Resources and Prediction for COVID-19 Pandemic in United States

Bingan Chen

February 17, 2022

Summary of Research Questions

1. How to show the ranking about which region are commenting about the medical resources during the pandemic?
   - I will display it using the map.

2. How the increased cases data relates to the commenting?
   - The machine learning skills will be critical.

3. What is the relationship between the commenting percentage and the severity of the pandemic NATIONALLY?
   - This may be generalize by the previous question.

4. How to use the comments to indicate when would the pandemic end?
   - Using the ML skills to predict the commenting 'density' by a certain amount of cases.

Motivation

While browsing on Twitters during the pandanmic, I noticed that lots of people dead and were sick but weren't treated carefully because of the shortage in medical resources. I want to organize people's comments regarding the medical resources during the pandemic to evaluate how they care about this situation and organize get a possible relationship between the indicated cases and comments. A geographic distribution by state about the comments-density in order to help distribute resources. Also, I want to predict when would the pandemic be not that severe based on the reflection on Twitter.

Dataset

- The data grasped by Twitter API (`https://developer.twitter.com/en/docs/twitter-api`) in real-time.
- `https://covid.cdc.gov/covid-data-tracker/#trends_dailycases`

Challenge goals

- Messy Data: I decided to use the API to get data on Twitter.
- New Library: Request may be useful to get the data online.
- Multiple Datasets: One grasped by API, the other is from CDC.

Method

1. The geopanda can be used to display the commenting density regarding the pandemic by state, so that the government can distribute the resources.
2. The machine learning will be applied to learn the relationship to predict the other.

3. The `.sum()` to get the national data from states data.

4. I would determine a level of cases to indicate the 'end' of this pandemic. In this way, there would be a method to reflect the severity by the comments.

Work Plan

1. Learn to use the Twitter API to filter the data by content. (20 hrs)

2. Write snippets to test the functions and filter. (15 hrs)

3. Learn to change the returned `.json` to usable `.csv`file. (10 hrs)

4. Determine which part of data is for test that wouldn't be trained. (5 hrs)

5. Integrate the functions and organize data. (20 hrs)

6. Finalize and test the codes. (15 hrs)

Dev Environment

Locally on my laptop.