Detecting Topics and Sentiments of Public Concerns on COVID-19 Vaccines with Social Media Trend Analysis

Michal Monselise, Chia-Hsuan Chang, Gustavo Ferreira, Rita Yang, Christopher C. Yang

Submitted to: Journal of Medical Internet Research
on: May 27, 2021

Table of Contents

Detecting Topics and Sentiments of Public Concerns on COVID-19 Vaccines with Social Media Trend Analysis

Michal Monselise[1]; Chia-Hsuan Chang[2]; Gustavo Ferreira[1]; Rita Yang[3]; Christopher C. Yang[1] PhD

[1]College of Computing and Informatics Drexel University Philadelphia US
[2]Department of Information Management National Sun Yat-sen University Kaohsiung TW
[3]Virtua Voorhees Hospital Voorhees Township US

Corresponding Author:
Michal Monselise
College of Computing and Informatics
Drexel University
3675 Market St
10th Floor
Philadelphia
US

Abstract

Background: As a number of vaccines for COVID-19 are given emergency use authorization by local health agencies and are being administered in multiple countries, it is crucial to gain public trust in these vaccines to ensure herd immunity through vaccination. One way to gauge public sentiment regarding the vaccines for the goal of increasing vaccination rates is by analyzing social media such as Twitter.

Objective: The goal of this research is to understand public sentiment towards COVID-19 vaccines by analyzing discussions about the vaccines on social media. Using the combination of topic detection and sentiment analysis, we identify some plausible causes for vaccine hesitancy of the public that appear in social media.

Methods: To better understand public sentiment, we collected tweets between December 16th, 2020 and February 13th, 2021 that contained hashtags or keywords related to COVID-19 vaccines. We detected and analyzed the different topics of discussion of these tweets as well as their emotional content. Vaccine topics were identified using non-negative matrix factorization (NMF) and emotional content was identified using the VADER sentiment analysis library as well as using sentence BERT embeddings and comparing the embedding to different emotions using cosine similarity.

Results: After removing all duplicates and retweets, 7,864,640 were collected during the time period. Topic modeling resulted in 50 topics of those we selected the 12 topics with the highest volume of tweets for analysis. Administration and access to vaccines are some of the major concerns in the pubic. Additionally, we classified the tweets in each topic into one of 5 emotions and found fear to be the leading emotion in the tweets followed by joy.

Conclusions: This research focuses not only on negative emotions that may lead to vaccine hesitancy but also on positive emotions toward the vaccine. By identifying both positive and negative emotions, we are able to identify the public's response to the vaccines overall and to news events related to the vaccines. These results are useful in developing plans for disseminating the authoritative health information and better communication to build understanding and trust.

(JMIR Preprints 27/05/2021:30765)
DOI: https://doi.org/10.2196/preprints.30765

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Detecting Topics and Sentiments of Public Concerns on COVID-19 Vaccines with Social Media Trend Analysis

Abstract

Background: As a number of vaccines for COVID-19 are given emergency use authorization by local health agencies and are being administered in multiple countries, it is crucial to gain public trust in these vaccines to ensure herd immunity through vaccination. One way to gauge public sentiment regarding the vaccines for the goal of increasing vaccination rates is by analyzing social media such as Twitter.

Objective: The goal of this research is to understand public sentiment towards COVID-19 vaccines by analyzing discussions about the vaccines on social media for a period of sixty days when the vaccines were started in US. Using the combination of topic detection and sentiment analysis, we identify different types of concerns regarding vaccines that are expressed by different groups of the public that appear in social media.

Methods: To better understand public sentiment, we collected tweets for exactly 60 days starting December 16th, 2020 that contained hashtags or keywords related to COVID-19 vaccines. We detected and analyzed the different topics of discussion of these tweets as well as their emotional content. Vaccine topics were identified using non-negative matrix factorization (NMF) and emotional content was identified using the VADER sentiment analysis library as well as using sentence BERT embeddings and comparing the embedding to different emotions using cosine similarity.

Results: After removing all duplicates and retweets, 7,864,640 were collected during the time period. Topic modeling resulted in 50 topics of those we selected the 12 topics with the highest volume of tweets for analysis. Administration and access to vaccines are some of the major concerns in the pubic. Additionally, we classified the tweets in each topic into one of 5 emotions and found fear to be the leading emotion in the tweets followed by joy.

Conclusion: This research focuses not only on negative emotions that may lead to vaccine hesitancy but also on positive emotions toward the vaccine. By identifying both positive and negative emotions, we are able to identify the public's response to the vaccines overall and to news events related to the vaccines. These results are useful in developing plans for disseminating the authoritative health information and better communication to build understanding and trust.

Keywords: Healthcare Informatics; Topic Detection; Unsupervised Sentiment Analysis; COVID-19; Vaccine Hesitancy

Introduction

In late 2020, the COVID-19 pandemic had approached the year mark when a number of pharmaceutical companies began to release their vaccine clinical trial results. The global sense of relief was felt when the results of the clinical trials looked promising. The first vaccine, developed by Pfizer and BioNTech, was given for emergency use authorization in December 2020 by the FDA [1]. While this timeline seemed too fast for some, most vaccines for COVID-19 relied on many years of previous scientific work. For example, mRNA based vaccines have been in development for over a decade at that point [2–4]. Despite efforts of the scientific community to assure the public that these vaccines are safe and effective, public sentiment has been mixed. There has been a significant amount of public hesitancy towards vaccination against COVID-19 [5]. At the same time, many have expressed excitement over the prospect of returning to a pre-pandemic world. Given this mixed reaction, it is essential to investigate the actual public sentiment regarding COVID-19 vaccines. Particularly, we are interested in learning about public sentiment for a period of sixty days when the vaccines were started in the United States. Social media provides a great data source for listening to the public on what they are thinking and what concerns and questions they have. We use Twitter as a proxy for public sentiment and are able to find the most important discussion topics that pertain to COVID-19 vaccines in the early days of the vaccine rollout. Additionally, we are able to classify public sentiment as it pertains to the vaccines and how this sentiment changes over time overall and in each topic as well. The goal of this research is to examine the discussion topics and public sentiment towards COVID-19 vaccines. By studying the topic and sentiment of the discussion on COVID-19 vaccines on Twitter, we may understand public concerns as they happen and learn more accurately about the source of vaccine hesitancy. By learning what drives vaccine hesitancy, we can better address it and formulate tailored and targeted communication. Conversely, we may also learn about excitement towards the vaccine and study what is going well and what resonates well with the public on social media. This research will use the results uncovered by the topic and sentiment analysis of the Twitter data and suggest actionable insights for practitioners to address COVID-19 vaccine hesitancy. This research will also address how to utilize positive sentiment towards the vaccine.

Previous Work

COVID-19 Vaccine - Public Sentiment

A number of studies about vaccine hesitancy on social media have been published during the pandemic. Before any vaccine was approved, research showed hesitancy on social media. Harrison and Wu [6] examined vaccine hesitancy at the start of the pandemic and discussed methods to reduce vaccine hesitancy in preparation for the vaccine that will eventually come. This paper critiques current approaches for combating vaccine hesitancy with a goal of improving on these approaches when the COVID-19 vaccines are authorized for emergency use. A study by Chou and Budenz [7] discusses both methods for reducing hesitancy as well as fostering positive emotions towards the vaccine. They propose acknowledging fear, anger and other negative emotions and addressing them to convince the public to get vaccinated. A study by Wilson and Wiysonge [8] showed the existence of organized disinformation campaigns against the vaccines for COVID-19. However, this study focused on exposing negative sentiment against the vaccine and did not measure the positive sentiment towards the vaccine on social media. While these studies discuss public sentiment, they do not measure both positive and negative sentiment and some just make recommendations rather than looking at empirical evidence.

Topic Detection in COVID-19 Related Tweets Sentiment Analysis

Due to the pandemic and quarantine policy, the social media such as twitter becomes the main channel for people to share thoughts and to express their opinions about any impacts caused by the COVID-19. The hidden topics underneath such massive textual contents on social media helps governments and health care units to understand the demand of the general public so as to make better decision and quick response. Cinelli et al. [9] extracted topics using Partitioning Around Medoids (PAM) algorithm on word vector representations and proposed a custom epidemic model for characterizing misinformation spreading speed in different social platforms. Since the temporal trends of the hidden topics reflect concerns of general public through time, Chang et al. [10] proposed two temporal models based on NMF, which help to identify the trends of several important themes such as government policy, economic crisis, COVID-19 case updates, COVID-19 urgent events, prevention, vaccines and treatments, and COVID-19 testing.

Sentiment Analysis

Sentiment analysis is a research area that involves the classification of text, images, or audio into a set of one or more sentiments [11]. In the context of this research, we will be classifying the sentiment of short snippets of text. When classifying text, we can classify at the word, sentence or document level. There are different classification methods including rule based [12–14], SVM [15,16], random forest [17], and Naive Bayes [18,19], embedding based [20,21], as well as sentiment analysis using neural networks [22–25]. Additionally, we may classify sentiment using unsupervised methods such as methods using rule based unsupervised sentiment analysis [26], embeddings like Word2Vec and Doc2Vec [27], and lexical resources for sentiment analysis [28].

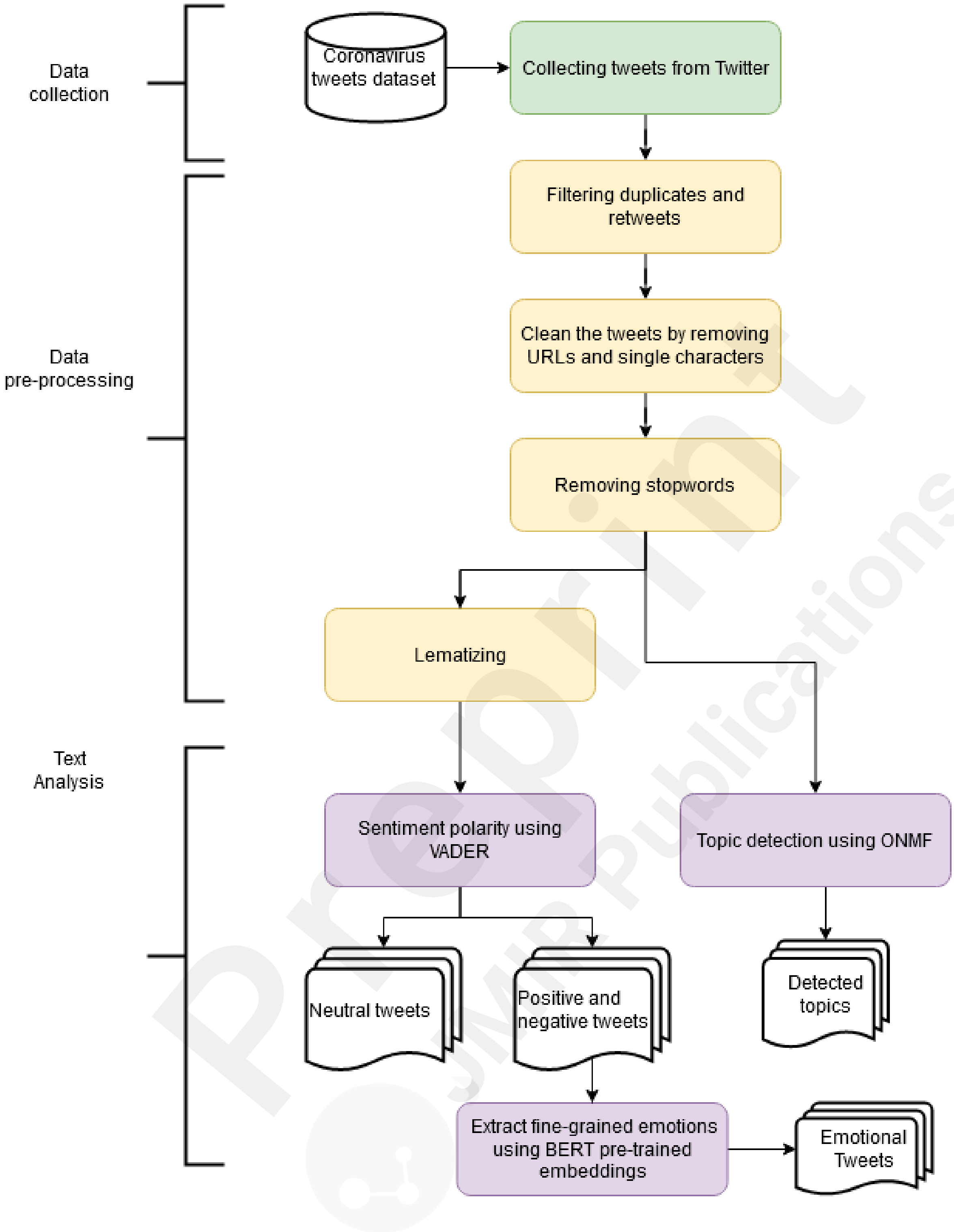
Sentiment Analysis in Twitter

Sentiment analysis is an established research field in the area of natural language processing. However, performing sentiment analysis on tweets is a slightly different task. Zimbra et al. [29] review a number of techniques for classifying sentiment in tweets. They find that due to factors like the brevity of tweets, their Twitter specific language[30], and a class imbalance [31], classification algorithms achieve an accuracy of around 70%. However, Adwan et al. [32] also reviewed a large number of techniques and they found a mix of accuracy scores with some papers passing 80% accuracy while others still perform below 80% even with new algorithms [33]. Amongst those who have improved their accuracy, some only focus on specific politics related datasets [34], some propose methods that require a large number of steps [35], while others address the issues with tweets, like Twitter specific language [36].

Methods

Our entire pipeline is described in figure 1. We first introduce the data collection and pre-processing. We then detail our topic detection algorithm and procedure of sentiment and emotion classification.

Figure 1. Pipeline of our text analysis.



Data Collection

We adopted the coronavirus tweets dataset [37] as our data source, which uses over 90 keywords and hashtags [38] to monitor the real-time coronavirus-related tweets from February 05, 2020 till present. Since the US Food and Drug Administration (FDA) authorized Pfizer–BioNTech COVID-19 vaccine and Moderna vaccine for emergency use in mid-December, we only kept tweets that were created during a 60-day period between December 16, 2020 and February 13, 2021 for the goal of extracting discussion topics and their sentiment from the general public about COVID-19 vaccine. Due to the data sharing policy from Twitter, the coronavirus tweets dataset only shares the ids of collected tweets. Therefore, we employed the Twitter's tweet lookup API [39] to retrieve the content and meta information of each retained tweet. In order to downsize the corpus and retain vaccine-related tweets, we only selected tweets that contain at least one keyword in our predefined keyword list: "vaccine", "vaccines", "#vaccine", "#vaccines", "corona vaccine", "corona vaccines", "#coronavaccine", "#coronavaccines", "pfizer", "biontech", "moderna", "Pfizer-BioNTech", "Pfizer/BioNTech", "Pfizer BioNTech", "#PfizerBioNTech", "COVAX", "COVAX", "Sinopharm", "Sinovac", "AstraZeneca", "Sputnik V" and "Gamaleya". The list of keywords was generated by the authors with the intention of collecting data on COVID-19 vaccines in general as well as the specific vaccines that were available to the public at the start of the data collection period. We also filtered out duplicated content, e.g., retweets, and non-English contents for providing more consistent data. As a result, we have 7,864,640 tweets for further text analysis.

Topic Detection

There are two types of models for topic detection: latent Dirichlet allocation (LDA) [40] and non-negative matrix factorization (NMF) [41]. In this study, we choose NMF because it has been proved its superiority in extracting topics from tweets [42]. NMF is a matrix factorization algorithm that learns and maps high-dimensional data into low-dimension representations. In this study, our tweet corpus $V \in R^{F \times N}$ is represented as a matrix with F rows (words) and N columns (tweets). After the pre-processing process detailed in figure 1, we construct the corpus using tf-idf weighting scheme:

$$tf-idf_{i,j} = \frac{n_{i,j}}{\sum_{i \in F} n_{i,j}} \times \log \frac{N}{N^{(i)}},$$

where $n_{i,j}$ is the count of word $i \in F$ appearing in tweet $j \in N$, and $N^{(i)}$ is the number of tweets containing word i. With such weighting scheme, the word has more weights as it is an important word for a tweet. After encoding the corpus, we apply NMF for extracting topics, whose objective of factorization is as follows:

$$argmin_{W,H} = \|V - WH\|_F^2, s.t. W, H > 0$$

We can exploit the topic word distribution using $W \in R^{K \times N}$ because each column represents a hidden topic, where the representative words will be encoded more weights. $H \in R^{K \times N}$ can be served as document topic distribution since each column indicated a topic weight distribution of each tweet. For coping with the large-scale tweets and the subsequent memory issue, we adopt online NMF (ONMF) [10,43] to solve both W and H in an online learning fashion. Specially, the whole tweet corpus will be divided into a set of small batches $\{V^q\}_{q=1}^{Q}$ and be sequentially used for updating W^q and H^q of each batch. The step for updating coefficient of current batch H^q is to fix the word dictionary of previous batch W^{q-1} and find a H that recovers W^q with least error (see line 6 in Algorithm 1). Similarly, to update dictionary of current batch W^q, H^q is then fixed, and the best W is solved using line 8 in Algorithm 1. The mathematical details of two updating forms could refer to Zhao and Tan (2017) [43]. As a result, Algorithm 1 is the whole procedure for topic detection.

Algorithm 1: Topic detection using ONMF

Input: Corpus *D*, number of topics *K*, and a size of a batch *s*

1. apply tf-idf weighting scheme on *D* to construct *V*
2. split *V* into a set of batches $\{V^q\}_{q=1}^{Q}$ based on *s*
3. foreach batch $q \in [1, Q]$ do
4.   if *q-1*=0 then

5. Randomly initialize $W^{q-1} \in R^{F \times K}$

6. $W^q = argmin_{H>0} \frac{1}{2} \left\| V^q - W^{q-1} H \right\|_2^2$

7. $A = \frac{1}{¿ V^q \vee ¿ H^q (H^q)^T ; B = \frac{1}{¿ V^q \vee ¿ V^q (H^q)^T ¿} ¿}$

8. $W^q = argmin_{W>0} \frac{1}{2} tr(W^T WA) - tr(W^T B)$

Output: The last dictionary $W^{q=Q}$

Note that we will use the final topic word dictionary $W^{q=Q}$ to infer topic weights of each tweet (i.e, *H*). The representative topic of a tweet j is determined by selecting the topic with maximum weight: $argmax_{k \in K} H_{k,j}$ and we record the representative topics of all tweets as $H^{rep} \in R^{1 \times N}$.

Sentiment Analysis

To detect the sentiment conveyed in the tweets we utilize a two-step approach. In the first step, we compute the polarity score of our tweets and based on this score classify the tweets as either positive, neutral, or negative. In the second step we classify the emotional content of the tweet into one of 5 emotions: anger, fear, joy, hopefulness, and sadness.

Polarity Classification

The first classification step is performed using the VADER (Valence Aware Dictionary and sEntiment Reasoner) Python library [14]. The VADER library is a rule-based model for general sentiment analysis. VADER is constructed using existing well established sentiment lexicons such as LIWC (Linguistic Inquiry and Word Count) and supplemented using lexical features commonly used to express sentiment in social media. After expanding using social media lexical terms, VADER was then human validated and is currently considered a gold standard in social media lexicons [44]. VADER evaluates the sentiment of each tweet by returning a compound sentiment score between -1 and 1. Based on the classification thresholds determined by the developers of the library, we assign a negative sentiment to a compound score less than or equal to -0.05, a positive sentiment to all compound scores greater than or equal to 0.05, and a neutral sentiment to a compound score between -0.05 and 0.05 [14]. Since VADER is more sensitive to expressions of sentiment in the social media context, it performs better than other rule-based classification algorithms in this context [45]. It is found that VADER outperforms individual human raters [14] in F1 score.

Emotion Classification

In the second step, we separate our data into positive, negative, and neutral and detect one of two emotions for positive polarity: joy and hopefulness, and one of three emotions for negative polarity: anger, fear, and sadness. Since VADER only includes positive, negative, and neutral sentiment, to detect more fine-grained emotions, we use zero-shot classification, an unsupervised method for discovering the applicable emotion for each tweet. Zero-shot classification is used in machine learning to classify things like images and text [46,47]. We detect the emotion by finding the BERT (Bidirectional Encoder Representations from Transformers)[48] embeddings of the tweets and of the emotion words (fear, joy, hopefulness, anger, and sadness) and then computing the cosine similarity of the emotion words and each tweet and selecting the emotion with the highest cosine similarity as the emotion associated with the tweet.

BERT [48] is a word representation model that uses unannotated text to perform various natural language processing tasks such as classification and question answering. By considering the context of a word using the words both before and after the word, we are able to produce embeddings for words that are more context aware. Our research uses the pre-trained sentence BERT [49] model to generate the embedding vectors for our emotion classification task.

Given our tweet corpus $V \in R^{F \times N}$, we represent our emotion results as $X \in R^{C \times N}$, where C is number of emotion categories and N is the number of tweets. For each emotion, we compute an embedding vector E_{c_i} where $i=1,\ldots,C-1$, and for each tweet, we compute an embedding vector E_{v_j} where $j=1,\ldots,N$ using the pretrained sentence BERT model. To populate our emotion matrix, we first compute the VADER sentiment score and assign the score to the neutral category in our matrix. We then compute the cosine similarity between each of the remaining *C-1* categories and each tweet using the following equation:

$$X_{i,j}=\cos\left(E_{c_i},E_{v_j}\right)=\frac{E_{c_i}E_{v_j}}{\left\|E_{c_i}\right\|\left\|E_{v_j}\right\|}$$

Where $i=1,\ldots,C-1$ and $j=1,\ldots,N$. We assign a representative emotion to each tweet by finding $argmax_{c \in C} X_{c,i}$ for $i=1,\ldots,N$, resulting into $X^{rep} \in R^{1 \times N}$, which records representative emotions of all tweets.

Combining Topic and Sentiment

We merge the detected topics $H^{rep} \in R^{1 \times N}$ and identified emotions $X^{rep} \in R^{1 \times N}$ using the unique ids of tweets, resulting into a matrix $O \in R^{2 \times N}$. By referring to the timestamp of each tweet, we are able to track the changes in sentiment and topics over time to see how the public responded to the different vaccines as time passed.

Results

Tracking Topic Over Time

We started by generating 50 topics (*K*=50) using the ONMF algorithm with 2,000 as batch size (*s*=2000). In order to only retain the representative topics about vaccines, we calculate the ratio of each topic *k* using the following equation:

$$\frac{\sum_{i \in N} 1_{\{H_i^{rep}=k\}}}{N}$$

With the topic ratio, we can estimate how many tweets belong to topic *k* and filtered out 38 insignificant topics whose topic ratios below the average, i.e., 2%. As listed in Table 1, the remaining 12 topics were then labeled by reviewing the most contributed keywords in each topic.

Table 1. The most significant 12 vaccine-related topics and the percent of tweets in each topic.

Topic ID	Topic Label	Topic Ratio
1	Vaccination of Front-line Workers	8.68%

2	Access to Vaccines - Signing Up Online	8.28%
3	South African Variant	6.79%
4	Biden Stimulus Plan	3.68%
5	mRNA Vaccines	3.15%
6	Complaints about pharmaceutical company profits	3.07%
7	Vaccine Conspiracy Theories online	2.93%
8	Trials in non mRNA vaccines	2.54%
9	Vaccine distribution in Canada	2.50%
10	Supply and Herd Immunity	2.45%
11	Genetic Concerns about Vaccines and Kids	2.19%
12	Low Distribution of AstraZeneca Vaccine	2.06%

Fig. 2 shows the trends for the six most important topics whose topic ratios are greater than 3%. The most important topic discusses the vaccination of front-line workers (topic 1), whose topic ratio stayed above 7% from mid-December to mid-February. Such a high attention of topic 1 indicates that people concerned about the eligibility of vaccination and relevant plans from governments, especially in the early roll-out phases (i.e., phase 1a and phase 1b). A discussion peak is observed on December 20, 2020, and December 21, 2020, as shown in Fig. 3 because some congress members got vaccines before front line workers, which trigger heated debates. The representative tweets of topic 1 during that period are:

- Speakers: Finding eligible #candidates for #COVID19Vaccine have to be ensured (12/20/2020)
- What makes Blumenthal and Murphy eligible for the vaccine. Are they front line workers? (12/20/2020)
- They are depriving front line workers of a vaccine. They are literally scum (12/20/2020)

The above tweets reported that the priority of accepting COVID-19 vaccines and justice are also critical concerns of people. The second-largest topic is about access to vaccines - signing up online (topic 2). After the early distribution of vaccines, we observed that people started to be concerned about the access to the vaccines, resulting in growth starting from the last week of 2020. The following relevant tweets of topic 2 indicate that governments and healthcare facilities began implementing online appointment for vaccination.

- Heads up Ottawa County-you can sign up for vaccine notifications online (01/05/2021)}
- @drharshvardhan Please implement Aadhaar based online appointment for Covid vaccine as applicable in case of appointment for passport and driving license (01/05/2021)}
- “A step-by-step guide for the online vaccine appointment process" cited from wenatcheeworld [50]

The third-largest topic is about the South African variant (topic 3), whose peak in late December is relevant to the announcement of South African variant from South African health officials [51], and the first variant case detected in US [52], resulting in a rising trend from the late January of 2021. The high ratio of topic 3 indicates the effectiveness of released vaccines is of great concern, and people are skeptical and conservative. Finally, comparing to the top 3 significant topics, topic 4-6 (i.e., Biden Stimulus Plan, mRNA Vaccines, and Complaints about pharm company profits) showed relatively steady discussion trends.

Figure 2. The topic trends for the most significant 6 topics that have a topic ratio above 3%.

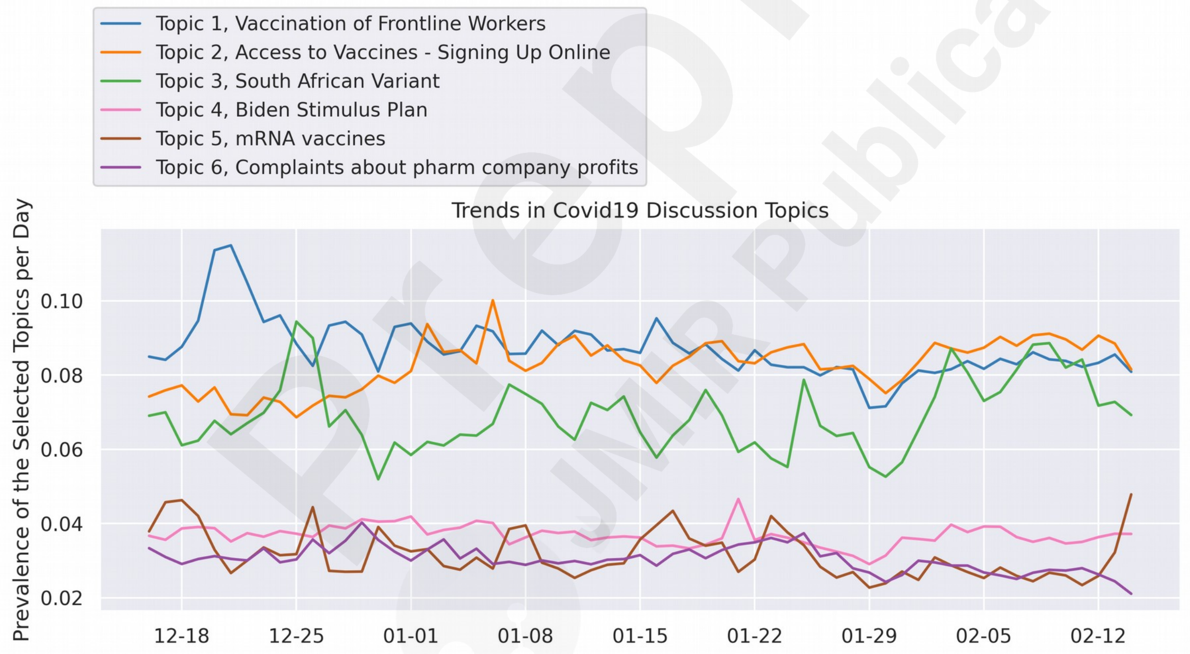


Figure 3. The daily and weekly trend of topic 1.

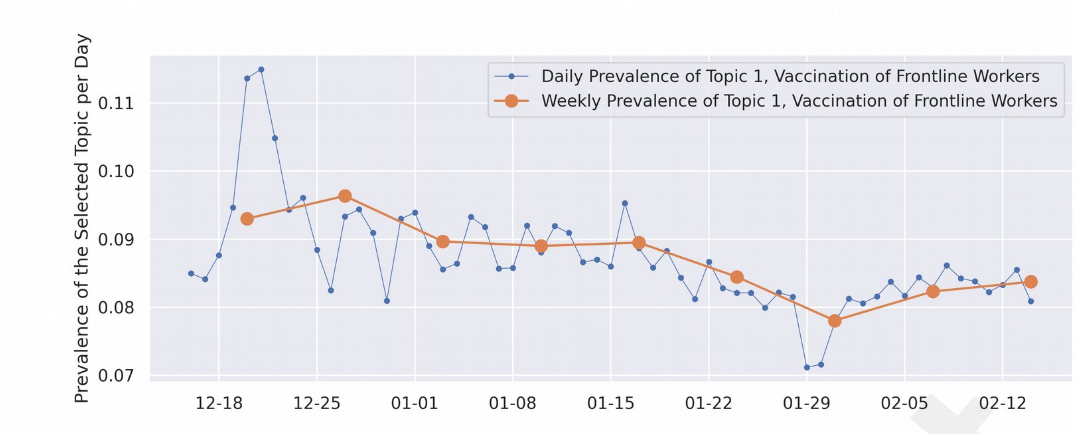


Figure 4. The topic trends for the rest of topics that have a topic ratio below 3%.

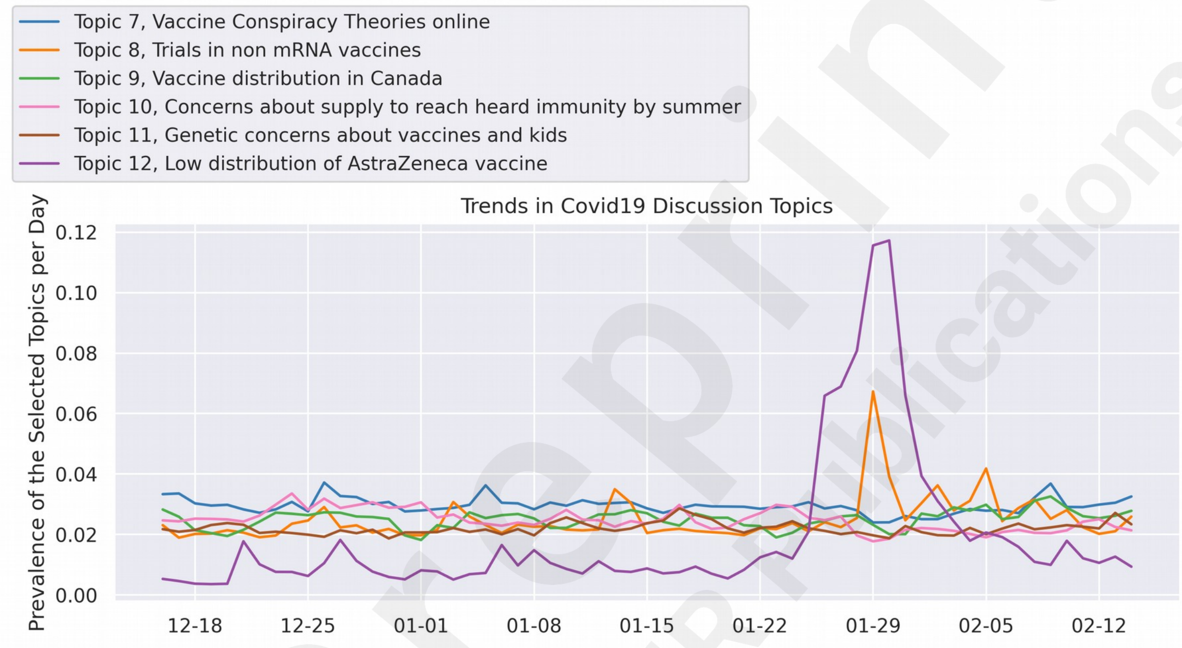


Fig. 4 presents the remaining six important topics. Topic 8 (Trials in non-mRNA vaccines) and topic 12 (Low distribution of AstraZeneca vaccine) had apparent spikes on January 29, 2021. For the peak of topic 8 (see Fig. 5), we found that the emerging event "the positive trail results of Johnson & Johnson's single-shot vaccine" caught the public's eye and stimulated discussion. The relevant content are tweeted frequently at that moment, and most of them cited news sources. The sample tweets are as follows:

- "Single-shot Johnson & Johnson vaccine 66 percent effective against moderate and severe illness" cited from washingtonpost [52] (01-29-2021)
- "Johnson & Johnson says its single-shot vaccine is 66% effective overall at preventing moderate to severe illness" cited from fox8live [53] (01-29-2021)
- "Johnson & Johnson's one-shot #COVID19 vaccine is effective against severe disease" cited from sciencenews [54] (01-29-2021)

The spike on Topic 12 (see Fig. 6) can be related to the dispute between EU and Astra Zeneca in the third week of January [55]. The citizen in the EU expressed their depression about the delay and inefficiency of vaccine ordering, and the representative tweets are as follows:

- EU vaccine delays prompt press frustration (01-28-2021)
- AstraZeneca is supplying EU vaccine at cost with zero profit. EU has a cheek to talk about suing AZ (01-29-2021)
- The actions of the EU to cover their abject failure to obtain vaccine (01-29-2021)

Figure 5. The daily and weekly trend of topic 8

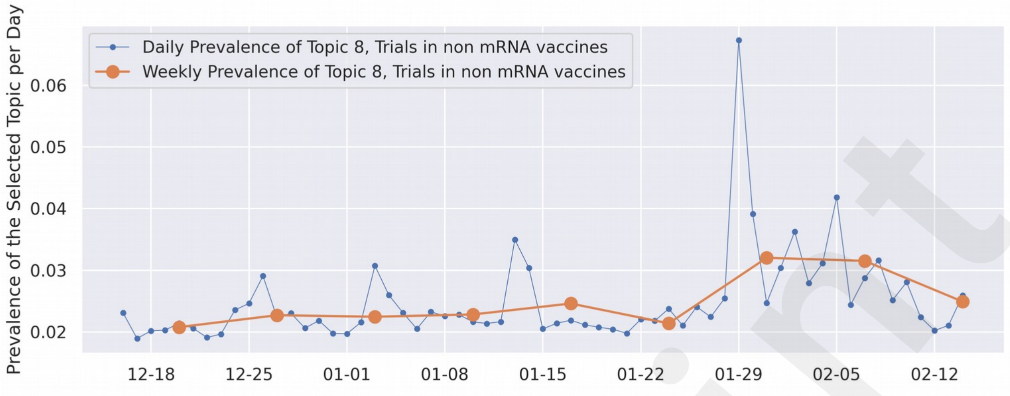
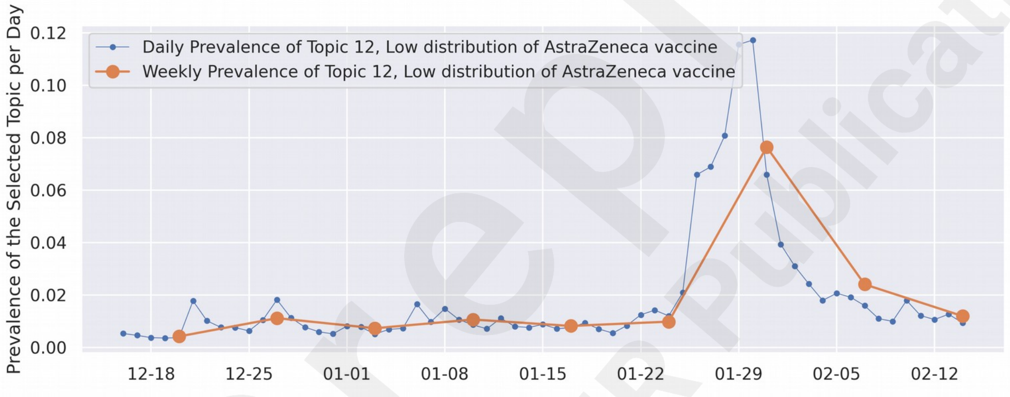


Figure 6. The daily and weekly trend of topic 12



Tracking Sentiment Over Time

When summarizing the sentiment in all 7,864,640 million tweets throughout the entire period, we observe that the top emotion that appears in our tweets is fear followed by joy. The percent of tweets containing each of the emotions from the tweets collected during the entire period is described in table 2.

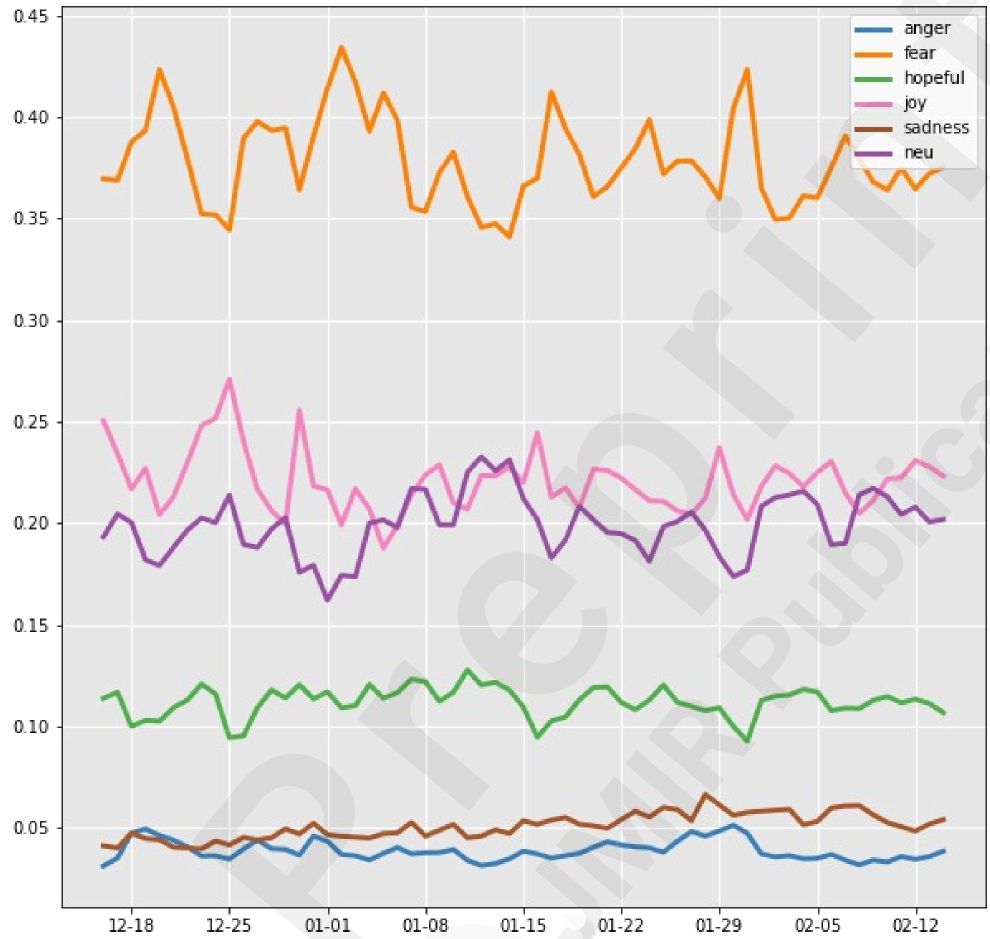
Table 2. Percent of Tweets by Emotion

Sentiment	**Emotion**	**Tweet Proportion (%)**
Negative	Fear	37.87
	Sadness	5.1
	Anger	3.93
Neutral		19.9
Positive	Joy	21.97

	Hopefulness	11.24

Figure 7 presents the trends of the 5 emotions during the 60 day period starting December 16th , 2020,. It shows that fear is consistently the most frequently detected emotion. Joy is the second most common emotion followed by neutral sentiment. Hopefulness, sadness, and anger have a lower proportion of tweets. The Augmented Dickey-Fuller test shows that all emotions except for sadness are stationary throughout the entire period, while sadness increases throughout the period.

Figure 7. Emotion trend over time



Sentiment Trends in the 12 Detected Topics

To analyze the sentiment of each of the top 12 topics, we plot the percent of each sentiment for each topic and observe how the percent changes over time. The percent of tweets in each sentiment is described in figure 8.

Negative Sentiment

Negative sentiment is the leading sentiment in our tweets with fear as the leading emotion.

Fear

Our graphs show that for the majority of topics fear is the top observed emotion. In topics 1, 2, 4, 6, 7, 10 and 11, it is the top observed emotion throughout the majority of the time period.

Topic 1 discusses the vaccination of frontline workers. Some representative tweets from this topic that contain fear are:

- @POTUS Mr. President, Im really worried about my state (GA) and the rollout with vaccines. There doesn’t seem to be a plan and we are being pushed to have school and teachers are not vaccinated and barely hospital workers and senior citizens have. (02/13/2021)
- @CTVNews Hi, I am an Ontario resident and my wife works at X-ray & Ultrasound clinic in Newmarket. I am worried about her and her Assosciates not getting the vaccine along with hospital workers, she sees patients everyday and I think they must be vaccinated ASAP. Thanks, Charlie (01/12/2021)

The main theme in these tweets is a fear that front-line workers would not be vaccinated soon enough and that they would not receive the highest priority in the vaccine rollout.

Topic 2 discusses access to vaccines and signing up online. The most prominent emotion in this topic throughout the period is fear. Below are some example tweets from this topic:

- I got vaccinated. I'm Latino. Making my appt was confusing and my 2nd appt kept getting cancelled even though I work in a hospital. Also lots of fear, distrust and misinformation, people saying the vaccine gives you the 666 sign of the devil, etc. Many people are scared of it. https://t.co/98pguyfuiJ (01/31/2021)
- I’m very concerned my 82-year-old mother must go online to a website; register for the vaccine in Nevada that is still not available until February 28? How do we solve this for our older generation with no computer knowledge to help them get vaccines quicker?🙏♡ (01/06/2021)

We can identify a struggle to obtain an appointment for vaccinations in many states. There were also technical difficulties with multiple websites that caused concern among many Twitter users.

Topic 4 discusses Biden's stimulus plan. The plan contains funding for COVID-19 vaccine distribution [56] . In January, the gap between fear and joy widens; however, after Biden takes office in January, joy increases and the gap between fear and joy becomes smaller.

Examples of tweets from topic 4 that convey fear are:

- @GovInslee I’m a fan Jay, but I’m worried Washington is going to screw up the vaccine distribution. (01/13/2021)
- @JoeBiden Please save Texas from @GovAbbott ‘s ignorance and massive logistical failures with respect to distribution of the vaccine (01/17/2021)

Many of the tweets in this topic convey fear with respect to not executing Biden's plan rather than fear of the plan itself.

Anger

While fear is the most prominent emotion followed by joy, some topics contain spikes of anger related tweets. Topic 5 contains a few spikes of anger.

Here are some examples of angry tweets from topic 5:

- Coronavirus: EU anger over reduced Pfizer vaccine deliveries. Why to rely on profiteering Pfizer ? There are other vaccine ! 😒 https://t.co/E27tWB71IJ (01/15/2021)
- @latimes Is that why its killing old people? 20+ dead in Norway alone. Global scientists calling for immediate stoppage of Pfizer drug. Btw it's not a vaccine by definition. Its mrna therapy. A vaccine uses a dead virus that's incubated and cultured. (01/16/2021)

There is anger due to lack of trust of the vaccine manufacturers as well as anger over rumors of deaths and injuries due to the vaccines.

Sadness

Sadness is one of the least prominent emotions in our data. It is highest in topic 10 which discusses concerns about vaccine supply that would enable reaching herd immunity by summer 2021. Here are some representative tweets from this topic containing sadness:

- My dad was so close to getting his vaccine. But he didn't make it" Meredith pays tribute to her father who died 4 days ago with Covid19. He was a Cumbrian farmer. She describes him as grumpy but in a charming way. https://t.co/OR5NsNVuZG (01/13/2021)
- @SHCGreen @NicolaSturgeon @jasonleitch @edinburghpaper @lothianlmc @NHS_Lothian @DrGregorSmith Glad to see some people getting the vaccine. Sadly my aunt didn't get to have hers. Died early hours from covid. Will miss her very much.

Many of the tweets in this topic containing sad emotion discuss deaths due to COVID-19 that could have been prevented by a quicker vaccine rollout.

Additionally, we can see below tweets from topic 1 containing sadness:

- Some of these are so painful. 65-year-old local pharmacist, kept working, hence couldn't social-distance like, well, a writer. Dead as a consequence. Why front-line workers should be further up in the vaccine queue than even 78-year-olds like me. https://t.co/EYZ6uUr5K8 (01/21/2021)
- An extended family member was a carer in a home, no vaccine, was in a coma for 2 weeks and passed last week. I didn't personally know her but her niece is heartbroken. Thought all care home staff had the vaccine according to the Govnt. (02/03/2021)

These tweets contain sadness and concern that front-line workers will not be vaccinated soon enough and might contract COVID-19.

Neutral Sentiment

Many neutral tweets contained information from news websites or from official sources. As a result, we observed that many of these tweets contained links or media. Neutral sentiment is not the leading emotion in any of the topics, however, we still detected many neutral tweets in all topics. Below are tweets from different topics containing neutral sentiment from the top 6 topics:

- Topic 1:

    - Westminster residents ages 65 and older are now eligible to receive the COVID-19 vaccine. Read the full press release below for instructions. #westminsterca #covidvaccine #orangecounty https://t.co/7cgiOLQLl5 (01/13/2021)
    - After 40 hours of work, the volunteers of Broadbent Arena, in Louisville, Ky., are eligible for their own vaccines. Every day, the oldest volunteers with 40 hours under their belts get the leftover doses. https://t.co/tB3NY2ECSE (02/04/2021)
- Topic 2:
    - #Healthcareworkers, anyone 70 years and older, and state/local government employees and contractors who perform #COVID_19 vaccinations and testing in SC can make appointments to get a #vaccine. https://t.co/65iyk1qJWi (01/15/2021)
    - “The fastest way to register into this system will be online,” WV rolling out new vaccine registration system https://t.co/gTzl9s54vq (01/22/2021)
- Topic 3:
    - Virus Updates: S. Africa Halts AstraZeneca Shot; COVID Reinfections May Be Overlooked https://t.co/VRvgEd0DDV (02/08/2021)
    - Moderna says it's working on Covid booster shot for variant in South Africa, says current vaccine provides some protection https://t.co/UQLInvRVoO (01/25/2021)
- Topic 4:
    - Covid-19 vaccine distribution ramps up for 20 million to be immunized by the start of the new year https://t.co/zWUVzjxTNw (12/21/2020)
    - The $900 billion stimulus package includes unemployment support of up to $300 per week. The bill also includes $45 billion in support for transportation, $82 billion for schools, $20 billion for coronavirus vaccine distribution and $25 billion in emergency assistance to renters. (12/20/2020)
- Topic 5:
    - "Sir Ian McKellen says he feels ‘euphoric’ after receiving the Pfizer/BioNTech vaccine" https://t.co/Jr4XvRUDlH (12/17/2020)
    - The @nytimes reported Pfizer announced that they will ship fewer vials of their coronavirus #vaccine to the US, in response to the FDA approving a change to the label saying the vials contain six doses rather than five: https://t.co/w8pmbwWBoB (01/25/2021)
- Topic 6:
    - Column: Pfizer, Moderna expect billions in profits from COVID vaccines. That's a scandal https://t.co/LIhZT0uTlB (01/04/2021)
    - The pharmaceutical company expects around $15 billion of revenue from sales of its Covid-19 vaccine this year, while Wall Street had anticipated $12.7 billion. https://t.co/KkjT4vur1d (02/02/2021)

As we can see, in all topics, there are a multitude of articles and opinion pieces from different media outlets. The articles typically follow the theme in the topic to which they were classified.

Positive Sentiment

Positive sentiment is the second most common in our data and contains two emotions: joy and hopefulness.

Joy

In topics 3, 5, 8, and 9, the leading emotion fluctuates throughout the time period. While joy is not the leading topic throughout the entire period, in these few topics, the expression of joy exceeds fear for at least some days during the period.

Topic 5 discusses mRNA vaccines. The vaccines discussed in this topic are only the Pfizer and Moderna vaccines since they were given emergency use authorization for use at the time of data collection.

Examples of tweets from topic 5 that contain joy are:

- Congratulations! Still wear your mask and wash those hands, keep yourself safe! 💞 I get my second one tomorrow. Moderna or Pfizer? I got the Pfizer, people I know who have gotten their second dose are having a rough couple days. Molly must be so happy! (02/06/2021)
- Pfizer and Moderna seem to be the clear vaccine winners (1/29/2021)
- Wow vaccine is looking awesome. I'm super impressed with moderna and pfizer-- and in record time:) (2/13/2021)

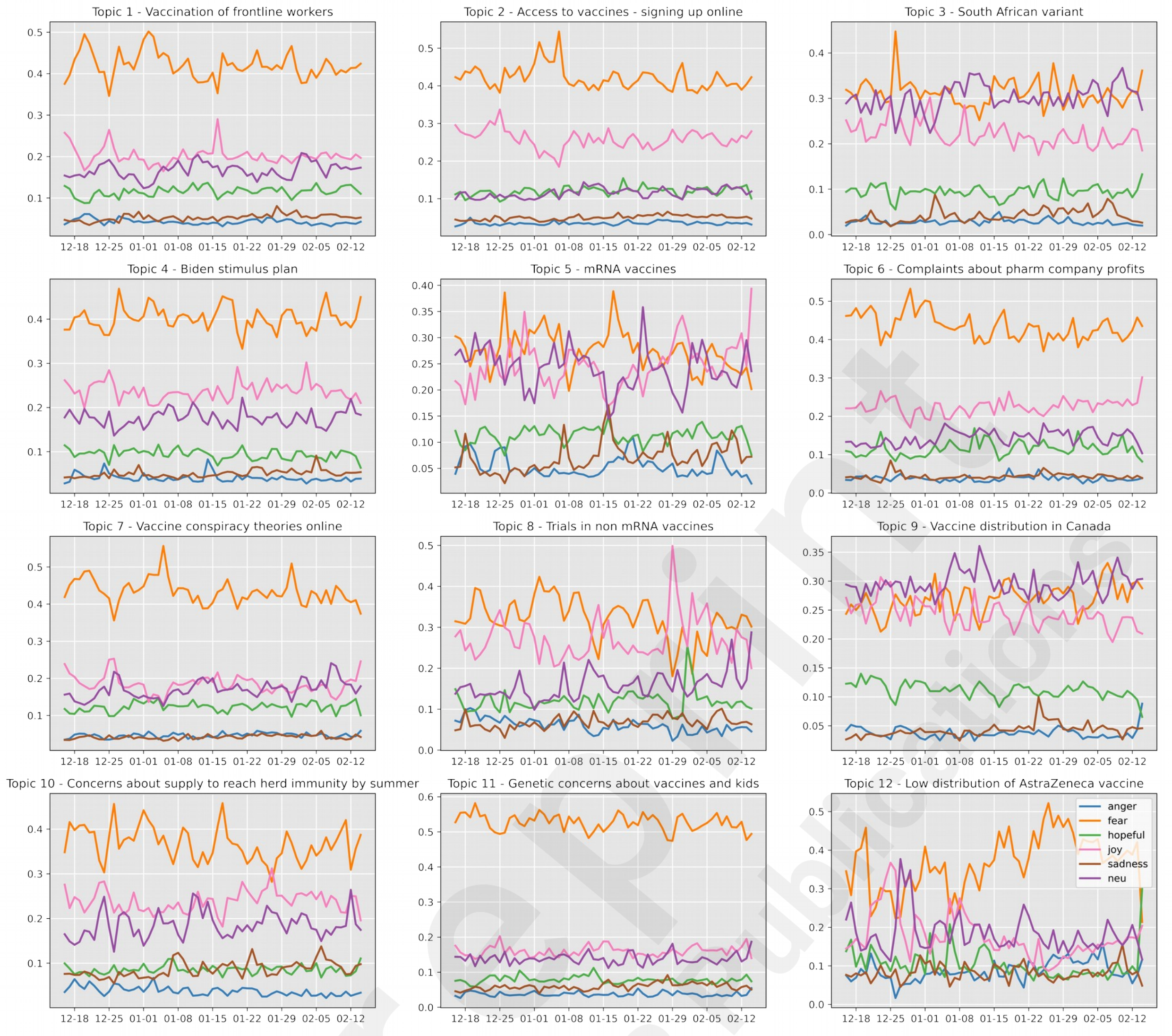
Topic 8 discusses trials in non-mRNA vaccines. While there are many days where fear is the top emotion in this topic, joy is a prominent emotion in the tweets discussing this topic since it is the leading emotion in some days during the time period.

Below are examples of tweets containing joy from topic 8:

- Waking up to great news on the Covid vaccines front: Novavax 89% efficacy ,Johnson&Johnson single dose, and 100% protected from death 28 days after single shot , Astrazeneca fully approved in EU. #VaccinesSaveLives (01/30/2021)
- I participated in the Janssen/Johnson & Johnson #ENSEMBLE2 Covid19 vaccine trial💉 Only time will tell whether I received vaccine or placebo. But so happy to be taking part. Thanks to all the amazing staff at St. Thomas’ Hospital London 🙌@GSTTnhs #janssen #covıd19 https://t.co/brHCDOJC6u (01/13/2021)

The possibility of having a variety of vaccines that are approved is a cause for joy in many Twitter users.

Figure 8. Emotion trend over time in topics 1 - 12



Hopefulness

Topic 12 contains a spike of hopefulness in late December. This topic discusses concern low distribution of the AstraZeneca vaccine. Below are examples of hopefulness in topic 12:

- Hopefully the Oxford vaccine can help out those countries, not just in EU, who don't have enough vaccines. https://t.co/BrC3dJ71tN (12/21/2020)
- @ChristinaSNP What a smashing day.Sun is shining, a British vaccine for COVID is approved. The EU approved #brexit deal is being flown in at the moment. When signed the @theSNP can surely let us know their plans for our future, not merely criticise others like #NoDealNicola #BetterTogether (12/30/2020)

We can see that there was some hopefulness regarding the distribution of the AstraZeneca vaccine. However, hopefulness was not the leading emotion during that time period. Additionally, by the end of the time period fear was by far the most prominent emotion.

Discussion

Our study aimed to detect the topics and sentiments of public concerns of COVID-19 vaccines using trend analysis on tweets collected for a period of sixty days when the vaccines were started in US and make practical suggestions to address the concerns of different groups in the public as expressed on social media. Approximately 8 million tweets related to COVID-19 vaccines were collected and 12 important topics were selected for analysis. The three most important topics with the highest topic ratio were “Vaccination of Front-line Workers”, “Access of Vaccines – Signing Up Online”, and “South African Variant”. The other topics were mostly related to the concerns about the vaccines as well as their supply and distribution. There were also topics related to the stimulus plan, profits of pharmaceutical companies, and conspiracy theories. Through the trend analysis, it was found that the peaks of the topics were impacted by the events reported in the news and spread through social media. The Sentiment analysis showed that 46.9% of tweets were negative with emotions of mostly fear and followed by sadness and anger. 33.21% of tweets were positive with emotions of joy and hopefulness. 19.9% of tweets were neutral. Fear and joy were the most detected emotions.

Our analysis examined the 6 different sentiments detected in the tweets and their change over time. We observe that the keywords in each topic don't change much over time; therefore, we are able to track our tweets using the same topics throughout the entire period. In some topics, sentiment is stationary throughout the period, while in others there are significant trends. For example, in topic 3 - South African variant, we see an increase in fear and neutral sentiment over the period and a decrease in joy at the same time. Similarly, we see an increase in fear and a decrease in joy in topic 12 - Low Distribution of the AstraZeneca Vaccine. Overall, fear is the top emotion followed by joy. Sadness and Hopefulness remain low in most topics throughout the entire period.

Identifying Specific Concerns in Each Topic Using Emotional Content

The most notable conclusion from the data is that the main reaction to the COVID-19 vaccines on social media is fear. However, we can identify every one of the emotions in each topic. In each topic, we can find tweets related to the topic containing each one of the emotions.

By looking at the representative tweets for each topic and each emotion, we are able to learn what specific concerns people may have that may lead to vaccine hesitancy. For example, from topic 1, we find that there is fear surrounding the vaccination of government officials prior to frontline workers. By addressing this publicly and assuring the public that the frontline workers will receive their vaccines as soon as possible, this will help to build public confidence in the vaccine rollout. We can also identify tweets that contain sadness to identify further concerns about the rollout to front-line workers and see Twitter users expressing sadness regarding front-line workers possibly dying due to lack of vaccines. This can be addressed by being more transparent about vaccination timelines or by advocating for more vaccine supply. By being aware of specific concerns as they happen (e.g., the vaccination of front-line workers), we will be better able to address the source of concern and reduce vaccine hesitancy.

Vaccine Administration

The very first dose of mRNA-COVID 19 vaccine by Pfizer and BioNTech was given to a healthcare worker on December 14, 2020. This can explain why the most significant topic at the start of the study is vaccination of front-line workers (topic 1). As more vaccines were administered, reports of anaphylaxis began to surface, especially with the Moderna vaccine[57]. In US alone, ten cases of anaphylaxis were reported after 4,041,496 (0.002%) were given between December 21, 2020, and January 10, 2021. This created fear as indicated in the trend and has dominated all other emotions throughout the course of the study period. It will be interesting to find out how many of these tweets are from healthcare personnel verses the general public. According to CDC recommendation, both healthcare personnel and residents of long-term care facilities were first to offered the COVID vaccine [58]. Healthcare personnel include both clinical and non-clinical staffs like those who work in food, environmental and administrative services. It can assume that clinical staff have adequate knowledge of vaccines not to be afraid of it. Therefore, public health authorities and healthcare systems can focus on educating the adverse effects of the vaccine to the non-clinical staff and the general public. For example, anaphylactic reactions occur mostly in people who have a similar reaction to other food and drugs, and it usually occurs within minutes after injection. Better understanding of the adverse effects will minimize fear of the vaccine and thus reduce vaccine hesitancy.

Access to Vaccines

Signing up online (topic 2), Vaccine distribution in Canada (topic 9), and Low distribution of AstraZeneca Vaccine (topic 12) can all categorized as accessibility of vaccines. A good amount of positive emotion all through the study period in topic 2 indicates that there is a sense a hope in the midst of daily raising COVID cases. There is still a large amount of fear in this topic. It maybe the fear of unable to obtain an appointment for the vaccine. Unlike the US, Canada does not have her own domestic manufacturers to produce vaccines. As a result, Canada is relying on international vaccine manufacturers. Advance purchase contract was signed but there was no specific date for delivery except “first quarter of 2021”. There was a shortage of supply of vaccines in Canada. It pushes the Canadian government to prioritize giving the first dose to the population first and the second dose to 16 weeks later [59] as supposed to 3 week or 4 weeks later. Across to the other continent, the European Union (EU) was furious when in early January AstraZeneca announced that there would be 60% fewer doses of vaccines than it had promised to deliver in the first quarter of 2021. The spikes of fear and anger emotions during this period in topic 12 are the direct reflection of this news. Being able to have access to the vaccines is important once COVID-19 vaccines are authorized for emergency use. Therefore, public health authorities must have plans to work with vaccine manufacturers to manufacture and deliver the vaccines in a timely manner. The transparency of the access information from the autoreactive units is helpful to reduce the fear and anger in the public.

Practical Implications

In December 2020, the WHO released a safety surveillance manual for COVID-19 vaccines. This manual addressed a number of topics with regards to vaccine administration including how to communicate information regarding the vaccine on social media [60]. Among other points, the report offers proposes to listen proactively and craft tailored messages to different audiences and address specific concerns of different groups. Using this research, we can take the WHO recommendations to provide more specific advice to clinicians and policy makers. To address specific concerns, we divide the 12 topics into 3 groups: favoring vaccines, vaccine hesitant, and vaccine opposed.

Favoring Vaccines

The topics that lean towards those who favor vaccines are topic 1 – vaccination of frontline workers, topic 2 – access to vaccines – signing up online, topic 9 – vaccine distribution in Canada, Topic 10 – Concerns about supply to reach herd immunity by summer, and topic 12 – Low distribution of AstraZeneca vaccine. While these topics also produce negative feelings of fear, anger, and sadness, these negative feelings are regarding concern about not having enough vaccines or not having access to vaccines fast enough. It is crucial to monitor topics that contain tweets from individuals who do want to get vaccinated and keep them informed. Here are some examples of tweets that convey fear or concern by individuals who want to get vaccinated:

“Anybody know what's going on with BAT 24-hour appts? Are they fully back up and running again after being shut down for lack of vaccine? My second shot is at 2:45 a.m. next week, and I'm wary of getting up in the middle of the night to go down there to find them closed.”

“To be honest, I'd rather risk my life / keep myself in lockdown, for younger key workers to have the vaccine. They are the ones keeping the country going after all.”

“Blocking access to a vaccine that could save my life is, oh I don’t know, attempted murder? So is exhaling their COVID breath around me, but the former is active and so much more egregious. Ain’t nobody got time for that mess.”

Identifying the topics that vaccine favoring individuals will discuss is crucial to reducing their concern. In accordance with the WHO document, communication on vaccine availability should be active and frequent. An example of using the analysis from this study to inform the public is looking at the visualizations in real time to produce the right messaging on social media. In figure 3, we observed a spike in the volume of topic 1 in the weeks of December 18th. When looking at figure 8, we see that the leading emotion for that week and topic is fear, not only that, but we see a spike in fear during that week for topic 1. Therefore, it is crucial to post messages on social media that week that address the public fear that healthcare workers will not have adequate access to vaccines. Another key component in keeping the public informed is updating official websites with vaccine information very frequently. During the early days of vaccination, there was a lack of information in many states about the timeline of vaccination for each risk group. Providing more information on the rollout schedule would help ease the concern of individuals in this group. It is crucial to look at the tweets that convey fear and anger in these topics to create the right messaging and address points that concern this group of the public.

Vaccine Hesitant

This group of individuals is the most crucial to reach since they can be persuaded to get vaccinated. Topics that discuss vaccine hesitancy are topic 3 – South African Variant, topic 5 – mRNA vaccines, topic 8 - Trials in non mRNA vaccines, and topic 11 - Genetic concerns about vaccines and kids. Below are examples of tweets of the vaccine hesitant from these topics:

“Just keep in mind that some small percentage of those who received the vaccine did not develop immunity, during the clinical trials. And its effectiveness against variant strains is still not fully known.”

“The fact that 3 vaccines all appeared to show lowered effectiveness against the variant from South Africa is not encouraging, and the results Novavax announced Thurs were the 1st to occur outside of a lab, testing how well a vaccine worked in people infected with a new variant.”

“’There were obviously several people in the United Kingdom who had had a severe allergic reaction to this vaccine and had a history of severe allergic reaction,’ said Offit Several people!!!!! #vaccine”

Like the vaccine favorable group, we should also target this group with facts and do so often. However, with this group we should focus on messages that can be detected in these topics such as related to side effects of the vaccine, the efficacy of different types of the vaccine for the original strain of COVID-19 as well as for variants, as well as why you can still contract COVID-19 even after being vaccinated. We can craft helpful messaging for this group by looking at the topic and emotion data for these topics. For example, we see an increase in the volume of topic 3 – South African variant towards the end of January. The most prominent emotion for that topic during that time is fear. Therefore, we can craft messaging on social media regarding the variant that will help with this fear. As the WHO recommends, we should mainly focus on facts and provide up to date information to the public through social media regarding the variant.

Vaccine Opposed

This group is the least likely to be persuaded by messaging on the vaccine but should not be ignored. This is because they produce messaging on social media that may convince others. Therefore, we should attempt to counter their messaging with up to date and correct information. Topics that contain a large number of tweets from individuals that are vaccine opposed are topic 6 - Complaints about pharm company profits but we can find a small number of tweets from this group in all topics, particularly in tweets that have been labeled angry or fearful. Examples of tweets from this group are:

“We have been here before with the Nazis and Thalidomide yet the whole world rushes to take an untested vaccine. People are dying after having the vaccine yet no enquiries into what happened just a rapid cremation and silence. We should all be very worried”

“I bind you up satan in the name of Jesus, no weapon formed against us shall prosper, and I mean this vaccine is satan here.. “Mark of the beast” read your bibles people..”

“He didn’t take the vaccine! He’s a Eugenics partner with Bill Gates they don’t take their own vaccines! How about some proof! He’s just trying to coverup the ill side effects and deaths that are already happening!”

Those who are opposed to vaccines are hard to persuade, but we must spread truthful messaging to counteract the messaging that they spread. Many of the tweets by these individuals don’t even discuss concerns that can be addressed but are more about vaccine refusal and the freedom to refuse vaccines. It is important to amplify stories of those who suffered severe consequences by refusing to take the vaccine. This is mostly for the sake of the vaccine hesitant rather than the vaccine opposed. An example of messaging can be obtained by looking at the patterns for topic 6. This topic is stable over time and does not experience any spikes. Therefore, we should stay consistent with our messaging over time and counteract any information on this topic with facts on a consistent basis as recommended by the WHO report.

Limitations of This Research

Limitations of Twitter

Twitter is a large social network with 353 million monthly active users [61]. While this is a significant number of users, there is no guarantee that Twitter users are representative of the global or the US population as a whole. Mislove et al. have investigated the ability of Twitter data to represent the US population and have found that areas that are more densely populated tend to be overrepresented in Twitter [62]. Additionally, Gore et al. [63] and Padilla et al. [64] have found geographic bias in their analysis of Twitter data. Both studies have found an overrepresentation of urban areas in the demographic data of Twitter users included in their studies. Given this prior research, we must assume that users from urban areas are overrepresented in this dataset as well.

Keyword Selection

The keywords that were chosen to generate this dataset were selected by the authors. The list of keywords described in the data collection section contains keywords that name the colloquial names for the available vaccines at the time of the study. The list also contains terms such as “vaccine” and “coronavaccine” that were included in order to capture a more general discussion regarding COVID-19 vaccines. The list is not meant to be exhaustive and represents the vaccines publicly available at the start of data collection in December 2020.

Duplicated Tweets

Bots posting on Twitter are well documented phenomenon [65–67]. One of the issues our study faced was the duplication of content due to bot activity on the topic of vaccines. Other research has documented bot activity on COVID-19 and COVID-19 vaccine misinformation as well [65,68,69]. The main issue this may cause in our analysis is that bot activity may overinflate the importance of certain topics. To combat this, we deduped the Twitter data as part of our analysis and reduced the number of tweets from approximately 20 million to approximately 8 million tweets.

Conclusion

We use topic detection and sentiment analysis as social media trend analysis to better understand the discourse on COVID-19 vaccines tweets. Using this methodology, we can identify the trending topics that reflect the public concerns on COVID-19 vaccines and their responses to the topics indicated by the polarity and emotions on the sentiments. We find that the administration and access to vaccine are some of the major concerns. While most of the information are received from the Internet, they are not directly coming from the autoreactive health organization. Misinformation may cause negative emotions. In some cases, conspiracy spreading in social media may cause substantial amount of fear. The findings in social media trend analysis are helpful for the health organizations to develop strategies for better communication to the target groups and assist them in coping with their concerns that cause negative emotions or vaccine hesitancy. Disseminating the accurate information of COVID-19 vaccines will reduce the negative emotion caused by misinformation or rumors. A report on COVID-19 vaccines by the WHO suggested careful examination of social media to detect specific concerns regarding the vaccines [60]. By understanding what drives different emotions regarding the vaccines, tailored and targeted communication can be developed to provide authoritative health information, which will be helpful to achieve herd immunity and end the pandemic.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant NSF-1741306 and Grant IIS-1650531. Any opinions and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Oliver SE, Gargano JW, Marin M, Wallace M, Curran KG, Chamberland M, McClung N,

Campos-Outcalt D, Morgan RL, Mbaeyi S, Romero JR, Talbot HK, Lee GM, Bell BP, Dooling K. The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine — United States, December 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 18;69(50):1922–1924. PMID:33332292

2. Midoux P, Pichon C. Lipid-based mRNA vaccine delivery systems. Expert Review of Vaccines 2015 Feb;14(2):221–234. [doi: 10.1586/14760584.2015.986104]

3. Schlake T, Thess A, Fotin-Mleczek M, Kallen K-J. Developing mRNA-vaccine technologies. RNA Biology 2012 Nov;9(11):1319–1330. [doi: 10.4161/rna.22269]

4. Pardi N, Hogan MJ, Weissman D. Recent advances in mRNA vaccine technology. Current Opinion in Immunology 2020 Aug;65:14–20. [doi: 10.1016/j.coi.2020.01.008]

5. Vergara RJD, Sarmiento PJD, Lagman JDN. Building public trust: a response to COVID-19 vaccine hesitancy predicament. J Public Health (Oxf) [Internet] 2021 Jan 18 [cited 2021 Apr 14]; PMID:33454769

6. Harrison EA, Wu JW. Vaccine confidence in the time of COVID-19. Eur J Epidemiol 2020 Apr 1;35(4):325–330. [doi: 10.1007/s10654-020-00634-3]

7. Chou W-YS, Budenz A. Considering Emotion in COVID-19 Vaccine Communication: Addressing Vaccine Hesitancy and Fostering Vaccine Confidence. Health Communication 2020 Dec 5;35(14):1718–1722. [doi: 10.1080/10410236.2020.1838096]

8. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206. [doi: 10.1136/bmjgh-2020-004206]

9. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F, Scala A. The COVID-19 social media infodemic. Sci Rep 2020 Dec;10(1):16598. [doi: 10.1038/s41598-020-73510-5]

10. Chang C-H, Monselise M, Yang CC. What Are People Concerned About During the Pandemic? Detecting Evolving Topics about COVID-19 from Twitter. J Healthc Inform Res 2021 Mar 1;5(1):70–97. [doi: 10.1007/s41666-020-00083-3]

11. Prabowo R, Thelwall M. Sentiment analysis: A combined approach. Journal of Informetrics 2013 Mar 4;143–157. [doi: 10.1016/j.joi.2009.01.003]

12. Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA. Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLOS ONE Public Library of Science; 2017 Feb 23;12(2):e0171649. [doi: 10.1371/journal.pone.0171649]

13. Tan LI, Phang WS, Chin KO, Patricia A. Rule-Based Sentiment Analysis for Financial News. 2015 IEEE International Conference on Systems, Man, and Cybernetics 2015. p. 1601–1606. [doi: 10.1109/SMC.2015.283]

14. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. ICWSM [Internet] 2014 May 16 [cited 2021 May 17];8(1). Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

15. Ahmad M, Aftab S, Ali I. Sentiment Analysis of Tweets using SVM. IJCA 2017 Nov

15;177(5):25–29. [doi: 10.5120/ijca2017915758]

16. Ahmad M, Aftab S, Salman M, Hameed N, Ali I, Nawaz Z. SVM Optimization for Sentiment Analysis. ijacsa [Internet] 2018 [cited 2021 May 4];9(4). [doi: 10.14569/IJACSA.2018.090455]

17. Fauzi MA. Random Forest Approach for Sentiment Analysis in Indonesian Language. undefined [Internet] 2018 [cited 2021 May 21]; Available from: /paper/Random-Forest-Approach-for-Sentiment-Analysis-in-Fauzi/a6decd239a552c62bfdadd0616dacc52f3e61d65

18. Tan S, Cheng X, Wang Y, Xu H. Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. In: Boughanem M, Berrut C, Mothe J, Soule-Dupuy C, editors. Advances in Information Retrieval Berlin, Heidelberg: Springer; 2009. p. 337–349. [doi: 10.1007/978-3-642-00958-7_31]

19. Dey L, Chakraborty S, Biswas A, Bose B, Tiwari S. Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier. International Journal of Information Engineering and Electronic Business 2016 Jul 8;8:54–62. [doi: 10.5815/ijieeb.2016.04.07]

20. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment Embeddings with Applications to Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering 2016 Feb;28(2):496–509. [doi: 10.1109/TKDE.2015.2489653]

21. Li Q, Shah S, Fang R, Nourbakhsh A, Liu X. Tweet Sentiment Analysis by Incorporating Sentiment-Specific Word Embedding and Weighted Text Features. 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI) 2016. p. 568–571. [doi: 10.1109/WI.2016.0097]

22. Arras L, Montavon G, Müller K-R, Samek W. Explaining Recurrent Neural Network Predictions in Sentiment Analysis. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis [Internet] Copenhagen, Denmark: Association for Computational Linguistics; 2017 [cited 2021 May 21]. p. 159–168. [doi: 10.18653/v1/W17-5221]

23. dos Santos C, Gatti M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers [Internet] Dublin, Ireland: Dublin City University and Association for Computational Linguistics; 2014 [cited 2021 May 21]. p. 69–78. Available from: https://www.aclweb.org/anthology/C14-1008

24. Wang X, Jiang W, Luo Z. Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers [Internet] Osaka, Japan: The COLING 2016 Organizing Committee; 2016 [cited 2021 May 21]. p. 2428–2437. Available from: https://www.aclweb.org/anthology/C16-1229

25. Sentiment Analysis Using Convolutional Neural Network | IEEE Conference Publication | IEEE Xplore [Internet]. [cited 2021 May 21]. Available from: https://ieeexplore-ieee-org.ezproxy2.library.drexel.edu/document/7363395

26. Vashishtha S, Susan S. Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Systems with Applications 2019 Dec;138:112834. [doi:

10.1016/j.eswa.2019.112834]

27. Chen Q, Sokolova M. Unsupervised Sentiment Analysis of Objective Texts. In: Meurs M-J, Rudzicz F, editors. Advances in Artificial Intelligence Cham: Springer International Publishing; 2019. p. 460–465. [doi: 10.1007/978-3-030-18305-9_45]

28. Pandarachalil R, Sendhilkumar S, Mahalakshmi GS. Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach. Cogn Comput 2015 Apr 1;7(2):254–262. [doi: 10.1007/s12559-014-9310-z]

29. Zimbra D, Abbasi A, Zeng D, Chen H. The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. ACM Trans Manage Inf Syst 2018 Sep 5;9(2):1–29. [doi: 10.1145/3185045]

30. Bermingham A, Smeaton AF. Classifying sentiment in microblogs: is brevity an advantage? Proceedings of the 19th ACM international conference on Information and knowledge management [Internet] New York, NY, USA: Association for Computing Machinery; 2010 [cited 2021 May 24]. p. 1833–1836. [doi: 10.1145/1871437.1871741]

31. Hagen M, Potthast M, Büchner M, Stein B. Webis: An Ensemble for Twitter Sentiment Detection. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) [Internet] Denver, Colorado: Association for Computational Linguistics; 2015 [cited 2021 Apr 25]. p. 582–589. [doi: 10.18653/v1/S15-2097]

32. Adwan OY, Al-Tawil M, Huneiti A, Shahin R, Zayed AA, Al-Dibsi R. Twitter Sentiment Analysis Approaches: A Survey. International Journal of Emerging Technologies in Learning (iJET) 2020 Aug 14;15(15):79–93.

33. Abbas AK, Salih AK, Hussein HA, Hussein QM, Abdulwahhab SA. Twitter Sentiment Analysis Using an Ensemble Majority Vote Classifier. Journal of Southwest Jiaotong University 2020;55(1):9. [doi: 10.35741/issn.0258-2724.55.1.9]

34. Wang M, Hu G. A Novel Method for Twitter Sentiment Analysis Based on Attentional-Graph Neural Network. Information Multidisciplinary Digital Publishing Institute; 2020 Feb;11(2):92. [doi: 10.3390/info11020092]

35. Zarisfi Kermani F, Sadeghi F, Eslami E. Solving the twitter sentiment analysis problem based on a machine learning-based approach. Evol Intel 2020 Sep 1;13(3):381–398. [doi: 10.1007/s12065-019-00301-x]

36. Chen Y, Yuan J, You Q, Luo J. Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. Proceedings of the 26th ACM international conference on Multimedia [Internet] Seoul Republic of Korea: ACM; 2018 [cited 2021 Apr 26]. p. 117–125. [doi: 10.1145/3240508.3240533]

37. Lamsal R. Design and analysis of a large-scale COVID-19 tweets dataset. Appl Intell 2021 May 1;51(5):2790–2804. [doi: 10.1007/s10489-020-02029-z]

38. Lamsal R. Coronavirus (COVID-19) Tweets Dataset [Internet]. IEEE; 2020 [cited 2021 Aug 31]. Available from: https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset

39. Tweet lookup introduction [Internet]. [cited 2021 Aug 31]. Available from: https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/introduction

40. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 1;3(null):993–1022.

41. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature Nature Publishing Group; 1999 Oct;401(6755):788–791. [doi: 10.1038/44565]

42. Chen Y, Zhang H, Liu R, Ye Z, Lin J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems 2019 Jan 1;163:1–13. [doi: 10.1016/j.knosys.2018.08.011]

43. Zhao R, Tan VYF. Online Nonnegative Matrix Factorization With Outliers. IEEE Transactions on Signal Processing 2017 Feb;65(3):555–570. [doi: 10.1109/TSP.2016.2620967]

44. Bonta V, Kumaresh N, Janardhan N. A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology 2019;8:6.

45. Elbagir S, Yang J. Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. Proceedings of the International MultiConference of Engineers and Computer Scientists Hong Kong; 2019. p. 5.

46. Rios A, Kavuluru R. Few-Shot and Zero-Shot Multi-Label Learning for Structured Label Spaces. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing [Internet] Brussels, Belgium: Association for Computational Linguistics; 2018 [cited 2021 May 18]. p. 3132–3142. [doi: 10.18653/v1/D18-1352]

47. Biswas S, Annadani Y. Preserving Semantic Relations for Zero-Shot Learning. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition [Internet] Salt Lake City, UT: IEEE; 2018 [cited 2021 May 18]. p. 7603–7612. [doi: 10.1109/CVPR.2018.00793]

48. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet] Minneapolis, Minnesota: Association for Computational Linguistics; 2019 [cited 2021 May 21]. p. 4171–4186. [doi: 10.18653/v1/N19-1423]

49. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv:190810084 [cs] [Internet] 2019 Aug 27 [cited 2021 Apr 30]; Available from: http://arxiv.org/abs/1908.10084

50. manager MH product. A step-by-step guide for the online vaccine appointment process [Internet]. The Wenatchee World. [cited 2021 Aug 31]. Available from: https://www.wenatcheeworld.com/news/coronavirus/a-step-by-step-guide-for-the-online-vaccine-appointment-process/article_25dcefa8-6725-11eb-bbf2-df282a071772.html

51. South Africa announces a new coronavirus variant. - The New York Times [Internet]. [cited 2021 Aug 31]. Available from: https://www.nytimes.com/2020/12/19/world/south-africa-announces-a-new-coronavirus-variant.html

52. JJ vaccine: Single-shot Johnson & Johnson covid vaccine 66 percent effective against moderate and severe illness - The Washington Post [Internet]. [cited 2021 Aug 31]. Available from: https://www.washingtonpost.com/health/2021/01/29/covid-vaccine-johnson-and-johnson/

53. J&J 1-dose shot prevents COVID-19, but less than some others [Internet]. [cited 2021 Aug 31]. Available from: https://web.archive.org/web/20210129132041/https://www.fox8live.com/2021/01/29/jj-dose-shot-prevents-covid-less-than-some-others/

54. One-shot COVID-19 vaccine is effective against severe disease [Internet]. Science News. 2021 [cited 2021 Aug 31]. Available from: https://www.sciencenews.org/article/covid-19-coronavirus-vaccine-johnson-and-johnson-results-variants

55. Covid: EU-AstraZeneca disputed vaccine contract made public - BBC News [Internet]. [cited 2021 Aug 31]. Available from: https://www.bbc.com/news/world-europe-55852698

56. Higgins T. Covid relief bill includes about $8 billion for vaccine distribution [Internet]. CNBC. 2020 [cited 2021 Aug 31]. Available from: https://www.cnbc.com/2020/12/21/covid-relief-bill-includes-about-8-billion-for-vaccine-distribution.html

57. CDCMMWR. Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Pfizer-BioNTech COVID-19 Vaccine — United States, December 14–23, 2020. MMWR Morb Mortal Wkly Rep [Internet] 2021 [cited 2021 May 21];70. [doi: 10.15585/mmwr.mm7002e1]

58. Dooling K. The Advisory Committee on Immunization Practices' Interim Recommendation for Allocating Initial Supplies of COVID-19 Vaccine — United States, 2020. MMWR Morb Mortal Wkly Rep [Internet] 2020 [cited 2021 May 21];69. [doi: 10.15585/mmwr.mm6949e1]

59. COVID-19 vaccination coverage in Canada - Canada.ca [Internet]. [cited 2021 May 21]. Available from: https://health-infobase.canada.ca/covid-19/vaccination-coverage/

60. Covid-19 vaccines: safety surveillance manual [Internet]. [cited 2021 May 25]. Available from: https://www.who.int/vaccine_safety/committee/Module_Communication.pdf?ua=1.

61. Digital Trends 2021 [Internet]. [cited 2021 Aug 30]. Available from: https://www.hootsuite.com/pages/digital-trends-2021

62. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist J. Understanding the Demographics of Twitter Users. Proceedings of the International AAAI Conference on Web and Social Media 2011;5(1):554–557.

63. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLOS ONE Public Library of Science; 2015 Sep 2;10(9):e0133505. [doi: 10.1371/journal.pone.0133505]

64. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. Emmert-Streib F, editor. PLoS ONE 2018 Jun 14;13(6):e0198857. [doi: 10.1371/journal.pone.0198857]

65. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health American Public Health Association; 2018 Oct 1;108(10):1378–

1384. [doi: 10.2105/AJPH.2018.304567]

66. Lee K, Eoff B, Caverlee J. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter. Proceedings of the International AAAI Conference on Web and Social Media 2011;5(1):185–192.

67. Gilani Z, Farahbakhsh R, Tyson G, Wang L, Crowcroft J. Of Bots and Humans (on Twitter). Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 [Internet] New York, NY, USA: Association for Computing Machinery; 2017 [cited 2021 Sep 1]. p. 349–354. [doi: 10.1145/3110025.3110090]

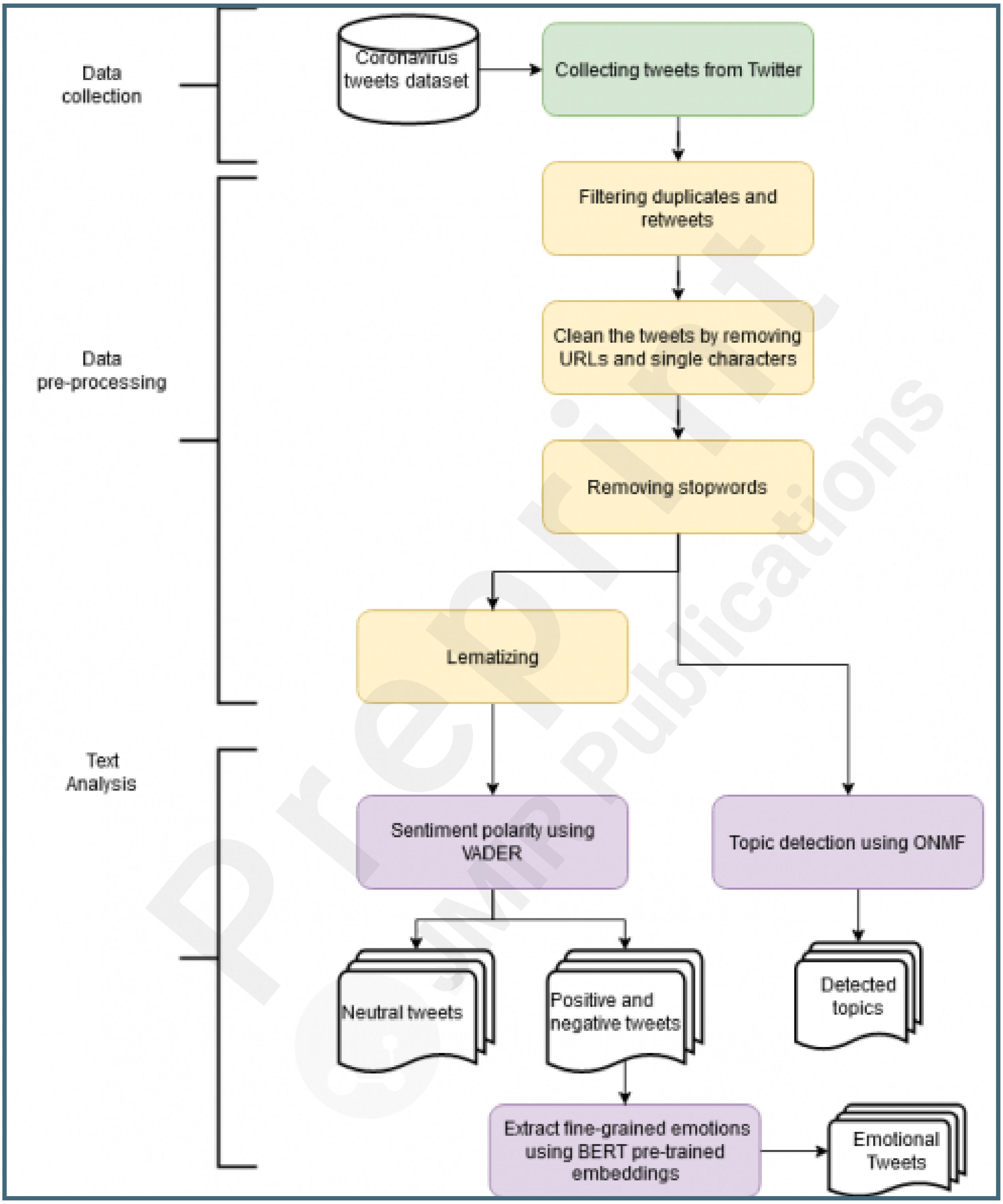
68. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, Rosner M, Mathieu J, Fulk M. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation. Journal of Health Communication Taylor & Francis; 2021 Aug 4;0(0):1–17. PMID:34346288

69. Egli A, Rosati P, Lynn T, Sinclair G. Bad Robot: A Preliminary Exploration of the Prevalence of Automated Software Programmes and Social Bots in the COVID-19 #antivaxx Discourse on Twitter. The International Conference on Digital Society 2021. p. 5.

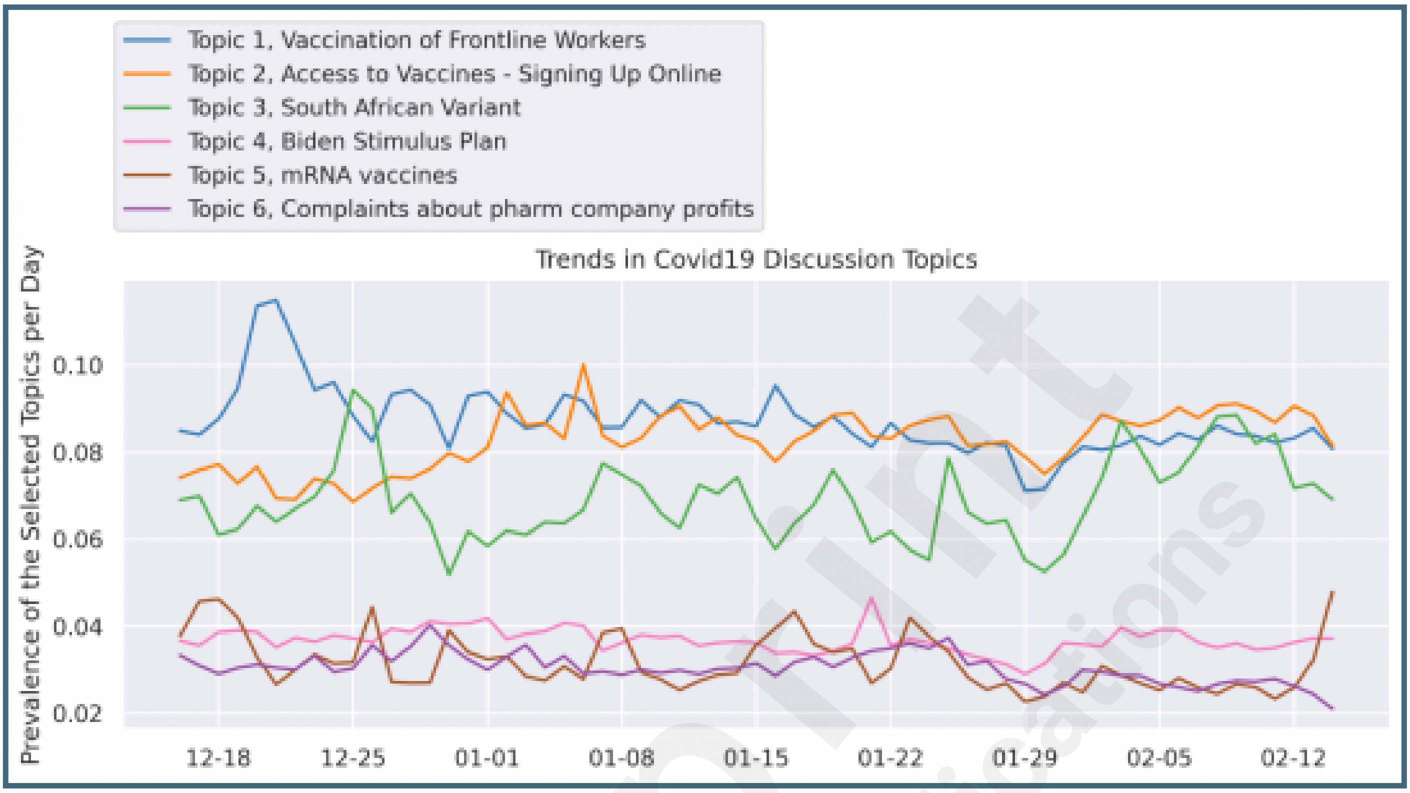
Supplementary Files

Figures

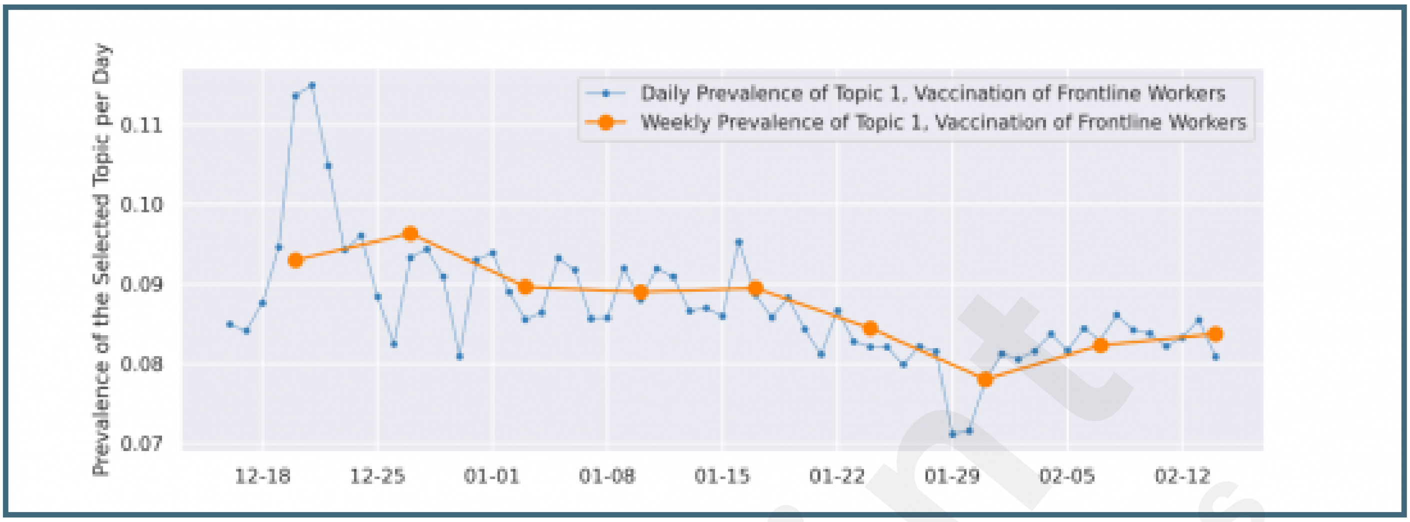
Pipeline of our text analysis.



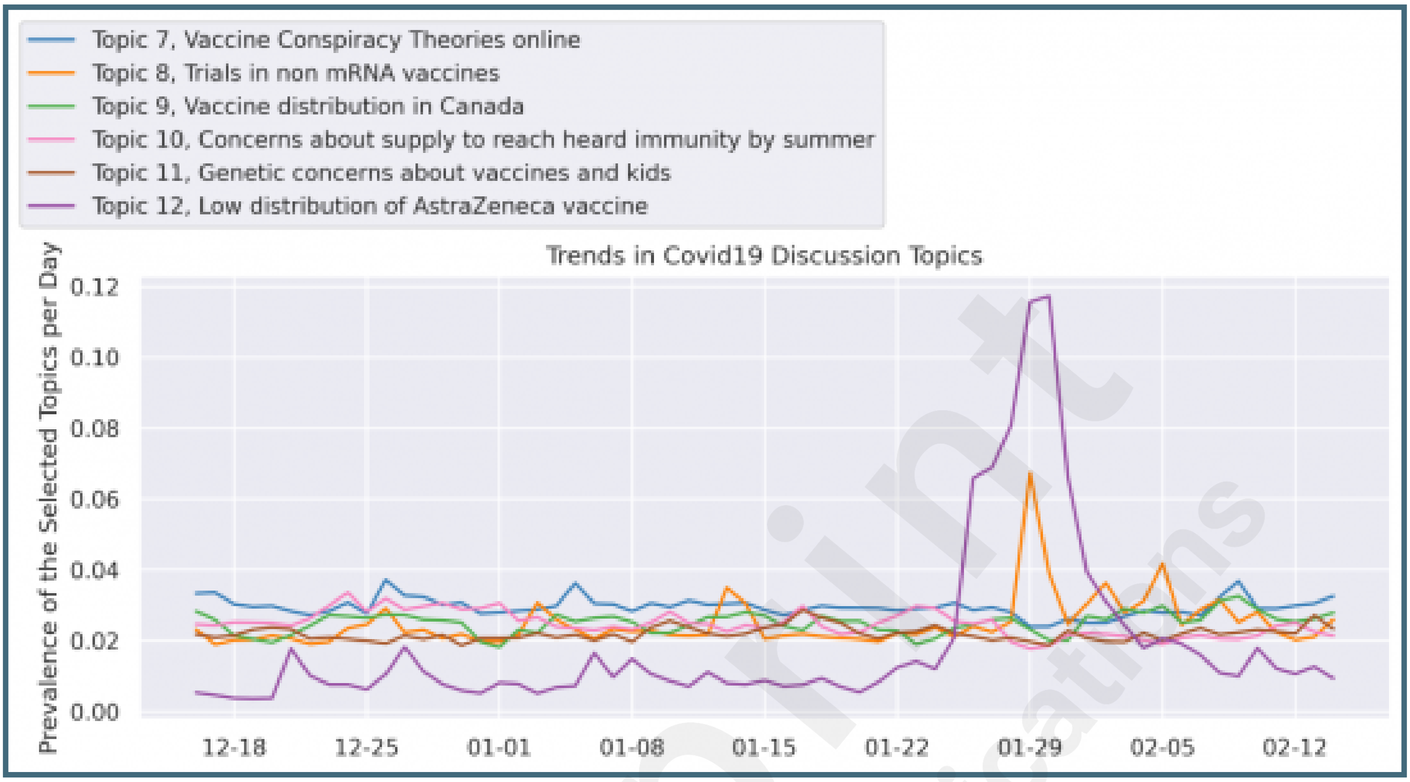
The topic trends for the most significant 6 topics that have a topic ratio above 3%.



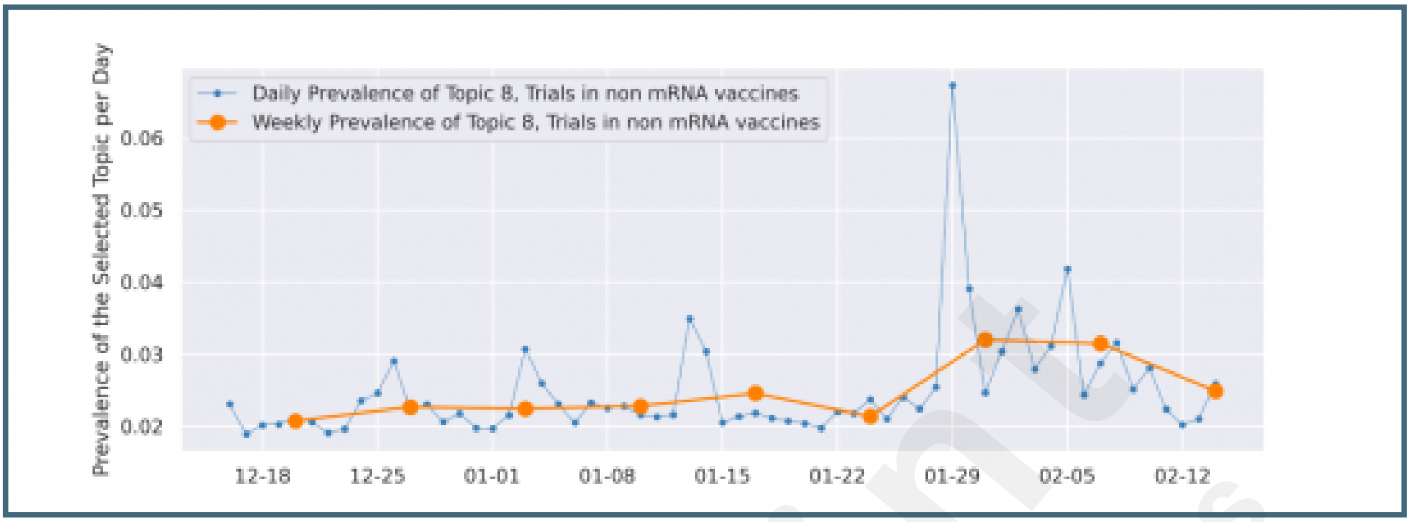
The daily and weekly trend of topic 1.



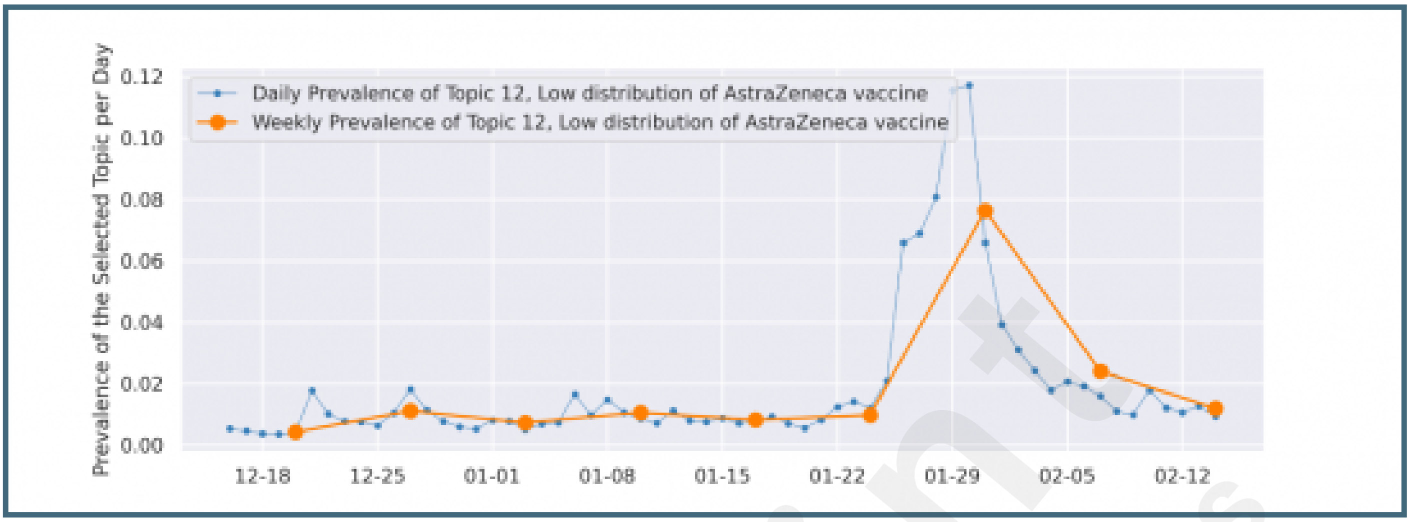
The topic trends for the rest of topics that have a topic ratio below 3%.



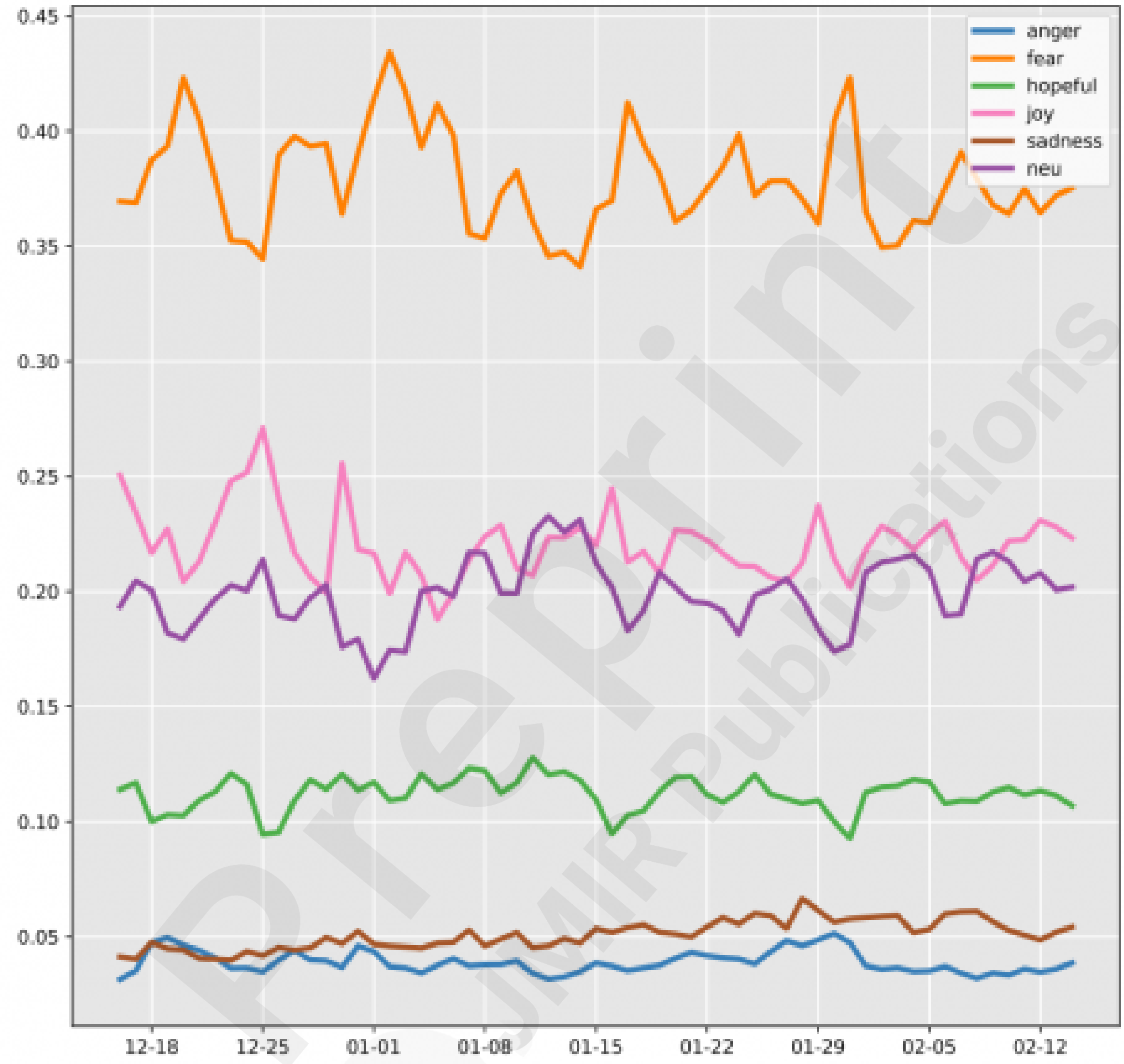
The daily and weekly trend of topic 8.



The daily and weekly trend of topic 12.



Emotion trend over time.



Emotion trend over time in topics 1 - 12.