
VA project - Presentation

Ismaili, Rava, Felici, Lombardo

Sentiment Analysis Covid Vaccine

The intent is to provide a dashboard that could visualize the sentiment extracted from natural language in tweets regarding the Covid pandemic with a particular focus on the topic of vaccinations. In order to do that the sentiment will be evaluated in relation with other variables such as time, geographical location, specifics related to Twitter (verified accounts, number of follower et cet.).

Task accomplished

First of all we have focused on the preparation of the dataset. The starting data are All COVID-19 Vaccines Tweets (Kaggle).

- (a) Preprocessing of the data
- (b) Compute geographical coordinates based on name of locations
- (c) Infer the sentiment of the tweets

Technology used: python geopy library for the latitude and longitude evaluation, textblob API and python nltk library for sentiment analysis.

References: <https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets>, <https://pypi.org/project/geopy>, <https://textblob.readthedocs.io/en/dev>, <http://www.nltk.org>

Next steps

Structure the dashboard. Insight and ideas to visualize:

- (a) Variation in sentiment with respect to the number of deaths (or cases or non-cases). Chart: x-axis will present the number of deaths per month and on the y-axis we will have the number of user divided by positive and negative sentiment
- (b) In analogous way we will show the variation in sentiment with respect to number of vaccinated
- (c) Sentiment with regards to the number of followers (we would like to test our hypothesis: those with more followers may not expose themselves extremely negatively). We will divide of the observations into classes based on the number of followers and visualize the sentiment with a barplot with the number classes on the abscissa and the ratio of positive to negative tweets on the ordinates (percentage of positive tweet - percentage of negative ones). Moreover we will provide tooltips to exactly show the sentiment score, the text of the tweet and the specifics of each instance
- (d) Variation of sentiment in the period of blocking of various vaccines (e.g. astrazeneca scandal) will be visualized in the charts providing a slider on time and legend showing the time intervals of the 'salient' events
- (e) A map with colour gradations will show in which states is sentiment higher and in which states is it lower
- (f) Variation in sentiment with respect to vaccine type over time will be showed with smiles that fill up based on the changing of the sentiment
- (g) Sentiment differences between verified and unverified accounts will be visualized applying a filter in other graphs or just in the map
- (h) Sentiment change with regard to number of retweets/likes with a slider on the map described above