

COVID-19 Sentiment and Social Network Analysis

Dr. Priyadarshini (FDA Faculty)



Introduction

Owing to their potential relevance to numerous diverse sectors, sentiment analysis and opinion mining have been playing an increasingly important role in both commercial and research applications. As a result, a great number of businesses have made consumer sentiment and opinion analysis a significant component of their purpose. The automatic analysis of social network messages based on the feelings and emotions expressed is one of the most intriguing applications of these methods

Problem Statement

The given challenge is to build a classification model to predict the sentiment of Covid-19 tweets. The tweets have been pulled from Twitter and manual tagging has been done. We are given information like Location, Tweet At, Original Tweet, and Sentiment.



We aim to obtain insights from textual data focused on COVID-19. Our project involves a sentiment and network analysis of data specific to COVID-19 using R. Since we do not have a specific dataset for this part, we will use data from multiple sources including Bing, COVID19 Daily Updates, etc. We also aim to visualize the primary emotions in the data (trust, fear, anticipation, etc.). We aim to have as many visualizations as possible in order to understand the data better. Social Network analysis would help us visualize the pattern between most frequent set of words with the help of graphs and networks

- it immensely helps in monitoring our brand's health in the market. From a single dashboard, you can quickly understand if our brand health is good, neutral, or depleting.
- It helps us manage our brand reputation better and quickly address ORM concerns and crises
- Supports the development of better marketing campaigns by letting us understand the pulse of your audience and tapping into it
- Competition analysis can be optimized through sentiment analysis to significant extents
- Most important of all, customer service can be improved for more satisfaction and quick turnarounds

Objective



Proposed Methodology

We propose an analytical approach where we apply multiple visualizations like bar chart, wordclouds, etc. depicting various patterns and hidden information behind the data. This also helps us to analyse primary emotions within the tweets and identify trends between various attributes spread across the datasets. Wordclouds also help us to analyse positive and negative words in the tweets further strengthening our analysis. Ours is a lexicon based approach, hence, we rely on manually defining stop words, etc used to clean the dataset to remove ambiguity at multiple instances. We also use the fundamental concept of tokenization of text into unigrams, bigrams or trigrams which would make handling and analysis easier and efficient.



Algorithms Used

— Rule/Lexicon based Approach

— Keyword Stopping

1. Read and explore data in order to find any anomalies 06
2. Prepare data for the forthcoming task- this step would involve data manipulation and cleaning
3. Clean the textual data- remove punctuations, extra white space, stopwords, etc.
4. Create N-grams from textual data to enable easier processing.
5. Use lexicon to predict the sentiments of words through semantic orientation.
6. Analyse primary emotions in the data and create word clouds to predict the sentiment.
7. Perform social network analysis to visualize the connection between the most frequent combination of words with the help of graphs and networks.
8. Draw conclusions from obtained results →

Algorithm Description

- Rule/Lexicon based Approach: This method uses manually created data classification rules to ascertain sentiment. This method determines a score by using dictionaries of words with positive or negative values to indicate their polarity and sentiment strength. Expressions can also be used to create more functionality. By creating even better rules, rule-based sentiment analysis algorithms can be modified according to context. How it works: It counts the amount of words in the given text that are both positive and negative. It will return a positive sentiment if there are more positives than negatives. It will give a neutral response if both are equal.
- Keyword Stopping: Keyword spotting is one of the simplest and most straightforward techniques and widely used by Sentiment Analysis algorithms. The input text is thoroughly scanned for the prominent positive and negative words like “sad”, “happy”, “disappoint”, “great”, “satisfied”, and such.
- Keyword spotting system detection performance degrades when substitution errors occur during the keyword recognition. Therefore, a keyword recovery algorithm was developed to reduce the detection degradation to a low false alarm rate to recover keywords after rejecting acoustic noise. Tests show that the detection performance improves from 78.1% to 85.3% for a false alarm rate of 10% with the upper range of the detection performance improved from 85.0% to 92.0%. Thus, the system performance is enhanced by the keyword recovery algorithm for both the average and maximum detection performance.



Parameter Used

08

Confirm: Number of confirmed covid cases

Recovered: Number of recovered cases

Death: Number of deaths

User Location: The location of the user when the tweet was posted

Text: The actual tweet posted by the user

Date: Date when tweet was posted

User followers: Number of followers of the user who posted the tweet

User favorites: Number of favorites user had at the time of posting

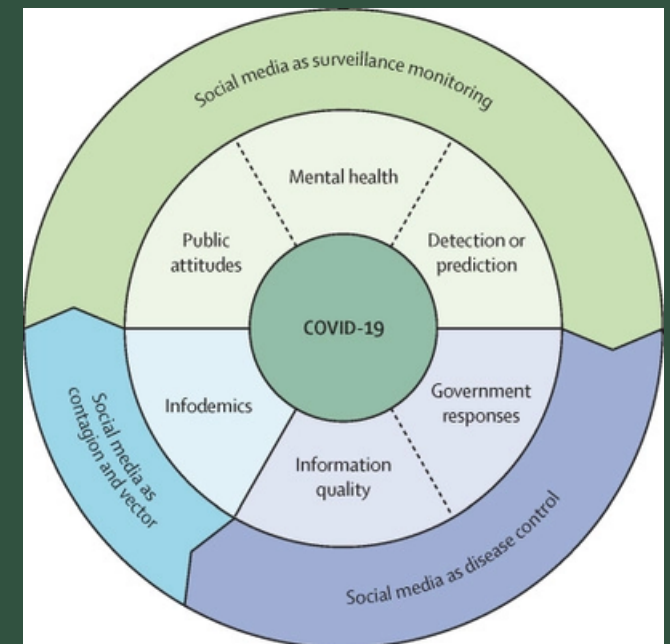
User friends: Number of friends user had at the time of posting



Findings

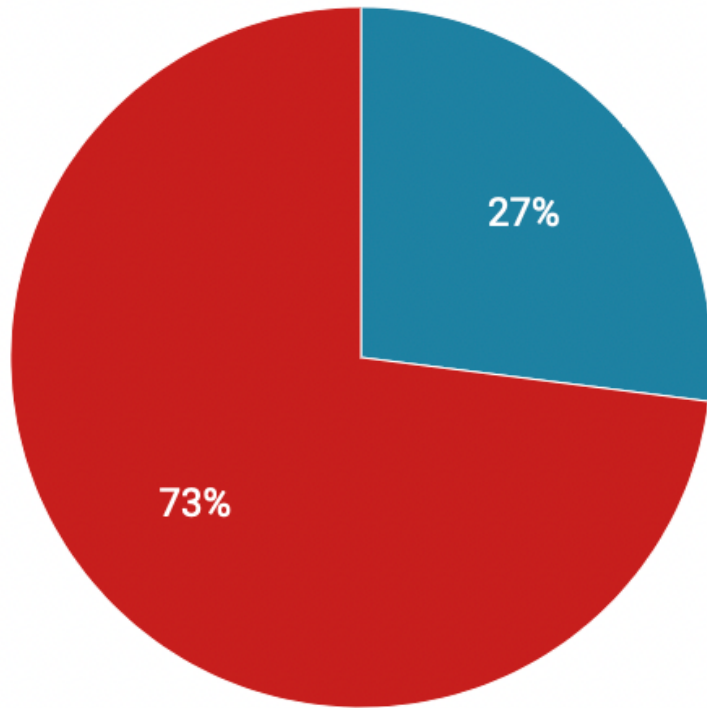
Our findings would include primary sentiments behind the textual data along with the intensity of the sentiments and overall mood behind the data. We would also include visualizations like radar chart, word clouds, histogram plots, etc.

1. The number of cases rose exponentially with time.
2. Brazil, India, Russia, Spain, and USA were the top five countries with respect to number of confirmed COVID cases. However, USA had the highest number of cases, standing at around 306 million.
3. In USA, New York was the worst affected state of all.
4. Wordcloud shows us some of the positive as well as negative words including protect, thank, etc. being positive and loss, symptoms, etc. being negative words.
5. The overall mood in tweets is analyzed to be trust as the primary emotion, and fear, anticipation, etc. being some of the other emotions.
6. Most frequent words by emotion shows multiple words like death, time, pandemic, etc. to be most frequent words for emotions like anger, anticipation, fear, etc.



Conclusion

Our Conclusion would be to find out the primary sentiments in the text and the intensity of those sentiments through visualization as well as natural language processing techniques and performing a short social network analysis to draw further conclusions.



Social media overall

OUR TEAM



HARNUR SINGH
20BCE1809



RISHAB SINGH
20BCE1777



RITIK SINGH
20BCE1384

Thank You!