MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220169

Vaccine Rollout and Shift in Public Sentiment: Twitter-Based Surveillance Study

Uzair Shah[a], Md Rafiul Biswas[a], Khalid Mahmoud Mohammed Dolaat[a], Mowafa Househ[a], Zubair Shah[a], Tanvir Alam[a]

[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha 34110, Qatar*

Abstract

This study aims to find out the variation of Twitter users' sentiment before and after the COVID-19 vaccine rollout. We analyzed all COVID-19 related tweets posted on Twitter within two timeframes: September 2020 (T1) and March 2021 (T2). A total of 3 million tweets from over 132 thousand users were analyzed. We then categorized the users into two groups whose overall sentiment shifted positively or negatively from T1 to T2. Our analysis showed that 27% of users' sentiment shifted from T1 to T2 positively and the users were more confident about vaccine safety and efficacy. Users reported positive sentiments about travelling and the easing of lockdown measures. Also, 20.4% of the users' sentiment shifted negatively from T1 to T2. This group of Twitter users were more concerned about the adverse side effects of the vaccine, the pace of vaccine development as well as the emerging novel coronavirus variants. Interestingly, over half of the users' overall sentiment remained the same in both periods of T1 and T2, indicating indifference about vaccine rollout. We believe that our analysis will support the exploration of public reaction to COVID-19 vaccine rollout and assess policy makers' decision to combat the pandemic.

Keywords:

COVID-19, Twitter, Vaccine

Introduction

People share their opinions and thoughts on recent issues through microblogging websites. Twitter is one of the most popular social media platforms for the public to express their opinions and share informaiton [1]. Twitter is also a great source of information to investigate people's interactions and sentiments [2]. To understand the fears, facts, and prevailing public concerns, Twitter could be considered as a useful platform that has often served as a source of communication during disease outbreaks worldwide [3] including the current COVID-19 pandemic. The number of affected people, outbreaks location and time of outbreak can also be discovered through Twitter data. More importantly, the public perception to take preventive measures against COVID-19 can be identified through Twitter data analysis and this valuable source of information could be used in a meaningful way to combat against pandemic.

Though the tweets are informative, the use of informal language challenges the downstream analysis and decision-making process. Twitter data is unstructured and short in length to explore. The tweets are usually expressed in informal language and the colloquial tones (i.e., it doesn't follow grammatical rules and its short form) and slang words make it more challenging to interpret [4]. Emoticons which are also used in tweets to express human facial expression are also difficult to interpret in all scanarios [5]. Moreover, the vast number of conspiracy theories regarding Coronavirus are moving into the social media sphere at the current pandemic situaton [6]. While the health organizations and the governments are working together to mitigate the spread of COVID-19, misinformation flowing in social media hinders the way COVID-19 is dealt with.

There exist multiple studies focusing on COVID-19 related tweets to summarize all the tweets and retweets to discover overall public sentiment. One study [7] determined the perspective of Twitter users towards coronavirus disease and relevant tweets were retrieved from Twitter. The authors found that most of the sentiments were positive during the COVID-19 pandemic and were related to respect for government, health workers, etc. To monitor public health's real-time discussion and concern, Twitter can be a promising social media to influence public trust by analyzing the sentiment [8]. However, people's sentiments change over time. The study reported that from Jan 2019 to 23 Mar 2020, the maximum number of the tweets were neutral or negative sentiments. Another study performed between December 2019 and May 2020 showed more positive sentiments [9]. The authors identified temporal factors like weather, day of the week, type of interaction, and locations of tweets that could play crucial role in sentiment variations [10].

Our current study aimed to identify the change in public sentiment globally regarding COVID-19 in two different timeframes: September 2020 'before vaccine rollout' (T1) and March 2021 'after vaccine rollout' (T2). We considred March 2021 as T2 because the vaccine rollout kicked off at January – February, 2021 in most parts of the world [11]. We hypothesized that after the vaccine rollout, the public sentiment would be impacted positively. However, because the misinformation and anti-vaccine campaigns in social media the public sentiment could shift negatively. Therefore, we collected the tweets data related to COVID-19 in the two timeframes; before and after vaccine rollout which will support the proper analysis of public sentiment as well as identify the variation in their sentiment. We leveraged sentiment analysis and topic modeling to explore the public perception and opinions in these two timeframes.

Methods

Study Data

To address the mentioned objective, the tweets related to COVID-19 were collected. Publicly available datasets related to COVID-19 tweets [12] were used in the current study. The authors are collecting COVID-19 related tweets since March 2020 and updating the GitHub repository on weekly basis. The authors used following terms to collects the tweets: ("coronavirus", "Coronavirus Pandemic", "COVID-19", "COVID19", "2019nCoV", "CoronaOutbreak", "WuhanVirus"). Due to Twitter privacy concern we were able to collect the tweet IDs, language, and created time. Before downloading the full tweets, we removed the non-English and retweet IDs. The September 2020 (T1) and March 2021(T2) tweet IDs were filtered out for the current study on time and resource constraints. Twarc (a command-line tool) and Python library were utilized to download the full tweets from T1 and T2.

Data Preprocessing

After downloading all the tweets, the verified users were identified using the tweet meta-data. We removed the tweets from verified users from the government or non-government organizations or public figures as their tweets usually remain positive and consistent. The text of tweets was normalized to lower case letters, and hyperlinks, hashtags, and user mentions were replaced by empty strings. After that, punctuations and stop words were removed from the text (Figure 1). Additionally, for topic modeling, the words of the tweets were converted to its lemma word using WordNetLemmatizer Python library. We filtered out those who had less than two tweets in both periods. Finally, all the tweets from the selected users were split into two time periods (T1: Sep 2020 and T2: Mar 2021).

Sentiment Analysis of a Tweet

To find the sentiment in the tweets, SentiStrength (Lexicon-based sentiment analysis) Python library was utilized [13]. In lexicon-based approaches, the text was represented by bag-of-words. The sentiment values from dictionary were assigned to all negative and positive words and ultimately calculated the overall score. The SentiStrength library reports the two sentiment strengths (positive and negative) of the text. Therefore, we categorized the tweets into positive (if the overall positive sentiment score is greater than the negative sentiment score), negative (if the overall negative sentiment score is greater than the positive sentiment score), and the rest were considered neutral.

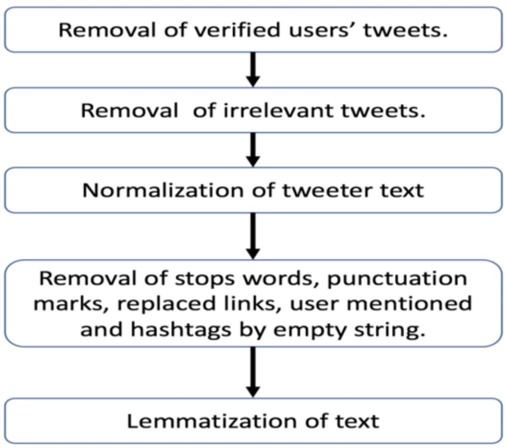

Figure 1– Workflow of data preprocessing

Computation of Users' Sentiment

To analyze the overall sentiment of a user in T1 or T2, we counted the sentiment (positive: P, negative: N, neutral: n) of each tweet from a user at T1 and T2. Then we decided the overall sentiment of a user at a timeframe (T1 or T2) based on the majority voting. However, if there is a tie between P and others (P-N, P-n, P-N-n) then we assigned positive sentiment for the user at a particular timeframe. In case of tie (N-n), negative sentiment (N) was assigned to the user for a timeframe.

Computing Change of Sentiment

Based on the overall sentiment of a user at T1 and T2, we categorized them into two groups: i) **G1**: The group of users, whose sentiments were positively changed from T1 to T2. We considered users whose sentiment either changed from negative (at T1) to positive/neutral (at T2); or neutral (at T1) to positive (at T2). ii) **G2**: The group of users, whose sentiments changed negatively from T21 to T2. This group of users' sentiment changed from positive (at T1) to negative/neutral (at T2); or neutral (at T1) to negative (at T2) (Figure 2).

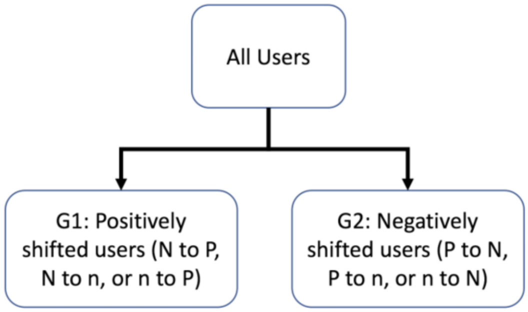

Figure 2– Workflow of splitting the tweets for topic modeling. G1 and G2 representing the group of users. T1 and T2 representing the periods. N, P, and n representing the negative, positive, and neutral sentiments, respectively.

Topic Modeling and Thematic Analysis

The topic modeling was performed on each group (G1 and G2) of users at two timeframes (T1 and T2) to identify the topics of discussion from the users. Latent Dirichlet allocation (LDA) was used from Python sklearn library for this purpose. LDA is an unsupervised learning algorithm mainly used to cluster the collection of documents. Before fitting LDA on the dataset, it requires a fixed numbers of topic sets. Based on our analysis, 15 topics were proposed by the LDA (Supplementary File S1). To extract the relevant themes from the identified topics, we manually analyzed topics represented by unigram and bigram words produced through LDA. Then we analyzed the tweets manually from each cluster. After manual analysis, we proposed three main themes for G1 and G2 at T1 and T2. We highlighted 20 random Tweets from each themes under Supplementary File S2.

Results

Tweets Collection

As shown in Figure 3, a total of 7,304,646 tweets were collected from T1. Of those tweets, 728,407 (10%) verified users' tweets were removed. Of the remaining 6,612,239 non-verified users' tweets, 1,895,774 (28.7%) were non-related to the search terms and were excluded. These tweets were captured in the search either because of search terms matches with user profiles or names. Furthermore, a total of 4,817,138 tweets were collected from T2. Of those tweets, 583,462 (12.11%) verified users' tweets were removed. Of the remaining 4,233,676 non-verified users' tweets, 1,895,774 (29.34%) were irrelevant to the search terms and were excluded. Finally, 4,716,465 unique tweets in T1 and 2,991,123 unique tweets in T2 were analyzed in the studies.

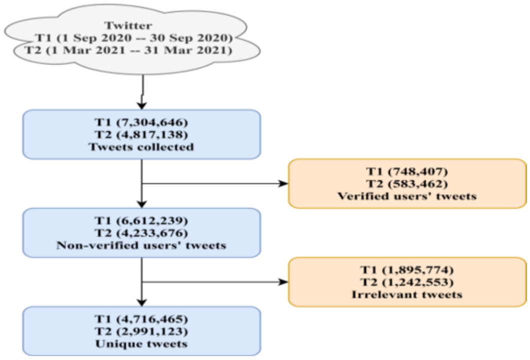

Figure 1 – Flowchart of selection of tweets. T1 representing the Sep 2020 tweets and T2 representing Mar 2021 tweets.

Sentiment Analysis

There was a total of 4,716,465 tweets in T1. Out of these, 800,173 (17%), 1,804,970 (38.3%) and 2,111,322 (44.7%) tweets were labelled as positive, neutral, and negative, respectively (Table 1). From the 2,991,123 tweets in T2, the sentiment analysis labelled 538,660 (18%) tweets as positive, 1,255,059 (42%) tweets as neutral, and 1,197,404 (40%) tweets as a negative sentiment.

Table 1– Tweet sentiment in period T1 and T2

Period	Positive	Neutral	Negative
T1	800,173	1,804,970	2,111,322
	(17%)	(38.3%)	(44.7%)
T2	538,660	1,255,059	1,197,404
	(18%)	(42%)	(40%)

Shift of Users' Sentiment between T1 and T2

A total of 132,802 (3,664,487 tweets) unique users (who had more than one tweet in two time periods T1 and T2) met our eligibility criteria. Out of those users, 35,918 (27%) and 27,041 (20.4%) users' were from G1 and G2, respectively (Table 2). In G1, 19.8%, 3.7%, and 3.5% of total users' sentiments were shifted from negative to neutral, neutral to positive, and negative to positive, respectively. In G2, 14.3%, 3.5%, and 2.5% of total users' sentiments were shifted from neutral to negative, positive to neutral, and positive to negative, respectively. Therefore we can conclude that in G1 majority of the users' sentiment shift was from negative to neutral and in G2 it was from neutral to negative. A vast 69,843 (52.6% of total users) number of users' sentiment remained unchanged between T1 and T2 (Table 2).

Table 2– Shift of users' sentiment from T1 to T2;. P: Positive; N: Negative; n: Neutral; PN: P to N; Pn: P to n; NP: Negative to Positive; Nn: Negative to Neutral; nP: Neutral to Positive; nN: Neutral to Negative; U: Unchanged;

User	NP	Nn	nP	nN	PN	Pn	U
Count	4,686	26,328	4,904	19,036	3,302	4,703	69,843
(%)	3.5	19.8	3.7	14.3	2.5	3.5	52.6
	G1: Positive shift (27.0 % users)			G2: Negative shift (20.4% users)			

Topic Modeling and Thematic Analysis

To understand the change in sentiment of the public, we performed topic modeling on both groups G1 and G2 in both timeframes T1 and T2 (Supplementary File S1). From the topics, we identified three main themes for each group (G1 and G2) at each timeframe as summarized in Table 3.

Table 3– Main themes of both group G1 and G2 in two different timeframes: T1 and T2

G1: T1 themes	G1: T2 themes
a) Worries and fears	a)Vaccine safety and efficacy
b) Mental health	b) Release of lockdown
c) Economic loss	c) Travel
G2: T1 themes	**G2: T2 themes**
a) Hopeful for vaccine rollout	a) Side effect of the vaccine
b) Precautionary measurements	b) Rapid development of the vaccine
c) Fundings for vaccine	c) New coronavirus variants

After analyzing the G1 users' tweets in period T1, we discovered three main concerns: (a) worries and fear, (b) mental health, and (c) economic loss (Supplementary File S2). We have highlighted a sample tweet from each theme below.

- Worries and fears: " *I wish that covid-19 never happened. sometimes i wonder, though, if it was because it's an election year in china, and here at home in the united states, for that matter i wonder if the rhetoric from the vaccine is being delayed for political reasons for trump to win?*"
- Mental health: "*defeating the **mental health** challenges of the coronavirus, the despair, is nearly as important as defeating the physical dangers…*".
- Economic loss:"… *Even with lighter lockdowns, Sweden has suffered **big economic losses***".

And we also discovered their (G1) joy and confidence in timeframe T2. We discovered three main themes from their tweets : (a) vaccine safety and efficacy, (b) release of lockdown, and (c) travels (Supplementary File S2). We have highlighted a sample tweet from each theme below.

- Vaccine safety and efficacy: "*the covid-19 vaccines have met strict standards of safety, quality and effectiveness set out by the independent medicines*".
- Release of lockdown: "*it's been a week since the gov announced the roadmap **out of lockdown** & we all excited for **freedom**! 💪make sure you are prepared to keep your staff…*"
- Travels: "*johor mentri besar datuk hasni mohammad said the reopening of the malaysia-singapore border in johor and the green travel bubble initiative can be considered upon …*"

Furthermore, we identify three themes for G2 users in period T1 (Supplementary File S2). They were following:

- Hopeful for vaccine rollout: "*pennsylvania's top health official has near daily conversations about the **progress on a series of covid-19 vaccines** with federal officials and is **hopeful** that a limited distribution of at least one vaccine will be ready by the end of 2020*".
- Precautionary measurements: *"be a patriot testing **mitigation, distancing, face masks** are key be 6' apart or 6' below ask americans to **stay inside at home** for 12 to 16 weeks to slow the spread …"*
- Fundings: *"we are pleased to have new @esrc **funding** to support changing research practice for #covid19. project led by @m_nind involves rapid evidence review, knowledge exchange workshops #copro of resources to **support the research community"**.*

But in T2, we found G2 were more concerned about three themes: (a) side effect of the vaccine, (b) rapid development of vaccine, and (c) coronavirus variants (Supplementary File S2).

- The side effect of Vaccine: *"yes, that's about right! **keep your unsafe vaccines** … #coronavirusastrazeneca **rejected in germany** but 4 **australians #morrison refuses** any other choice…"*.
- Rapid development of vaccine: *"**no vaccine was produced in 12 months in human history** except for covid 19 which i think even them …"*.
- Coronavirus variants: *"we've conveniently once again found a **new variant** just in time to tell you, we will not be stopping lockdowns as you thought' 'vaccines useless against newly mutated b.1.1.7. coronavirus strain"*.

Discussion

In this study, we revealed the change of Twitter users' sentiment between T1 (before vaccine rollout) and T2 (after vaccine rollout). About 27% of users' in (G1) sentiments changed positively, and 20% of users' in (G2) sentiments changed towards negative sentiment. This indicates that after the vaccine rollout most users from G1 than G2 were more optimistic about the fight against COVID-19. Our results reflect that the overall sentiment of users from G1 shifted positively as they were more optimistic about vaccine safety and efficacy. They were also hopeful about travel issues and the release of lockdown (Table 3). On the other hand, the sentiment of users from G2 shifted negatively from T1 to T2 as they were more concerned about the side effects of the vaccine and the pace of vaccine development or rollout (Table 3). G2 users were also concerned about the emerging of novel coronavirus variants. Interestingly, over half of the users' (52% users) overall sentiment remained the same in both periods T1 and T2, indicating their indifference about vaccine rollout. Otherwise, misinformation and public unawareness about vaccines might be the reason for this reluctance.

There are several limitations contributed to this study analyzing tweets related to COVID-19. In this study, no geographical restrictions were imposed on tweets collection and analyzed the worldwide COVID-19 related tweets. However, we only considered the tweets in the English language, which may not be generalized to our findings worldwide. Moreover, we collected data from two months only, a month before and after the vaccine roll-out, limiting the generalizability of our findings over longer periods. We skipped users who made less than two tweets in both periods; therefore, topics discussed may not be representative for all Twitter users. Moreover, we only analyzed the Twitter data, limiting our findings to all other social media platforms.

Conclusions

After a safe and effective vaccine rollout, the world was excited and hopeful of normal life back. However, our analysis showed over half of the Twitter users' were indifferent when we compared their tweets in two timeframes, before and after the vaccine rollout. We believe our analysis would help the policymakers and healthcare professionals in understanding the public perceptions and sentiments towards specific events (i.e., vaccine rollout) and help to explore and identify the factors affecting the variation in their sentiments.

Supplementary Materials :

Source code and Supplementary Files are available in GitHub:

https://github.com/rafiulbiswas/Vaccine-Rollout-and-Sentiment-Shift-in-Twitter-Population-A-Surveillance-Study

References

[1] David Sayce. (n.d.). Social media archives. Retrieved May 15, 2021, from https://www.dsayce.com/category/social-media/

[2] A. Agarwal, B. Xie, I. Vovsha, O.Rambow and R. Passonneau. "Sentiment analysis of twitter data, " In Proceedings of the workshop on language in social media, pp. 30-38, 2011.

[3] C. Bartlett and R. Wurtz, "Twitter and public health," J. Public Heal. Manag. Pract., vol. 21, no. 4, pp. 375–383, 2015, doi: 10.1097/PHH.0000000000000041.

[4] Y. Zhou and Y. Fan, "A sociolinguistic study of American slang," Theory Pract. Lang. Stud., vol. 3, no. 12, pp. 2209–2213, 2013, doi: 10.4304/tpls.3.12.2209-2213.

[5] M. Comesaña, A. P. Soares, M. Perea, A. P. Piñeiro, I. Fraga, and A. Pinheiro, "ERP correlates of masked affective priming with emoticons," Comput. Human Behav., vol. 29, no. 3, pp. 588–595, 2013, doi: 10.1016/j.chb.2012.10.020.

[6] W. Marcellino, T. C. Helmus, J. Kerrigan, H. Reininger, R. I. Karimov, and R. A. N. N. Lawrence, Detecting Conspiracy Theories on Social Media. 2021 doi: 10.7249/RR-A676-1

[7] M. Bhat, M. Qadri, N. ul A. Beg, M. Kundroo, N. Ahanger, and B. Agarwal, "Sentiment analysis of social media response on the Covid19 outbreak," Brain. Behav. Immun., vol. 87, pp. 136–137, 2020, doi: 10.1016/j.bbi.2020.05.006.

[8] Xue J, Chen J, Hu R, Chen C, Zheng C, Zhu T. Twitter discussions and concerns about COVID-19 pandemic: Twitter data analysis using a machine learning approach. arXiv preprint arXiv:2005.12830. 2020.

[9] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and 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. J., vol. 97, 2020, doi: 10.1016/j.asoc.2020.106754.

[10] Z. Shah, P. Martin, E. Coiera, K. D. Mandl, and A. G. Dunn, "Modeling spatiotemporal factors associated with sentiment on twitter: Synthesis and suggestions for improving the identification of localized deviations," J. Med. Internet Res., vol. 21, no. 5, 2019, doi: 10.2196/12881.

[11] M. Shrotri, T. Swinnen, B. Kampmann, and E. P. K. Parker, "An interactive website tracking COVID-19 vaccine development," Lancet Glob. Heal., vol. 9, no. 5, pp. e590–e592, 2021, doi: 10.1016/s2214-109x(21)00043-7.

[12] J. M. Banda, R. Tekumalla, G. Wang, Y. Jingyuan, T. Liu, Y. Ding, K. Artemova, E. Tutubalina, and G. Chowell. "A large-scale COVID-19 Twitter chatter dataset for open scientific research--an international collaboration." arXiv preprint arXiv:2004.03688 (2020).

[13] Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. Journal of the American Society for Information Science and Technology, 63(1), 163-173

Address for correspondence

Dr. Tanvir Alam, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar.

Email: talam@hbku.edu.qa; Phone: +974-4454-2277