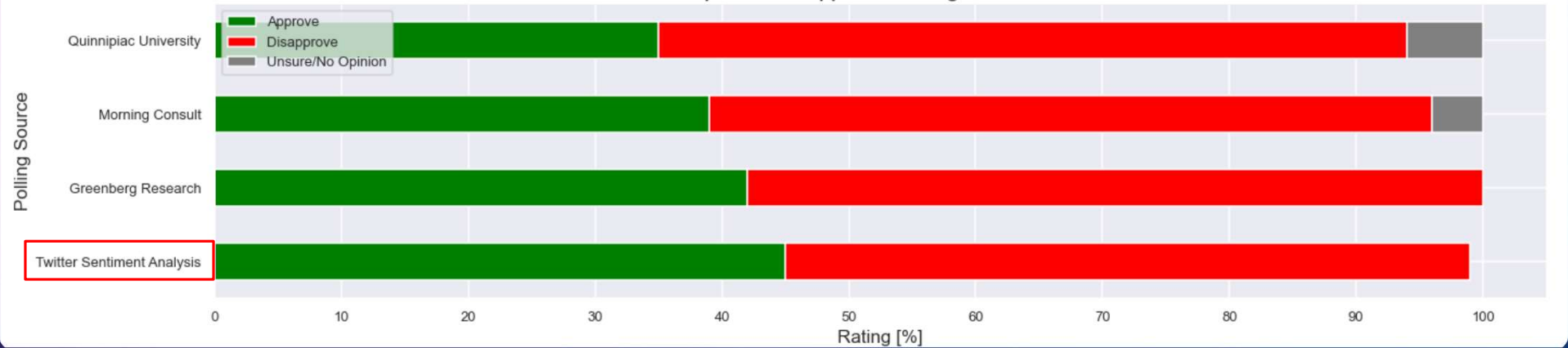
MODEL

- ❖ Model was developed from Training Data
 - *Recurrent Neural Network (RNN)*
- ❖ Model produced sentiment labels for tweets regarding President Biden
- ❖ NEUTRAL tweets were discarded
- ❖ Sentiment labels were converted into approval ratings:
 - ***Approval Rating*** → *percentage of tweets that were POSITIVE*
 - ***Disapproval Rating*** → *percentage of tweets that were NEGATIVE*
 - ***Net Approval Rating*** → *Approval Rating – Disapproval Rating*

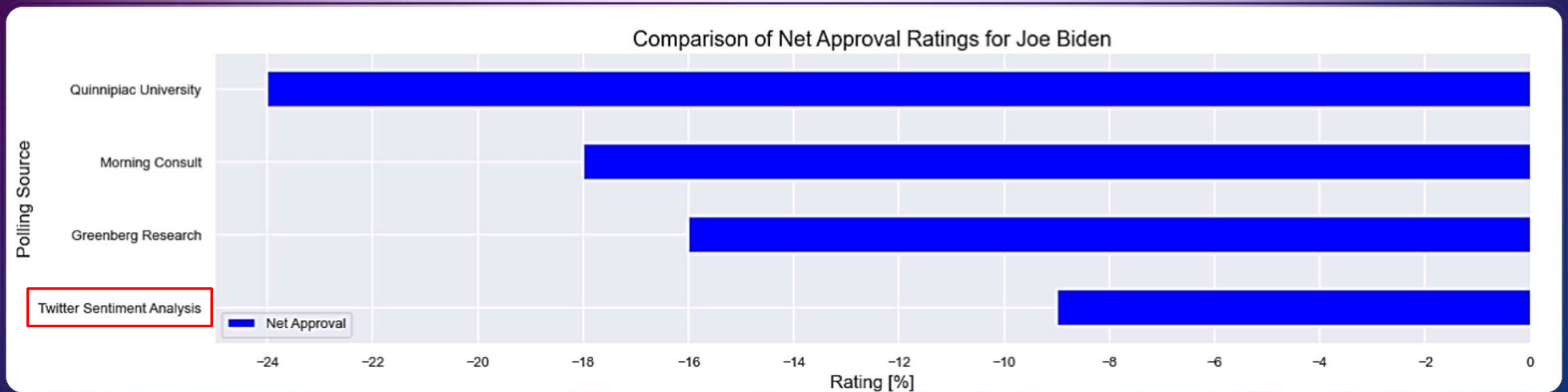
RESULTS

Overall Approval/Disapproval

Comparison of Approval Ratings for Joe Biden

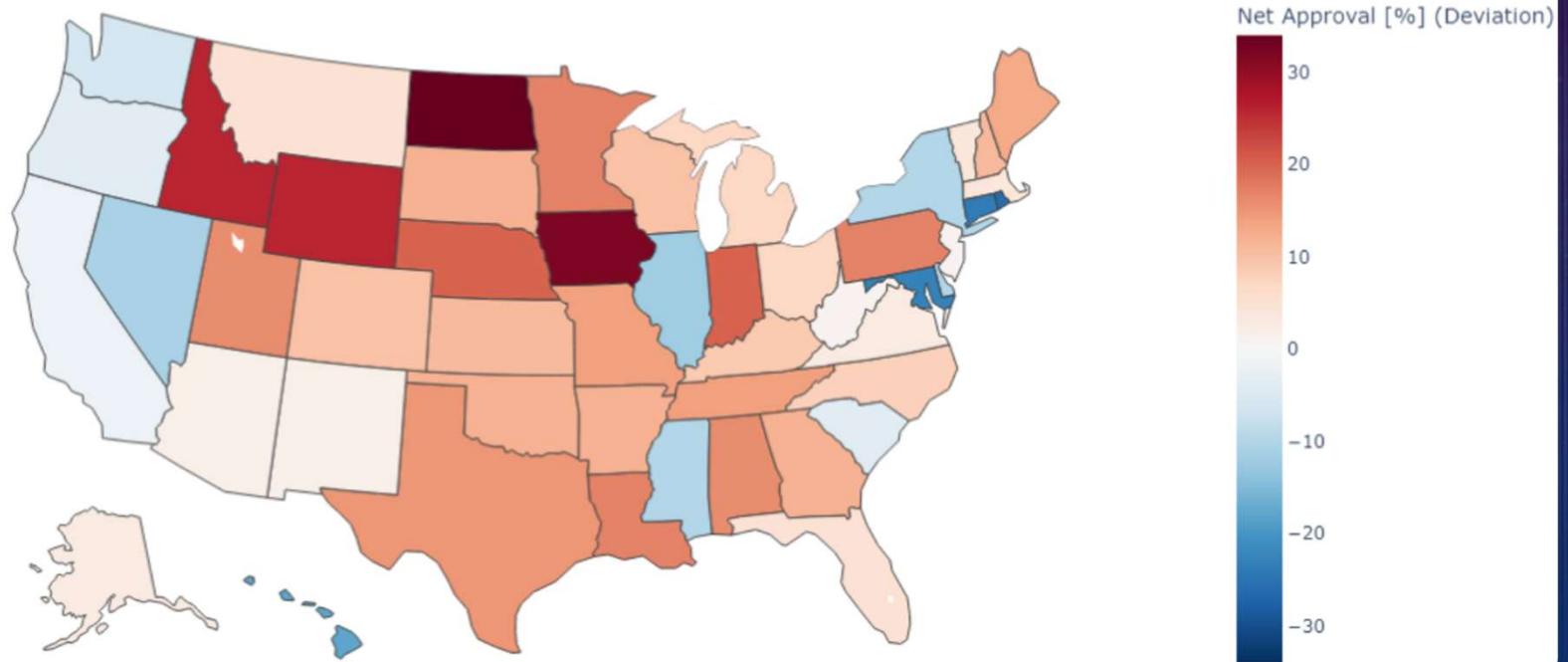


Overall Net Approval



Statewide Net Approval (Deviation from Polling Data)

State Approval Map for Joe Biden (Sentiment Analysis Deviation from Polling Data)



EVALUATION

❖ Model is biased

- *tends to overestimate Biden's approval rates*

❖ Comparison of overall approval ratings shows promise

- *always overestimated Biden's approval rating*
- *predicted a net approval rating that differed by 10% or less for 7/10 sets of polling data*

❖ Model predicted statewide net approval ratings surprisingly well

- *average of 435 labeled tweets per state*
- *predictions deviated from polling data by*
 - *$\leq 10\%$ in 27 states*
 - *$\leq 15\%$ in 37 states.*

LIMITATIONS & NEXT STEPS

- ❖ Rate limits imposed by Twitter API
- ❖ Issues pertaining to location data
- ❖ Fine-grained sentiment analysis
- ❖ Identifying topic of tweets
- ❖ Utilizing emoticon data
- ❖ Utilizing other social media platforms to collect sentiment data

THANK YOU

Email: sqayum33@gmail.com

Github: [@sqayum](https://github.com/sqayum)

