How do Americans perceive their government's response to the Covid pandemic and how will it affect the 2020 election?

Covid & 2020 in Tweets

JAMAAL SMITH SEPTEMBER 2020 FLATIRON SCHOOL

Executive Summary

- Twitter has been a useful tool for companies and organizations to monitor how they are perceived by their customers and the public at large in real-time
- With the American presidential election this year and ongoing Covid-19 pandemic, analysis of tweets provides policy-makers and political campaigns opportunity to make adjustments to reduce Covid transmissions or improve their candidate's electoral prospects
- The primary goal of this analysis will be to ascertain topics expressed in Covid tweets to determine whether American public is assigning any blame for the pandemic.

Flow of Presentation

- Data Collection Methodology
- Topics Present in Tweet Data
- Deceptive Text Analysis
- Classifying Tweets by Response Type
- Response Type and Poll Spreads
- Take Aways and Next Steps

Methodology for Gathering Tweets

- Tweets were gathered from January 1, 2020 until August 21, 2020
  - 4 subsets were created to promote a diversity of viewpoints. Further by having these separate ‘buckets’ of tweets, we will be able to examine and compare text features to assist with determining whether or not a tweet is deceptive or not
  - Final dataset comprised of 27,437 tweets

TRUMP TWEETS

POTUS TWEETS THAT CAN BE CONSIDERED POLE1

CUOMO TWEETS

NY GOV TWEETS CONSIDERED POLE2

BASELINE TWEETS

TWEETS BY NEUTRAL PUBLICATIONS

PUBLIC TWEETS

TWEETS FROM AMERICAN PUBLIC

Topics Present in Tweets

TOPIC 1: PUBLIC HEALTH RESPONSE TO PANDEMIC

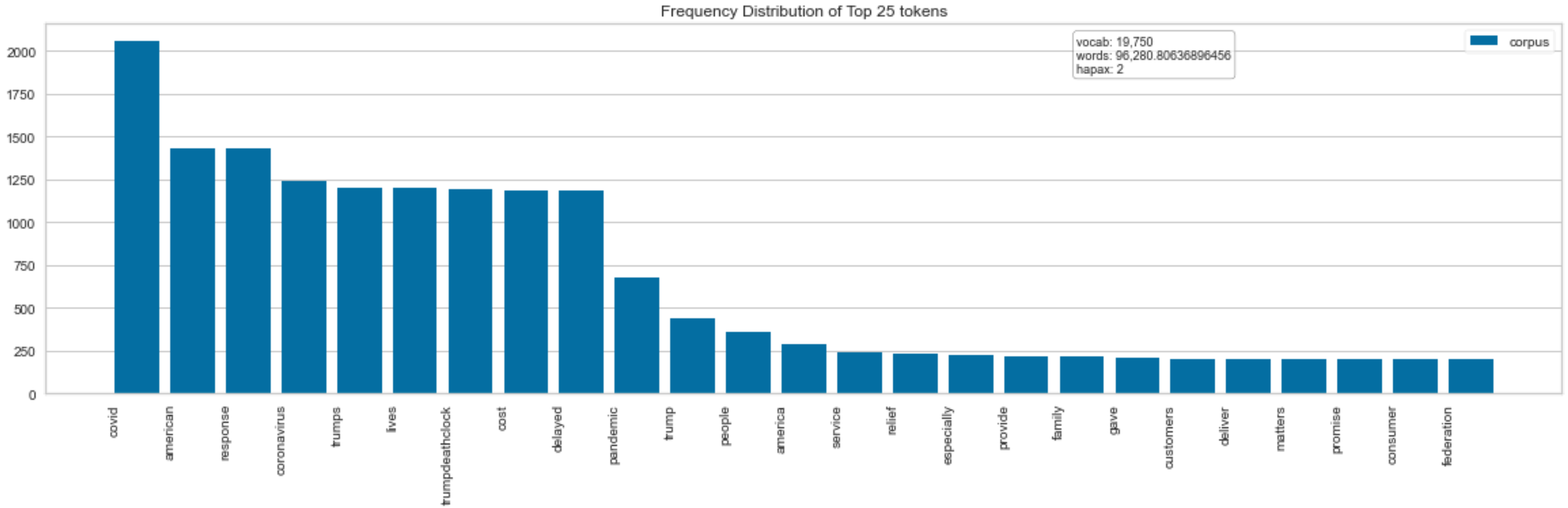
TOPIC 2: DEATHS/HEALTH IMPACT OF COVID

TOPIC 3: FINANCIAL AND FOOD RELIEF NEEDED DUE TO COVID

TOPIC 4: TRUMP RESPONSE TO PANDEMIC AND COSTS OF DELAY

TOPIC 5: NATIONAL RESPONSE TO PANDEMIC

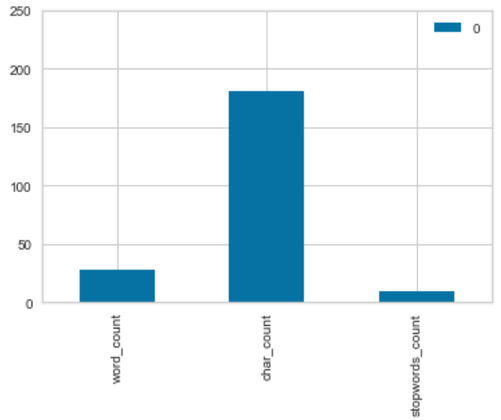
What are people talking about?



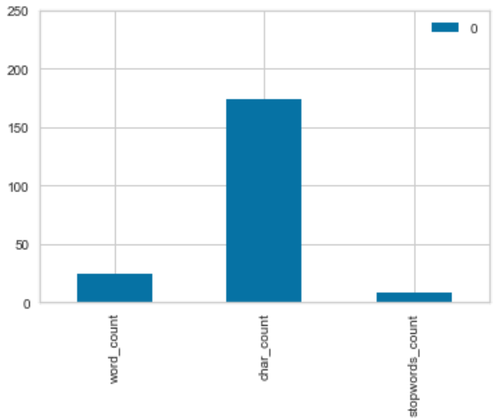
Detecting Deception

- Deception can be identified through analyzing text features of a tweet. According to linguistic research, deceptive texts are often longer, contain more filler words, and have more verbs than nouns

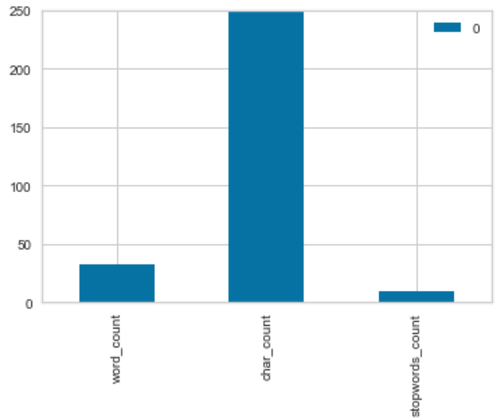
Trump Tweets



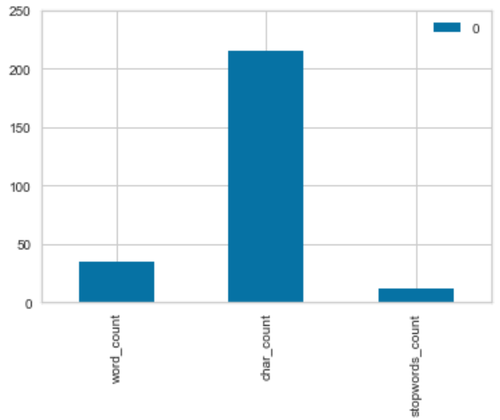
Baseline Tweets



Public Tweets

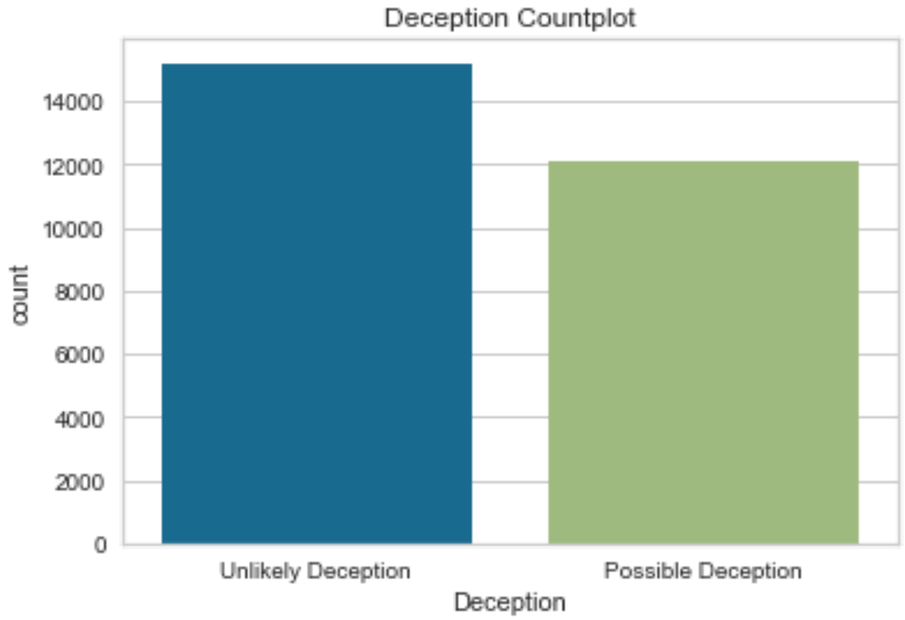


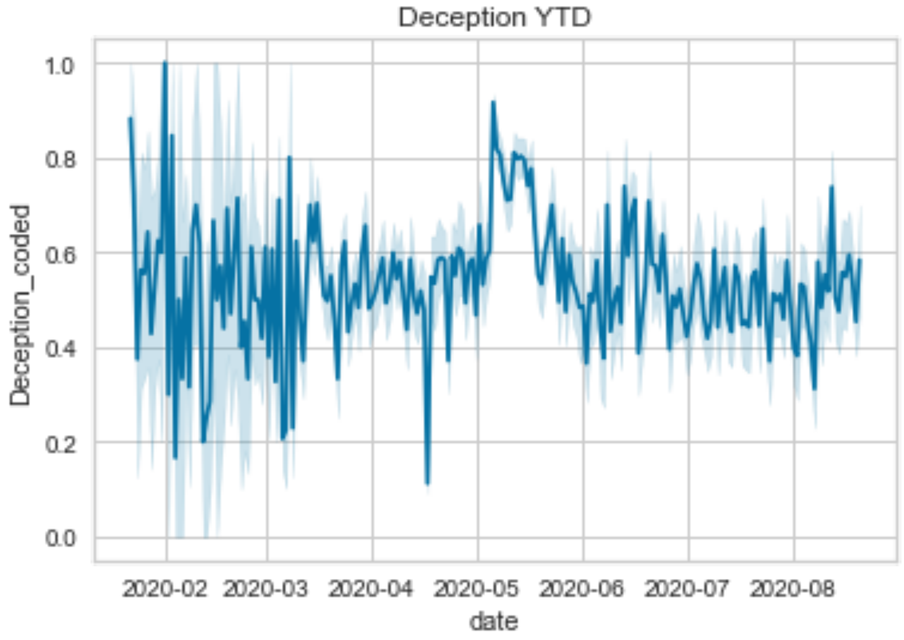
Cuomo Tweets



- For the purposes of this project, tweets will be highlighted as possibly deceptive if they have a stop word count greater than 10 and a word count greater than 40

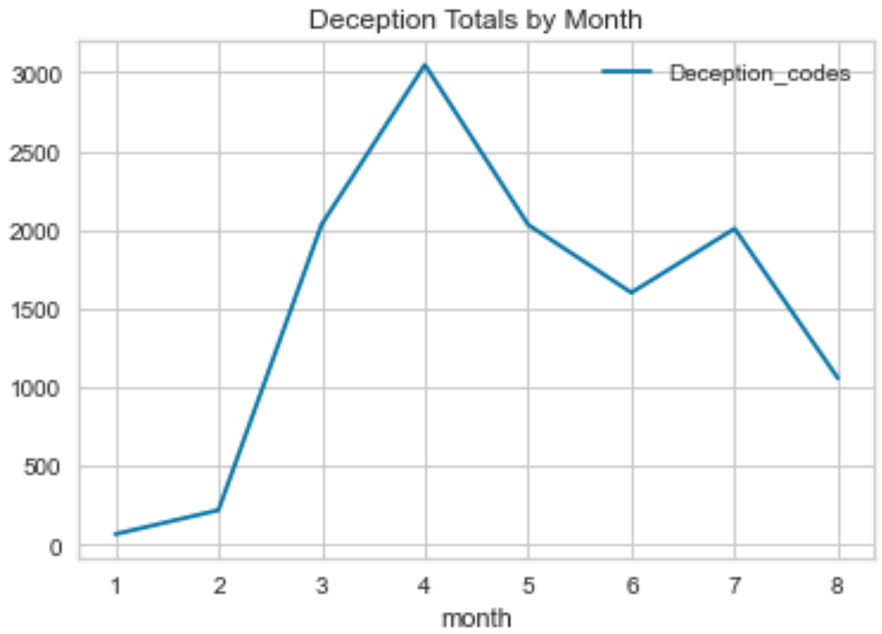
Dataset Deception Detail



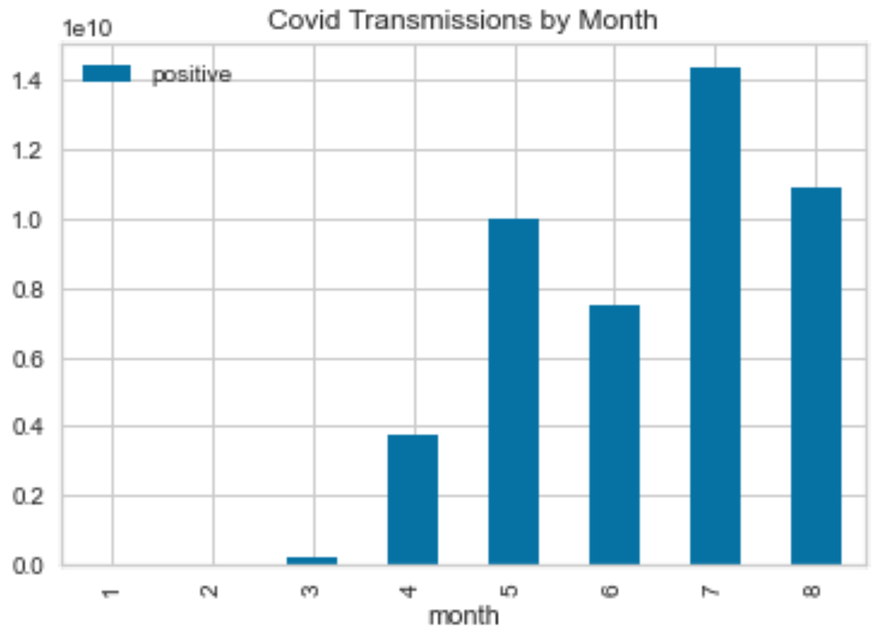


Deception as a Covid Forecaster

Monthly Deception Tracker

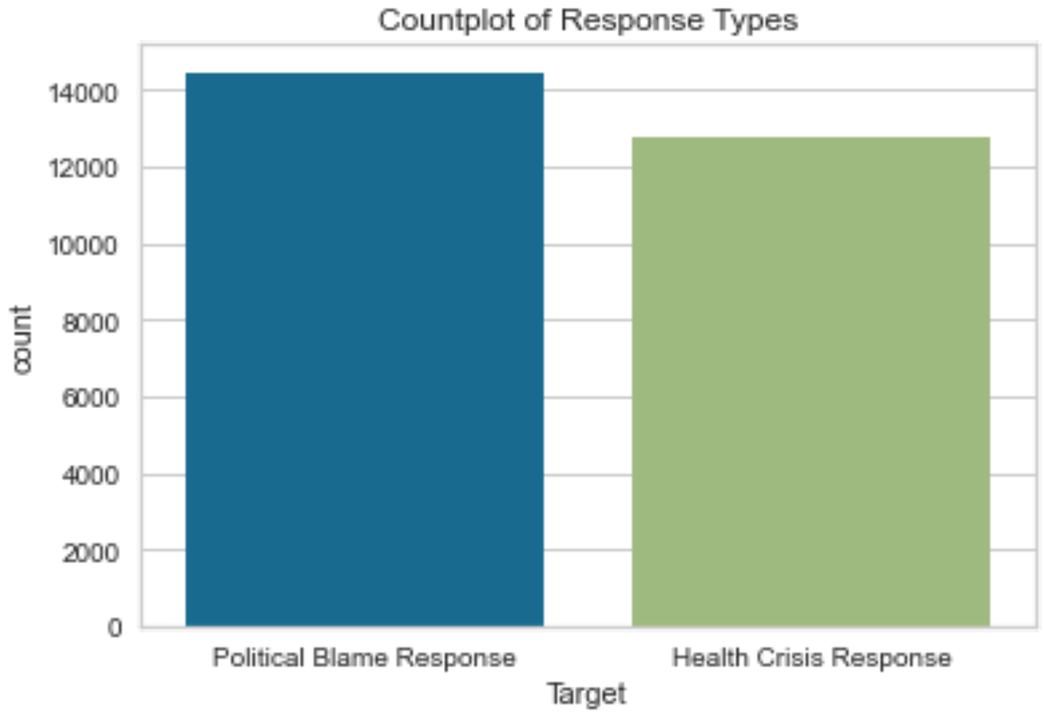


Monthly Covid Tracker



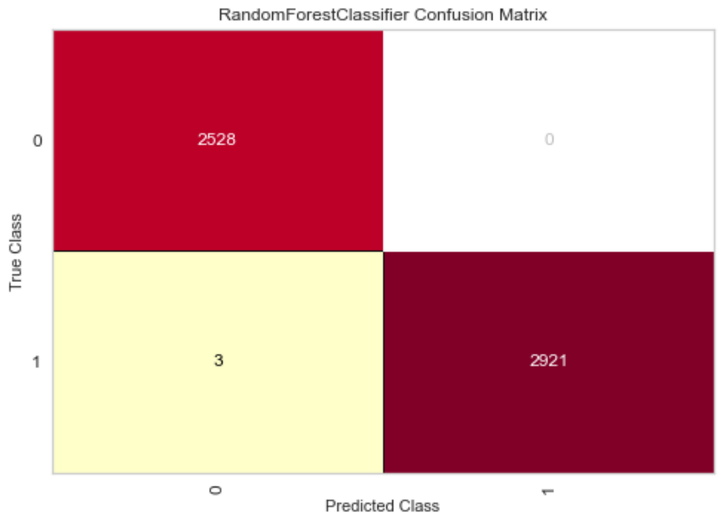
The deception tracker appears to forecast movements one month in advance in the Covid Tracker. The Pearson coefficient between both variables is .02 which is pretty low. However, as deceptive tweets related to Covid increase in a month, one can expect growth in Covid cases in the next month. Likewise, a decrease in deceptive texts leads to a decrease in cases the following month.

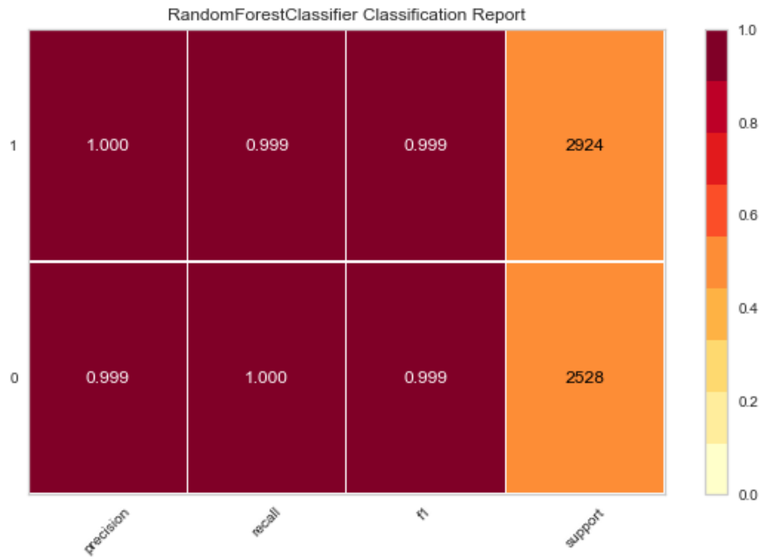
Classifying Response to Pandemic

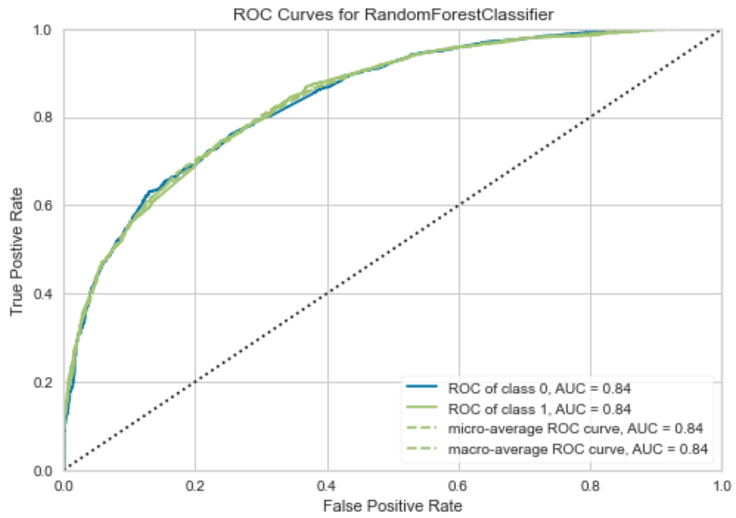


- Based on the 5 topics that were produced by our LDA model, two types of responses were prominent in the dataset
  - Political Blame Response: Tweets where a political figure is blamed for the current plight faced by the country
  - Health Crisis Response: Tweets where focus is on problems created by pandemic and no noticeable blame is evident

Classifying Future Tweets

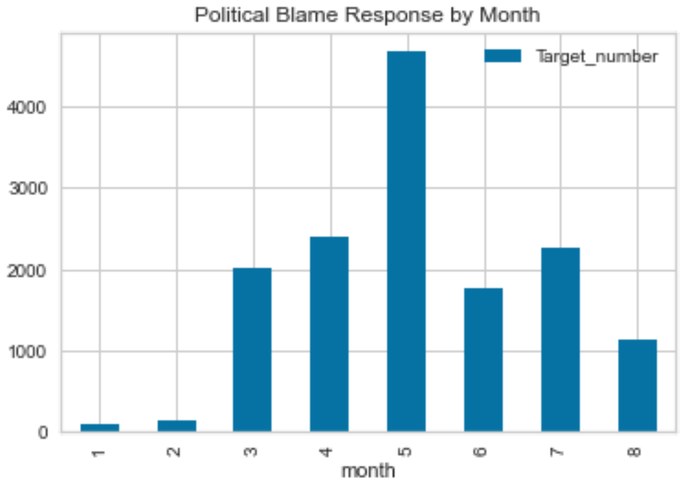


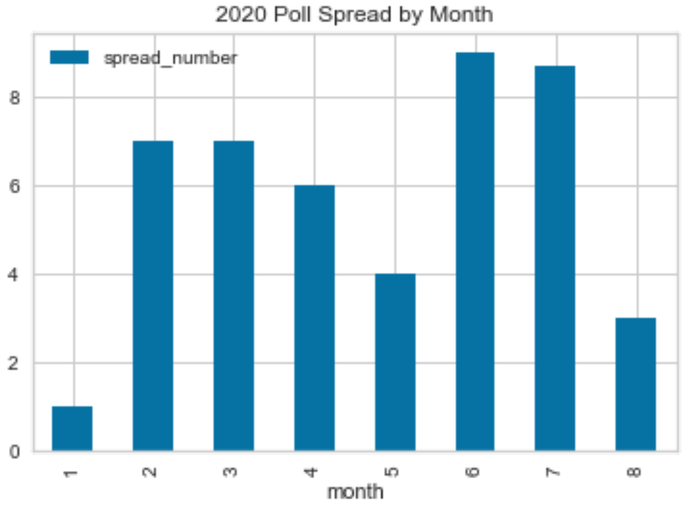


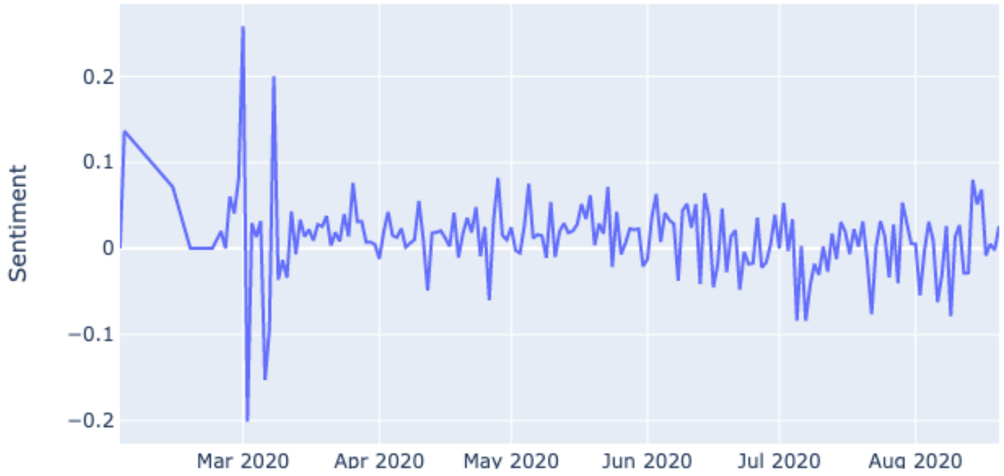


A Random Forest Classifier provides the perfect combination of accuracy with the training/test data. The ROC Curve also indicates that new tweets will be classified well and won't simply replicate findings from training data - no overfitting.

Response Type And Poll Spreads







The immediate impact of political blame on poll spreads has waned over time. As the sentiment graph suggests, many Americans were initially deeply disturbed by Covid in March, but have relatively neutral with a trend toward slightly negative in the months post-June

Recommendations/Take Aways

POLITICIANS: HIGHLIGHT PLANS TO REDUCE HEALTH COSTS AND PROVIDE RELIEF TO THOSE IN NEED

IDEAL TWEET: AT OR BELOW 175 CHARACTERS AND MORE VERB PHRASES THAN NOUN PHRASES

SCHOOL OFFICIALS: MONITOR LEVEL OF DECEPTION IN TWEETS TO FORECAST FUTURE TRANSMISSIONS

COVID FATIGUE: SENTIMENT ANALYSIS HOVERING AROUND NEUTRAL SUGGESTS PUBLIC ISN'T MOVED BY CASE/DEATH MILESTONES LIKE THEY WERE IN FEB/MARCH

Future Work

- Current analysis focused on text classification of tweets related to Covid-19 response in America
    - Additional modeling work to potentially quantify how deception impacts change in positive cases
    - Additional modeling work to quantify how assigned blame impacts change in spread in presidential poll numbers
    - Predictive text generator that can construct tweets based on chosen author to see difference in styles