

What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

We are group GUGUGU, and we have four team members:

Jingze He (jingzeh2)

Qiao Jiang (qiao2)

Xinyi Fang (xinyif5)

Zejie Zhou (zejie2)

Zejie Zhou will be the captain.

What is your free topic?

Vaccine Sentimental Analysis with Tweets

Please give a detailed description.

We would perform sentiment analysis with a collected COVID-19 vaccines related tweets dataset to analyze the sentiment of the overall public discussion. A comparative analysis with the vaccine progress and the result of sentiment analysis of public discussion on Twitter will also be conducted to provide a better understanding of how the public discussion is correlated with vaccine progress.

What is the task?

Our task is to perform sentiment analysis on tweets about COVID-19 vaccines by using the dataset of tweets sentiments from all over the world. The dataset will include tweets about people's perspectives towards different kinds of vaccines, the number of different vaccinations per day, week, month and country over the number of people being vaccinated. By performing sentiment analysis, we can get an overall prediction of the relationship between vaccination rates and people's discussions on Twitter.

Why is it important or interesting?

It can help us predict the vaccination rate from twitter posts then take measures according to the predicted vaccination rates. The release of official statistics of vaccination rate need multiple steps and may cause delay from days to weeks. Predicting the vaccination rate from twitter posts can help the government prepare for future policies with shorter delay. To some extent, this project can be a small step towards the goal of herd immunity!

What is your planned approach?

We will firstly clean the dataset and then apply models to it to find a best model that enables analyzing sentiment of the target tweets. Next, we will add time information about world vaccination progress to get an overall sense of the relationship between public sentiment towards COVID-19 vaccine and the vaccination rate.

What tools, systems, or datasets are involved?

We will use two datasets found from Kaggle. The first dataset contains tweets discussing vaccines from December 2020 to March 2021. The second dataset contains vaccination progress for countries all over the world.

What is the expected outcome?

1. Collect data then process and extract texts from these data(twitter posts, vaccine related articles)

2. Get sentiment score/label by our sentiment prediction model in a specific time range
3. Find some relations between twitter posts sentiment and vaccination progress

How are you going to evaluate your work?

We will evaluate our work based on:

1. Properly process the text data(twitter posts, vaccination rate related articles)
2. The accuracy of our sentiment prediction model
3. Get meaningful conclusions from analyzing the relation between twitter posts sentiment and vaccination rate

Which programming language do you plan to use?

python

Please justify that the workload of your topic is at least $20 \cdot N$ hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

We have four teammates: $N = 4$, $20 \cdot N = 80$

Data collection: 10h

Coding and Testing: 40h

Analysis: 15h

Writing: 15h