Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Spatial and sentiment analysis of public opinion toward COVID-19 pandemic using twitter data: At the early stage of vaccination

Shaghayegh Jabalameli a,*, Yanqing Xu b, Sujata Shetty a

a Department of Geography and Planning, The University of Toledo, Toledo, OH, 43606, USA
b School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, 430079, China

ARTICLE INFO

Keywords:
Coronavirus
Social media
Public opinion
Spatial analysis
Topic modeling
Sentiment analysis

ABSTRACT

During the crisis of Coronavirus pandemic, social media, like Twitter, have been the platforms on which people have been able to share their opinions and obtain information. The present study provides a detailed spatial-temporal analysis of the Twitter online discourse (approximately 280 thousand tweets) in Ohio and Michigan at the early stage of vaccination rollout (January 2021, till March 2021). This work aims to explore how people were feeling about the pandemic, the most frequent topics people were talking about, and how the topics spatially were distributed. Moreover, state government responses and important news were gathered to analyze their impacts on public opinion based on the temporal analysis of the tweets. In this project, Natural Language Processing using the LDA method was employed to identify 11 topics and 8 sub-topics in the Twitter data. The temporal analysis of topics shows the sensitivity of the online discourse to the significant state news and the local government’s reactions to the pandemic. Moreover, the spatial distribution of Coronavirus-related tweets and sentiments demonstrates concentrations in the more populated urban areas with a high rate of COVID-19 cases in Ohio and Michigan. The government’s economic and financial policies taken during this time, the vaccination timeline phases specified by each state, and the pandemic-related information can contribute to public opinion and sentiment trends. The findings of this study can help explore public demands, and reactions, follow the impacts of the local authorities’ policies at the county level and manage their future responses to such a pandemic.

1. Introduction

The outbreak of COVID-19 has brought various challenges to the whole world from the time of its advent in December 2019, without any specific insights on treatment, through becoming a pandemic in March 2020, until the first vaccination in the U.S. after approximately one year, and beyond. (World Health Organization) Many social, economic, political, and public health disturbances have accompanied this pandemic [1]. It prompted extensive discourse on social media, like Twitter regarding various aspects of the disease and its prevalence. Social media provides an opportunity to share personal experiences and feelings. It also helps the policymakers to monitor public opinion and perceptions in each stage of responding to the Coronavirus pandemic. The analysis of public sentiment and opinions toward the present crisis in various periods with different policies taken by the U.S. government and local authorities can help communities deal with the pandemic’s challenges and according to decision making [2]. Investigating spatio-temporal social media information along with public opinion analysis can have a vital role in responding to the COVID-19 pandemic.

* Corresponding author.
E-mail address: shaghayegh.jabalameli@rockets.utoledo.edu (S. Jabalameli).

https://doi.org/10.1016/j.ijdrr.2022.103204
Received 25 January 2022; Received in revised form 26 June 2022; Accepted 19 July 2022
Available online 1 August 2022
2212-4209/© 2022 Elsevier Ltd. All rights reserved.
With the proliferation of using location-based social media (e.g., Twitter or Facebook) with time-stamped data to share the information on widespread happenings, researchers have an opportunity to address a range of social phenomena in different areas like urban food environment, public health, business, psychology, etc [4]. The implementation of internet surveillance tools was previously reported for the prediction of other epidemics including influenza [5,6], dengue fever [7], and Middle East respiratory syndrome [8]. X. Ye et al. show that social media can be used to detect and analyze the temporal-spatial spread of the infectious disease (Dengue fever in China) by exploring the relationship between the number of posts and the infected cases [7]. Also, another study by Lee et al. indicated that social media can provide a real-time surveillance system to automatically track flu and cancer disease activities [9]. Therefore, location-based social media provides the methods to identify and detect the outbreak of epidemic or pandemic diseases. Based on a review on using Twitter in health research, six categories of content analysis, surveillance, engagement, recruitment, intervention, and network analysis were identified to delineate Twitter platform opportunities for health researchers [10].

On the other hand, investigating the public attitudes, behaviors, and perceptions from social media regarding health of the community have been articulated in various domains including food consumption and choices [11–13], physical activity [14], mental disorder (depression) [15], non-infectious disease (heart disease, cancer) [9,16] and contagious one (flu) [9]. In research by Gibbons. J et al. [17], the sentiment-based lexicons inferred from social media were examined in order to evaluate the effectiveness of this strategy in predicting health outcomes such as self-rated mental health, sleep quality, and heart disease on a neighborhood scale. The authors concluded that the overall well-being of a place can be measured by leveraging sentiment data from social media that both reflect and influence health outcomes including physical activity, obesity, diabetes, heart disease, and mortality [17]. Yang & Mu, (2014) [15], focused on building a procedure to detect the MDD (Major Depressive Disorder) users and the spatial distribution of this disease, also its association with the SES (Socioeconomic Status) using Twitter dataset. The most important issue in these kinds of studies using the large datasets of conversational or textual data of social media is to identify the relevant topics in public conversation [12]. In this regard, one research by Ghosh & Guha was focused on the topic modeling and using GIS to map the tweets about obesity. The purposes of this study were to find the common obesity-related themes, and their spatial pattern, also the challenges of using the large dataset from SNS (Social Network Service). With this background in social media data mining and analysis, many researchers have started to work on social media data to analyze public opinion.

Using social media as a tool of information dissemination and consumption in investigating the Covid-19 pandemic has been started since the beginning of the crisis. Based on a review by Tsao et al. on the intersection of social media and the Coronavirus pandemic, Twitter was identified as a prominent platform used in those reviewed articles published in the first year of the disease outbreak (Nov2019-Nov2020) [18]. In this review paper, they explored six general themes and topics from the extracted articles including the public opinion and attitudes, mental health, predicting the trend of COVID-19 cases, the government responses to the pandemic, investigating the misinformation dissemination, and assessing the quality of health information [18]. Investigating the public interest, concerns and feelings has been one of the major topics of the research in different stages of pandemic and various geographical locations. The public sentiment on social media regarding the coronavirus pandemic has notably varied over time and geography [19]. Han et al. [20] investigated public opinion from social media data, Sina-Weibo, (a Chinese popular microblogging platform like Twitter) in China during the early stages of the COVID-19 outbreak. In that study, the spatial-temporal analysis was conducted on the extracted COVID-19 related Weibo posts to explore the distribution of data and find any relation between the temporal trends of social media data and the confirmed coronavirus cases. In addition, the common topics of the coronavirus-related Weibo texts were extracted and analyzed in terms of time and space [20]. Furthermore, in recent studies related to the COVID-19 and social media, the focus is on public opinion about the vaccination [21–23] and real-time surveillance based on social media monitoring to explore the trends of COVID-19 cases [24] and vaccinations [25] to support decision making and timely interventions. Villavicencio et al. [26] studied the sentiments of people in Philippines regarding the government efforts after delivering COVID-19 vaccines. In this research, other than using Natural Language Processing techniques to explore the general sentiment, the Naïve Bayes model was employed to classify the tweets into positive, neutral, and negative polarities with an 81.77% accuracy. Another study by Lyu et al. [23] investigated the sentiment and opinions in the public COVID-19 vaccine-related discussion on social media. It provided the discourse patterns on social media regarding vaccine efficacy and safety. This study concluded that public opinion was greatly mirroring the news topics in mainstream media. We also want to detect this effect of governmental response in the form of news on social media discourse at the local level using topic modeling and temporal analysis.

In 2020, most of the news focused on how COVID-19 spread across the world, but 2021 represented the distribution of vaccination. Based on a report by Food and Drug Administration, the vaccination became available in the United States in December 2020 [27]. This study aims to work on social media (Twitter) data during the first period of vaccination phases from the 12th of January to the March 10, 2021 in two states of the United States, Ohio, and Michigan. Social media has been considered as a way of collecting data instead of using other costly and time-consuming methods. Although vaccinations for healthcare workers began on December 14th, the start date of this study is during the time of phase 1 B vaccinations (available to those who are 65 and older or those with special medical conditions), and after the new year holiday to obtain Twitter posts more related to the Coronavirus pandemic. This study aims to detect public opinion and sentiments and their spatial and temporal distribution. In addition, by building a topic extraction model, the main topics discussed in Covid-19-related Twitter data were identified; also, the spatial-temporal analysis for the tweets related to each topic was conducted to discuss their trends during this period in these two states. This provides the opportunity to compare the impacts of Ohio and Michigan government responses toward the Coronavirus pandemic on the related social media discourse topics and to find any meaningful relationships between them.
2. Data and methods

2.1. Data collection

Twitter is one of the most popular social media platforms in the United States and even in the whole world. Due to the Coronavirus pandemic and its consequences on social distance, quarantines, and business closures, people rarely used public spaces to communicate with each other. Thus, much of their conversation about various happenings occurred online on the social media platforms like Twitter. This study acquired Twitter text related to COVID-19 for approximately two months (12th of January 2021 until the 10th of March 2021) using the Standard Search API since it allows to access an already existing Twitter data set from the past 6–9 days. The library “rtweet” which is the most recent package in software program R regarding Twitter data analysis was implemented, and the “search_tweets” module from this package was used to extract tweets based on some criteria. By using this module, the tweets related to coronavirus were collected with “covid”, “coronavirus” and “pandemic” keywords. In order to cover the timestamps of two months, extracting data was repeated 6 times every nine days, also, the “number of tweets (n)” parameter was specified as 100,000 for each time of extraction. In addition, for the “location” parameter, three circles were specified based on one point in the center (measured in Latitude and Longitude) and the radius specified in miles for Ohio (40.37, −82.99, 115mi) and Michigan (44.18, −84.51, 140mi, and 46.41, −86.65, 95mi).

By this search function, the geographic coordinates of limited numbers of generated tweets were provided but most of the tweets have the location parameters being geocoded using Address Locator in ArcGIS.

The generated data, after eliminating duplicated tweets and retweets, was 67,983 tweets, including 3284 texts with geographical location information for Michigan. For Ohio, this was 212,860 tweets in general and 5699 tweets with the geographic coordinates. In Fig. 1, the distribution of the tweets with geographic coordinates in Ohio & Michigan is displayed using software environment R.

As expected, most of the tweets aggregated around the big cities like Cleveland, Columbus, Cincinnati, and Dayton in Ohio, and Detroit, Lansing, and Grand Rapids in Michigan. Although this map contains only a small number of tweets, the pattern of distribution could be in some way too close to the map of the coronavirus confirmed cases.

2.2. Method

2.2.1. Word Frequency and Word Clouds

Word Frequency and Word Cloud analysis can be the start point of the text mining analysis. They represent the overview of the tweets’ text by providing a statistical summary of the most frequent isolated words. Therefore, the analysis is not based on any linguistic knowledge, they just summarize the text with only limited interaction capabilities [28].

After generating the Twitter data and importing it into the programming languages Python, the dataset was changed into the corpus, and retweets were discarded to avoid repeated results. The preprocessing of the data was continued by removing biases and flaws and cleaning the data by eliminating some factors like punctuation characters, numbers, white spaces, URL, non-English words, and stop words. Besides, the tokenization method to break a piece of text into smaller components was applied to give useful phrases that lead to some additional insights. Then, the text document changed into a matrix to extract the most frequent words in tweets of Ohio & Michigan.

2.2.2. Time series analysis

A time-series analysis was provided to identify any correlation between the trend of the number of coronavirus-related tweets and the prevalence of Covid-19. Nagel et al. [29], suggested that the trend of the social media streams can be related to how the events developed and even used to make predictions.

In this study, the time series analysis of the tweets in that specific period for Ohio and Michigan was investigated to explore the temporal diversification of the Covid-19 tweets and the outbreak of this disease. To find the temporal trend, the time series can be

![Fig. 1. The distribution of collected COVID-19-related tweets with geographic coordinates in Ohio & Michigan.](image)
decomposed into three components using the Seasonal- Trend decomposition procedure based on Loess (STL), using a package (statsmodels.tsa.seasonal.seasonal_decompose) in the program language Python. As shown in the following equation, time series is the combination of, trend, seasonality, and noise (remainder) components.

\[ X_t = T_t + S_t + R_t \] 

where \( X_t \) is the original time series. \( T_t \) is the trend component. \( S_t \) is the seasonal component. \( R_t \) is the residual component [20].

2.2.3. Sentiment analysis and emotional analysis

The importance of social media in data acquisition and controlling many outbreaks and pandemics has been emphasized in various studies [30]. The sentiment analysis of social media to gain human feelings is extremely helpful and completely affects the recognition of public trend for business decisions, and policy approaches [31].

Sentiment analysis is the systematic process of identifying, extracting, and classifying subjective information and affective states in text data. This might be an opinion, a judgment, or a feeling about a particular topic [31].

This analysis can be classified in different ways. One of them is text categorization or classification referring to the computation of the sentiment analysis based on the number of occurrences of positive and negative words in each text document and calculating their scores. In this method, based on the previous step, the cumulative score was provided indicating the polarity of each tweet denoting positive (> 0), negative (< 0), and neutral (= 0) [32].

Emotional analysis, on the other hand, refers to investigating the mental situation of humans. This analysis can help to perceive human feelings in more detail [31].

The difference between sentiment analysis and emotional one is regarding how deeply each of them investigates the dominant feeling of the text. Sentiment analysis focuses on the polarity of the expression (as positive, negative, and neutral), but emotional analysis dives deeper into the subtlety of the emotions that appeared in the comments, so provides more categorization (such as sadness, happiness, disgust, fear, …) [31].

In this study, the expert pre-defined lexicon was utilized to detect the emotional effects of a text. The “NRCLex” package in Python was the tool that its affect dictionary consists of 27,000 words based on the National Research Council Canada (NRC) and the NLTK library’s WordNet synonym sets [33]. Through this method, each tweet based on the containing words was scored with eight categories of emotions of fear, anger, anticipation, trust, surprise, sadness, disgust, and joy. Specific words (containing the words in the dictionary of this package) in each tweet can represent emotions. The emotion count reflects how many times any emotion was recognized by the words in the tweets. For comparison purposes, the percentage of emotions was calculated for OH and MI twitter data.

Sentiment analysis can be performed by two methods – lexicon-based and machine learning-based. For this study, the lexicon-based approach which is unsupervised performing analysis using lexicons and scoring methods was used [34]. The “nltk.sentiment.vader” package in python was employed using a “SentimentIntensityAnalyzer” module. This tool helps to perform sentiment analysis by implementing NLTK features and classifiers; also, it provides not only the polarity scores but also the intensity of negativity or positivity [35]. Its outcome contains four columns for each tweet representing positive, negative, and neutral sentiment scores along with a compound sentiment score or “normalized, weighted composite score”. The compound score is the most useful metric for a single dimension of the document (one sentence) to gain the sentiment. It normalized to be between −1 (most extreme negativity of the tweet) and +1 (most extreme positivity of the tweet). Thus, each tweet has its specific score, based on the valence score of each word in the lexicon and adjusted according to the rules [36].

2.2.4. Topic modeling

One of the important steps in this research is opinion mining which is approached by topic modeling analysis. In topic modeling, a group of words is analyzed together to gain the meaningful relations between containing words upon the broader context [32]. Many techniques can be used to achieve topic modeling such as Latent Dirichlet Allocation (LDA), Latent Semantic Allocation (LSA), and Non-Negative Matrix Factorization (NNME). Based on the existing studies, the best way to identify latent topic information from the text is the LDA model [37,38].

2.2.4.1. LDA model. In order to obtain more relevant topics and content related to public health and the Coronavirus pandemic, the Latent Dirichlet allocation (LDA) model can be implemented [39]. The main structure of this model is based on some assumptions stating that each document in the corpus contains a probabilistic mixture of topics, and each topic is a probabilistic mixture of words [32]. In general, LDA is a generative probabilistic model of a text and other collections of data to model the topical structure of documents. In fact, each document consists of some latent topics and subsequently, each topic is characterized by words; therefore, it is a useful model to generate a topic distribution for every document and to identify multiple topics used to classify the individual tweets, based on the words in each tweet [40]. In this model, it needs to determine the number of topics and the words in each topic. Some researchers suggested running LDA with different numbers of K (number of topics) to achieve the suitable one for an effective result.

2.2.4.2. Topic classification. In this study, the package “Gensim” in Python was used to apply the LDA model. After examining different numbers of k for this model, we come up with the optimal number of 20 topics and 10 words for each topic. The topic-terminology list gained through running the LDA model contains the vocabularies for each topic and their probability of associating with the topic. Based on the topic-terminology list, these 20 topics are classified into 11 main topics namely “Sentiment & Opinion”, “Personal Response”, “News & Reports”, “Government Response”, “Lockdown and Its Impacts on Businesses”, “Monetary Issues”, “Seeking Help”, “Popularization of Prevention and Treatment”, “School Closure”, “Healthcare Situation”, and “Scientific Research”, also 8 sub-topics as “Sentiment & Opinion-Fear & Worry”, “Sentiment & Opinion -Back to Work”, “Sentiment & Opinion- Staying at Home & Wearing a Mask”, “Sentiment & Opinion-Questioning the Government”, “Personal Response-The Future of Pandemic”, “Personal
Response-Adapting to the new Situation”, “News & Reports-Vaccination”, “News & Reports-Updated COVID-19 cases and Events Notification”. Based on literature review [18,20,41], this topic classification was generated by discarding improper topics and combining the relevant ones, each tweet in this procedure was then assigned to the associated topic.

2.2.5. Kernel Density Estimation

To evaluate the density of the tweets and identify their hotspots in both states, first, they were geocoded using Address Locator in ArcGIS, so the number of geocoded tweets in Ohio and Michigan became 108,386 and 56,607 respectively, then the Kernel Density Estimation with the search radius of 20 km was implemented using ArcGIS software. This parameter is generally used to explore the intensity of the events on a smooth surface using a quadratic kernel function [42]. Suppose we have a series of point events; the kernel density can be estimated as followed:

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{\tau} k\left(\frac{s - s_i}{\tau}\right)$$

where \(k\) is the kernel function, \(\tau\) is a smoothing parameter (called bandwidth, specifying by the search radius), and \(s - s_i\) is the distance between \(s\) and \(s_i\) [20].

In this model, the numerical statistical algorithms are used to calculate the density and measure the distribution pattern of features and hot spot detection. Kernel density estimation fits a window function on each point based on an assumption that it is continuously distributed within the fitted kernel window. It calculates the fraction of an observation point at a location “\(s\)” based on kernel function \(k\) and the distance between location \(s\) and the observation point specified by the search radius. Therefore, for each location, the density can be computed by summing up the fractions of all the observation points at that location [42].

3. Results

3.1. Word frequency analysis

In this section, the word frequency of both Ohio and Michigan twitter data were provided to give us an overview of what people talk about and help us start the analysis.

From the bar charts, it is inferred that most of the coronavirus-related conversations in both states are about the vaccine, health, the rate of deaths and positive cases, quarantine, wearing a mask, staying at home and back to work. Among some differences between these two charts, the presence of the words like “school” and “student” in Fig. 2-b shows the concern about school closures among people in Michigan, while those words do not appear in the list of Ohio’s most frequent words. On the other hand, words like “storm” and “winter” in Fig. 2-a refer to winter storm of Ohio in February. Although the storm also hit Michigan, it was not a frequently used word in Michigan online discourse.

Although the word frequency cannot be used to prove any statement, it reveals some insight into the data. For instance, the different ranking of the two words “Biden” and “Trump” in these 2 bar charts among other words, can be referred to as their different political orientations (Democrat or Republican party).

In Fig. 3, the word clouds for the two datasets visually represent word frequency. The more commonly a word is repeated within the text, the larger it appears in the image. It is mostly used as a tool to identify the focus of the tweets and acquire insights on trends and patterns, though it has some limitations.

3.2. Time series

In Fig. 4, the confirmed COVID-19 cases, the frequency of the Coronavirus-related tweets, and the new daily cases each day were represented in the chart line for both Ohio (in red) and Michigan (in blue). The numbers (frequencies) in these three trends were converted into the log scale numbers to be more comparable. In addition, in this chart the important news and events in Ohio [43] and
Michigan [44] regarding COVID-19 are provided, to be analyzed in more detail [45]. As can be seen, the trend of the confirmed cases was growing steadily and gently without fluctuations for both states; however, the line graphs of the related tweets fluctuated during that period in which the patterns of changes were almost the same for both states despite their different trends. The rising trend of Ohio tweets followed Ohio confirmed cases trend, but Michigan Twitter data represented the declining trend opposing its confirmed cases trend.

On the other hand, the new daily Coronavirus cases have started to decline for both states from the beginning of this period till Feb 20, 2021, reflecting the growth in the vaccination; after that, the delay in delivering vaccines due to the severe storm (on 16th of Feb) in both states has caused an increase in the new daily cases for approximately one week. Ohio’s new cases trend returned to its decreasing position, but Michigan’s remained constant and then increased due to identifying new COVID-19 variants and other factors like the return to school in-person (also happened in Ohio), and inefficiency in the vaccination programs. Although Ohio started vaccination a week later than Michigan, it became among the top five states for delivering vaccines. Also based on the trend of the new cases, it can be inferred that Ohio has acted appropriately in providing vaccinations. One of the reasons might be related to their different vaccination programs. In Ohio, each phase of the vaccination plan was divided into several steps taking place on various dates, but in Michigan, in this period, one phase in one step was conducted for people with age 65 and older, frontline essential workers, and childcare and pre K-12 staff.

Although the trend of Ohio Twitter data was opposed to the Ohio new cases trend, their fluctuations followed almost a similar
pattern. For Michigan, the trend of Twitter data acted also in almost the same way (with decreasing trend) with the daily new cases trend until March 1st, the time that in-person school was opened, so the trend of new cases started to grow in contrary to the declining trend in the number of tweets.

The important news related to how these two states of Ohio and Michigan responded to the COVID-19 pandemic was shown in Fig. 4. For Ohio, from January 12th until Feb 3rd, the trend of Twitter data remained almost constant with slight fluctuations due to the announcement of the vaccination timeline and curfew extension. After this time, the conversation related to Coronavirus has grown steadily with lots of changes. For instance, in the first week of February, due to some economic policies like the formation of a new Ohio Department of Job and Services (ODJFS) to help with unemployment services or rent assistance, the number of tweets suddenly increased. However, by Feb 11th, after the announcement of the order on reopening the food service stations, it declined for three days. From mid-February to March 10th, the trend of Ohio Twitter data has continued to grow with some significant fluctuations like delay in vaccine shipment due to the severe storm on 16th of February for three days (which also impacted the new COVID-19 daily cases), reopening sporting and environment venues on 26th February, back to school on March 1st, and opening several regional mass vaccination sites on March 5th.

On January 12th, Michigan issued the Pandemic Unemployment Compensation (PUC) and started vaccination for people aged 65 and older, frontline essential workers, and PreK-12 teachers. Therefore, as shown in Fig. 4, the conversation on Twitter suddenly increased. From January 13th till February 3rd, like for Ohio, the trend of Michigan coronavirus related Tweets remained constant with slight changes which were related to some reports such as the announcement of starting unemployment payment, the recovery plan for the economy and ending the pandemic, reopening indoor dining and indoor group exercise. After that on February 4th and 5th, the announcement of a mask order for public transportation and allocating of a grant for the innovative projects to reduce the outbreak of COVID-19 on public transit caused a growth in the rate of related tweets. After this period until March 9th, Michigan Twitter data has experienced a downward trend, though it has faced some peaks and valleys. Significant peaks were related to the announcements requiring mask-wearing even after receiving the vaccine (Feb 13th), severe storm weather and delay in vaccine shipments (Feb 19th), returning to school in person (Feb 24th), updated restaurant orders regarding more restrictions, and expanding access to the vaccine on March 3rd. The reason for the declining trend in Michigan tweets might be related to reopening the indoor dining on February 1st and backing to school in person on March 1st which caused many public conversations to take place in the real world.

In general, most of the changes in Covid-19 related twitter data trends of both states can be explained by the policies regarding Coronavirus taken by the state government authorities. In addition, these policies also affected the daily new case trends in Ohio and Michigan.

In the following section, the result of the time series analysis of Twitter data for Ohio and Michigan was provided to compare their trends through time in more detail. Fig. 5-a showed their original time series of the number of tweets, which was also presented in the previous chart line. Fig. 5-c illustrated the cyclic change in the number of Coronavirus-related tweets repeated every week. For Ohio (Fig. 5-I-c), the lowest point of the seasonal Twitter trend occurred on the weekend and has raised during the week with one drop in the middle. For Michigan (Fig. 5-II-c), this cyclic change acted slightly differently with the lowest level on the weekend remaining for more than three days, then reaching a peak in the middle of the week. In general, this might reflect the fact that people do not spend their time on social media on the weekend. After eliminating the seasonal effects on the original temporal trend of tweets, the trend component can be achieved to show the general trend of Twitter data. In Fig. 5-b, the trend components of the Coronavirus-related tweets in both states were depicted. For Ohio (Fig. 5-I-b), this trend first has raised slightly for about three weeks from the start point, then it reached the peak on the 15th and 16th of February because of the winter storm at that time and it caused business shut down for two days in some areas, so the conversation on Twitter rose. After that time, the Twitter trend decreased a little and remained constant for the rest of the study period due to the start of other vaccination phases. The corresponding trend for Michigan (Fig. 5-II-b) illustrates a different pattern. Twitter activity has moved steadily at a high level until February 1st then declined and reached the lowest level because the first day of February was the time the restaurants, stadiums, and gatherings were reopened. After that, from February 16th, the Twitter activity trend significantly raised due to the winter storm, the reports of vaccination and encouraging
people to vaccinate, and the announcement of returning to school in person on March 1st.

The residual graphs, Fig. 5-d, depict the location of each day based on the frequency of the tweets from the trend line, so some of the outliers can be detected. For Ohio (Fig. 5-I-d), on most of the days, the number of tweets is close to the trend line with just slight differences, except for February 16th, February 26th, and March 4th, which are the outliers with high frequencies. For Michigan, Fig. 5-II-d, the number of tweets in each day from the beginning of the period till February 1st, have had variations from the trend line, but after this time most observations fluctuated around the trend line.

3.3. Sentiment analysis

In Fig. 6, the extracted tweets from Ohio & Michigan were examined to detect dominant emotions in the Twitter conversations. These were classified into eight types, namely anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. However, in order to compare the sentiment scores between these states, the scores were changed into percentages based on the total tweets for each of the two places. As shown in these 2 bar graphs, the percentage of negative emotions regarding sadness, fear, disgust, and anger in Ohio are more than in Michigan, though they are close. Part of the interest is the balance between fear and trust in these two states. In Ohio, fear overcomes trust, but in Michigan, they are almost the same.

The sentiments from Twitter data of Ohio and Michigan were evaluated to understand public perception of the coronavirus pandemic. In this regard, the polarity of each tweet based on the compound sentiment score was specified through the analysis. Fig. 7-b shows that the percentage of positive tweets in both states was greater than the percentage of negative and neutral ones. The proportion of Ohio negative posts was greater than the proportion for Michigan, while the frequency of its positive tweets was less than the frequency in Michigan.

Fig. 7-a displays the distribution of sentiment scores for both states. According to this line chart, most of the tweets had a sentiment score between (−0.3, 0.2). The higher rate of negative tweets in Ohio mostly happened in the [−0.5, 0) range, while Michigan’s positive tweets were more strongly positive in [0.5,1] range.

3.4. Spatial analysis

Fig. 8-a shows the spatial distribution of coronavirus-related tweets in Ohio and Michigan. In order to specify the aggregation of the data, Kernel density was used to calculate the density in a neighborhood of each point (Fig. 8-b). The Kernel density map with a search radius of 20 km, shows the high-density areas of COVID-19-related tweets in the Southeast and Southwest of Michigan, and the Northeast, Center, and Southwest of Ohio. It shows the aggregation of Twitter data around large cities and populated counties, so the counties of Oakland (Pontiac), Washtenaw (Ann Arbor), Kent (Grand Rapids), and the northern part of Wayne (Detroit) are the hot points in Michigan. In Ohio, the great number of Tweets are mainly concentrated in three counties, Cuyahoga (Cleveland), Franklin (Columbus), and Hamilton (Cincinnati). In the second level, Saginaw, Genesee, Muskegon, Macomb, Marquette, and Grand Traverse show the high density of Twitter data in Michigan, also Lucas, Montgomery, Summit, Stark, and Lorain counties illustrate the trend of hot spots in Ohio.

Fig. 9, depicts the Kernel density of Negative (Fig. 9-a), Neutral (Fig. 9-b), and Positive (Fig. 9-c) sentiment of the COVID-19-related tweets. In general, all three maps represent the aforementioned hot spots; however, their differences are regarding other low-populated counties. in Michigan, the positive tweets are mostly concentrated in populated counties, but the negative sentiments are distributed among other counties, too. On the contrary, Ohio’s negative tweets are mainly aggregated in the aforementioned hot spots, but its positive tweets are spread all over the state in addition to the concentrated areas. Besides, the Neutral sentiment density map (Fig. 9-b) for both states shows the high dispersion compared with the maps of positive and negative sentiment.

3.5. Topic modeling analysis

3.5.1. Topic description

Topic modeling analysis was implemented on coronavirus-related tweets of Ohio and Michigan. Fig. 10-b displays the classification and the statistical percentage of main topics for Michigan. It can be inferred that “Opinion and Sentiments” contribute the most at
Fig. 7. Sentiment polarity of COVID-19-related tweets in Ohio & Michigan.

Fig. 8. The spatial distribution of COVID-19-related Twitter data in Ohio & Michigan. (a) Distribution of Tweets related to Coronavirus. (b) Kernel density of Tweets related to Coronavirus.

Fig. 9. The kernel density of the COVID-19 related public sentiment from Twitter data in Ohio & Michigan. (a) Negative Sentiment. (b) Neutral Sentiment. (c) Positive Sentiment.

Fig. 10. The percentage of topics discussed in COVID-19-related Twitter data. (a) Ohio. (b) Michigan.
34.5%. “News and reports” and “personal responses” are the second and third most frequent, at 18.2% and 14.1%, respectively. The other seven topics come after with less proportion. “Government responses” and “lockdown” topics account for 7.2% and 6.3%, respectively. The proportions of “monetary issues”, “feeling help”, “school closure” and “popularization of prevention and treatment” are in the range of 4%-4.7%. Also, the last two topics with the least contribution at 2.1% and 0.1% respectively, are “scientific research” and “healthcare situation”.

Fig. 10-a displays the percentage of the main topics in Ohio. The contributions of topics are in the same order as in Michigan, with only one difference that the “School closure” topic is less than “popularization of prevention and treatment”. Also, the “healthcare situation” topic has been more mentioned in Ohio with 1% compared with 0.1% in Michigan.

Fig. 11, illustrates the result of topic modeling analysis on sub-topics. It shows that the percentage of “fear and worry” in both states is in first place with approximately 20%. In Fig. 11-b, “to work”, “vaccination” and “the future of pandemic” are the second, third and fourth most frequent sub-topics (with slight differences) in Michigan. However, for Ohio, in Fig. 11-a, this order changes to “vaccination”, “to work” and “the future of pandemic” with 18.4%, 16.3%, and 15.5%, respectively. The contributions of “staying at home and wearing a mask” and “reports on updated cases” are more than 10% for both states with the same order. However, the proportions of the last two sub-topics namely, “adapting to the new situation” and “questioning the government” are less than 5% again for both states.

Fig. 12 displays the proportion of the main topics along with the sub-topics for both states to provide the opportunity for comparison and analysis. Five topics of “reports on vaccination”, “sentiment & opinion-fear & worry”, “sentiment and opinion-back to work”, “personal response on the future of pandemic” and “sentiment & opinion-staying at home and wearing a mask” are the most frequent both in Ohio and Michigan. The percentage of most of the topics is approximately at the same level for both states. However, “school closure”, “lockdown and its impact on businesses” and “personal response to the new situation” are mentioned more often in Michigan conversations than in Ohio. On the other hand, the proportions of “reports on vaccination” and “sentiment-staying at home & wearing a mask” in the Ohio tweets were more than those in Michigan.

3.5.2. Temporal topic trend

This section is dedicated to displaying the proportion trends of each extracted topic from Twitter data of Ohio and Michigan over time. The temporal analysis of each graph provided here is based on the information and reports presented in Fig. 4. From Fig. 13, the Ohio temporal trend of three topics, seeking help”, “healthcare situation” and “school closure” acted differently compared with the corresponding topic trends of Michigan. However, in other topic diagrams, the trends of both states followed almost the same pattern. The topic of “seeking help” in Ohio conversations dropped after mid-January then inconsistently experienced an upward trend to reach the start point at the end of the period with two peaks in the first and third weeks of February, respectively, due to unemployment and rent assistance, and the severe weather and its impact on the vaccination plan. However, the aforementioned topic for Michigan slightly fluctuated without considerable changes over time, except for January 20th and the second week of February because of unemployment issues and food shortages. Like Ohio during the days of the storm, the rate of “seeking help” related tweets increased.

“Healthcare situation” was a common topic in Ohio twitter posts in January on account of vaccination for healthcare staff and nursing homes, but remained constantly low in Michigan. Michigan Twitter posts related to the “school closure” topic were more frequent in January reaching the highest with two peaks in the time of vaccination for pre-K teachers and childcare providers and the announcement of returning to school in person, but its frequency declined over time. The mention of this topic in Ohio tweets was also high from January 22nd till February 5th with some fluctuations due to the same reasons mentioned for Michigan. Also, it dropped on March 1st for a few days (since the schools were reopened) and during the time of severe weather in which the main topic of conversation tended to focus more on the delay in vaccination.

“Personal responses” and “sentiment & opinions” temporal topic trends were nearly similar, remaining approximately constant with slight fluctuations and two sharp drops on March 1st – 3rd again owing to back to school in-person and 14th-15th of February (before the storm) repeated for almost all the topics. “Scientific research” acted the same as the two aforementioned topics but tended to decline over time which might because of the availability of COVID-19 vaccination causing extensive discourse on Twitter to focus more on the policies of distributing vaccines rather than the research on vaccinations. Moreover, this topic for Michigan tweets had a higher rate during the first five days of that period due to the existence of new cases of the COVID-19 variant.

Fig. 11. The percentage of sub-topics discussed in COVID-19-related Twitter data. (a) Ohio. (b) Michigan.
“News & reports” and “government responses” topic trends were almost like each other seesawing over time with some abrupt increases on specific days. This similarity of these two trends indicates the immediate government reactions in both states toward the important Covid-19-related happenings or people’s expectations of government responses. For Ohio, the dramatic rises in the aforementioned topic trends are regarding vaccination policies and their timelines, curfew extension, planning for equity in vaccination, the delay in vaccine shipments, and opening the regional vaccination sites. Michigan’s trends of these two topics also experienced some increases related to government policies of vaccination like plans for the timeline of its phases (in January), mask orders for public transportation (in February), expanding the response forces for vaccination, and planning for vaccine equity, reopening public centers, and delay in vaccine shipments due to severe weather.

The similarity between “lockdown & businesses” and “popularization of prevention & treatment”, also “monetary issues” in some ways were thought-provoking, showing the relations between these three topics. They rose gradually but simultaneously experienced
some peaks in mid-January, at the time of providing the pandemic unemployment compensation payment and the recovery plan for the economy during the Coronavirus pandemic in Michigan and curfew extension in Ohio. Besides, these trends for both states faced increases in the first and last weeks of February due to economic policies. In Ohio, the plans for rent assistance and forming a new department to help with the unemployment, the negative outcomes of the storm and severe weather, and providing nursing home visitation influenced the numbers of tweets related to these three topics. In Michigan, other than the direct impacts of economic policies like food assistance, expansion of public transit, and teacher grants on the frequency of related tweets, the government provision of free consultations for businesses and infection control training caused the increase in related topic trends.

3.5.2.1. Spatial distribution of topics. The spatial distribution of Twitter data for each topic was analyzed by using Kernel density (radius = 20 km) of distribution. Oakland, Wayne, Washtenaw, Ingham, and Kent in Michigan, and Cuyahoga, Franklin, and Hamilton in Ohio are the common high-density counties in all 11 topics density maps though with some differences. The topics can be categorized into four groups based on their similarities in spatial density. As illustrated in Fig. 14, the spatial distributions of “sentiment & opinion” and “personal responses” topics in the first group follow the same pattern showing the hot spots similar to the ones in Fig. 8-b.
In addition to the main counties, these two topic density maps are followed by Saginaw, Genesee, Macomb, Muskegon, Gladwin, and Livingstone in Michigan and Stark, Summit, Lucas, Montgomery, Clark, and Morrow in Ohio. They are the most distributed maps with high-value density, also these two topics are responded to by other areas between the main nodes. In the next step, the similarity of the four topics “lockdown & its Impact on Businesses”, “government Response”, “news & reports” and “scientific research” are completely recognizable. They reflected the same areas as the previous two topics just with a lower density, except for Muskegon, Gladwin, and Livingstone in Michigan and Stark, and Morrow in Ohio. They also are not as distributed as the density map of the two topics in the first group. From Fig. 14, it is inferred that the distribution in “seeking help”, “popularization of prevention and treatment”, “monetary issues” and “school closure” followed the same pattern showing the mentioned common areas supplemented by Genesee in Michigan and Lucas, Summit, and Montgomery in Ohio. “Healthcare situation” has the lowest density of distribution demonstrated just in the main areas which are common among all the topics. It takes Livingstone and Washtenaw in Michigan and Cuyahoga in Ohio as the center points with high-value density.

Fig. 15 shows the spatial distribution of Kernel density analysis (radius = 20 km) for 8 sub-topics. These are also can be grouped into four categories based on the similar distribution density. “Personal response-adapting to the new situation” has a distinct pattern to compare with other subtopics following the main nodes in Washtenaw, Ingham, Kent, and Muskegon in Michigan and Cuyahoga,
Franklin, and Hamilton in Ohio, albeit with less density. “Sentiment & opinion-back to work” and “sentiment & opinion-questioning the government” have a similar distribution trend. They form the hot spots in Washtenaw, Ingham, Kent, and Oakland in Michigan and Lucas, Summit, Montgomery, Cuyahoga, Franklin, and Hamilton in Ohio. The distributions of these two subtopics mostly focus on the aforementioned main centers with less density for the areas between them. The spatial distribution of “news & reports-vaccination” is similar to “news & reports-updated confirmed cases” with the same high-values areas mentioned for the previous category complemented by Saginaw, Genesee, and Macomb in Michigan and Clark in Ohio. Moreover, the distribution pattern continues to the surrounding areas, showing more density and distribution. The last group contains “sentiment & opinion-fear & worry”, “sentiment & opinion-staying at home & wearing mask”, and “personal response-future of the pandemic” with the most distributed and dense spatial pattern among other subtopics. These subtopics were addressed by most of the counties in Ohio and Michigan. They are led by the high-value areas mentioned for the previous group with additional areas in Livingstone, Wayne, Gladwin, and Bay counties in Michigan, and Stark, Morrow, Athens, and Ashland counties in Ohio.
4. Conclusion

In this study, social media data related to the Coronavirus pandemic at the early stage of vaccination (January 12th, 2021, till March 10th, 2021) was comprehensively analyzed to understand public opinions and sentiments in Ohio and Michigan. This research supplemented other studies on exploring public sentiments and opinions during the vaccination rollout plans [19,22,23,26] by providing a detailed analysis of the public discourse on social media at the county level. A mostly similar approach was taken in the previous studies during the vaccination period to investigate the topics and themes discussed on social media. They have focused on vaccine hesitancy and antivaccine messages [46], vaccine efficacy and safety [23], and vaccination real-time surveillance [25]. This study provided a comparison of the public opinions and sentiments regarding the COVID-19 pandemic at the early stage of vaccination between two near states in the US, and it also showed how the local government responses reported on the news have shaped the social media discussion on different topics.

Although the sentiment analysis and investigation of the dominant emotions in Twitter data represented the predominance of the positive sentiment in both states during the vaccination phases, Ohio had more negative sentiments with the “fear”, and “sadness” emotions in comparison with Michigan. In this work, 11 topics and 8 sub-topics were obtained and compared with each other using spatial-temporal analysis. In addition, the daily policies taken by the local governments in Ohio and Michigan during the vaccination period were investigated, to find their impacts on the daily conversations on Twitter. Four topics of “sentiment & opinion”, “news & reports”, “personal response”, and “government response” covered approximately 75% of the whole topics in both two states. For sub-topics, almost 70% proportion contains “fear & worry”, “vaccination”, “back to work”, and “the future of pandemic” being the same for Ohio and Michigan with just the difference in their orders. The temporal analysis of topics showed that their general trends in both states during this period acted in a similar way except for three topics, “school closure”, “seeking help”, and “healthcare situation”. Besides, this analysis showed how public opinions on different topics, specifically economic-related ones, were vulnerable to the local government responses to the pandemic each day. In other words, the analysis demonstrated that most of the changes in COVID19-related tweets trends can be explained by related policies implemented by local authorities. In addition, how each government releases its vaccination timeline, and some of its economic and lockdown policies were the areas where these two states acted in different ways. The different governments’ reactions are reflected in the number of daily confirmed Coronavirus cases and public opinions on social media.

The spatial distribution analysis of COVID19-related tweets represented the highest density of Twitter data in the counties of Oakland (Pontiac), Washtenaw (Ann Arbor), Kent (Grand Rapids), and the north of Wayne (Detroit) in Michigan, and Cuyahoga (Cleveland), Franklin (Columbus), and Hamilton (Cincinnati) in Ohio. These counties comprise the most populated cities with a high
rate of COVID-19 cases in Ohio and Michigan. In Michigan, the negative sentiments were more dispersed over the state than the positive ones (though the positive sentiment rate is more), while in Ohio the reverse happened. Thus, the Michigan policies regarding the Coronavirus pandemic needed to be more focused on less populated areas. Moreover, the spatial distribution of public opinions depended on the urban agglomeration, population, and key pandemic areas. A detailed investigation of the spatial analysis of the topic at the county level revealed the meaningful spatial distribution similarities between some of the topics, like “seeking help” and “monetary issues”, “lockdown” and “government responses”, and “fear & worry” and “the future of pandemic”. Such dependencies between the spatial distribution of the topics can help give direction to local government interventions and help manage a crisis.

In general, the topics and themes identified in this work using Latent Dirichlet Allocation can be used for further studies regarding COVID-19 pandemic-related topics on social media. The findings of this research on public opinions generated from social media and their vulnerability to the local government responses highlight the power of social media analysis for policymakers to understand people’s demands and the reflections of their policies at the county level. It is essential for the public health agencies to explore public interests, interactions, and reactions by analyzing the trends of any new local Coronavirus-related policies. It helps them to develop their immediate responses based on the dominant topics and emotions on social media and their geographical distributions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] G. Mertens, L. Gerritsen, S. Duijndam, E. Salemink, I.M. Engelhard, Fear of the coronavirus (COVID-19): predictors in an online study conducted in March 2020, J. Anxiety Disord. 74 (2020). 102528.

[2] X. Han, J. Wang, Using social media to mine and analyze public sentiment during a disaster: a case study of the 2018 Shougou City Flood in China, ISPRS Int. J. Geo-Inf. 8 (4) (2019) 185.

[3] Z. Wang, X. Ye, Social media analytics for natural disaster management, Int. J. Geogr. Inf. Sci. 32 (1) (2018) 49–72.

[4] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, D.S. Ebert, Public behavior response analysis in disaster events utilizing visual analytics of microblog data, Comput. Graph. 38 (2014) 51–60.

[5] J. Gimbse, M.H. Mosehbi, R.S. Patel, L. Brammer, M.S. Smolinski, L. Brilliant, Detecting influenza epidemics using search engine query data, Nature 457 (7232) (2009) 1012–1014.

[6] R. Nagar, et al., A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives, J. Med. Internet Res. 16 (10) (2014) e236.

[7] X. Ye, S. Li, X. Yang, C. Qin, Use of Social Media for the Detection and Analysis of Infectious Diseases in China, 2016.

[8] S.-Y. Shin, et al., High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea, Sci. Rep. 6 (1) (2016) 1–7.

[9] K. Lee, A. Agrawal, A. Choudhary, Real-Time disease surveillance using twitter data: Demonstration on flu and cancer, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 11288 (2013) 1474–1477.

[10] L. Sinnenberg, A.M. Bottenheim, K. Padrez, C. Mancheno, L. Ungar, R.M. Merchant, Twitter as a tool for health research: a systematic review, Am. J. Publ. Health 107 (1) (2017) e1–e8.

[11] L.N. Lopez-Lopez, et al., Using social media to identify sources of healthy food in urban neighborhoods, J. Urban Health 94 (3) (2017) 429–436.

[12] D. Ghosh, R. Guha, What are we ‘tweeting’ about obesity? Mapping tweets with topic modeling and Geographic Information System, Cartogr. Geogr. Inf. Sci. 40 (2013) 90–102.

[13] X. Chen, X. Yang, Does food environment influence food choice? A geographical analysis through ‘tweets’, Appl. Geogr. 51 (2014) 82–89.

[14] Q.C. Nguyen, et al., Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity, Appl. Geogr. 73 (2016) 77–88.

[15] W. Yang, L. Mu, GIS analysis of depression among Twitter users, Appl. Geogr. 60 (2015) 217–223.

[16] J.C. Eichstaedt, et al., Psychological language on Twitter predicts county-level heart disease mortality, Psychol. Sci. 26 (2) (2015) 159–169.

[17] J. Gibbs, et al., Twitter-based measures of neighborhood sentiment as predictors of residential population health, PLoS One 14 (7) (2019) 1–19.

[18] S.-F. Tsao, H. Chen, T. Tiseversanghe, Y. Yang, L. Li, Z.A. Butt, What Social Media Told Us in the Time of COVID-19: a Scoping Review, Lancet Digit. Heal., 2021.

[19] S. Liu, J. Liu, Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis, Vaccine 39 (39) (2021) 5499–5505.

[20] X. Han, J. Wang, M. Zhang, X. Wang, Using social media to mine and analyze public opinion related to COVID-19 in China, Int. J. Environ. Res. Publ. Health 17 (8) (2020).

[21] A. Benis, A. Seidmann, A.S. Ashkenazi, Reasons for taking the COVID-19 vaccine by US social media users, Vaccines 9 (4) (2021) 315.

[22] A.A. Mir, S. Rathinam, S. Gul, Perception of COVID-19 Vaccines from the Digital Footprints Left on Twitter: Analyzing Positive, Neutral and Negative Sentiments of Twitterati, Libr. Hi Tech., 2021.

[23] J.C. Lyu, E. Le Han, G.K. Luli, Covid-19 vaccine-related discussion on twitter: topic modeling and sentiment analysis, J. Med. Internet Res. 23 (6) (2021).

[24] N. Rosillo, et al., Real-time surveillance of COVID-19 space and time clusters during the summer 2020 in Spain, BMC Publ. Health 21 (1) (2021) 1–11.

[25] X. Zhou, et al., Monitoring global trends in Covid-19 vaccination intention and confidence: a social media-based deep learning study, medRxiv (2021).

[26] C. Villavicencio, J.J. Macrohon, X.A. Inbaraj, J.H. Jeng, J.G. Hsieh, Twitter sentiment analysis towards covid-19 vaccines in the Philippines using Naive Bayes, J. Med. Internet Res. 23 (6) (2021).

[27] A.C. Nagel, et al., The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets, J. Med. Internet Res. 15 (10) (2013) e2705.

[28] F. Heimerl, S. Lohmann, S. Lange, T. Ertl, Word cloud explorer: text analytics based on word clouds, in: 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1833–1842.

[29] A.C. Nagel, et al., The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets, J. Med. Internet Res. 15 (10) (2013) e2705.

[30] R. Singh, R. Singh, A. Bhata, Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics, Int. J. Adv. Sci. Res. 3 (2) (2018) 19–24.

[31] K. Chakraborty, S. Bhata, S. Bhattacharya, J. Platos, R. Bag, A.E. Hassanien, Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—a study to show how popularity is affecting accuracy in social media, Appl. Soft Comput. 97 (2020), 106754.
[33] No title.“ [Online]. Available: http://sentiment.nrc.ca/lexicons-for-research/.

[34] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, P. Badhani, B. Tech, Study of Twitter sentiment analysis using machine learning algorithms on Python, Int. J. Comput. Appl. 165 (9) (2017) 29–34.

[35] No title.” [Online]. Available: https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/.

[36] No title.” [Online]. Available: https://jackmckew.dev/sentiment-analysis-text-cleaning-in-python-with-vader.html.

[37] B. Dahal, S.A.P. Kumar, Z. Li, Topic modeling and sentiment analysis of global climate change tweets, Soc. Netw. Anal. Min. 9 (1) (2019) 1–20.

[38] X. Ye, S. Li, X. Yang, C. Qin, Use of social media for the detection and analysis of infectious diseases in China, ISPRS Int. J. Geo-Inf. 5 (9) (2016) 156.

[39] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.

[40] A.A. Aslam, et al., The reliability of tweets as a supplementary method of seasonal influenza surveillance, J. Med. Internet Res. 16 (11) (2014) e250.

[41] Y. Su, A. Venkat, Y. Yadav, L.B. Puglisi, S.J. Fodeh, Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities, Comput. Biol. Med. 132 (2021), 104336.

[42] A. Okabe, T. Satoh, K. Sugihara, A kernel density estimation method for networks, its computational method and a GIS-based tool, Int. J. Geogr. Inf. Sci. 23 (1) (2009) 7–32.

[43] No title.” [Online]. Available: https://coronavirus.ohio.gov/wps/portal/gov/covid-19/resources/news-releases-news-you-can-use.

[44] No title.” [Online]. Available: https://www.michigan.gov/coronavirus/0,9753,7-406-98158—Y,00.html.

[45] No title.” [Online]. Available: https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021.

[46] N. Puri, E.A. Coomes, H. Haghbayan, K. Gunaratne, Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases, Hum. Vaccines Immunother. 16 (11) (2020) 2586–2593.