International Journal of Information Management Data Insights 1 (2021) 100021

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore

Khairiyah Mohamed Ridhwan, Carol Anne Hargreaves*

Department of Statistics and Applied Probability, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
COVID-19
Singapore
Sentiment analysis
Topic modeling
Social media
VADER classifier
Latent Dirichlet Allocation (LDA)
Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM)

ABSTRACT

Microblogging has become one of the most useful tools for sharing everyday life events and news and for expressing opinions about those events. As Twitter posts are short and constantly being generated, they are a great source for providing public sentiment towards events that occurred throughout the COVID-19 period in Singapore. In this project, we perform sentiment analysis and topic modeling on the tweets about COVID-19 in Singapore, from 1 February 2020 to 31 August 2020. We accomplished this by collecting tweets discussing about COVID-19 and geolocated as 'Singapore', using the Python library 'SNSCRAPE'. We used the sentiments returned from the VADER lexicon-based classifier and emotions from pre-trained recurrent neural networks to find correlations between real-life events and sentiment changes throughout the whole period. From our analysis, we discovered an increase in tweets about COVID-19 during key periods such as the circuit breaker and found that the overall sentiment polarity was dominantly positive. However, emotion analysis revealed that there were changes in the prevalence of fear and joy emotions over time, due to real-life COVID-19 developments in Singapore. Additionally, sentiment polarity was found to differ from topic to topic.

Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. It has spread to numerous countries across all continents since its initial discovery in Wuhan, China, and was declared as a pandemic by the World Health Organization (WHO) on March 11, 2020. In the local context of Singapore, there have been a few key developments with regards to the effects of COVID-19. More specifically, on 7 February 2020, the Disease Outbreak Response System Condition (DORSCON) level was raised from Yellow to Orange, with a few local cases emerging without links to previous cases or travel history to China. On 3 April 2020, a "circuit breaker", a nationwide partial lockdown, was announced, where only essential services could continue to keep their premises open, all schools were closed, and students had to shift to home-based learning. A study (Tulshyan, Sharma & Mittal, 2020) indicated that, during the lockdown phase, while having strict measures in place, there were fewer cases predicted and the situation was stable and controlled. Singaporeans were encouraged to stay home as much as possible and the wearing of masks became mandatory when they left their homes. Three phases of planned reopening were announced on 19 May, namely "Safe Reopening" (Phase 1), "Safe Transition" (Phase 2) and finally "Safe Nation" (Phase 3); the third phase will last until an effective treatment or vaccine is found to stop the spread of COVID-19. Phase 1 started on 2 June, while Phase 2 started on 19 June. A general timeline of key developments in Singapore pertaining to COVID-19 is shown in Fig. 1.

Monitoring public health-related issues expressed on the internet through digital channels have been particularly useful to study how the public perceives several diseases, including H1N1 influenza (Chew & Eysenbach, 2010), Ebola virus (Lent, Sungur, Kunneman, Velde & Das, 2017; Odlum & Yoon, 2015), and Zika virus (Stefanidis et al., 20; Daughton & Paul, 2019; Pruss et al., 2019). Several studies have relied on data from the Twitter platforms and analysed data from early periods of the COVID-19 pandemic. A study (Chhetri, Goyal, Mittal & Battineni, 2021) focused on the prevalence of stress among students due to the COVID-19 pandemic and the closure of educational institutions. This study found that students were worried about their studies during the pandemic and the confinement at home, and these factors resulted in psychological pressure including depression, stress, phobia, fear, social disconnection, and so on.

Effective communication during a health crisis can ease public concerns and promote the adoption of important risk-mitigating behaviors. A study (Slavik, Buttle, Sturrock, Darlington & Yiannakoulias, 2021) examined the content and engagement of COVID-19 tweets authored by Canadian public health agencies and decision makers. They concluded that public health agencies and decision makers should examine what messaging best meets the needs of their Twitter audiences to maximize sharing of their communications. Further, they noted that public health

* Corresponding author.
E-mail address: carol.hargreaves@nus.edu.sg (C.A. Hargreaves).

https://doi.org/10.1016/j.jjimei.2021.100021
Received 25 March 2021; Received in revised form 7 June 2021; Accepted 7 June 2021

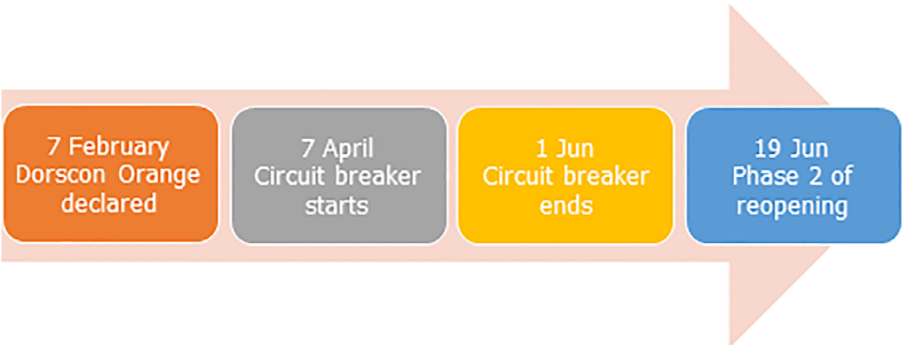

Fig. 1. Timeline of key events related to COVID-19 in Singapore.

accounts that do not currently employ risk communication strategies in their tweets may be missing an important opportunity to engage with users about the mitigation of health risks related to COVID-19.

There is, however, limited understanding of the changes in public sentiments and discourse about COVID-19 over time focusing on Singapore alone. Thus, in this study, we specifically aim to we perform sentiment analysis and topic modeling on the tweets about COVID-19 in Singapore, from 1 February 2020 to 31 August 2020 and answer the following research questions:

- What are the sentiments and emotions expressed in Singapore tweets about COVID-19?
- What are the dominant topics discussed on Twitter about COVID-19 in Singapore?
- How do public health-related topics vary over time?
- For topics discovered, what are the sentiments associated with it?

We organize the succeeding sections as follows. Section 2 presents the literature review, covering social media and its challenges and past studies; Section 3 describes our methodology. In Section 4, we present the results of our study. In Section 5, we discuss the theoretical contributions, implications to practice and future research directions. Finally, we conclude our findings in Section 6.

Literature review

Overview of social media and the twitter platform

Social media is considered as a rich digital channel for increased access to information with more transparency and liability (Andersen, Medaglia & Henriksen, 2012) and makes citizens that consult it more likely to be more informed and aware of health problems. For example, Huh et al. (2013) classified various online health communities, such as diabetes patients' community, using posts from WebMD.com to provide information. Tuarob, Tucker, Salathe and Ram (2014) also analysed user generated content related to health to detect various topics from patients' perspective.

Rathore, Kar and Ilavarasan (2017) showed that Twitter, a microblogging platform, is the most popular social media platform among other platforms because of its wide range of applications. A closer investigation on the dominance of Twitter data revealed specific reasons, for example, Twitter data can be easily extracted based on a keyword or hashtag search and can also be extracted based on a profile-specific timeline search. This ease of data collection using APIs facilitates a deeper analysis of Twitter data, which is often lacking in other platforms.

Several studies have relied on data from the Twitter platforms and analyzed data from early periods of the COVID-19 pandemic. The volume of data used in these studies ranges from a few hundred tweets to a few million. These studies have documented how Twitter users have reacted to the pandemic and their concerns in the early stages of the outbreak.

Sentiment analysis and topic modeling

Social media and its challenges

With the striking development of social media platforms, more and more people post online texts on different platforms to express their opinions on social issues (Zhang, Xu & Wan, 2012). The analysis of health event data posted on social media platforms not only provides first-hand evidence of health event occurrences but also enables faster access to real-time information that can help health professionals and policy makers frame appropriate responses to health-related events proactively. Although many studies analyze COVID-19 tweets on a worldwide scale, one of the challenges was that only a few zoom in on Singapore alone.

Sentiment analysis

Infectious diseases have a huge social and economic impact. They are caused by pathogenic microorganisms such as bacteria, viruses, parasites or fungi and they can be transmitted, directly or indirectly, from one person to another or from animals to humans (Zoonoses). Nowadays it is very important to detect the infectious diseases as soon as possible to prevent critical problems for the society. Garcia-Dias et al., (2018) proposed an approach for the sentiment classification of tweets related to infectious diseases.

To date, there are numerous studies surrounding sentiment analysis of tweets pertaining to COVID-19 on a worldwide scale. In fact, our project initially drew inspiration from the study by Medford, Saleh, Sumarsono, Perl and Lehmann (2020), who conducted sentiment analysis and topic modeling on tweets related to COVID-19 for a period of two weeks; one week before and after the Centers for Disease Control and Prevention (CDC) activated its Emergency Operations Center and the World Health Organization (WHO) released its first situation report about COVID-19. In terms of methodology, they conducted sentiment analysis using R's Syuzhet library (Jockers, 2015), and emotion analysis using pre-trained recurrent neural networks (Colnerič & Demšar, 2018). Their study revealed that the negative sentiment and fear emotion was increased after the announcement by WHO about the COVID-19 situation.

Many studies on sentiment analysis of tweets (Chandrasekaran, Mehta, Valkunde & Moustakas, 2020; Valdez, Thij, ten, Bathina, Rutter & Bollen, 2020) used the Valence Aware Dictionary for sEntiment Reasoning (VADER), a lexicon-based method specifically attuned to social media texts such as tweets (Hutto & Gilbert, 2014). Further, many studies that analyzed tweets adopted (Medford et al., 2020; Wang & Wei, 2020) pre-trained recurrent neural networks to identify Ekman's six basic emotions (joy, surprise, fear, anger, disgust, and sadness). In this study, we also adopted the pre-trained recurrent neural networks to identify Ekman's six basic emotions (joy, surprise, fear, anger, disgust, and sadness).

Further, while most studies focused on the general sentiment analysis of tweets, Jang, Rempel, Roth, Carenini and Janjua (2021) conducted a more focused study approach which monitored the COVID-19 discourse and found that there were negative sentiments related to the overall outbreak, but positive sentiments related to physical distancing. This indicated that different topics were likely to be related to different sentiment polarity, and thus conducting solely sentiment analysis cannot reveal much without coupling the results with topic modeling. We therefore decided to include topic modeling in our study.

Topic modeling

Gosh, S., et.al (2017) demonstrated that temporal topic trends extracted from disease-related news reports successfully captured the dynamics of multiple outbreaks such as whooping cough in U.S. (2012), dengue outbreaks in India (2013) and China (2014). Their observations also suggested that, when news coverage was uniform, efficient modeling of temporal topic trends using time-series regression techniques can estimate disease case counts with increased precision before official reports by health organizations.

A topic modeling study by Colnerič et al. (2018) revealed that topic modeling produced ten general topics pertaining to COVID-19, ranging from economic and political impact to the prevention efforts. Further, a notable study that analyzed Singapore's COVID-19 conversations on a social media platform was conducted by Shorey, Ang, Yamina and Tam (2020). This study performed a qualitative analysis on 2075 comments of 29 local Facebook news articles relating to COVID-19 in Singapore, dated from 23 January 2020 (first case of COVID-19 detected and announced in Singapore) to the 3 April 2020 (announcement of circuit breaker). Their analysis revealed that fear and concern were the main reasons behind the public's responses, and the emerging themes were about panic buying, hoarding, reality, and expectations about the situation, staying positive amid the 'storm', worries about the future. However, it is important to note that these findings are specific to the topics of the local news articles at hand and hence may be biased based on the articles selected by the authors for the study.

Thus, in our study, instead of limiting it to selected local news articles, we examined tweets that mentioned COVID-19 keywords throughout the study period of seven months, from 1st February 2020 (a week before Singapore was declared to be DORSCON Orange) to 31st August 2020 (a few months after Phase 2 of reopening). To focus the scope of the project on Singapore tweets, tweets were filtered according to their location as Singapore. Topic modeling was coupled with sentiment analysis to discover more hidden details about the COVID-19 tweets and to provide more context to better explain the sentiment patterns.

For topic modeling, Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) was used to discover topics within the dataset, while Latent Dirichlet Allocation (LDA) was used to choose the number of topics to set in the GSDMM model.

Methodology

In this section, we collect the relevant tweets, filter, clean and preprocess tweets, perform an overall sentiment analysis using VADER and an overall emotion analysis using pre-trained RNN. Thereafter, we perform topic modeling using GSDMM to obtain the topics that are present in the corpus, making use of LDA to find an upper bound of the optimal number of topics. Finally, we examine the VADER sentiments of each topic produced from the GSDMM model and analyze the sentiments that are associated with each topic. Fig. 2 presents an overview of the methods of the project.

Data collection, cleaning and pre-processing

We first collected the relevant tweets, filter, clean and preprocess tweets. We utilized SNSCRAPE, a Python 3 library that can be used to access old tweets (JustAnotherArchivist 2021). To obtain tweets related to the coronavirus, the keywords used for the query are "Wuhan OR Wuhan virus OR Wuhan coronavirus OR coronavirus OR nCoV2019 OR 2019nCoV OR COVID OR SARS-COV-2 OR circuit breaker". The term "circuit breaker" represents Singapore's version of a nationwide partial lockdown. We set the language parameter "Lang" as "en" to pick up English tweets, and the parameter "near" to the value "Singapore", which will pick up tweets that have geo-location coordinates within or near Singapore, as well as tweets whose user is from Singapore.

To prepare the data for analysis, we removed duplicated tweets based on their Tweet.Id. Tweets repeated by the same user consecutively, with the only difference in the tweets being the user mentions or hyperlinks were removed by first removing user mentions (@), hyperlinks and general stop words, leaving the tweet with only its text content. Thereafter, tweets with duplicate text content are eliminated. Tweets that were a mixture of English with other languages, spam or advertisement tweets, and tweets from Malaysia that had been picked up due to tweet's location coordinates proximity were all removed.

We applied separate pre-processing techniques for sentiment analysis and topic modeling. To pre-process the tweets for sentiment analysis, we were conservative in our approach. More specifically, we only removed hyperlinks, URLs, user mentions and excessive spaces. We did not remove emoticons, hashtags and retained the punctuations and original case of the words as they may affect the results of the sentiment and emotion classification.

For topic modeling, we conducted a more extensive and rigorous procedure for data pre-processing. We first removed 'RT' (representing re-tweet) and user information, web links, hashtags, audio/video tags or labels, lower case the tweet, stripped punctuations, removed double spacing and numbers. Next, we tokenized the tweet using a regular expression tokenizer from Python's Natural Language Toolkit (NLTK). Then, we use the WordNet lemmatizer from NLTK to lemmatize the tokens. Thereafter, we remove tokens that were in NLTK's stop words list, drop tokens of 3 or less characters, and form bigrams using the remaining tokens. We then removed custom stop words and phrases (Appendix A). We finally obtained a total of 48,057 unique terms. Upon filtering out terms that appeared at least 10 times, this reduced to 1028 terms. Only for topic modeling, we discarded tweets that have length 0 and 1, resulting in a total of 4511 tweets, and then transformed these tweets into bag-of-words vectors.

Sentiment and emotion analysis

Text mining implementation on seafarers' medical documents to generate better knowledge of medical issues was performed by Chintalapudi, N., et.al (2021), where both the Lexicon-based method and the Naïve Bayes' algorithms were adopted to perform sentimental analysis. To identify the sentiment and emotion of the tweets in our dataset, we present two main types of methods for sentiment or emotion analysis, namely:

1 Lexicon-based method
2 Deep learning-based method

An overview of the steps used for this section in shown in Fig. 3 below.

Lexicon-based method: VADER

A popular lexicon-based method to detect sentiment polarity (positive, negative, neutral) in tweets is using VADER, a tool which analyses the sentiments of the tweets and classifies the tweets based on a dictionary of words. Unlike usual dictionaries, the dictionary used by VADER include emoticons, slangs, contractions, negations, and acronyms commonly used in casual online conservations. VADER also takes care of word-order sensitive relationships between terms and consider degree modifiers which will impact sentiment intensity.

For our project, we classified the sentiment of each tweet as positive, negative, or neutral based on the VADER compound score. The compound score considers the value of each word in the tweet and is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between −1

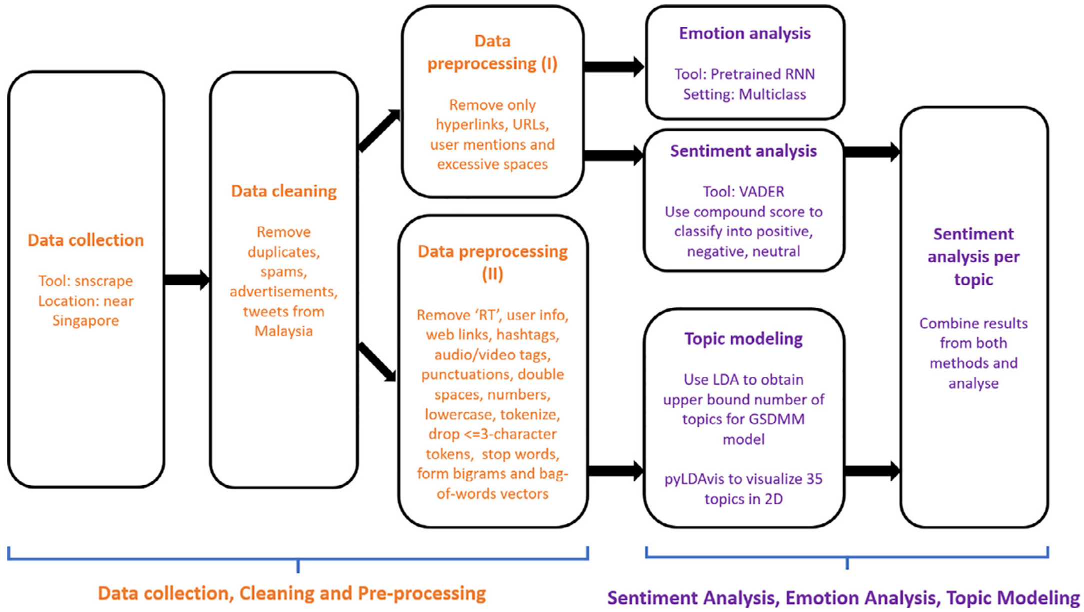

Fig. 2. Overview of project.

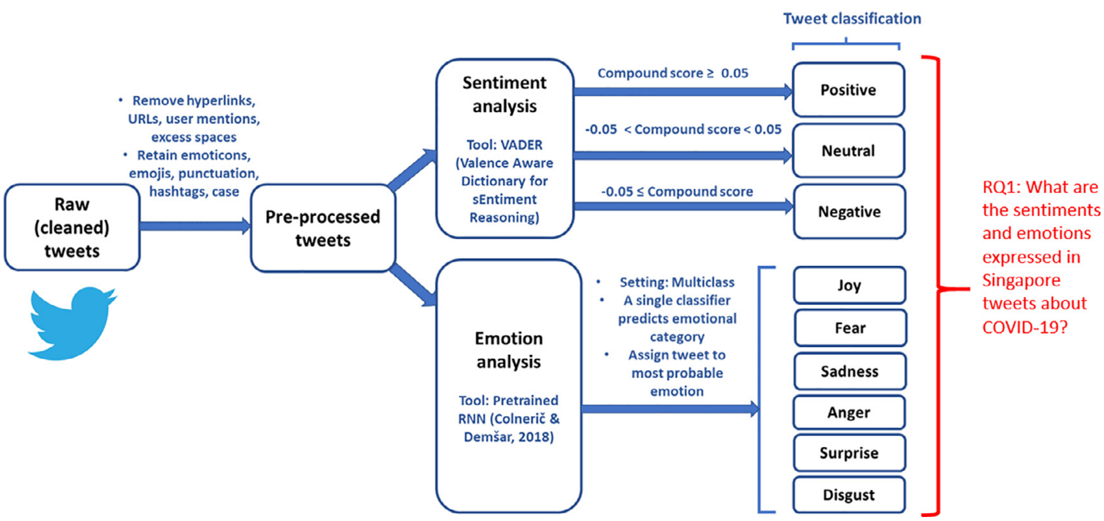

Fig. 3. Diagram depicting the Sentiment and Emotion Analysis steps.

(most extreme negative) and +1 (most extreme positive). A tweet with a compound score greater than or equal to 0.05 was classified as positive, a tweet with a score of between −0.05 and 0.05 was classified as neutral, and a tweet with a score less than or equal to −0.05 was classified as negative. These are the standardized thresholds based on literature (Hutto & Gilbert, 2014). For example, this threshold was also adopted by other studies (Botchway, Jibril, Kwarteng, Chovancova & Oplatková, 2019; Chandrasekaran et al., 2020; Pano & Kashef, 2020) who conducted sentiment analysis on tweets using VADER.

Deep learning based method: Recurrent Neural Network (RNN)

An example of a deep learning-based method to predict emotions from English tweets is the use of a trained recurrent neural network (RNN) model by Colnerič and Demšar (2018). Based on their experiments on a data set of 73 billion tweets, the best classification result was obtained by the Ekman emotion classification which achieved a micro-averaged F1-score of 73.0 percent using the character-based RNN. It implements Long-Short-Term Memory (LSTM) layers (Hochreiter & Schmidhuber, 1997). The model has been successfully used to detect emotions in tweets by other studies (Medford et al., 2020; Wang & Wei, 2020) and have shown good results when doing emotion analysis of tweets. Thus, we employ this model in our study to classify the tweets into their prevailing emotions. This model was conducted in Python using Keras (Chollet, 2015) library.

A single-labeling approach was utilized on our tweets, where we assign a single primary emotion for each tweet. This was done by using the multiclass setting in the model for the classification of emotion, where a single non-binary classifier was built to predict the first emotional category, disregarding any other emotional hashtag present later in the tweet. Classifiers in this setting only must pick the most probable emotion from a set of all possible emotions. Thus, each tweet will be assigned only one emotion which is the most probable. This resulted in a collection of tweets as well as their emotional categories.

Topic modeling

To identify the topics of the tweets in our dataset, we present two main types of topic modeling methods, namely:

1 Latent Dirichlet Allocation (LDA)

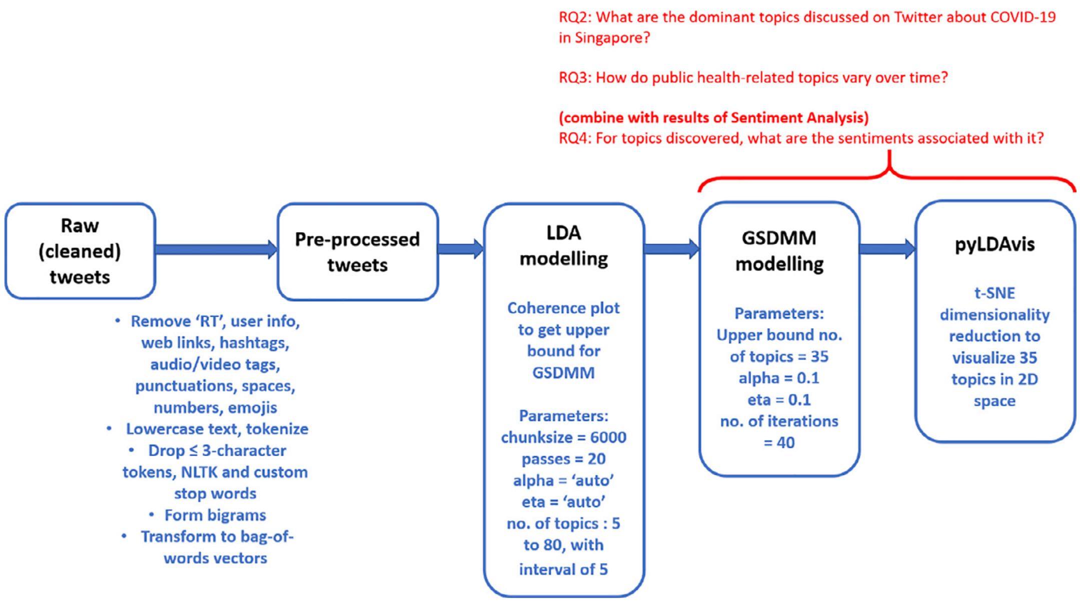

Fig. 4. Diagram depicting Topic Modeling steps.

2 Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM)

The number of topics is a crucial parameter in topic modeling and capable of making these topics human interpretable. In this project, we use the coherence score (Röder et al., 2015) as an indicator to choose the best number of topics for the model. Since the current implementation of GSDMM using the MovieGroupProcess package (Yin & Wang, 2014) does not allow for the calculation of a coherence score to determine the optimal number of topics like LDA. We make use of LDA to first derive an estimate of the upper bound for the number of topics for the GSDMM topic model. An overview of the steps used for this section in shown in Fig. 4 below.

Latent Dirichlet Allocation (LDA)

LDA was introduced by Blei et al. (2003), and it is a three-level hierarchical Bayesian model which is one of the most popular generative probabilistic topic modeling algorithms. LDA has been cited by many works (Medford et al., 2020; Valdez et al., 2020) that conducted topic modeling on COVID-19 tweets. LDA assumes that each tweet is modelled as a finite mixture over an underlying set of topics. Each topic is, in turn, modelled as an infinite mixture over an underlying set of topic probabilities.

To estimate an upper bound number of topics to set for the GSDMM model, we tune a LDA topic model over several topics, and then choose the number of topics that gives the higher coherence score in an elbow plot. To perform LDA, we followed the guidelines set by Řehůřek (2020) to select four main parameters: chunksize, passes, alpha, and eta. We set chunksize as 6000, which is larger than the total number of tweets, thus taking in all tweets in one go. We set passes to 20; chosen by setting an argument eval_every = 1, and then checking if most tweets have converged by the final passes. We finally set alpha = 'auto' and eta = 'auto', which will automatically learn these parameters. Using these settings, we trained LDA models with number of topics ranging from 5 to 80 with an interval of 5 and compute the coherence score of each topic model. Below we display the coherence measure against the number of topics in LDA (Fig. 5). Based on Fig. 5, a plausible upper bound estimate for the number of topics to be set for the GSDMM model are 30, 35, 40 or 45 topics.

Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM)

GSDMM, a short-text topic modeling method proposed by Yin and Wang (2014), is similar to LDA, except instead of assuming each tweet belongs to multiple topics, it assumes that each tweet only belongs to one topic and the words in each tweet are dependent on the topic. Since tweets are only a maximum of 280 characters long (Twitter, 2021), it is more likely that each tweet will discuss about only one topic rather than multiple topics. Thus, we will explore and perform topic modeling on the tweets using GSDMM.

We experimented setting the upper bound of GSDMM model with 30, 35, 40, and 45 topics. We finally chose a model with 35 topics among all models because 35 topics showed diverse and less redundant topics when manually examined. The parameters alpha and beta are both set to 0.1 as used in the original paper, and the number of iterations is set to 40 as GSDMM does not require a high number of iterations for the algorithm to converge (Yin & Wang, 2014).

We visualized the topics of the 35-topics GSDMM model interactively using a pyLDAvis visualization, where 35 topics is displayed into a graphable two-dimensional space and similar topics are grouped together. This was done by setting the 'mds' parameter as 'tsne', as we used the dimensionality reduction method, t-SNE (Van der Maaten & Hinton, 2008) for clustering of the topics which had been similarly used by Medford et al. (2020) in his COVID-19 tweets study.

Findings

What are the sentiments and emotions expressed in Singapore tweets about COVID-19?

An overall sentiment analysis on the tweets using the VADER lexicon-based method and was performed and Fig. 6 presents the overall sentiment distribution of the tweets collected during the study period.

Among the tweets in our dataset, the proportion of positive tweets is the highest, forming 50% of the whole dataset. This is followed by the proportion of negative tweets which makes up 26% of total tweets and lastly, neutral tweets with the lowest proportion of 24%. This indicates that despite the spread of COVID-19, the community showed a dominant positive sentiment during the study period. While these results may seem surprising, this conclusion had also been derived by other researchers who have conducted country-level sentiment analysis (Yin, Yang & Li, 2020). To examine the trends of the sentiment over time during the study period, we will also plot the number of tweets for each sentiment over time (Fig. 7).

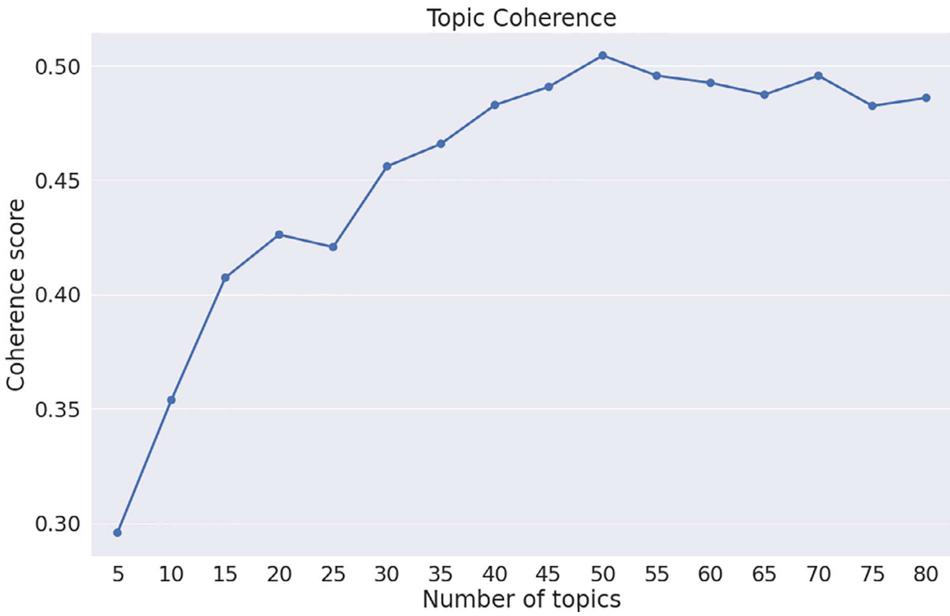

Fig. 5. Coherence score against number of topics for LDA.

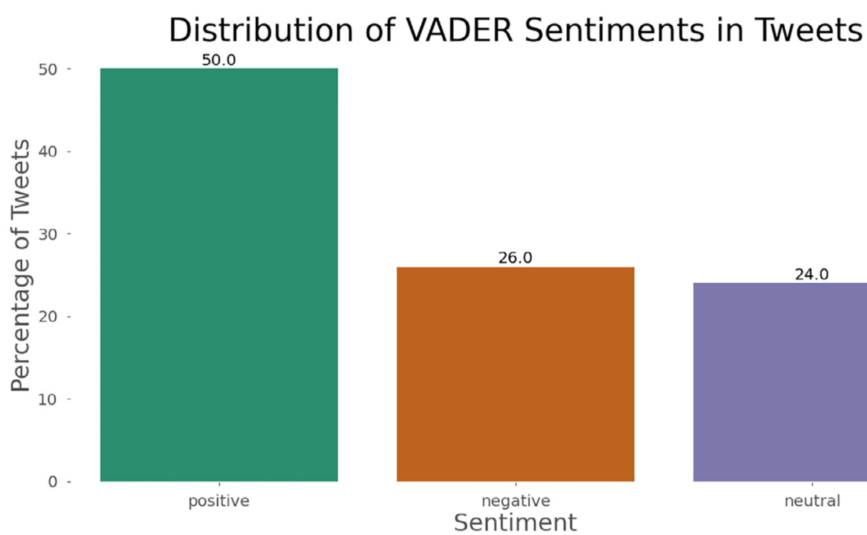

Fig. 6. Proportion of each VADER sentiment (positive, negative, neutral) in the dataset.

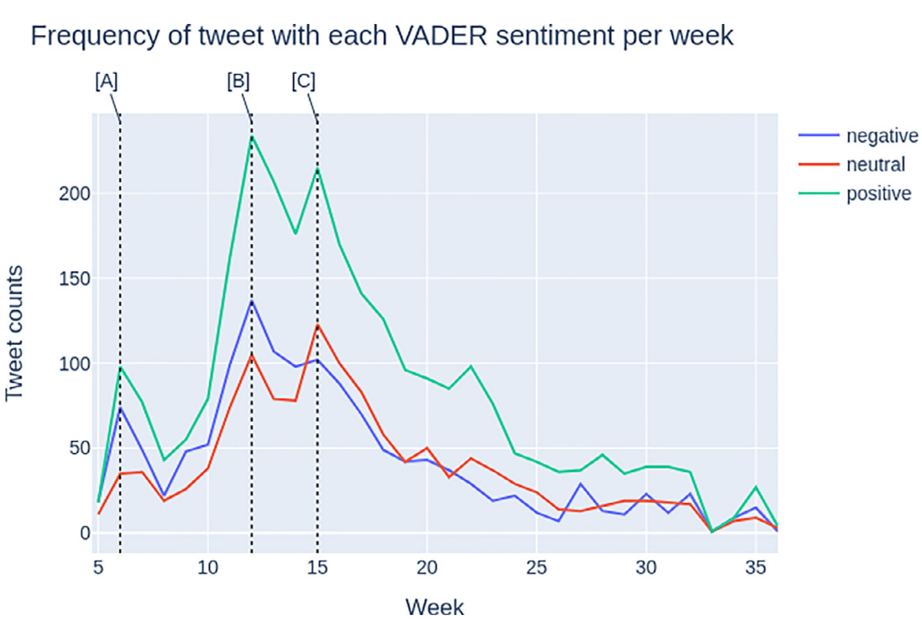

Fig. 7. Number of tweets with each VADER sentiment by week, where [A]. Week 6 (3 Feb to 9 Feb), [B]. Week 12 (16 Mar to 22 Mar), [C]. Week 15 (6 Apr to 12 Apr).

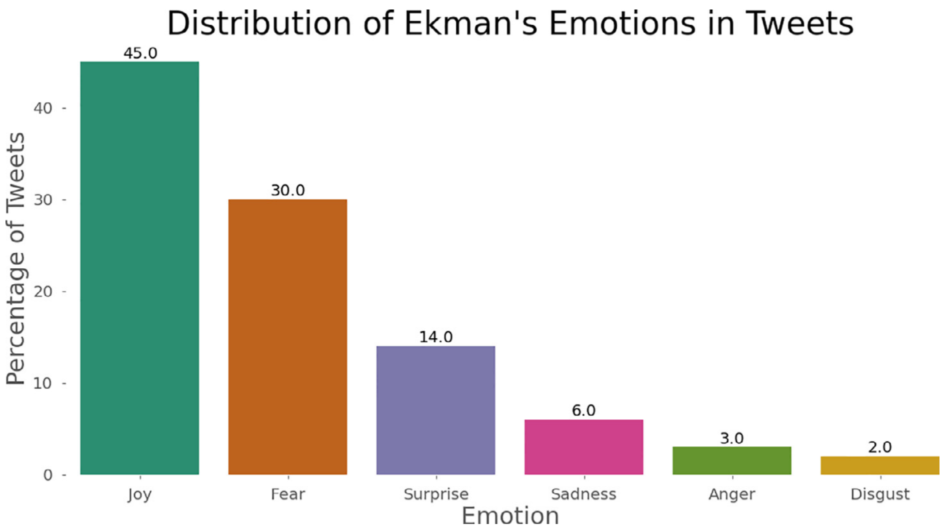

Fig. 8. Ekman's emotions analysis results.

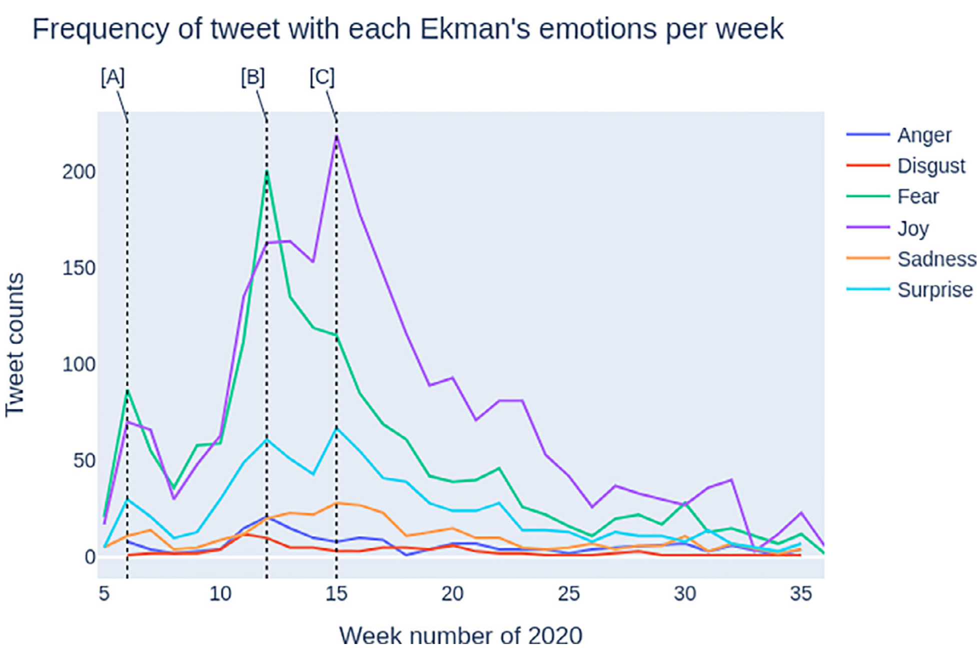

Fig. 9. Ekman's emotions over by week, where [A]. Week 6 (3 Feb to 9 Feb), [B]. Week 12 (16 Mar to 22 Mar), [C]. Week 15 (6 Apr to 12 Apr).

The plot of the frequency each VADER sentiment over the study period (Fig. 2) showed that although in general positive tweets made up the dominant percentage of total tweets, there are certain time periods where the frequency of positive and negative tweets was closer to each other, such as in the early weeks of the study. This indicated that there was a higher than usual proportion of negative tweets in the earlier period of the study. More specifically, we observed that the negative sentiments have some peaks where we observe a large spike in negative tweets, such as on Week 6 and Week 12. We focused less on the neutral tweets as they did not indicate a strong leaning towards either side of the sentiment spectrum.

We derive a possible explanation for these changes in sentiment by examining the results of emotion analysis using the pre-trained RNN by Colnerič and Demšar (2018). We classified the emotion dominant in each tweet as either anger, disgust, fear, joy, sadness, or surprise, based on the Ekman's emotions scale.

Fig. 8 above shows that the dominant emotion among all tweets is Joy, which is expressed in 45% of all tweets. This is then followed by Fear (30%), Surprise (14%), Sadness (6%), Anger (3%) and lastly, Disgust (2%). We plot the evolution of emotions by weeks (See Fig. 9) to examine the trends of emotion over time.

The frequency of Ekman's emotions by week (See Fig. 9) reveals that Joy dominated most of the time, except for in Week 6 (3 February to 9 February) and Week 12 (16 March to 22 March) where Fear is dominant. We examined these anomalies closer and derived a possible explanation for these changes in emotion for the time periods of Week 6 and Week 12. For the weekly trends of Ekman's emotion in terms of Fear, we saw a spike in Week 6 (3 February to 9 February) and Week 12 (16 March to 22 March). In Week 6, multiple important events with regards to COVID-19 have unfolded. Within this week on 4 February, Singapore saw its first case of infection in the community that was not linked to Wuhan, while on 7 February, DORSCON Orange was declared in Singapore. This had raised alarm among people in Singapore and thus resulted in the spike of Fear in Week 6. [Examples: "Oh no! It's getting too close for comfort. #WuhanCoronavirus #singapore #sg" and "Oddest thing about the panic buying in Singapore due to the Coronavirus? No spring onions! #coronavirussingapore"].

This is further supported by the word cloud on tweets classified as Fear in Week 6 (Fig. 10), which showed a higher frequency of words such as 'Singaporean' and 'confirm', referring to the first confirmed COVID-19 case in the community as well as advice such as 'stay safe', and the usage of 'mask' as prevention, and also 'panic buy' which was

Fig. 10. Word cloud for top 100 words among tweets with Fear emotion in Week 6.

an event that resulted as a reaction from Singapore's declaration as DORSCON Orange.

Similarly, within Week 12 (16 March to 22 March), there were significant events related to COVID-19. On 16 March, Singapore saw the biggest single-day spike of cases and on 21 March, Singapore saw its first two deaths from the virus. It is evident that the Fear emotion was present in tweets [Examples: "#coronavirus comes full circle in Singapore - after a period of calm, 6 weeks on from a short burst of panic-buying, we're back at it due to fears of imported cases" and "With #COVID—19 touching the new heights, let us stay home to be safe and protect others at this challenging times! #StayAtHome #coronavirus #CoronavirusPandemic"] as a reaction from the unraveling of these events. This is further supported by the word cloud on tweets classified as Fear in Week 12 (Fig. 11), where words such as 'die' and 'death' increased in frequency, as well as preventive measures such as staying 'home', and 'social distance'.

Meanwhile, the Joy emotion peaked in Week 15 (6 April to 12 April), were 7 April, which marked the start of circuit breaker and on 12 April, Easter celebrations. We also plot the word cloud (Fig. 12) for tweets with Joy sentiment in Week 15 of 2020 to corroborate our findings, where the Joy emotion stems from staying at home, working from home, and celebrating Easter celebrations. [Examples: "Almost a week into the #circuitbreaker!" and "How's everyone holding up? Anyway, have a great Easter Sunday!"].

What are the dominant topics discussed on Twitter about COVID-19 in Singapore?

A total of 35 topics was obtained from the final GSDMM model. Within the pyLDAvis visualization of topics (Fig. 13), the bubbles represent clusters of topics, and a larger 'bubble' represents a higher frequency of the topic while hovering each 'bubble' allows us to examine what are top words for each topic cluster.

We made use of our own judgement to assign labels to each of the 35 topics. This was done by checking the list of words that have a high probability of belonging to each topic and manually inspecting samples of tweets which weigh strongly on each topic to get a better sense of the content of actual tweets with a high topical content. We finally consolidated our results for all topics and assigned a label to each topic. The entire set of labelled 35 topics is listed in Appendix B. Although there are a total of 35 topics, some topics can be grouped under main themes (Table 1).

The discovered themes, such as Themes 2 to 7, were related to public health promotions and interventions such as circuit breaker, staying home and staying safe, social distancing, wearing face masks, and travel restrictions. Other topics included news about former US President Donald Trump and China, showing support for healthcare and frontline workers, and updates on new COVID-19 cases and clusters. To examine which topics were dominant in the tweets, we plot each topic with their expected topic proportions in terms of percentage of total tweets (Fig. 14).

The top six most prevalent topics in Singapore tweets are Topics 18 (Gratitude for time at home with loved ones), 10 (Migrant workers and healthcare workers), 33 (Stay home), 21 (Activities amidst COVID), 28 (Lockdown news), 34 (Counts of cases and deaths worldwide), as shown in Table 2 above.

We can group Topics 18 and 21 into the common theme of stay home activities, while Topic 33 (Stay home) is also a related topic. Hence, the tweets about staying home refer to the top Topics 18, 21, and 33. In total, these three topics this amounts to 18% of total tweets. This indicates that a significant portion of the tweets were discussing about staying at home and their activities while doing so.

Topic 10 regarding migrant workers and healthcare workers, appeared to be discussed frequently Singapore's Twitter users, showing that there is concern regarding the foreign worker dormitory cases. Meanwhile, Topics 28 and 34 about lockdown news and counts of cases and deaths worldwide are among the top six topics, indicating concerns about worldwide cases and the disease outbreak.

How do public health-related topics vary over time?

While Fig. 14 showed which topics were dominant in the conversation on twitter about COVID-19 in Singapore, we were also interested to assess changes in public health related topics of discussion over time. Thus, we plot the distributions of months for each of the 10 public health

Fig. 11. Word cloud for top 100 words among tweets with Fear emotion in Week 12.

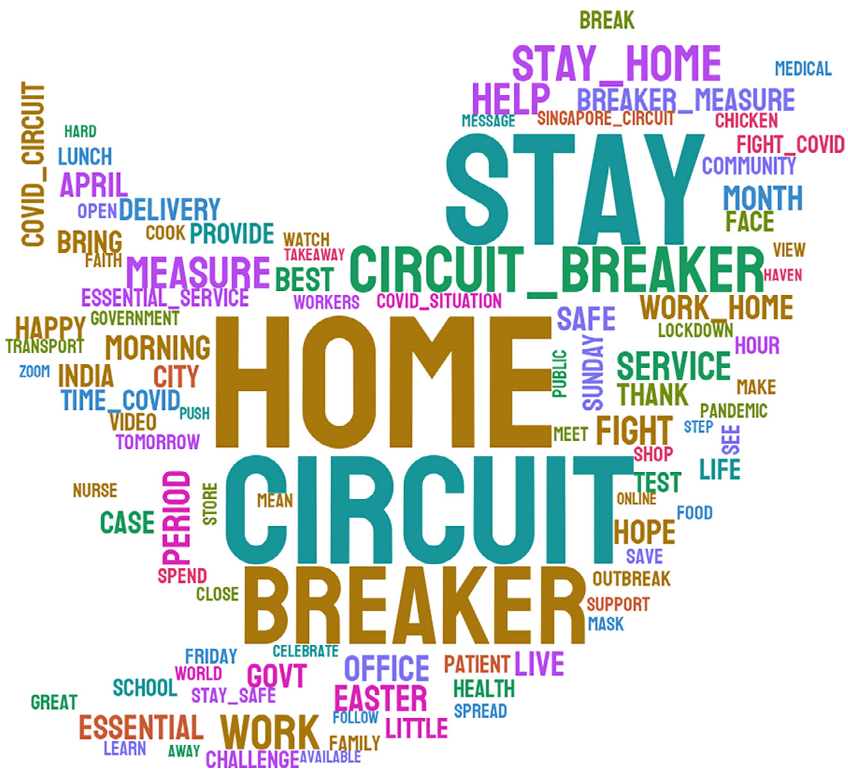

Fig. 12. Word cloud for top 100 words among tweets with Joy emotion in Week 15.

related topics that fall under Themes 2 to 7 during the study period, in Fig. 15.

We observed that the topic distribution across the months during the study period was not uniform.

Topics that peaked only in March were Topics 12 (Contact tracing), 27 (Social distance) and 30 (Cancelled flights and trips due to COVID). The increasing discussion on contact tracing parallels the launch of the Trace-Together contact tracing mobile app in Singapore (Kit & Mahmud, 2020), and the increasing discussion on cancelled overseas trips parallels Singapore's move to tighten border restrictions and travel (Menon & Goh, 2020). Additionally, social distancing was also actively promoted in March.

Topics that peaked only in April were Topics 8 (Stay home and stay safe), 9 (Closures due to circuit breaker), 19 (Circuit breaker), and 33 (Stay home). More specifically, Topic 19 (Circuit breaker), had the highest peak, which was expected as the circuit breaker in Singapore com-

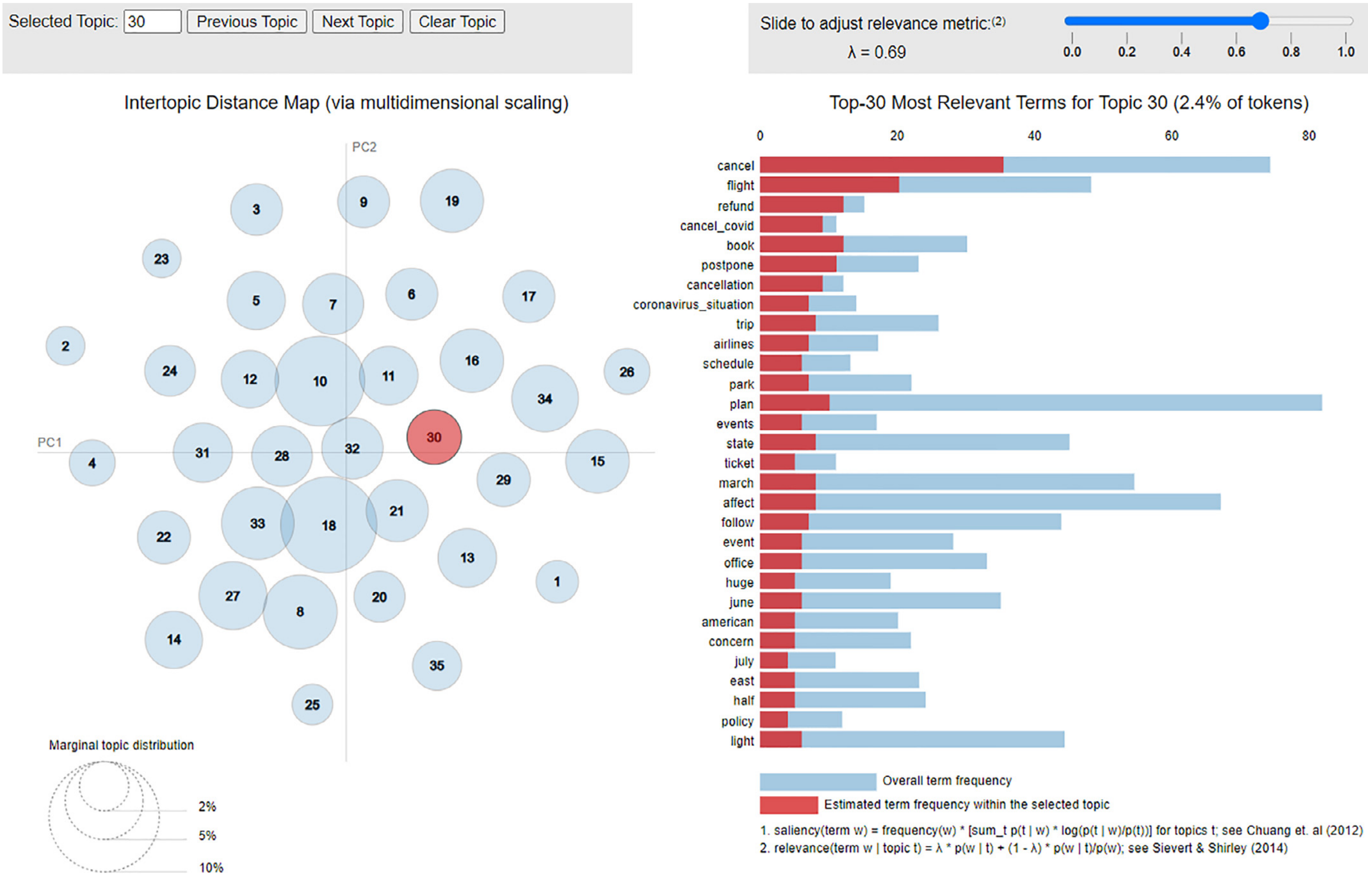

Fig. 13. A snapshot of pyLDAvis interactive visualization of all 35-topics GSDMM model, hovering over Topic 30 (Canceled flights and trips due to COVID) in red.

Table 1
Examples of possible themes with their corresponding topic numbers and the corresponding percentage of tweets.

Theme No.	Themes	Topic No.	Total percentage of tweets
1	Stay home activities during COVID-19	18, 20, 21	15.3%
2	Circuit breaker and closures	9, 19	4.7%
3	Stay home and stay safe	8, 22, 33	10.8%
4	Social distancing	27	2.8%
5	Mask wearing, sanitizers, handwashing	13	2.6%
6	Contact tracing	12	2.7%
7	Travel and border restrictions	1, 30	3.6%
8	Lockdown news and counts of cases and deaths worldwide	28, 34	7.1%
9	News about former US President Donald Trump and China	4, 7	5.4%
10	Showing support for healthcare and frontline workers	3	2.3%
11	Updates on new COVID-19 cases and clusters	15, 26	4.0%
12	Migrant workers and healthcare workers	10	6.3%

Table 2
Top 6 topics (in terms of percentage) in the corpus.

Topic No.	Topic label	Frequency count	Percentage of total tweets
18	Gratitude for time at home with loved ones	414	9.2%
10	Migrant workers and healthcare workers	282	6.3%
33	Stay home	233	5.2%
21	Activities amidst COVID	161	3.6%
28	Lockdown news	160	3.6%
34	Counts of cases and deaths worldwide	159	3.5%

menced on 3 April. Singapore's version of a partial lockdown, the circuit breaker, was announced on 3 April and started on 7 April, and this explained the rise of this topic in April. Similarly, the other two topics were in general related to staying home.

Topics that peaked both in March and April were Topics 1 (Travel restrictions), 13 (Wearing mask) and 22 (Stay safe). For these topics, their prevalence was almost the same for both March and April.

For topics discovered, what are the sentiments associated with it?

To discover which topic is contributing to the dominant positive sentiment revealed in tweets of COVID-19, we examined the proportion of VADER sentiment of tweets in every topic from our GSDMM model. We plot the number of tweets for each topic with their respective sentiment polarity in Fig. 16 below.

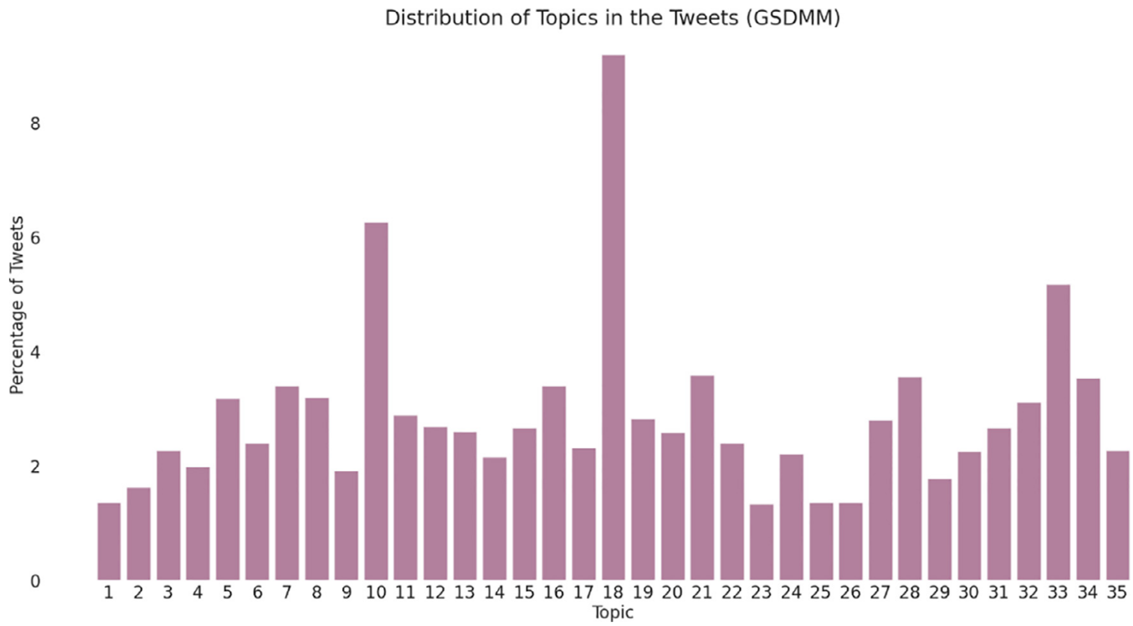

Fig. 14. GSDMM topic distribution among tweets in terms of percentage.

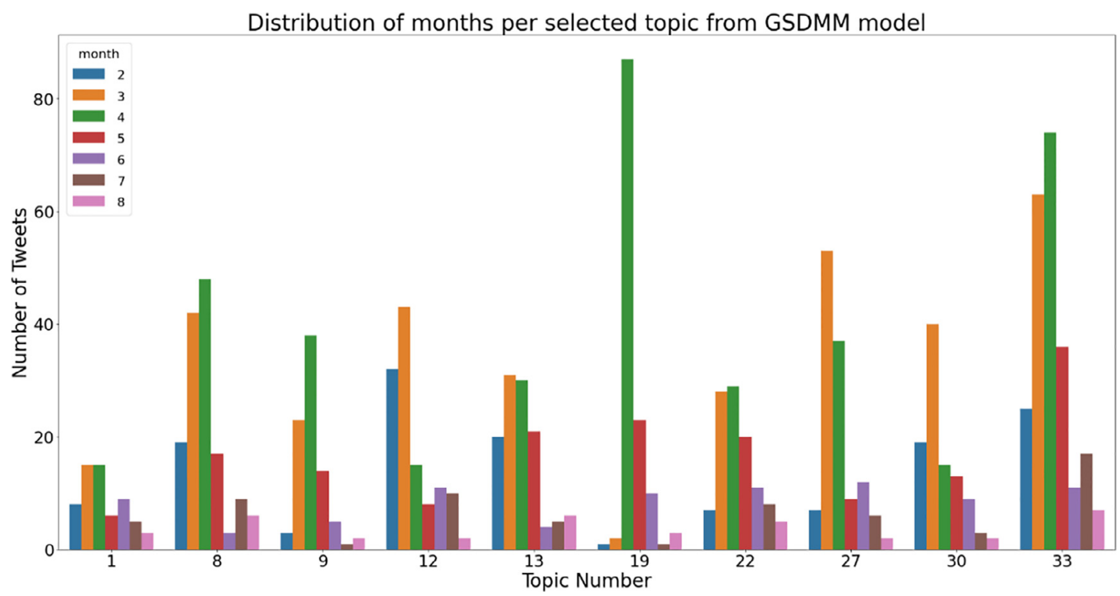

Fig. 15. Topic number of 10 public health related topics against raw count of tweets, with each color representing different month's.

Majority of the topics contain the positive sentiment as the dominant sentiment, except for Topics 7 (News about Trump and racist remarks), 30 (Cancelled flights and trips due to COVID) and 34 (Updates on number of deaths from COVID) where negative sentiments dominate within each topic. We also observe that the topic contributing the most to the overall positive sentiment is Topic 18 (Gratitude for time at home with loved ones), indicating that the public are appreciating staying home and spending time with their loved ones amidst the pandemic [Examples: "See the good in everything, because of Covid there's so many silly, fun videos of family celebrating Raya. Spread love not hate" and "Happy #520 everyone, may stronger relationships be forged amidst this #Circuit Breaker, bringing better understand between you and your love ones."].

We zoom in into public health topics that fall under Themes 2 to 7 during the study period in Fig. 17 to investigate people's opinion (positive, negative, or neutral) towards these topics. This can be viewed as feedback by the public about the COVID-19 measures used to contain the spread of COVID-19 in Singapore.

We observe that the distribution of sentiments varies from topic to topic. For all topics except Topic 30 (Cancelled flights and trips due to COVID), the proportion of positive sentiment is higher than that of the negative sentiment, indicating that the public generally has a positive response to these measures of COVID-19. Thus, for the topics on social distancing, circuit breaker, wearing mask, and staying home and staying safe, positive sentiment was dominant. However, Topic 30 regarding cancelled flights is met with mostly negative sentiments, demonstrating that the public is upset with the travel restrictions [Examples: "I have to cancel my Vietnam trip coz of covid-19, I'm so sad, we've been planning this trip since last year September,:(" and "unfortunately, we're going to cancel our flight this coming May. I'll let you know when this coronavirus is over"].

Thus, overall, this indicates that the public is supportive of most preventive measures for COVID-19 public health intervention. This acts a good feedback for the intervention as it possibly implies a strong level of compliance among Singaporeans for the preventive measures implemented to curb the spread of COVID-19.

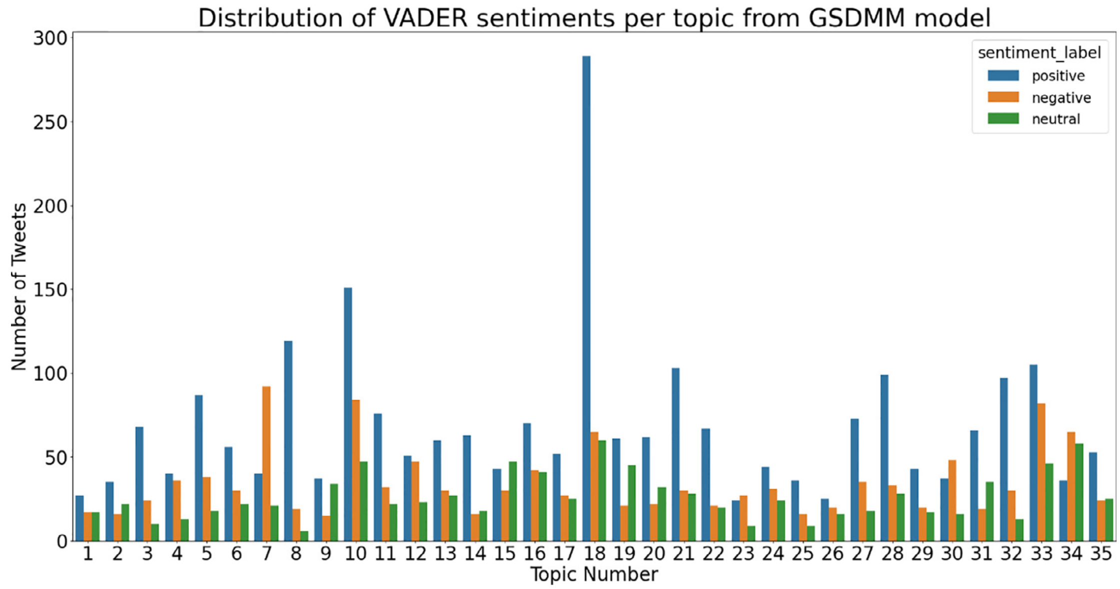

Fig. 16. Proportion of each VADER sentiment category for all topics in the corpus.

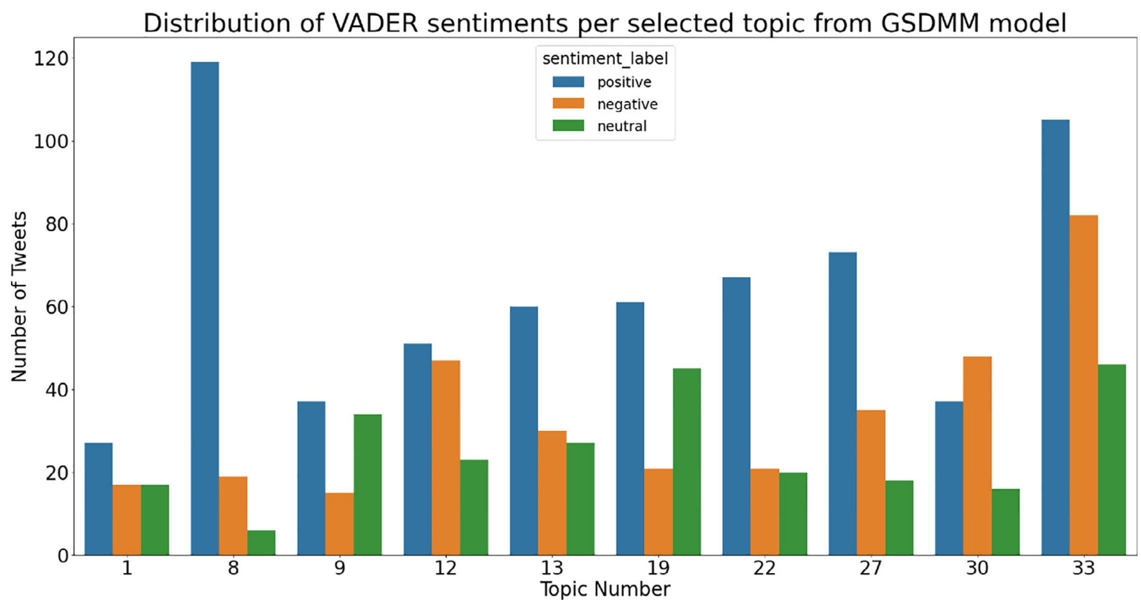

Fig. 17. Proportion of each VADER sentiment category for selected public health related topics.

Discussion

Theoretical contributions

This study analysed Singapore tweets for a period of seven months from 1 February 2020 to 31 August 2020, which to the best of our knowledge, had not been done so far.

Results of the sentiment analysis showed that positive sentiments made up about half of the tweets in the dataset, while negative and neutral sentiments each made up about a quarter of the dataset. While this may give us the idea that the sentiment is generally positive during the COVID-19 outbreak in Singapore, emotion analysis revealed that the emotion prevalence throughout the study period is not uniform and there are certain time periods where negative emotions such as Fear peaks. Upon our further inspection, these were mainly triggered by developments in Singapore regarding the COVID-19 outbreak, such as the DORSCON Orange declaration, to the first local community cases, and the first death in Singapore that have created fear and anxiety amongst Twitter users in Singapore.

On the other hand, emotions such as Joy was also in abundance when users tweeted about staying at home and spending time with family. Topic modeling also revealed that the dominant topics present in the conversation on Twitter regarding COVID-19 was about staying home, and these were dominantly positive, which had contributed to the dominant positive sentiment during the study period.

Implications to practice

Although results of this study have shown that the overall sentimental polarity was positive, sentiment polarity varied from topic to topic. This shows the importance of analysing the sentiment per topic to understand the tweets better. In terms of topics related to public health interventions, topics about social distancing, the encouragement to stay home and stay safe as well as the wearing of masks was met with a positive response from Twitter users. This serves as a feedback in analysing how the community is perceiving the public health efforts as it is an indicator that these preventive measures are well-received and hints at a strong compliance amongst Twitter users in Singapore. However, tweets pertaining to travel and border restrictions due to COVID-19 demonstrated more negative sentiments. This indicates that users were in general less positive with the travel and border restrictions.

Overall, our results demonstrated that the proactive and carefully calibrated measures taken by the Government Multi-Ministry Taskforce during the COVID-19 period were well supported by the community as evident by the positive sentiments in Twitter. In general, Singapore's public health communication for COVID-19 was seen as rational, transparent, and frequent. The multiple communication approaches that were adopted in both traditional and social media channels by multiple government agencies seems effective, evident by the positive sentiments in Twitter. Nonetheless, there is always room for improvement, and we hope that our recommendations and insights can contribute to shaping policy as well as implementation of measures for future pandemics. We also hope that our methodology will aid in analysing social media trends for public health issues in Singapore.

Future research directions

We believe that there are more ways to explore the tweet dataset at hand. For example, a Twitter user network analysis can be performed to analyze which users are the central and more influential characters in the conversation about COVID-19 outbreak in Singapore. Researchers may also be interested in adopting our topic-based sentiment analysis methodology for analysis on texts from other social media platforms such as Facebook and Reddit, whose users may be formed by age groups different from that of Twitter users' demographics. Overall, our results demonstrated that the COVID-19 measures implemented in Singapore has been effective in managing the pandemic. Nonetheless, there is always room for improvement, and we hope that our recommendations and insights will aid in the future implementation of policies for future pandemics to come.

Conclusion

We extracted English tweets dated from 1 February 2020 to 31 August 2020 discussing about COVID-19 and geo-located as 'Singapore'. Sentiment analysis showed that half of all tweets were of positive sentiments. Emotion analysis revealed that nearly half (45%) of all tweets expressed joy while 30% expressed fear. A major topic present in the conversation on Twitter regarding COVID-19 in Singapore was about staying at home. Public health topics such as social distancing, the encouragement to stay at home and stay safe, as well as the wearing of masks was met with a positive response from Twitter users, while travel and border restrictions due to COVID-19 was dominated by negative sentiments. Overall, our results demonstrated that measures taken by the Singapore government during the COVID-19 period were well supported by the community as evident by the positive sentiments on Twitter.

Declaration of Competing Interest

The authors report no conflict of interest.

Acknowledgements

The authors would like to thank Arpan Kumar Kar, (PhD) editor-in-chief of Science of the International Journal of Information Management Data Insights and the reviewers, who have provided us with valuable feedback and helped to improve the quality of our paper.

Funding Acknowledgement

Nil Funding

Appendix A

Custom Stop-words for Text Pre-processing in Text Analysis

"pm","t.co","http","https","amp","t","t.c","c","rt","pl","sg","singapore", "covid19","coronavirus", "covid","covid19","wuhan", "s", "p", "like", "im","new", "day", "days","year", "ur", "ve", "la", "ive", "cos", "guys", "didnt", "time", "people", "dont", "today", "thing", "week", "months", "post","yesterday", "man", "wont", "uk","st", "lets", "don", "feel", "feeling", "gonna","isnt", "pls", "share", "wait", "wanna", "na", "back", "means","fuck", "fucking", "lah", "good","due", "sa", "ingat", "sa covid", "just", "will", "can", "now", "get", "go", "us","can", "one", "even", "just", "ada", "ke", "got", "going", "last", "etc.", "kaypo", "still", "chibai", "woodlands","singapore", "woodlands_singapore", "say", "know", "virus","situation", "need", "want", "take", "come", "look","think", 'actually', 'especially', 'later', 'guess', 'note', 'dear', 'road', 'start', 'stop', 'things', 'give', 'try', 'tell', 'shit', 'maybe', 'keep', 'right','time', 'march_march', 'stacy', 'time_coronavirus', 'amid_covid','ingat_covid', 'iwas','mall','bedok', 'control', 'lawrence', 'wong', 'lawrence_wong', 'feel_like'

Appendix B

Topic No.	Topic label	Top words/phrases with highest probability
1	Travel restrictions	Travel, Temperature, Restrictions, Baby, Screen, Check, Travel_Restrictions, Airport, Flight, Help
2	Swab test	Test, Free, Swab, Changi, Airport, Changi_Airport, Home, Covid_Free, Live, Afternoon
3	Thanking frontline workers	Fight, Help, Fight_Covid, Support, Thank, Nurse, World, Frontline, Workers, Fight_Coronavirus
4	Spread news	News, Spread, China, Media, Viruses, Message, Alcohol, Blame, Fake, Trump
5	Hope to find vaccine	Outbreak, Soon, Hope, Vaccine, World, End, Economy, Travel, Cancel, Coronavirus_Outbreak
6	World post-covid	World, Post_Covid, Travel, Outbreak, Covid_World, Learn, Coronavirus_Outbreak, Online, Spread, Meet
7	Trump and China world news	Chinese, China, Trump, World, Deaths, Happen, Cure, Die, Number, Death
8	Stay home and stay safe	Stay, Safe, Home, Stay_Safe, Stay_Home, Fight, Healthy, Pray, Home_Stay, Safe_Healthy
9	Closures due to circuit breaker	Close, Work, Service, Circuit, Breaker, Circuit_Breaker, Measure, Month, Essential, Food
10	Migrant workers and healthcare workers	Workers, Health, World, Work, Pandemic, Crisis, China, Company, Help, Care
11	Impact of COVID	Impact, Lose, Help, Job, Hotel, Support, Service, Room, Bring, Impact_Covid
12	Contact tracing	Contact, Trace, Contact_Trace, Government, Outbreak, Fear, Better, Panic, Level, Change
13	Wearing mask	Mask, Wear, Wear_Mask, Government, Fight, Case, Face, Hand, Fight_Covid, Protect
14	Preventive measures	Spread, Safe, Measure, Government, Distance, Spread_Covid, Covid_Spread, Circuit, Breaker, Circuit_Breaker
15	Updates on COVID cases and clusters in Singapore	Case, Covid_Case, Cluster, Foreign, Case_Singapore, Workers, Work, Link, Include, Local
16	Vaccine trials	Patients, Vaccine, Covid_Patients, Case, Work, News, Study, Health, Report, Test
17	Positive test result for COVID	Test, Positive, Test_Positive, Positive_Covid, Covid_Test, Work, Test_Covid, Weeks, Doctor, India
18	Gratitude for time at home with loved ones	Thank, Love, Happy, Home, Celebrate, Period, Birthday, Family, Away, Month
19	Circuit breaker	Circuit, Breaker, Circuit_Breaker, Measure, Covid_Circuit, Singapore_Circuit, Breaker_Measure, April, Home, Stay,
20	Enjoying time at home during COVID	Home, Chicken, Food, Work, Cook, Work_Home, Video, World, Delivery, Speak
21	Activities amidst COVID	Best, School, Meet, Home, Work, Love, Pandemic, Read, Parent, Friends
22	Stay safe	Stay, Safe, World, Stay_Safe, Live, Life, Season, Second, Place, Gotta
23	A mixture of topics related to COVID	Work, Water, Drink, Miss, Close, Talk, Outside, Fight, Hard, Run
24	COVID pandemic	Pandemic, Covid_Pandemic, Crisis, Hold, Call, Mask, Handle, Health, Covid_Crisis, Cause
25	A mixture of topics related to COVID	Paper, Toilet, Toilet_Paper, Stay, Morning, Forward, Walk, Look_Forward, Check, Good_Morning
26	Updates on cases and death count	Case, Total, Update, March, Deaths, Recover, Count, Death, Source, Covid_Update
27	Social distance	Distance, Social, Social_Distance, Stay, Home, Hand, Safe, Mask, Help, Practice
28	Lockdown news	Lockdown, News, Live, Home, Life, Covid_Lockdown, Love, Soon, Talk, Play
29	Spread of cases	Spread, Case, Novel, Novel_Coronavirus, Prevent, Test, Help, Hong, Ncov, Prevent_Spread
30	Cancelled flights and trips due to covid	Cancel, Flight, Refund, Book, Postpone, Plan, Cancel_Covid, Cancellation, Trip, Affect
31	Updates on Singapore's COVID situation by Prime Minister	Minister, Covid_Situation, Current, Live, Prime, Long, Prime_Minister, Nation, Current_Covid, Address
32	Expressing gratitude amidst COVID	Thank, Join, Watch, Period, Social, Friends, Media, Covid_Situation, Pandemic, China
33	Stay home	Home, Stay, Work, Stay_Home, Live, Soon, Life, Away, Work_Home, Save
34	Counts of cases and deaths worldwide	Case, Deaths, Korea, Covid_Case, China, Report, Italy, Confirm, South, Test
35	A mixture of topics related to COVID	Open, Update, Happen, China, Online, Watch, Thank, Malaysia, Movie, Safe

Topics and their labels from GSDMM model k= 35.

References

Andersen, K. N., Medaglia, R., & Henriksen, H. Z. (2012). *Social media in public health care: Impact domain propositions* October. Aalborg University's Research Portal https://vbn.aau.dk/en/publications/social-media-in-public-health-care-impact-domain-propositions.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*. https://dl.acm.org/doi/10.5555/944919.944937.

Botchway, R. K., Jibril, A. B., Kwarteng, M. A., Chovancova, M., & Oplatková, Z. K. (2019). A review of social media posts from UniCredit bank in, Europe: a sentiment analysis approach. In *Proceedings of the 3rd International, Conference on Business and Information Management* https://dl.acm.org/doi/abs/10.1145/3361785.3361814.

Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. *Journal of Medical Internet Research*. https://www.jmir.org/2020/10/e22624/.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PloS one*. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0014118.

Chhetri, B., Goyal, L. M., Mittal, M., Battineni, G., et al. (2021). Estimating the prevalence of stress among Indian students during the COVID-19 pandemic: a cross-sectional study from India. *Journal of Taibah University Medical Sciences, 16*(2), 2060–2267. https://www.sciencedirect.com/science/article/pii/S1658361221000032.

Chintalapudi, N., Battineni, G., Di Canio, M., Sagaro, G. G., & Amenta, F. (2021). Text mining with sentiment analysis on seafarers' medical documents. *International Journal of Information Management Data Insights, 1*(1), Article 100005. https://www.sciencedirect.com/science/article/pii/S2667096820300057.

Chollet, F. (2015) https://keras.io.

Colnerič, N., & Demšar, J. (2018) https://ieeexplore.ieee.org/document/8295234.

Daughton, A. R., & Paul, M. J. (2019). Identifying protective health behaviors on Twitter: observational study of Travel Advisories and Zika Virus. *Journal of medical Internet research*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6535980/.

Garcia-Dias, J. A., Apolinario-Arzube, O., Medina-Moreira, J., Luna-Aveiga, H., Lagos-Ortiz, K., & Valencia-García, R. (2018). *Sentiment Analysis on Tweets Related to Infectious Diseases in south america. Digital library*. ACM https://dl.acm.org/doi/abs/10.1145/3293614.3293647.

Ghosh, S., Chakraborty, P., Nsoesie, E. O., Cohn, E., Mekaru, S. R., & Brownstein, J. S. (2017). Temporal topic modeling to assess associations between news trends and infectious disease outbreaks. *Scientific Reports, volume 7*, 40841. Article number https://www.nature.com/articles/srep40841 .

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*. https://dl.acm.org/doi/10.1162/neco.1997.9.8.1735.

Huh, J., Yetisgen-Yildiz, M., & Pratt, W. (2013). Text classification for assisting moderators in online health communities. *Journal of Biomedical Informatics*. https://www.sciencedirect.com/science/article/pii/S1532046413001391.

Hutto, C., & Gilbert, E. (2014). VADER: a parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

Jang, H., Rempel, E., Roth, D., Carenini, G., & Janjua, N. Z. (2021). *Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study using Topic Modeling and Aspect-Based Sentiment Analysis* February 10 https://www.jmir.org/2021/2/e25431/.

Jockers, M. L. (2015). *Syuzhet: Extract Sentiment and Plot Arcs from Text*. GitHub. https://github.com/mjockers/syuzhet.

JustAnotherArchivist. (2021). *JustAnotherArchivist/snscrape* (n.d.). https://github.com/JustAnotherArchivist/snscrape.

Kit, T. S., & Mahmud, A. H. (2020a). *Singapore Launches Tracetogether, Mobile App to Boost COVID-19 Contact Tracing Efforts* March 21.

Lent, L. G. G. van, Sungur, H., Kunneman, F. A., Velde, B. van de, & Das, E. (2017a). Too far to care? measuring public attention and fear for Ebola using Twitter. *Journal of Medical Internet Research*. https://www.jmir.org/2017/6/e193/.

Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020a). "Infodemic": Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. *OUP Academic*. https://academic.oup.com/ofid/article/7/7/ofaa258/5865318.

Menon, M., & Goh, T. (2020a). Tighter border restrictions as coronavirus cases in Singapore go past 800. *The Straits Times*. https://www.straitstimes.com/singapore/health/tighter-border-restrictions-as-cases-here-go-past-800.

Odlum, M., & Yoon, S. (2015a). What can we learn about the Ebola outbreak from tweets? *American Journal of Infection Control*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4591071/.

Pano, T., & Kashef, R. (2020a). *A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets During the Era of COVID-19* November 9. MDPI https://www.mdpi.com/2504-2289/4/4/33.

Pruss, D., Fujinuma, Y., Daughton, A. R., Paul, M. J., Arnot, B., & Szafir, D. A. (2019a). Zika discourse in the Americas: a multilingual topic analysis of Twitter. *PloS one*. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0216922.

Rathore, A. K., Kar, A. K., & Ilavarasan, P. V. (2017a). Social media analytics: literature review and directions for future research. *Decision Analysis*. https://econpapers.repec.org/article/inmordeca/v_3a14_3ay_3a2017_3ai_3a4_3ap_3a229-249.htm.

Řehůřek, R. (2020a). *Gensim: Topic Modelling for Humans* November 4. LDA Model - gensim https://radimrehurek.com/gensim/auto_examples/tutorials/run_lda.html.

Röder, M., Both, A., & Hinneburg, A. (2015a). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* https://dl.acm.org/doi/10.1145/2684822.2685324.

Shorey, S., Ang, E., Yamina, A., & Tam, C. (2020a). *Perceptions of public on the COVID-19 outbreak in singapore: A qualitative content analysis* July 10. OUP Academic https://academic.oup.com/jpubhealth/article/42/4/665/5869607.

Slavik, C. E., Buttle, C., Sturrock, S. L., Darlington, J. C., & Yiannakoulias, N. (2021a). Examining tweet content and engagement of Canadian Public Health Agencies and decision makers during COVID-19: mixed methods analysis. *Journal of Medical Internet Research, Vol.23*(Iss.3), 1–18 https://www.jmir.org/2021/3/e24883//pdf.

Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L., & Jacobsen, K. H. (2017a). *Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts*. JMIR Public Health and Surveillance https://publichealth.jmir.org/2017/2/e22/.

Tuarob, S., Tucker, C. S., Salathe, M., & Ram, N. (2014a). An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. *Journal of biomedical informatics*. https://pubmed.ncbi.nlm.nih.gov/24642081/.

Tulshyan, V., Sharma, D., & Mittal, M. (2020a). *An Eye on the Future of COVID'19: Prediction of Likely Positive Cases and Fatality in India Over a 30 Days Horizon Using Prophet Model* (pp. 1–7). Disaster Medicine and Public Health Preparedness.

Twitter. (2021a). *Counting characters | twitter developer*. Twitter https://developer.twitter.com/en/docs/counting-characters.

Valdez, D., Thij, M., ten, Bathina, K., Rutter, L. A., & Bollen, J. (2020a). Social media insights into US Mental health during the COVID-19 pandemic: longitudinal analysis of Twitter Data. *Journal of Medical Internet Research*. https://www.jmir.org/2020/12/e21418/.

Van der Maaten, L., & Hinton, G. (2008a). Visualizing data using t-SNE. *Journal of machine learning research, 9*(11).

Wang, J., & Wei, L. (2020a). Fear and hope, bitter and sweet: emotion sharing of Cancer Community on Twitter. *SAGE Journals* https://journals.sagepub.com/doi/full/10.1177/2056305119897319#_i7.

Yin, H., Yang, S., & Li, J. (2020). *Detecting Topic and Sentiment Dynamics, due to COVID-19 Pandemic Using Social Media* July 5. arXiv.org https://arxiv.org/abs/2007.02304.

Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* https://dl.acm.org/doi/10.1145/2623330.2623715.

Zhang, W., Xu, H., & Wan, W. (2012a). Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications, 39*(11), 10283–10291.