

Final Project Proposal

Jeremy Tsai

Research Question:

Can social media pertaining to COVID-19 like Tweets be used to accurately gauge the perception/emotion of the general public? More specifically, if more Tweets are negative during a period of time, does this cause people to react in a certain way (indicators such as greater food purchases, oil purchases, overall consumption, etc.)?

Outline of Methods:

The first step will be to gather the relevant data. For my research question, I will need a large set of Tweets that pertain to the Coronavirus in the last 3-4 months. I will also need data on food purchases, oil purchases, overall consumption, and any other variables that represent human reaction to the virus within the same time frame. Then, I will perform sentiment analysis, specifically polarity detection, on the Tweets grouped by week. This will allow me to see how positive/negative the Tweets are over time. I will then have to combine the Tweet positivity/negativity data and compare it over time with the indicators to see if there are any trends. I would think that if Tweets *do* represent the perception/emotion of the general public, then consumption and similar indicators will spike when the Tweets related to the Coronavirus are more negative. There are two possible reasons for this: perhaps the Tweets themselves are causing more fear thus causing people to try and stock up and prepare for the worst, or the Tweets *are* a representation of how fearful people are and because of this, people are consuming more in the time frame. I will also have to figure out a way to test significance for these results.

Sources:

How Twitter Represents the General Population:

<https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

Sentiment Analysis:

<https://monkeylearn.com/blog/sentiment-analysis-of-twitter/#What-is-Sentiment-Analysis>

<https://towardsdatascience.com/basic-binary-sentiment-analysis-using-nltk-c94ba17ae386>

Data Sets for Tweets:

<https://ieee-dataport.org/open-access/corona-virus-covid-19-tweets-dataset>

<https://www.kaggle.com/smld80/coronavirus-covid19-tweets>