The Effect of Social Media Posts on Country-wise Vaccination Rates

Background Study Performed

We performed a background study using the following research questions as guidance: How do social media posts correlate with vaccination rates? How different is the sentiment about vaccination by region? How accurately can the vaccination rate be predicted from social media data of a particular region? Because our project depends so heavily on a correlation between social media and vaccine rates, we investigated whether there is evidence of this phenomenon or not. Further we needed to determine what factors are evidence of this so we know what to look for when constructing our network analysis of these questions. So far, our background research indicates that there is indeed a correlation between vaccine rates and social media posts regarding the vaccine. Based on this preliminary research, we should move forward with conducting our network analysis after cleaning the data to isolate sentiments and location from social media posts about the COVID-19 vaccine.

Data Cleaning Completed

Our initial plan was to explore the social media datasets that can be generated. We tried to generate user data from Facebook, Reddit and Twitter. The main goal of our project was to correlate vaccination status data with vaccination sentiment data. Unfortunately, reddit does not provide the location of the users. Therefore, we were left with Twitter. Only a small period of twitter data can be fetched but it includes the location data. Finally we found a **Vaccination sentiment dataset in Kaggle**.
https://www.kaggle.com/hassanhshah/covid-vaccine-sentiment-and-time-series-analysis/data
However, the location data was not uniform and we had to aggregate the locations to a specific country. We did it using the **Geopy library** in python.
https://colab.research.google.com/drive/1wPH7O1pC_wlehqIYXC68EF0YX19_w7Nr?usp=sharing

Sentiment Analysis Complete

We used the LSTM model which is a variant of Recurrent Neural Network to generate the sentiment score labeled to be positive, negative and neutral as labeled with an accuracy score of about 70%. We plan to explore more options on how we can do sentiment to generate a better result if we have the time.

Graphs Plotted

From our cleaned data set we were able to develop a few preliminary graphs that help us understand our data. These include a bar graph that displays sentiment analysis counts of positive, negative, and neutral tweets with the posters location. This has given us an idea of where the majority of posts occur and as well as an idea of which countries had more support for vaccinations or vice versa. We still plan on making more graphs to better understand our data and answer our research questions, as well as comparing this data to statistics on vaccination data from these countries to see if we can spot any trends. We are excited to showcase these visuals together in our final report to support and illustrate our eventual conclusions.

Interview Questions made

Link to questionnaire made with some sample interview questions. More questions will be added later on based on results we get.
https://docs.google.com/forms/d/1R7T18cUyxi1B3aGvA8jkxrhSDiNSbKnSkj4s3dJLWTs/edit?usp=sharing

Pending Tasks

- Collect answers to the questionnaire
- Perform Social Network Analysis using Gephi and NodeXL
- Finish editing the final manuscript