Evaluating public options of COVID-19 vaccine using natural language processing

**Problem description**

2020 is an unprecedented time and the whole world were almost shut down during the pandamic. Recently, two types of vaccines have been approved by FDA and they are like the light in the dark. Hopefully, vaccine can help end the pandemic and we can go back to normal life in 2021. However, there are different options about vaccines. People may welcome, oppose, or be skeptical about Vaccinations. It is important to find what people think.

**Data sets**

My dataset is obtained from twitter API using stream filter method. I will utilize filters two times: one is to filter vaccine related tweets and the keywords utilized is ‘vaccine’, ‘vaccinated’ and ‘vaccination’; the other is to filter covid related tweets and the the keywords is 'covid' , 'covid-19', 'coronavirus', 'pandamic', 'covid19', 'social distance', 'cov2', 'quarantine', 'moderna', 'oxford', 'pfizer', 'reopen', 'social distance'. By filtering the tweets twice, I am able to obtain the datasets related to covid vaccine.

**Approach**

The topic in this project is related to twitter sediment. To obtain the tweet related to covid vaccine, I utilized tweepy API function Cursor. For the sediment analysis, I utilized textbolb module to evaluate polarity. People’s options about covid vaccine can be evaluated by evaluating the polarities.

**Applications and clients**

1. It would help government/companies to decide how many vaccines are needed
2. According to WHO, Community immunity requires at least 80% of the population to be vaccinated. As a result, knowing percentage of people who welcome, oppose, be skeptical is important.
3. It would help the government to decide whether they need to put more effort to persuade people to get vaccination.

**Deliverables**

Code, a final report