CS5830 Final Project Proposal

Quinton Oliver, Chetan Birthare, and Zion Steiner

OVERVIEW

COVID-19 has abruptly changed our world in many ways. Our political, economic, and social institutions are being tested by the effects of virus prevention measures. As these systems try to adapt to survive, information about how they are responding to the pandemic are being identified and communicated. The goal of our project will be to use natural language processing tools to analyze text related to COVID-19. We are interested in how COVID-19 is affecting two things: the communication of the general public, and corporate response.

We will use social media data from Twitter and Reddit to analyze how people are talking about the pandemic.To determine how companies are responding to COVID-19, we will analyze recent SEC earnings filings. We will segment this analysis by sector to compare and contrast industry-specific responses. If we are able to find recently updated economic datasets, we may also integrate this in with our filings analysis.

QUESTIONS

- SEC Earnings Filings
    - How is corona impacting company performance?
    - How do earnings with positive, negative conclusions impact share price?
    - What common themes exist between filings? Are there any groupings?
    - What are the companies saying that were bailed out in the CARES act?
- Social Media
    - Common themes
    - Hashtag-specific analysis (what hashtags are trending?)
    - Top level comment analysis of top reddit posts about covid19 (medical, political, social, etc)
- Other
    - Ideas on how data might be clustered?
    - Ideas on what economic data we can relate to SEC earnings analysis?
    - spaCy built-in BERT? Question answering?

DATA

We will be collecting data from Twitter, Reddit and SEC earnings filings.

TOOLS

- spaCy (NLP package)
- Tweepy (Twitter API wrapper)

- PRAW (Reddit API wrapper)
- Python-edgar (SEC filings retriever)

SCHEDULE

Week 1 - Data collection/Cleaning
Week 2 - Create models(?) and gain insights
Week 3 - Finalize models and insights

TEAM ROLES

To make sure all team members are able to practice all steps of the project lifecycle, we will all be working on related parts of the project simultaneously. We will keep in contact to make sure that we are all on the same page.

POTENTIAL PROBLEMS

Experience with NLP - Between the three of us, we have very little experience using natural language processors. This will require us to research tools and work together to learn how to use them.

In [79]:
```
df22= pd.DataFrame(data=texts2)
df22.head(20)
```

Out[79]:

	0
0	RT @dbongino: As PREDICTED in my prior tweet, ...
1	RT @ArthurSchwartz: Don’t ever forget that @ny...
2	RT @GA_peach3102: Left Wing Media endlessly co...
3	White House Will Not Reopen Obamacare Enrollme...
4	RT @WhiteHouse: "Our future is in our own hand...
5	@realDonaldTrump \nNeeds to stick you in a roo...
6	@seanhannity @TrumpLaney He’s a virus himself!!
7	RT @NYGovCuomo: We are all vulnerable to this ...
8	RT @bennyjohnson: Democrats need to be held cr...
9	RT @dfriedman33: Sen. Kelly Loeffler sold more...
10	Hey JUNIOR! How is it that South Korea got the...
11	I wonder white people will find a way to gentr...
12	Franklin Graham’s Charity Treats Virus Patient...
13	RT @nytimes: 2 weeks ago, amid the coronavirus...