CS122 Final Project Final Report

CNetID:
andresc-gdmorrison-jblehr-kyy

Group Members:
Andres Crucetta Nieto
Gabe Morrison
Jacob Lehr
Kelly Yang

A brief overview of the final project

On January 6th, 2021, violent insurrectionists stormed the US Capitol. Congress members were forced to halt their ratification of the 2020 election results and flee for safety. This riot was the culmination of President Donald Trump's efforts to delegitimize the election, and the result of President Trump and his supporters' political disinformation "stopthesteal" campaign.

From the November election up to the insurrection, speech from constituents and their representatives became increasingly polarized and inflammatory. We explored two topics during this period:

1. The extent to which Congressional Representatives' tweets between November 7th, 2020 and January 6th, 2021 were reflective of their actual support of President Trump's delegitimization campaign.
2. If representatives' tweets reflected those of their constituents.

We used the Twitter API to analyze each Representative's tweets in the established timeframe by assessing common keywords and performing a sentiment analysis. We conducted similar analyses on constituent tweets that provide geographic data, allowing us to match accounts to Congressional Districts. Finally, we used ProPublica's API to assess how representatives voted on the ratification of Arizona's and Pennsylvania's election results and President Trump's impeachment. Ultimately, our website allows users to explore and compare constituents' and Representatives' tweets, sentiments and common keywords.

The overall structure of the software

Running the software

To run the software, write the following lines of code in the linux command line:

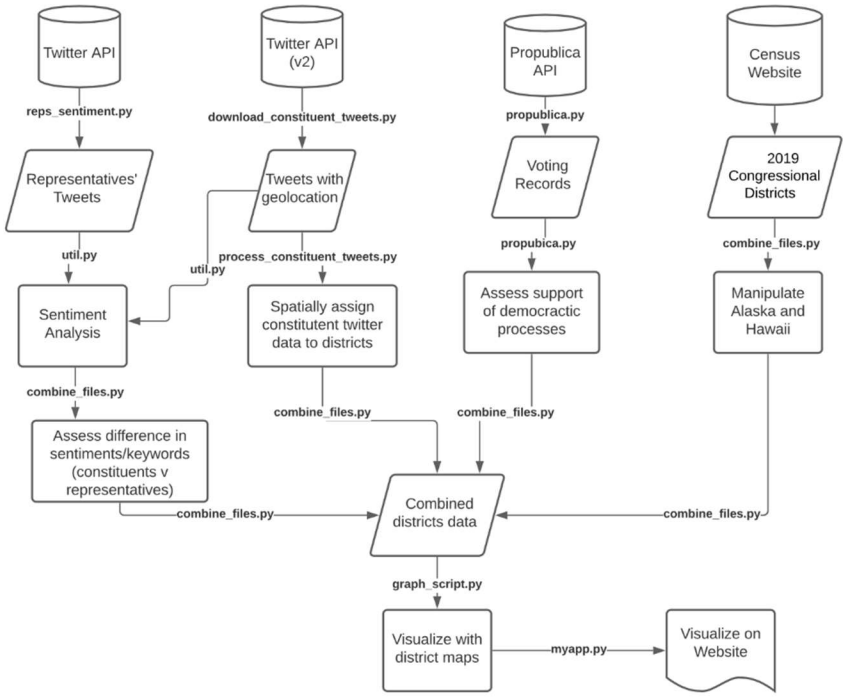
```
./install.sh
source projenv/bin/activate
./run_program.sh
```

Interacting with the software

Within the **run_program.sh** file, the user will receive several command-line prompts for user input that will ask whether the user would like to run intermediate analysis steps, or to continue to the final visualization. The user can also interact with the final visualization by zooming into a specific district and hovering to reveal more district-specific information.

Organization of the software

The diagram below shows the organization of the software. Please note that the arrows illustrate the script that moves the data form one data-frame or process to the next either type of data or to the next process in the data manipulation. All of the processes below are run in **main.py**, which is run within **run_program.sh**, with a couple exceptions. Several files use API's and are discussed below. **myapp.py** is also called from within **run_program.sh**.



Note on APIs:

This project uses three APIs: (1) Twitter API, (2) Twitter API v2, and (3) ProPublica API. Within the diagram above, several files were run separately to query the APIs using private API key information (**download_constituent_tweets.py** and **propublica.py**). One exception is the Representatives tweets (from Twitter API), which are both queried from the API and analyzed in one script (**reps_sentiment.py**). As noted above, the user has an option to run this process within **./run_program.sh**.

What the project tried to accomplish and what it actually accomplished

Our initial project goal was to compare voting records centered around the November Election and January 6th Insurrection with a sentiment analysis of US Representatives and their constituents. For voting records, we planned to utilize the ProPublica API. For the tweets of representatives and their constituents, we planned to utilize the Twitter API to pull tweets over the specified timeframe. By analyzing the data, we hoped to determine the extent to which Representatives actually represented their constituents, both in speech and voting record.

As we navigated, we decided to limit the constituent tweets to tweets that provided geographical data so we could properly map to districts. We were able to pull the US Representative Twitter data over that timeframe. A few Representatives' accounts were suspended or deleted, but we were able to collect tweets from 98% of the representatives. With that, we conducted a sentiment analysis on polarity and subjectivity using the textblob package.

Our final product consists of the mapped-out districts containing gradients that illustrate the use of #StopTheSteal by location, how congressmen voted in the election ratification, the sentiment behind representative tweets along with their most common keywords, and the subjectivity score of representatives compared to their constituents.