

Exploring Sentiment Analysis on Arabic Tweets about the COVID-19 Vaccines

Wesam H. Alsabban

Abstract: The COVID-19 pandemic has imposed a public health crisis across the world. The global efforts lead to the development and deployment of multiple vaccines. The success of ending the COVID-19 pandemic relies on the willingness of people to get the vaccines. Social media platforms prove to be a valuable source to perform experiments on sentiment and emotion towards COVID-19 vaccination in many languages, mainly focusing on English. The people express their opinions and emotion on Twitter briefly, which can have tracked almost instantaneously. This helps the governments, public health officials, and decision-makers to understand public opinions towards vaccines. The goal of this research is to investigate public sentiment on COVID-19 vaccines. Twitter social media extracted all Arabic-language tweets mentioning seven vaccines in 7 months from 1 November 2020 to 31 May 2021. A set of Arabic sentiment lexicons were prepared to perform the sentiment analysis. The tweets' monthly average sentiment were calculated from the collected dataset and evaluated comparatively for each vaccine throughout the 11 months. Out of 5.5 million tweets that have been retrieved using the most frequent keywords and hashtags during the COVID-19 pandemic, 202,427 tweets were only considered and included in the monthly sentiment analysis. We considered tweets that mentioned only one vaccine name of the text. The distribution of tweets shows that 47.5% of the considered tweets mentioned the Pfizer vaccine. It is reported that 64% of the total tweets are non-negative while 35% are negative, with a significant difference in sentiment between the months. We observed an increase of non-negative tweets in parallel with increasing negative tweets on May 2021, reflecting the public's rising confidence towards vaccines. Lexicon-based sentiment analysis is valuable and easy to implement the technique. It can be used to track the sentiment regarding COVID-19 vaccines. The analysis of social media data benefits public health authorities by monitoring public opinions, addressing the people's concerns about vaccines, and building the confidence of individuals towards vaccines.

Keywords: COVID-19; Sentiment Analysis; Social Media; Text Analysis; Twitter; Vaccine

1 INTRODUCTION

By mid-August 2021, more than 180 million global COVID-19 cases and 3.5 million deaths have been confirmed [1]. People were encouraged to stay at home, apply social distancing, and do their work remotely. This isolation led to many posts using social media to express people's opinions towards the pandemic [2]. In 2020 and 2021, most of the top hashtags and posts on Twitter were related to the COVID-19. Its variations [3] not just by people but also governments and health authorities share policies and news related to the pandemic [4]. Recently and after the approval of several COVID-19 vaccines worldwide, most discussions on social media are about the wide availability, people acceptance/rejection of taking vaccines, and side effects of each vaccine [5].

Utilizing computing power in several technologies helps to study different areas of the COVID-19, its impact, and people's response to it effectively [6-8]. Many research trends argue that social media play an important role in public health [9]. Surveillance systems are one component of public health that are used to monitor, identify, and evaluate health issues [10]. Social media offers a great chance for such research. Monitoring infectious diseases [11] and disease outbreaks [12], tracking public response to health issues [13], detecting target areas for intervention efforts [14], and spreading fake news about epidemics [15, 16] are examples of public health surveillance.

As the Twitter social media platform has over 187 million active users, it serves as a powerful medium to better understand public perception about the COVID-19 vaccines. Twitter data is used in many types of research to understand public attitudes and discussions about the COVID-19 pandemic. Some researchers have investigated many topics related to COVID-19, including the qualitative content

analysis to understand public communication [17], sentiment analysis and word frequency [18], misinformation detection [19], topic modeling [20-22], and social distancing measurement [23].

The key success to control the spreading of the COVID-19 pandemic is the development of vaccines [24-25], while the refusal to uptake the vaccine harms herd immunity [26]. The traditional surveys that have been conducted to understand people's opinions are more expensive because of some issues, including the time-consuming, addressing short health topics, and obtaining from small-scale data. Therefore, we need to look at the community, not just individuals. Social media, including Twitter, can be a great opportunity to infer public sentiment about the COVID-19 vaccines.

Multiple studies have discussed vaccine issues on Twitter data. Twitter is used to detect the community hesitancy regarding other vaccines such as MMR, HPV, and Rdp [26]. It was used to analyze vaccine images [27], detect sentiment about HPV [28], and understanding of vaccine debate of Russian trolls [29]. Since the starting of the COVID-19 pandemic in 2020, Twitter data has been utilized in many research to understand different issues related to pandemic vaccines, such as the detection of people opinions about vaccine hesitancy [30, 31], sentiment analysis [32], content analysis of COVID-19 tweets [33], fake news and misinformation [34-36], pro-vaccine campaign on Twitter [37], anti-vaccination tweets [38] and people emotion during the pandemic [39].

The previous studies offer valuable insights about COVID-19 vaccine problems with some limitations. First, most studies focused on the analysis of Tweets in English and some other languages. This research proposes a method that analyzes collected tweets to bridge the gap of covering some under-resourced languages such as Arabic. It identifies the content sentiment of tweets and the major discussed topics to

provide insight into the evaluation of public attitudes about the COVID-19 vaccine over time. This is the first article assessing sentiment towards seven vaccines in Arabic to the best of our knowledge.

2 METHODOLOGY

This part provides details on the method used in the paper. We propose a methodology framework containing five components, as depicted in Fig. 1.

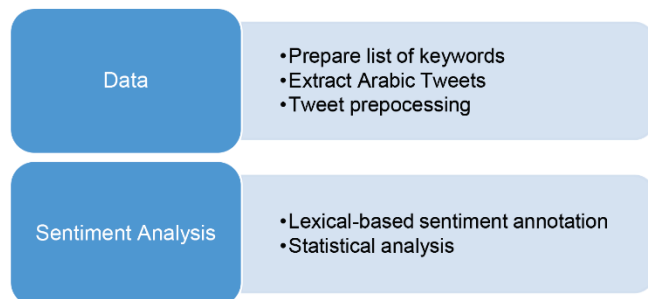


Figure 1 Research methodology

2.1 Data Collection

The Twitter streaming API (Application Programming Interface) retrieved Arabic tweets from November 1, 2020, to May 31, 2021. We used different keywords to search and retrieve relevant tweets such as, "لقاح، اللقاح، تلفيح، لقاحات" and names of each vaccine in Arabic. All Arabic tweets posted in that time were collected and filtered to be included in the sentiment analysis. Tweets that were posted from accounts related to spammer, advertising, and pornographic were excluded. Duplicated tweets from the same users and short tweets that contained less than three words were removed. This process yielded 180,000 tweets out of 5 million tweets.

2.2 Date Pre-processing

The daily tweets were measured, and completed descriptive statistics were conducted on the collected tweets. The natural language processing techniques were applied to process, analyze, and visualize the text from tweets. To preprocess the tweets, several steps are used to clean the text, such as removal of URLs, user mentions, non-Arabic words, and punctuations. All analysis was conducted using Java (Apache NetBeans IDE 12.4).

2.3 Sentiment Analysis

In our sentiment analysis, we used multiple lexicons from the www.saifmohammed.com website. The lexicons were compound and classified into negative and non-negative (positive and neutral) classes. The final lexicon contains 14,000 words. We calculated the difference between each tweet's negative and non-negative sentiment scores. If

the difference was less than zero, we assign a negative class to the tweet. Otherwise, the sentiment class would be non-negative. A rule-based approach was developed to extract the sentiment polarity of each tweet in our dataset.

To evaluate the usefulness of this approach, we randomly selected a sample of 1000 tweets representing all vaccines categories. Each tweet was tagged manually by two persons to show if a tweet had a negative or non-negative sentiment. Out of the manually annotated tweets, we used 745 tweets tagged by humans with results obtained by the rule-based approach.

The comparison showed that 78% is the highest agreement between human and rule-based annotations.

2.4 Statistical Analysis

This paper conducted a statistical analysis using Tableau¹ to compare each vaccine's sentiment changes over time. It also proposes a comparative comparison between vaccines each month. As mentioned above, we classified the tweets in our dataset into negative and non-negative classes only. Figure 2 shows the distribution of tweets over time concerning the date. The tweets distribution ranges for 13 months, from May 2020 to May 2021. We notice that the increase in the number of tweets in both classes starts in November 2020, which is the starting period for the production of COVID-19 vaccines.

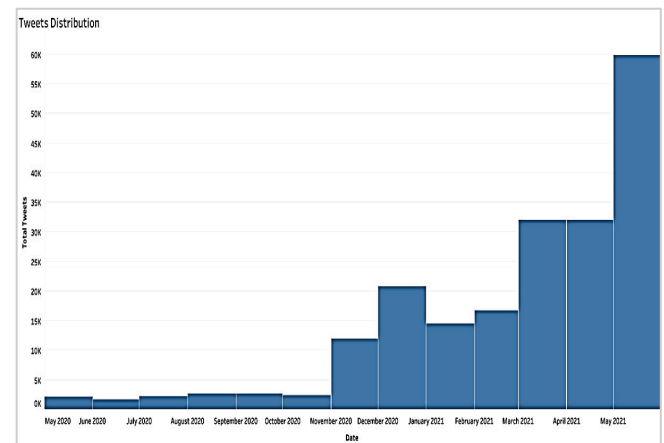


Figure 2 Distribution of tweets and sentiment classes over time

Fig. 3 shows the vaccine's tweets distribution over time. It depicted that the Pfizer vaccine has the highest number of tweets (96,156 tweets), while the Novavax has the lowest number of tweets (1,204 tweets).

In Fig. 4, the distribution of sentiment of each vaccine is shown. The total non-negative tweets (130,326 \approx 64.4%) are higher than the negative tweets (72,101 tweets \approx 35.6%) on each vaccine.

In Fig. 5, users mostly interact with the Pfizer vaccine's tweets (Likes = 763,230, Replies = 141,889 and Re-tweets = 257,250), as most of tweets in the dataset belongs to the Pfizer vaccine.

¹ <https://www.tableau.com/>

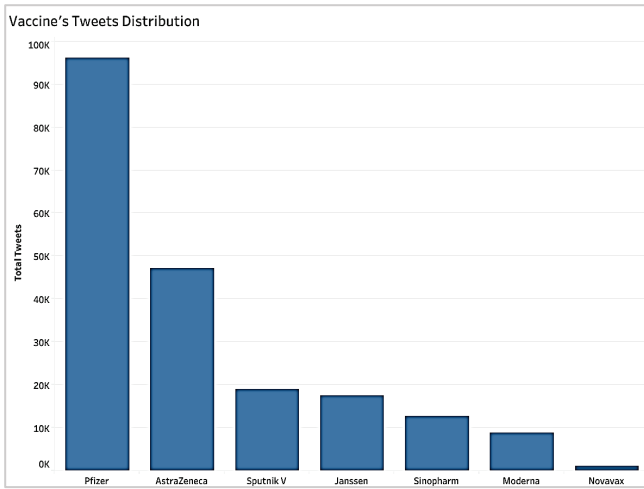


Figure 3 Vaccine's tweets distribution over time

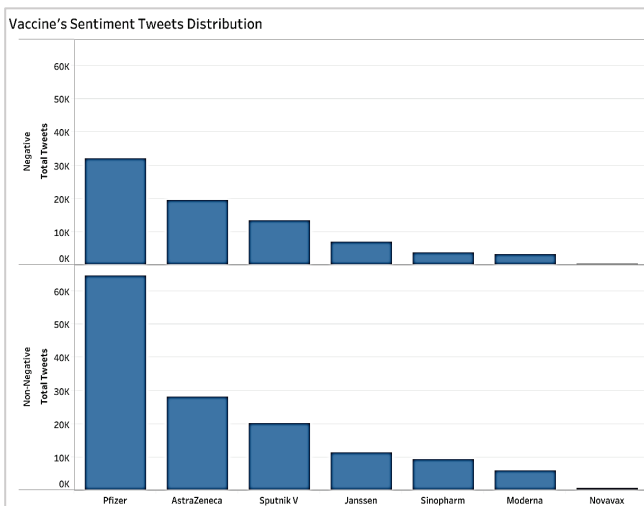


Figure 4 Vaccine's sentiment tweets distribution

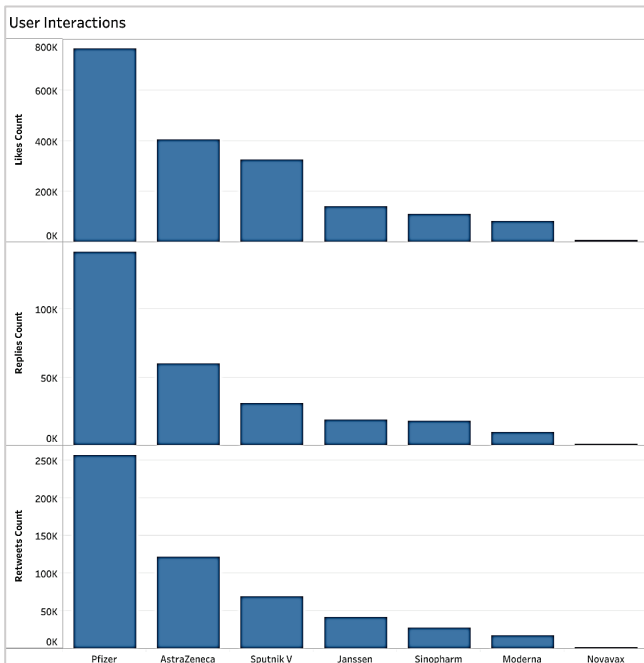


Figure 5 User interactions

3 RESULTS

A total of 202,427 tweets have been included in this study from a total of 5.5 million tweets that were retrieved. Sentiment analysis was performed and the selected tweets were annotation into two classes: negative and non-negative. The percentage of each vaccine tweets is as follow: Pfizer 47.5% ($n = 96,156$), AstraZeneca 23.3% ($n = 47,218$), Sputnik V 9.4% ($n = 18,948$), Janssen 8.6% ($n = 17,503$), Sinopharm 6.2% ($n = 12,605$), Moderna 4.3% (8,793) and Novavax 0.6% (1,204).

4 CONCLUSION

Lexical-based sentiment analysis is a valuable and easy way to implement. It helps to track the sentiment about COVID-19 vaccines. The results in this paper show that Twitter can help in giving suggestions about sentiment towards the COVID-19 vaccines. The methodology in this research can be applied in a similar data from other social media networks such as Facebook.

5 REFERENCES

- [1] Johns Hopkins University. COVID-19 Dashboard. (2021). Available: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> Retrieved: 10 May 2021. <https://doi.org/10.29121/ijesrt.v10.i5>
- [2] Samet, A. (2020). How the Coronavirus Is Changing US Social Media Usage. In: Insider Intelligence [Internet]. Available: <https://www.emarketer.com/content/how-coronavirus-changing-us-social-media-usage> Retrieved: 11 May 2021
- [3] McGraw, T. (2020). Spending 2020 Together on Twitter. Available: https://blog.twitter.com/en_us/topics/insights/2020/spending-2020-together-ontwitter.html Retrieved: 11 May 2021
- [4] Rufai, S. R. & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *Journal of Public Health*, 42, 510-516. <https://doi.org/10.1093/pubmed/fdaa049>
- [5] Haynes, B. F. (2021). A New Vaccine to Battle Covid-19. *N Engl J Med.*, 2021, 384, 470-471. <https://doi.org/10.1056/NEJMe2035557>
- [6] Asgary, A., Najafabadi, M. M., Karsseboom, R., & Wu, J. (2020). *A drive-through simulation tool for mass vaccination during COVID-19 pandemic*. Healthcare. Multidisciplinary Digital Publishing Institute, p. 469. <https://doi.org/10.3390/healthcare8040469>
- [7] Asgary, A., Valtchev, S. Z., Chen, M., Najafabadi, M. M., Wu, J. (2021). Artificial Intelligence Model of Drive-Through Vaccination Simulation. *International journal of environmental research and public health*, 18, 268. <https://doi.org/10.3390/ijerph18010268>
- [8] Asgary, A., Cojocaru, M. G., Najafabadi, M. M., & Wu, J. (2021). Simulating preventative testing of SARSCoV-2 in schools: policy implications. *BMC Public Health*, 21, 1-18. <https://doi.org/10.1186/s12889-020-10153-1>
- [9] Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: a systematic literature review through text mining. *IEEE Access*, 8, 67698-67717. <https://doi.org/10.1109/ACCESS.2020.2983656>

- [10] Paul, M. J. & Dredze, M. (2017). Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 9, 1-183. <https://doi.org/10.2200/S00791ED1V01Y201707ICR060>
- [11] Harris, J. K., Hawkins, J. B., Nguyen, L., Nsoesie, E. O., Tuli, G., Mansour, R. et al. (2017). Research brief report: using twitter to identify and respond to food poisoning: The food safety STL project. *Journal of Public Health Management and Practice*, 23, 577. <https://doi.org/10.1097/PHH.0000000000000516>
- [12] Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research*, 15, e85. <https://doi.org/10.2196/jmir.1933>
- [13] Chew, C. & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, 5(11), e14118. <https://doi.org/10.1371/journal.pone.0014118>
- [14] Salathé, M. & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol.*, 7, e1002199. <https://doi.org/10.1371/journal.pcbi.1002199>
- [15] Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control*, 38, 182-188. <https://doi.org/10.1016/j.ajic.2009.11.004>
- [16] Karami, A., Lundy, M., Webb, F., Turner-McGrievy, G., McKeever, B. W., & McKeever, R. (2021). Identifying and Analyzing Health-Related Themes in Disinformation Shared by Conservative and Liberal Russian Trolls on Twitter. *International Journal of Environmental Research and Public Health*, 18, 2159. <https://doi.org/10.3390/ijerph18042159>
- [17] Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2020). Communication related health crisis on social media: a case of COVID-19 outbreak. *Current issues in tourism*, 1-7. <https://doi.org/10.1080/13683500.2020.1733497>
- [18] Zhao, Y., Cheng, S., Yu, X., & Xu, H. (2020). Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. *Journal of medical Internet research*, 22, e18825. <https://doi.org/10.2196/18825>
- [19] Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on twitter. *Online Social Networks and Media*, 22, 100104. <https://doi.org/10.1016/j.osnem.2020.100104>
- [20] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of Tweeters during the COVID-19 pandemic: infoveillance study. *Journal of medical Internet research*, 22, e19016. <https://doi.org/10.2196/19016>
- [21] Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y. et al. (2020). Twitter Discussions and Emotions about the COVID-19 Pandemic: Machine Learning Approach. *Journal of medical Internet research*, 22, e20550. <https://doi.org/10.2196/20550>
- [22] Karami, A. & Anderson, M. (2020). Social media and COVID-19: Characterizing anti-quarantine comments on Twitter. *Proceedings of the Association for Information Science and Technology*, 57, e349. <https://doi.org/10.1002/pr2.349>
- [23] Anderson, M., Karami, A., & Bozorgi, P. (2020). Social media and COVID-19: Can social distancing be quantified without measuring human movements? *Proceedings of the Association for Information Science and Technology*, 57, e378. <https://doi.org/10.1002/pr2.378>
- [24] Moehring, A., Collis, A., Garimella, K., Rahimian, M. A., Aral, S., & Eckles, D. (2021). Surfacing norms to increase vaccine acceptance. Available at SSRN 3782082. <https://doi.org/10.2139/ssrn.3782082>
- [25] Anderson, R. M., Vegvari, C., Truscott, J., & Collyer, B. S. (2020). Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination. *The Lancet*, 396, 1614-1616. [https://doi.org/10.1016/S0140-6736\(20\)32318-7](https://doi.org/10.1016/S0140-6736(20)32318-7)
- [26] Ruiz, J., Featherstone, J. D., & Barnett, G. A. (2021). Identifying Vaccine Hesitant Communities on Twitter and their Geolocations: A Network Approach. *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 3964. <https://doi.org/10.24251/HICSS.2021.480>
- [27] Chen, T. & Dredze, M. (2018). Vaccine images on twitter: analysis of what images are shared. *Journal of medical Internet research*, 20, e130. <https://doi.org/10.2196/jmir.8221>
- [28] Massey, P. M., Leader, A., Yom-Tov, E., Budenz, A., Fisher, K., & Klassen, A. C. (2016). Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. *Journal of medical Internet research*, 18, e318. <https://doi.org/10.2196/jmir.6670>
- [29] Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A. et al. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health*, 108, 1378-1384. <https://doi.org/10.2105/AJPH.2018.304567>
- [30] Griffith, J., Marani, H., & Monkman, H. (2021). COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework. *Journal of medical Internet research*, 23, e26874. <https://doi.org/10.2196/26874>
- [31] Eibensteiner, F., Ritschl, V., Nawaz, F. A., Fazel, S. S., Tsagkaris, C., Kulnik, S. T. et al. (2021). People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. *Journal of Medical Internet Research*, 23, e28973. <https://doi.org/10.2196/28973>
- [32] Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K. et al. (2021). Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter toward COVID19 Vaccines in the United Kingdom and the United States: Observational Study. *Journal of medical Internet research*, 23, e26627. <https://doi.org/10.2196/26627>
- [33] Liu, J. & Liu, S. (2021). Understanding Behavioral Intentions toward COVID-19 Vaccines: A Theory based Content Analysis of Tweets. *Journal of Medical Internet Research*, 19. <https://doi.org/10.2196/30251>
- [34] Hernandez RG, Hagen L, Walker K, O'Leary H, Lengacher C. (2021). The COVID-19 vaccine social media infodemic: healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, 1-3. <https://doi.org/10.1080/21645515.2021.1912551>
- [35] Atehortua, N. A. & Patino, S. (2021). COVID-19, a tale of two pandemics: novel coronavirus and fake news messaging. *Health promotion international*, 36, 524-534. <https://doi.org/10.1093/heapro/daaa140>
- [36] Mahlous, A. R. & Al-Laith, A. (2021). Fake news detection in Arabic tweets during the COVID-19 pandemic. *Int J Adv Comput Sci Appl*. <https://doi.org/10.14569/IJACSA.2021.0120691>
- [37] Herrera-Peco, I., Jiménez-Gómez, B., Peña Deudero, J. J., Benitez De Gracia, E., & Ruiz-Núñez, C. (2021). *Healthcare Professionals' Role in Social Media Public Health Campaigns: Analysis of Spanish Pro Vaccination Campaign on Twitter*. Healthcare. Multidisciplinary Digital Publishing Institute, p. 662. <https://doi.org/10.3390/healthcare9060662>
- [38] To, Q. G., To, K. G., Huynh, V-A. N., Nguyen, N. T., Ngo, D. T., Alley, S. J. et al. (2021). Applying Machine Learning to Identify Anti-Vaccination Tweets during the COVID-19 Pandemic. *International journal of environmental research and public health*, 18, 4069.

<https://doi.org/10.3390/ijerph18084069>

- [39] Al-Laith, A. & Alenezi, M. (2021). Monitoring people's emotions and symptoms from Arabic tweets during the COVID-19 pandemic. *Information*, 12(2), 86.

<https://doi.org/10.3390/info12020086>

Author's contacts:

Wesam H. Alsabban

College of Computers and Information Systems,
Umm al-Qura University,
Al Taif Road, Makkah al Mukarramah, Saudi Arabia
whsabban@uqu.edu.sa
<https://orcid.org/0000-0001-9984-0621>