International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2022; 3(2): 56-59
Received: 01-06-2022
Accepted: 05-07-2022
www.computersciencejournals.com/ijcai

Karmanya Upadhyay
Research Scholar,
City International School,
Pune, Maharashtra, India

Corresponding Author:
Karmanya Upadhyay
Research Scholar,
City International School,
Pune, Maharashtra, India

Topic modelling and sentiment analysis architecture for social networks in the COVID pandemic

Karmanya Upadhyay

DOI: https://doi.org/10.33545/27076571.2022.v3.i2a.55

Abstract
The goal of this research is to identify the topics and feelings in the community COVID-19 vaccine-related debate on social networks and to decipher the pertinent adjustments in subjects and feelings atop gray moment in order to study the public's perception, concerns, and feelings that may meet a variety of vaccination rates goals. This will be carried out in order to track down the themes and opinions expressed in social media conversations around the COVID-19 vaccination. The World Health Organization declared COVID-19 a pandemic on March 11, 2020, and from that day until January 31, 2021, messages were retrieved from a sizable COVID-19 Twitter buzz data collection. We sorted through the tweets and only preserved the ones that contained the phrases immunisation, vaccination, vaccine, injections, and inoculated. And used the Dictionary, we carried out latent Dirichlet allocations for language models as well as emotion and emotional assessment. Both of these studies were completed with R's assistance. Outcomes Theme modelling was used to generate a total of 16 topics from twitter about the COVID-19 vaccine, which were then classified into 5 broad categories. People's views on vaccinations these were biggest trending topic on Twitter (227,840 out of 1,499,421 tweets, or 15.2%), and it stayed so for the most of the period we were examining the information. Additionally, the conversation included presentations with a broader perspective. Due to the increasing optimism around these vaccinations and the dominantly sense of trust displayed in the debate on social networks, it's probable that COVID-19 vaccinations are more widely accepted than previous vaccinations.

Keywords: Social media data, COVIDE-19, vaccination, topic modelling

Introduction
According to reports, China experienced a COVID-19 epidemic in December of 2019 [1]. The World Health Organization (WHO) has declared a state of crisis due to the SARS-CoV-2 virus' quick wide impact. According to new findings, the epidemic can spread to 2.2 additional persons on median each person, and the quantity of those afflicted with can quadruple each 7 days [2]. In Latin America, there were a total of 49,139 confirmed cases from COVID-19 as of May 31st, 2020, and 937,974 reported cases [3]. In this region of the world, Brazil is the country with the highest number of instances of the illness. According to the earlier-mentioned data [3], Brazil saw a number of 465,166 illnesses and 27,878 deaths.
The exchange of data during earlier epidemic epidemics was relatively slow. However, due to the internet's widespread use, 3.7 billion people worldwide (or around 49.7% of the world's population) regularly access web-based content [4]. These folks frequently rely on social networking sites and press media as their principal sources of online information. Through these channels, people keep constantly updated on remedies, preventative measures, and instances. They also participate in discussions about how the COVID-19 outbreak may affect daily life.
Reports on emergency situations around the globe are published on the sites of different news outlets. Because reporters and other people who are authorities in respective professions write the posts on these websites, individuals trust within those sources of data. These outlets, though, failed to keep up with the COVID-19 outbreak's rapid expansion [5], and many media outlets falsely reported that the epidemic still wouldn't travel outside of China or that the disease was less harmful than influenza [6]. After the intensity and scope of the cancer's spreading were clear, the way the epidemic was covered had an impact [7].
On the other extreme, in the latest media climate, social networking is a popular source of news and knowledge. Given that two-thirds of individuals use the computer and that one in 3 people use social networks globally [8], social networking has established itself as a credible

source of information and news in the current news industry. This is particularly true when it comes to health-related topics, as one-third of users claim that social media sites are an useful source of data [9]. Recent research suggests that social internet has also turned into a breeding site for false info on COVID-19, too. This related research focused on how conventional media sources describe its reportage of health-related occurrences and how they respond to them. Our research adopts a novel strategy by focusing on the examination of social medium and comparing it to conventional news media because we're looking to demonstrate how the COVID-19 epidemic has influenced people's lives.

Whether or not there is a connection between pandemic incidents and the knowledge that individuals communicate on social media, particularly Twitter, is another question that the science community is interested in answering. There are several instances that show how pertinent data may be gleaned through social media to help organisations take action that will enhance peoples' quality of life as well as aid in properly studying the dynamics of pandemics.

Related Works

Twitter has become one of the most important sources, as data has become a valuable resource in recent years. Particularly during the time of the pandemic [10]. This study intend to compare the sentiment of two different approaches to sentiment analysis (lexicon-based and machine learning) by using the tweets that are related to the COVID-19 vaccine in two different cultures. Specifically, I will be focusing on the United States and China.

The COVID-19 outbreak has been declared a pandemic, which represents a significant risk to the general population's health all over the world and has resulted in a flood of online social media. On social networking sites like Twitter and Facebook, users regularly share their thoughts, feelings, and perspectives regarding the pandemic's unfolding events. A great number of researchers are attempting to evaluate the attitude conveyed by COVID-19-related content found on these social networks. On the other hand, they have hardly ever concentrated on the vaccine. In [11], we investigate the topic of the COVID-19 vaccine that was discussed on Twitter. To be more specific, all of the tweets pertaining to the COVID-19 vaccine that were posted on Twitter between December 15th, 2020 and February 10th, 2021 were gathered with the help of the Twitter API. After that, an unsupervised learning VADER model was utilised to judge the emotion categories (positive, neutral, and negative), and it was used to calculate the sentiment value of the dataset. The construction of a communication topological network and the investigation of the emotional direction are both based on the interaction that occurs between users. We found that people in other countries had different opinions about the Chinese vaccine compared to those in China. The number of cases and fatalities reported in the daily news, as well as the nature of the most pressing problems in the communication network, may have an effect on the sentiment value. And demonstrating how the most influential nodes in a social network can spread negative feelings to other nodes in the network.

At this point in time, social media platforms are absolutely necessary for the purpose of informing people about global issues such as the ongoing COVID-19 pandemic. Twitter has proven to be an excellent medium for people to share their various perspectives on the COVID-19 vaccine. These perspectives have been shared by a variety of individuals. The objective of [12] study is to analyse the opinions expressed on Twitter in relation to the COVID-19 vaccine by proposing a text vectorized neural network (NN) model and contrasting it with long short-term memory (LSTM) and bidirectional long short-term memory (BiLSTM), respectively. The raw data from Twitter is collected, and then three different raters annotate the data as to whether or not it is positive or negative. The kappa value is used as the finishing touch on the label. Then, Natural Language Processing (NLP) methods are used to process the twitter data. According to the findings of the research, the proposed text vectorized NN model achieves a higher level of accuracy than other models, such as LSTM and BiLSTM, which only achieve 75% and 74%, respectively, in the test dataset. This finding leads the researchers to draw the conclusion that the proposed model outperforms other models. The Matthews' correlation coefficient (MCC) assigns a score of 51% to the text vectorized NN model, while the LSTM and BiLSTM models receive scores of 37% and 39% respectively. In order to validate the models in the most efficient manner possible, additional performance metrics including precision, recall, f1-score, and confusion matrix were utilized.

In spite of the evidence that demonstrates the benefits and safety of immunizations, a general reluctance toward vaccinations has been fueled by the widespread dissemination of misinformation and conspiracy theories related to vaccines on the internet. Coronavirus disease (COVID-19) vaccinations are not an exception to this trend. A lack of confidence in the effectiveness of the COVID-19 vaccine is recognised as a global risk to public health that undermines efforts to bring the COVID-19 pandemic under control. Twitter and other social media platforms give users the ability to communicate with one another and share information, as well as their concerns and feelings, regarding matters pertaining to COVID-19. Using information gathered from Twitter, the purpose of [13] study is to gain a better understanding of how people feel about COVID-19 vaccines. The authorities are able to improve their decision-making processes and achieve the desired level of herd immunity by conducting research into the public's perspective on vaccinations. In this paper, we classify people's opinions regarding COVID-19 vaccinations into positive, negative, and neutral categories using the state-of-the-art transformer-based classification models RoBERTa and BERT, along with multiple task-specific versions of each model. The models that are presented have been trained and evaluated with the help of data taken from Twitter and consisting of people's opinions on vaccinations. There are two techniques for ensemble learning that are presented here for further performance improvement. These techniques are majority voting and stacking with Support Vector Machine (SVM) as meta-learner. Both of these techniques aggregate the individual classifiers. The results also demonstrate that using ensemble learning achieves significantly better results than using individual classifiers when evaluating against all of the criteria. In addition to this, we discovered that assembling with stacking has an advantage over voting by simple majority.

It is possible to protect a significant number of people from contracting COVID-19 by administering the vaccination at

the appropriate time [14]. Despite this, some people are hesitant to get vaccinated because of how new the disease is and how quickly vaccines are being developed. The decisions that people make regarding vaccinations may be influenced by the vast amounts of information that are produced by online social networks such as Twitter. As a result, it is essential to examine the discussion regarding the COVID-19 vaccination through the prism of social media. The purpose of this study was to determine the nature of a larger conversation taking place on Twitter about the COVID-19 vaccine and investigate the interaction patterns that occur between users of Twitter who are participating in such a conversation. The findings of this study indicate that reactions characterised by a range of emotions, as well as discussions about the possible adverse effects of vaccines and their overall safety, predominated the conversation that took place online. Different groups of conversation stakeholders were highlighted by each of the four primary network clusters. The findings of this study highlight the significance of Twitter surveillance and draw attention to the conversational patterns and prevalent sentiments that predominate in online social networks during a significant public health emergency.

Proposed work

Extraction of Data and Preprocessing of It

Using the data set that is managed by the Panacea Lab at Georgia State University [22], we were able to obtain the IDs from a total of 1,499,421 tweets that were sent out between March 11, 2020 and January 31, 2021. These tweets did not include any retweets. These tweets were gathered by the Panacea Lab utilising the following 13 keywords: COVID19, Coronavirus Pandemic, COVID-19, 2019nCoV, Corona Outbreak, coronavirus, Wuhan Virus, covid19, coronavirus pandemic, covid-19, 2019ncov, corona outbreak, coronavirus, and wuhan virus. The only thing that the Panacea Lab can provide is the tweet IDs [23], and these tweet IDs need to be hydrated in order to get all of the tweet data.

Modeling of Topics

Data modeling can help organise massive material libraries via classifying according to its relevant themes. Theme modelling is also usually referred to as probability grouping. It is more dependable and often produces findings which are more real when opposed to fuzzy clustering, such as k-mean clustering [25]. Most methods provide one topic to each article after assuming that there is a connection among subjects. Learning algorithm, on the other hand, does not start with the presumption of a distance function among topics and instead opens by allocating a material to a group of themes with varied weight or possibilities. The latent Dirichlet allocation (LDA), though there are many distinct topic models, is the one that is most usually utilised. In order to extract prevalent subjects from the astronomically vast amount of twitter which were accessible, we employed the LDA method for recommender systems.

Methodology

The study identifies the key topics and viewpoints discussed in relation to the COVID-19 vaccine on social networks. It also examines how these topics and emotions have evolved over time to help understand the larger pattern. Out of 16 various topics, attitudes on vaccinations were the most common and remained so throughout time. Additionally, as worldwide vaccine development improved, the predominant themes shifted. The most often discussed subject at the start of 2021 was advise on getting the vaccine. The major news headlines on COVID-19 vaccinations that controlled the discussion about immunizations on media sites mirrored the popular topics in the press. A studies indicating is also included in the topic. The predominant emotion is optimism, which grew more positive over time. This suggests that social media conversations may reflect more knowledge of COVID-19 vaccinations than those for earlier doses. Due to the temporality of our data collection, we did not analyse the views about specific immunisation brand in this study. You may assume that the conversation would vary while looking for tweet opportunities and conducting sentiment using various major brands. There'd be more discussions on the subject and concerns about the negative impacts of vaccinations, particularly in the wake of the CDC's forced to cancel the supply of the Johnson & Dow medicine. Therefore, further study in this field is highly suggested.

Result Analysis

Twitter dataset is collected as tweets and analyzed topics involved. The main keywords used are,

- Covid-19
- Vaccination
- Vaccine Camp
- SARS-Covid
- Corono Virus
- Covid Vaccine

By using above keywords, the tweets are collected and used as dataset. Then, we applied sentiment analysis to analyze the sentiment score. The main sentiments involved are,

- Hesitation
- Anger
- Fear
- Happy
- Hope
- Pleasing
- Helpless
- Awareness

By applying topic modelling, major topics are detected and are summarized in table.1.

Table 1: Topics Collected

Topic	Description
Topic 1	Opinions on Vaccination
Topic 2	Global Opinion
Topic 3	Vaccine Rollout
Topic 4	Economic Report
Topic 5	Immunity and Vaccination
Topic 6	Critical Review
Topic 7	WHO Announcement
Topic 8	Vaccine Progress

In figure.1, average polarity of sentiment score is plotted. The score shows that the sentiment score varies with time. That is opinion about vaccination varies over a period. After some time, people starts to supports vaccination.

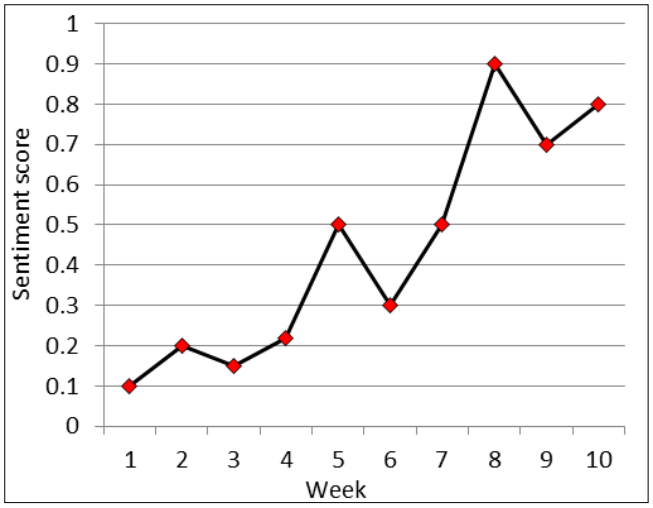

Fig 1: Analysis on sentiment score

Conclusion

The research highlights the main themes and opinions expressed on social media around COVID-19 vaccine-related issues. In order to better comprehend the bigger trend, it also looks at how these subjects and feelings have changed throughout time. Opinions on immunisation were the most prevalent as among 16 different issues and stayed so over time. The predominate subjects also changed as vaccine development advanced globally. Around the beginning of January 2021, advice on getting the vaccine emerged as the most talked-about topic. The big news stories concerning COVID-19 immunizations that dominated that conversation regarding immunisation on media platforms reflected the trending themes in the mass press. The topic also includes a global viewpoint. Confidence was the main feeling, and the general feeling became more favourable throughout period. This indicates that social media discussions may indicate high awareness of COVID-19 vaccines compared to prior vaccines. In this investigation, we did not thereby investigate the attitudes toward certain vaccination brands because of the timeframe of our data set. When searching for tweeting and performing sentiment utilizing various big brands, you might anticipate that the discussion will vary. Especially following the CDC's decision to halt the Johnston & Jones drug's distribution, there may have been an increase in debates surrounding the matter, and the adverse effects of the immunizations may have come up. Consequently, further research in this area is strongly advised.

References

1. Wang Kai, *et al*. Relational Graph Attention Network for Aspect-based Sentiment Analysis. ACL; c2020.
2. Basiri Mohammad Ehsan, *et al*. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. Future Gener. Comput. Syst. 2021;115:279-294.
3. Onan Aytuğ. Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. Computer Applications in Engineering Education. 2021;29(3):572-589.
4. Jang Hyeju, *et al*. Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. Journal of Medical Internet Research, 2021 Feb 10;23(2):e25431.
5. Costola M, Nofer M, Hinz O, Pelizzon L. Machine Learning Sentiment Analysis, Covid-19 News and Stock Market Reactions. SSRN Electronic Journal; c2020.
6. Muriyatmoko Dihin, *et al*. Sentiment Analysis Covid-19 Vaccination on Twitter Social Media Using Naïve Bayes Method. Procedia of Engineering and Life Science; c2021. n. pag.
7. Rustam Furqan, *et al*. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS ONE, 2021, 16. n. pag.
8. Nemes László, Attila Kiss. Social media sentiment analysis based on COVID-19. Journal of Information and Telecommunication. 2021 Jan 2;5(1):1-5.
9. Naseem Usman, *et al*. COVID Senti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. IEEE Transactions on Computational Social Systems. 2021 Jan 29;8(4):1003-15.
10. Alsabban M. Comparing two sentiment analysis approaches by understand the hesitancy to COVID-19 vaccine based on Twitter data in two cultures. 13th ACM Web Science Conference. 2021 Jun 21 143-144.
11. Xu H, Liu R, Luo Z, Xu M, Wang B. COVID-19 Vaccine Sensing: Sentiment Analysis from Twitter Data. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC); c2021. p. 3200-3205.
12. Ferdous Z, Akhter R, Tahsin A, Mustafina SN, Tabassum N. Sentiment Analysis on COVID-19 Vaccine Twitter Data using Neural Network Models. Proceedings of the 2nd International Conference on Computing Advancements. 2022 Mar 10;435-441.
13. Ismail Q, Obeidat R, Alissa K, Al-Sobh E. Sentiment Analysis of COVID-19 Vaccination Responses From Twitter Using Ensemble Learning, 2022 13th International Conference on Information and Communication Systems (ICICS). c2022. p. 321-327, Doi: 10.1109/ICICS55353.2022.9811132.
14. Ekaterina Malova. Understanding online conversations about COVID-19 vaccine on Twitter: vaccine hesitancy amid the public health crisis, Communication Research Reports. 2021;38(5):346-356.
DOI: 10.1080/08824096.2021.1983424