

SAPIENS

PROJECT 4: DISSECTING TWITTER DATA TO ANALYZE GOVERNMENT AND PUBLIC ATTITUDE TOWARDS COVID GOVERNANCE

Requirement 1 – Social Network Analysis

Influencer Score for each tweet has been calculated in the direct method using the formula as below

$$\text{Influencer Score Count} = \text{Number of Re-Tweets} + \text{Number of Followers}$$

$$\text{Influencer Score} = \text{Verified: True?} (\text{Influencer Score Count})^2: (\text{Influencer Score Count})$$

That is, when the tweet is from a verified account, the field Influencer Score Count is squared or else it remains the same. This formula provides emphasis to authority by giving preference to authorized sources.

Requirement 2 – Content/Topic Analysis

Latent Semantic Analysis has been used for performing Topic Analysis. By giving number of topics (E.g. 10) the whole dataset is grouped to 10 different topics. By manually going through all 10 the topics pertaining to “covid” have been selected. The total number of tweets posted by each POI from 12th September to 18th September, pertaining to covid have been calculated and this data was plotted against covid curve between the same dates.

For each country the topics pertaining to all tweets posted, are represented using word cloud.

Word Cloud



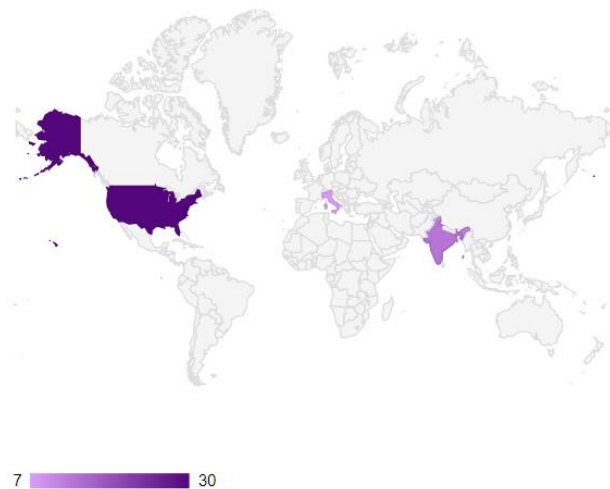
Requirement 3 – Insights/Analytics

When a keyword is searched, sentiment analysis, location analysis, verified tweets have been calculated and displayed on the home page.

Sentiment Analysis: For each tweet returned from Solr, the average sentiment score has been calculated using the python library of SentimentIntensityAnalyser and presented on the page.

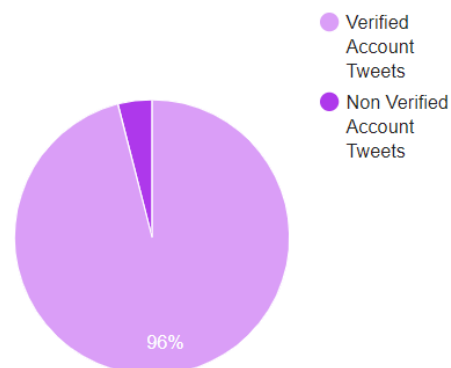
Location Analysis: For the results returned from the search, location analysis has been done for the top 50 tweets by calculating the number of tweets from each country and plotted in a geo chart using google charts API.

Geographical Distribution of Tweets



Verified vs Non-Verified Tweets: A pie chart has been plotted by calculating the number of tweets from verified accounts as well as from non-verified accounts, from the top 50 tweets retrieved.

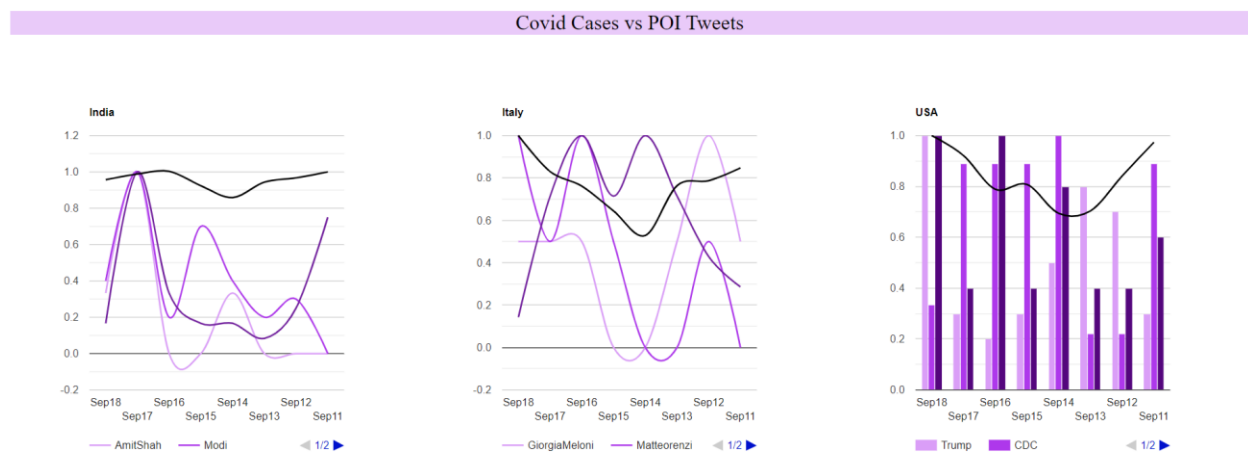
Verified vs Non Verified Account Tweets



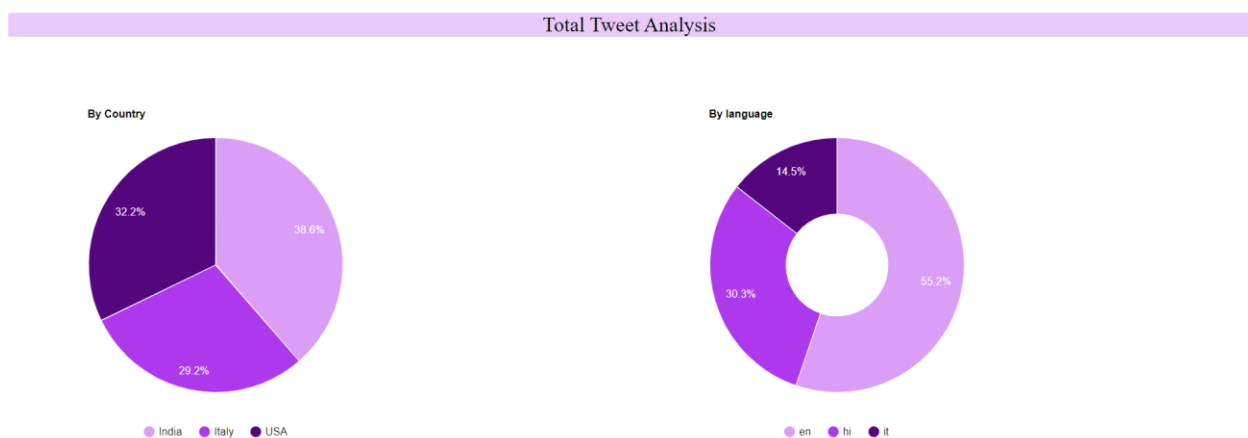
In addition to the tweets, news articles are also included based on the searched keyword, poi name, country and language. News-API has been used to scrape the data. The link to the news article is also provided when clicked on “Link” button, for full information.

Graphs:

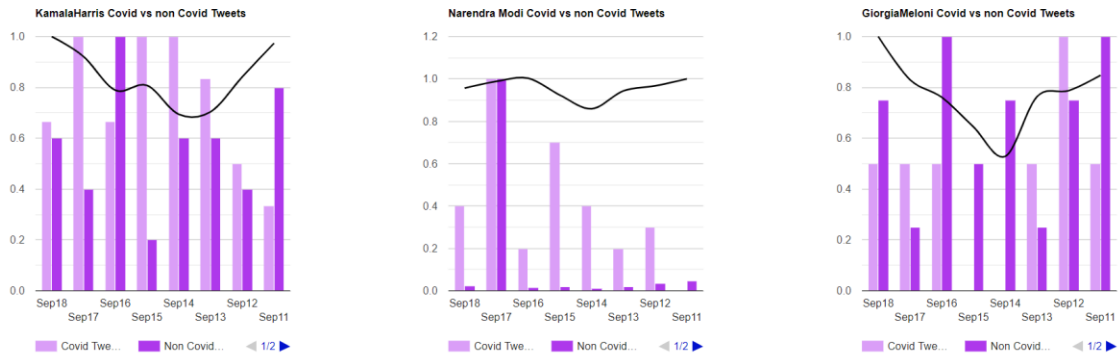
1. After performing content analysis as mentioned above, poi tweets have been plotted against covid cases between Sep 12 to Sep 18 for each country. The graphs for the same are shown below, from which we can infer the rate at which poi tweets have changed with respect to newly registered covid cases.



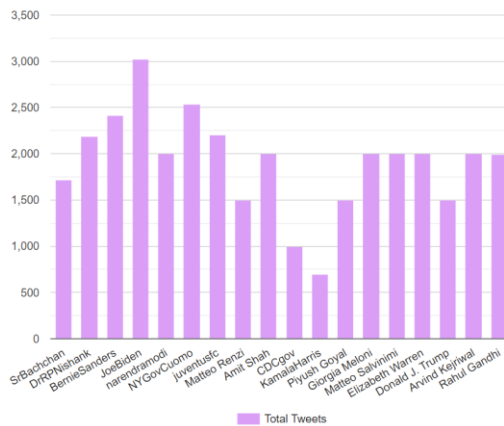
2. Pie Chart depicting number of tweets per country and a donut chart representing number of tweets by language has been drawn as shown below.



3. The correlation between covid/ non-covid tweets per poi from each country with new covid cases is depicted below with the help of combination chart. Bars represent the covid/non-covid tweets posted by poi on a particular day, whereas curve represents the number of new covid cases on a particular day.



4. A bar graph representing the total number of tweets per each POI.



Requirement 4 – Faceted Search

The website is provided with a navigation bar to navigate between the “Home” page and “Analytics” page.

The home page contains a text box to type in the search word and drop-down menus for POI name, country, language to select from along with a submit button. The page is divided into left, middle and right columns, to fit in tweets, news and query related analytics respectively. The total number of tweets retrieved for that particular search are also shown.

The Influencer Score has been used on Solr data to provide the results in the descending order of the score.

Flask served as an interface to render the html pages and to communicate to the back-end python.

The analytics page contains graphs that have been populated using static data collected by all the team members from project1.



Team Members and their contributions

Anjana Tejaswini Kalava (50338176): UI, News Scraping, Sentiment Analysis

Sindhuja Parnam (50337026): Topic Modeling, Word Cloud

Sindhu Sola (50336646): Indexing, Charts

Resources

<https://newsapi.org/docs/client-libraries/python>

<https://developers.google.com/chart>