

Political Tweet Search Engine

William Harer

Creating the data set (initial ideas)

1. Use wikipedia to generate list of politicians
2. Use Twitter API (tweepy) to confirm users as politicians (election labels)
3. Accumulate tweets using Tweepy api
 - 18,000 tweets max a day
 - Each user can only have most recent 3,200 tweets scraped

Creating the data set (Reality)

1. X Use wikipedia to generate list of politicians (names are weird)
2. Use Twitter API (tweepy) to confirm X users as politicians (election labels)

Election labels not used in all cases

3. Accumulate tweets using Tweepy api :)
 - Works really well! Pretty intuitive, lots of Data!
 - 15 minutes per 2.5k tweets to prevent lockout
 - Each user can only have most recent 3,200 tweets scraped

Scraping Continued

- 1) Manually Make list of names, attach to State, attach to Photo
- 2) Scrape most recent tweets of all names add relevant info
- 3) Create Json of Username ->(description,followers)
- 4) Go through each senators scraped tweets to get statistics

Info about the DataSet

- Made of every current US Senator
- Contain info about State, UserName, Likes, Retweets, TweetID
- Secondary DataSet contains newest tweet, description, followers, most liked tweet, most retweeted tweet, avg likes, avg retweets, total likes, total retweets, total interaction, avg interaction

Search Methods

- Basic search (BM25)
- Filter by State (sorry Kansas :()
- Sort by likes and retweets
- Do both at once

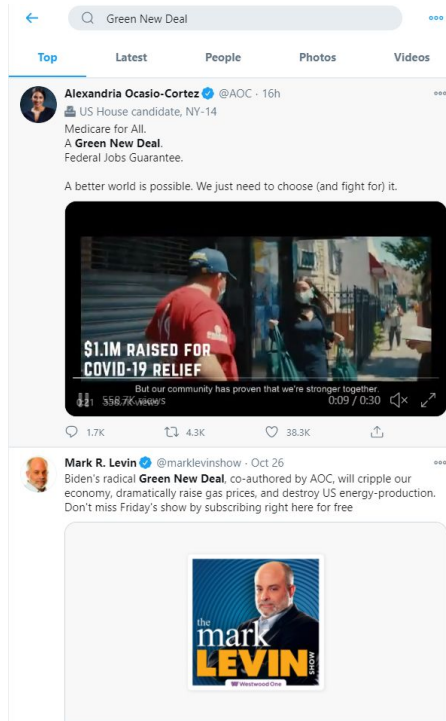
Additional features and metrics

- Allow user to use modes to:
 - Weight location as more important/input location
 - Which limits results per politician
 - Weight results based on tweets/comments
- Display individual statistics per politician:
 - Ratio of followers to average comments/likes
 - HeatMap of twitter activity
 - Maybe look at twitter usage vs election result statistics (Did active tweeting benefit the campaign?)

Before

- A ton of results, bad format
- Results come from secondary sources instead of primary

From visual clutter/autoplay



After:

Condensed

Primary sources only

Lack of visual disorientation

