

Opinion Mining of Saudi Responses to COVID-19 Vaccines on Twitter

A Computational Linguistic Approach

Fahad M. Alliheibi^{1*}

Department of Arabic Language, Faculty of Arts
King Abdulaziz University, Jeddah, Saudi Arabia

Abdulfattah Omar²

Department of English, College of Science and Humanities
Prince Sattam Bin Abdulaziz University, Saudi Arabia
Department of English, Faculty of Arts
Port Said University, Egypt

Nasser Al-Horais³

Department of Arabic Language and Arts
College of Arabic Language and Social Studies
Qassim University, P.O. Box 6611, Buraidah 51425
Saudi Arabia

Abstract—In recent months, many governments have announced COVID-19 vaccination programs and plans to help end the crises the world has been facing since the emergence of the coronavirus pandemic. In Saudi Arabia, the Ministry of Health called for citizens and residents to take up the vaccine as an essential step to return life to normal. However, the take-up calls were made in the face of profound disagreements on social media platforms and online networks about the value and efficacy of the vaccines. Thus, this study seeks to explore the responses of Saudi citizens to the COVID-19 vaccines and their sentiments about being vaccinated using opinion mining methods to analyze data extracted from Twitter, the most widely used social media network in Saudi Arabia. A corpus of 37,467 tweets was built. Vector space classification (VSC) methods were used to group and categorize the selected tweets based on their linguistic content, classifying the attitudes and responses of the users into three defined categories: positive, negative, and neutral. The lexical semantic properties of the posts show a prevalence of negative responses. This indicates that health departments need to ensure citizens are equipped with accurate, evidence-based information and key facts about the COVID-19 vaccines to help them make appropriate decisions when it comes to being vaccinated. Although the study is limited to the analysis of attitudes of people to the COVID-19 vaccines in Saudi Arabia, it has clear implications for the application of opinion mining using computational linguistic methods in Arabic.

Keywords—Computational linguistics; COVID-19 vaccines; lexical semantics; opinion mining; vector space classification

I. INTRODUCTION

The COVID-19 pandemic, which emerged in late 2019, has had and continues to have grave health, economic, social, and political consequences for almost all countries around the world [1-4]. So far, more than three million people have died of this coronavirus. In response, many governments have enforced lockdown laws, placing restrictions on the movement of citizens globally. This has resulted in serious economic problems for many countries, including the Arab Gulf states, which have suffered due to the lower demand for oil. As

pointed out by Kesar, et al. [5], the coronavirus pandemic has resulted in a large-scale humanitarian crisis leading to misery and suffering for all of humanity, and pushing social and economic well-being to the brink of collapse.

To alleviate the dangers of the COVID-19 pandemic, several vaccines have been developed, including the AstraZeneca, Johnson & Johnson's Janssen, Moderna, Novavax, and Pfizer-BioNTech vaccines. Given the sophisticated development and long authorization process for vaccinations, the COVID-19 vaccines were given temporary authorization for emergency use. Nevertheless, the manufacturers and researchers stressed the safety of the vaccines and their effectiveness in reducing the risks of severe illness associated with catching the virus. They confirmed that the new vaccines had undergone rigorous testing to verify their safety before they were recommended for widespread use.

In light of these developments, many governments have announced COVID-19 vaccination delivery plans to help end the crises the world has been facing since the emergence of the coronavirus pandemic. In Saudi Arabia, the Ministry of Health called for citizens and residents to take up the vaccine as an essential step in returning life to normal. However, these appeals were met by skepticism from many Saudis who had various fears about the vaccines. Furthermore, many activists fueled concerns about vaccine safety on social media platforms, especially Twitter. In response, officials and professionals stressed the safety of the vaccines, and called for citizens and residents to get the vaccines for the good and safety of everyone.

Twitter in particular was the forum for controversy and heated debates about the effectiveness and usefulness of the COVID-19 vaccines in Saudi Arabia. Millions of users voiced their opinions concerning the uptake of the vaccines on the platform. Apart from reflecting the issue of the effectiveness and safety of the vaccines, this clearly indicated the growing influence of social media platforms and networks on public opinion in Saudi Arabia, as noted by Gunter, et al. [6]. The

*Corresponding Author

increasing numbers of users of social media platforms and networks in Saudi Arabia is having a direct effect on their ever greater role and effectiveness in shaping public opinion concerning a range of social, economic, and political issues [6, 7].

Thus, it is important for policymakers and those in government to understand people's attitudes towards certain issues as represented on social media. In this regard, they could exploit the potential of opinion mining, defined as a computational linguistics science that uses text analysis and natural language processing (NLP) to understand the drivers of public sentiment through extracting people's opinions and sentiments from the Web in general, and social media networks and platforms in particular [8, 9]. Despite the importance of opinion mining applications in helping governments and organizations to make appropriate decisions concerning critical issues, especially at times of crisis, their potential has not yet been fully explored in Arab nations [10]. In response to this lack of research, our study has sought to explore Saudi responses and sentiments regarding the COVID-19 vaccination take-up calls.

The study addresses two key research questions. First, what were individuals' responses and sentiments concerning the COVID-19 vaccines in Saudi Arabia as expressed on Twitter? Second, can the opinions, attitudes, and responses of individuals regarding given issues be analyzed conceptually through computational linguistic methods in Arabic?

The remainder of the article is structured as follows. Section 2 reviews the development of opinion mining and outlines its theoretical underpinnings and its relationship with other disciplines, including sentiment analysis, NLP, and computational linguistics. Section 3 sets out the data collection methods and procedures and the corpus design. Section 4 details the computational linguistic analysis based on a vector space classification (VSC) of the selected tweets. Section 5 concludes, summarizing the main findings and highlighting the implications of the study for further research.

II. PREVIOUS WORK

Opinion mining is a rising field of study that is gaining prominence largely due to the growing influence of social media platforms and networks in today's world. Indeed, there has been an extraordinary increase in the number of social media networks, providing easy access to millions of users around the world to express their opinions freely on different issues. Given the daily growth in numbers of users around the world, social media networks have become indispensable in contemporary life, and are now integral to the economic, social, and political marketing mechanisms that are employed in many fields.

With the increasing influence of social media networks, opinion mining was developed to analyze human behavior through computational tools to help manufacturers, marketing agencies, branders, government institutions, organizations, and policymakers make appropriate decisions concerning their products, brands, services, and policies [11]. This caters to the growing need for businesses and organizations to take account

of public opinion expressed on social media networks and platforms in their decision making [12].

In the literature, opinion mining is commonly related to sentiment analysis, and the two terms are often used interchangeably [13-15]. They are both concerned with the study and analysis of human behavior as expressed on social media networks and platforms using computer modeling approaches. As Liu and Zhang [12] point out, they have in common that they entail the "study of people's opinions, appraisals, attitudes, and emotions regarding entities, individuals, issues, events, and topics and their attributes." Our study, however, considers that they differ in terms of depth, arguing that opinion mining goes to the next level in understanding the conversational drivers underpinning the sentiment.

The literature reports extensive studies that have been carried out to measure public opinion on different issues and policies through computational-based opinion mining methods. There is growing interest in opinion mining applications to analyze political issues due to the already vast and constantly increasing creation of political content on the Web and its exchange, particularly on social media networks and platforms [16-18]. The underlying principle in such studies is that the analysis of public opinions expressed on social media networks and platforms, if well done, "can be quite useful for government agencies, as it can significantly assist them to understand the needs and problems of society, and the perceptions and feelings of the citizens, and to formulate effective public policies" [16].

Despite the prolific and increasing literature on opinion mining applications, there have been very few studies based on Arabic content to date. This can be attributed to several reasons. First, as noted by Badaro et al. [19], "there are fewer freely [available] resources for Arabic opinion mining in terms of large clean sentiment lexica and annotated collections of opinions" compared to English and other western languages. In other words, most opinion mining methods have been developed for English texts. It is difficult to use these methods with Arabic, which is completely different from English [20]. Second, the morphological system of Arabic poses serious challenges for NLP applications with Arabic data [21-24]. As pointed out by Ryding [25], Arabic is a highly inflectional language. Unlike English, words in Arabic have a unique morphological system, which typically has negative effects on the reliability and effectiveness of NLP processes and applications. Finally, the diglossic nature of Arabic, and its different vernacular and colloquial dialects make it difficult for the classification systems of opinion mining applications to work properly [26-28].

It has only been over the last five years that opinion mining in Arabic has started to receive considerable interest from the research community [29]. In a recent study, Althagafi, et al. [30] studied and analyzed the attitudes and responses of people to the shift in schools and universities towards online learning in 2020 due to the COVID-19 pandemic. The authors used Arabic tweets to gain an understanding of the sentiments and opinions of Twitter users regarding online learning to help government, educational institutions, and policymakers

recognize the emotional effects of online learning on students, teachers, parents, and individuals in the Saudi community.

In another study, Omar and Hamouda [10] analyzed the responses of citizens to a new real estate law in Egypt through computational tools. They aimed to explore Egyptians' sentiments, opinions, and attitudes regarding the new Real Estate Registration Law based on Facebook posts, comments, and replies. The study concluded that the citizens' views of the new law were predominantly negative. The authors highlighted the effectiveness of applying opinion mining to examine responses to regulations and policies through citizens' responses on social media. Accordingly, they suggested that opinion mining applications should be viewed as presenting valuable opportunities for policymakers and legislators to gauge people's sentiments, attitudes, and perceptions.

Despite the importance of such studies in terms of introducing researchers to opinion mining applications in Arabic, there was no focus on the manipulation of computational linguistic tools or lexical semantic analysis in the interpretation of the data and users' opinions.

III. METHODS AND PROCEDURES

This study used Twitter posts to identify the responses of citizens and residents to the COVID-19 vaccines in Saudi Arabia since social media networks and platforms have become an integral part of the lives of the Saudis today and Saudi Arabia is one of the leading countries in terms of Internet penetration. According to a report released by the World Bank, 95% of the total population of Saudi Arabia was using the Internet in 2019. In 2021, the percentage increased to 95.7%, with mobile connections up to 112.7%. Furthermore, the number of active social media users amounted to 27.80 million, individuals constituting around 79.3 % of the total population [31], as shown in Fig. 1.

In light of these statistics, it can be claimed that social media networks can usefully be employed to identify the responses of individuals in Saudi Arabia to specific issues. Of the platforms available, Twitter is one of the most reliable and suitable sites for the application of opinion mining and sentiment analysis [32, 33], and as of January 2021, Saudi Arabia has been ranked one of the top countries in terms of Twitter users [34], as shown in Fig. 2.

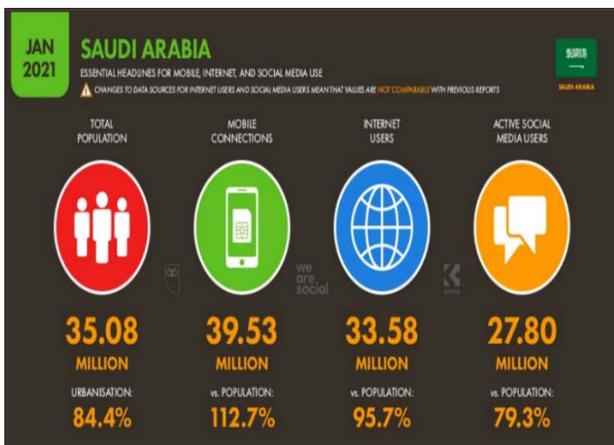


Fig. 1. The use of the Internet and Social Media in Saudi Arabia.

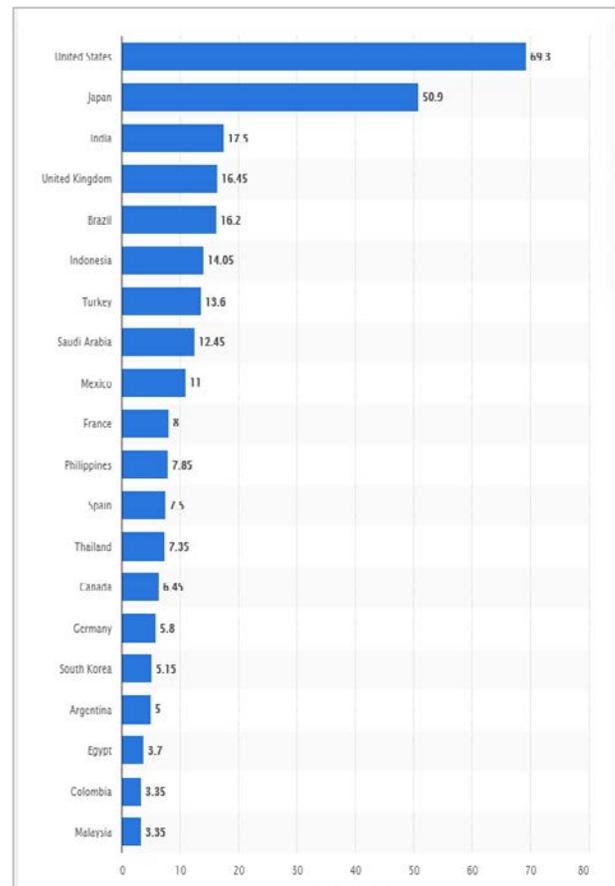


Fig. 2. Countries with the Most Twitter users in 2021.

Furthermore, Twitter was ranked topmost for information and posts on COVID-19 in Saudi Arabia. Finally, the Twitter account of the Saudi Ministry of Health has the highest number of followers, as shown in Fig. 3.

To extract information about users' polarity and responses concerning the issue of COVID-19 vaccines, four key terms were defined, as presented in Table I.

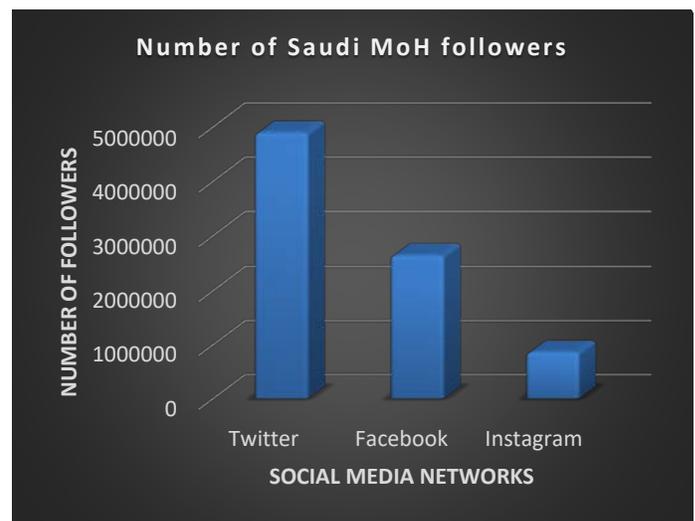


Fig. 3. Numbers of the Ministry of Health followers on Social Media Networks.

TABLE I. LIST OF THE DEFINED KEY TERMS

Key Terms	Transliteration	English Translation
لقاح كورونا	Liqah kooruna	Corona Virus vaccine*
تطعيم كورونا	Tat'eem kooruna	Corona virus vaccine*
التحصين	Altahsin	Vaccination
الحصول علي اللقاح	Alhusul eali alliqah	Vaccine uptake

* In Arabic, unlike English, coronavirus is expressed as two words.

Keywords were chosen rather than hashtags to ensure the data sample was as inclusive and representative as possible for generalizability and reliability purposes, since hashtags are generally used by more proficient social media users.

We built a corpus of 37,467 tweets from 23,748 users drawing on posts, comments on posts, and replies to posts on the topic of COVID-19 vaccines. The selected tweets covered the period from December 15, 2020, when the Ministry of Health called for the citizens and residents of Saudi Arabia to register for the COVID-19 vaccine, until May 25, 2021. Fig. 4 shows the first two citizens in Saudi Arabia who took the vaccine on December 17, 2020.

The selected tweets were cleaned and all non-alphabetical characters were removed. A stop-word list was also generated and all function words were removed. As a next step, a vector space matrix containing only the content or lexical words was designed and comprised 9,348 lexical types. There were fewer lexical types than tweets as many of the selected posts were repeated.

One critical problem with the matrix was the high dimensionality of the data, which commonly has adverse impacts on the reliability of classification [35]. High dimensionality results from a high volume of irrelevant information, which makes the classification process problematic and leads to poor, potentially unreliable, performance. As the number of vectors or attributes increases, the texts become increasingly homogeneous. For a data matrix in which the rows are the data items and the columns the variables, the matrix defines a manifold (the shape of the data in space) in \mathcal{N} -dimensional space. Let us assume a data matrix with 1,000 3-dimensional vectors. Plotting these vectors in 3-dimensional space will result in a cloud of points, as shown in Fig. 5.

Consequently, it becomes increasingly difficult to find meaningful and distinct groups based on such a dataset. The retention of unimportant variables in analysis gives invalid results [36]. It is thus important to retain only those variables that are most distinctive within a data matrix through dimensionality reduction, improving the reliability of classification performance [37]. To do this, two tests were carried out: variance and term frequency-inverse document frequency (TF-IDF). The rationale is that the classification is based on only the most distinctive lexical variables in the corpus.

First, all 9,348 lexical types were sorted in terms of variance based on the assumption that there must be variations in the characteristics of documents, distinguishing one from another, to enable classification. It follows that measuring variance within datasets can be useful in identifying the most

distinctive features within a matrix. The variance test indicated that the first 600 lexical types were the most significant variables in terms of variance, as shown in Fig. 6.



Fig. 4. The Initiation of the Vaccination Program in Saudi Arabia.

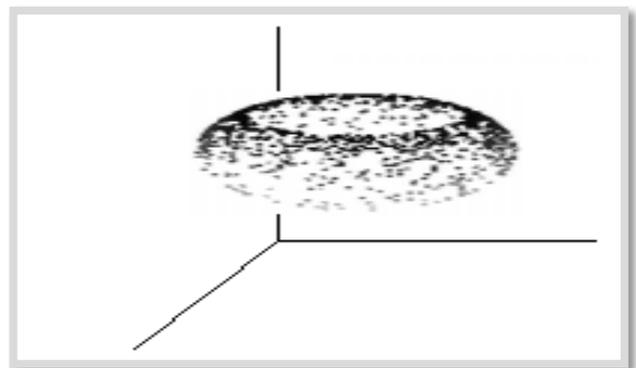


Fig. 5. A Manifold in 3-dimensional Space.

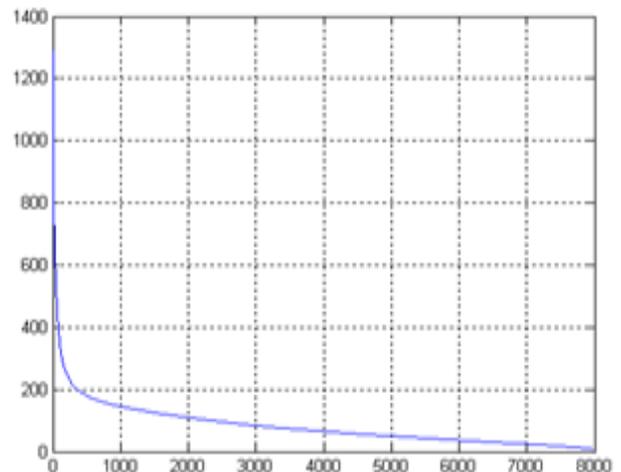


Fig. 6. Measuring Variables in Terms of Variance.

As such, only the lexical types presenting significant variation were retained and those with low or no variation were deleted. This resulted in a reduction in lexical types within the matrix to just 600 variables. As a final step in the data dimensionality reduction process, a TF-IDF analysis was carried out. The 600 lexical types were sorted in descending order in terms of TF-IDF using the following function:

$$tfidf(t_i) = tf(t_i) \log_2 \left[\frac{M}{df_1} \right]$$

where $tf(t_i)$ is the frequency of term t_i across all documents in the data matrix M . The result is shown in Fig. 7.

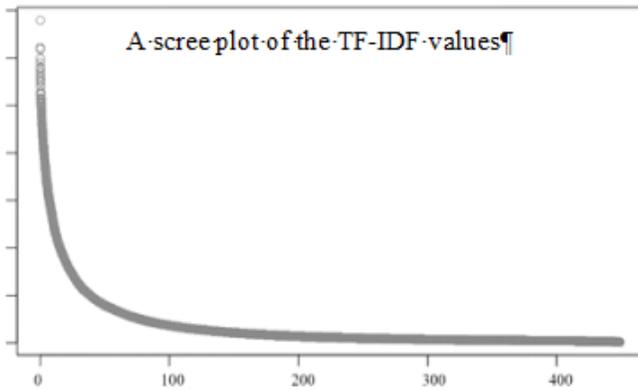


Fig. 7. Term Weighting using TF-IDF.

In this process, only the 60 lexical types with the highest TF-IDF were retained. These are assumed to be the most distinctive features within the data matrix.

IV. ANALYSIS AND DISCUSSION

To classify each opinion concerning the issue as positive, negative, or neutral, a computational lexical semantic analysis was conducted based on vector space classification (VSC). VSC is one of the most reliable classification systems in data mining applications, including information retrieval (IR), annotation and summarization, sentiment analysis, and opinion mining. VSC is based on measuring the relative distances between row vectors. The size of the angle between the lines that joins vectors to the origin of the space's coordinate system and the lengths of the lines jointly determine the distance between any two vectors in a space [38]. In VSC, for each selected variable, its semantics are used to obtain an interpretation of the domain of interest, the domain thus being measured based on the semantics [39-41].

The classification is formed using opinion words and phrases, these being the dominant indicators for opinion classification. Liu and Zhang [12] define opinion words as "words that are commonly used to express positive or negative sentiments. For example, beautiful, wonderful, good, and amazing are positive opinion words, and bad, poor, and terrible are negative opinion words. Although many opinion words are adjectives and adverbs, nouns (e.g., rubbish, junk, and crap) and verbs (e.g., hate and like) can also indicate opinions. Apart from individual words, there are also opinion phrases and idioms, e.g., cost someone an arm and a leg." The 37,467

tweets were classified into three categories or classes: positive, negative, and neutral. The results are shown in Fig. 8.

It is clear that most responses to the COVID-19 vaccine take-up calls are negative. The negative responses account for approximately 54%, positive responses 34%, and neutral responses 12%.

Positive responses were reflected in words such as (فعال) effective, (آمن) safe, (طبيعية) normality, (يحمي) protect, and (وقاية) protection. These focused on the concept of reducing the risks associated with COVID-19 infection, especially for the elderly, as shown in the examples given in Fig. 9.

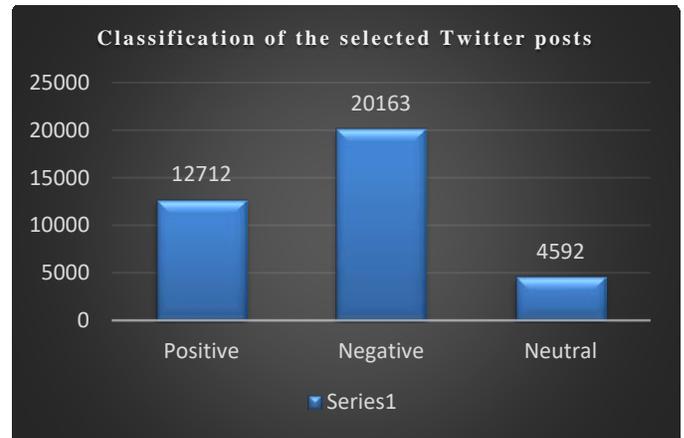


Fig. 8. Results of the Classification of the Selected Twitter Posts.



Fig. 9. Examples of Positive Responses to the COVID-19 Vaccines.

Negative responses, in turn, included words such as (أرفض) reject, (لا) No, (العقم) infertility, (جيش) Gates, (مؤامرة) conspiracy (الجلطات) clots, and (الحامض النووي) DNA. These responses focused on the ineffectiveness of the vaccines and that they had not been carefully evaluated in clinical trials. They also linked the COVID-19 vaccines to infertility and making changes within the DNA of individuals, as seen in the examples given in Fig. 10.

Furthermore, many posts claimed that COVID-19 vaccination was cover for a plan devised by Bill Gates to implant trackable microchips to control people. It is therefore not surprising that many hashtags rejecting vaccination were generated in response to the Saudi government's call for vaccine uptake. These included:

#لا_لربط_السفر_بالتطعيم (No vaccine certificates for travel).

#لا_للتطعيم_الاجباري (No for mandatory vaccination).

#نرفض_اللقاح_الاجباري (We reject mandatory vaccination).

#نرفض_التطعيم_الاجباري (We reject mandatory vaccination).

Based on these findings, it can be claimed that computational lexical semantic and opinion mining methods can usefully be employed to define people's attitudes and responses to given issues. It was clear that the classification of the Twitter posts based on their lexical semantic properties was meaningful. In particular, it was a straightforward matter to define polarity concerning the issue of COVID-19 vaccine uptake in Saudi Arabia.



Fig. 10. Examples of Negative Responses to the COVID-19 Vaccines.

V. CONCLUSION

This study analyzed the responses of individuals to COVID-19 vaccines in Saudi Arabia through their Twitter posts. To do so, computational lexical semantic methods were used for the classification of people's responses. The results clearly indicate that most people in Saudi Arabia hold negative attitudes concerning the COVID-19 vaccination take-up programs established by the government. The findings of the study illustrate the importance of understanding people's sentiments, opinions, and attitudes regarding government decisions and policies. The results argue for the benefits of government institutions and departments making use of the potential of data mining applications to identify trends that may have adverse impacts on policies and practices, as well as to help government institutions make appropriate decisions, and adopt reliable and workable policies and procedures. In view of the significance of social media networks and platforms in modern societies, opinion mining applications should be an integral part of all government processes. Although the study is limited to analysis of the attitudes of people to the COVID-19 vaccines in Saudi Arabia, it has clear implications in terms of the value of opinion mining applications using computational linguistic methods in Arabic.

ACKNOWLEDGMENT

The corresponding author would like to acknowledge the support provided by the Deanship of Scientific Research (DSR) in King Abdulaziz University, Jeddah, Saudi Arabia.

REFERENCES

- [1] Z. Allam, Surveying the Covid-19 Pandemic and Its Implications: Urban Health, Data Technology and Political Economy. Amsterdam: Elsevier Science, 2020.
- [2] M. Nicola et al., "The socio-economic implications of the coronavirus and COVID-19 pandemic: a review," International journal of surgery, 2020.
- [3] S. Gössling, D. Scott, and C. M. Hall, "Pandemics, tourism and global change: a rapid assessment of COVID-19," Journal of Sustainable Tourism, vol. 29, no. 1, pp. 1-20, 2020.
- [4] D. Bhuiyan and A. Bashar, "The covid-19 pandemic: conceptual model for the global economic impacts and recovery," The Covid-19 Pandemic: Conceptual Model for the Global Economic Impacts and Recovery (July 20, 2020), 2020.
- [5] S. Kesar, R. Abraham, R. Lahoti, P. Nath, and A. Basole, "Pandemic, informality, and vulnerability: Impact of COVID-19 on livelihoods in India," CSE Working Paper, vol. 2020, 2020.
- [6] B. Gunter, M. Elareshi, and K. Al-Jaber, Social media in the Arab world: Communication and public opinion in the Gulf states. Bloomsbury Publishing, 2016.
- [7] H. T. Alsudairy, The Role of the Social Media in Empowering Saudi Women's Expression. Newcastle, UK: Cambridge Scholars Publishing, 2020.
- [8] B. Pang and L. Lee, Opinion Mining and Sentiment Analysis. Delft, the Netherlands: Now Publishers, 2008.
- [9] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. New York: Cambridge University Press, 2015.
- [10] A. Omar and W. Hamouda, "A Sentiment Analysis of Egypt's New Real Estate Registration Law on Facebook," International Journal of Advanced Computer Science and Applications, vol. 12, no. 4, pp. 656-663, 2021.
- [11] S. Choudhary, "Opinion Mining Techniques and Its Applications: A Review.," in Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019), vol. 121, P. Singh, W. Pawłowski, S. Tanwar, N. Kumar, J. Rodrigues, and

- M. Obaidat, Eds. (Lecture Notes in Networks and Systems, Singapore: Springer, 2020, pp. 549-559.
- [12] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data New York; London: Springer, 2012, pp. 415-463.
- [13] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent systems, vol. 28, no. 2, pp. 15-21, 2013.
- [14] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," Knowledge-based systems, vol. 89, pp. 14-46, 2015.
- [15] P. K. Singh and M. S. Husain, "Methodological study of opinion mining and sentiment analysis techniques," International Journal on Soft Computing, vol. 5, no. 1, p. 11, 2014.
- [16] Y. Charalabidis, M. Maragoudakis, and E. Loukis, "Opinion Mining and Sentiment Analysis in Policy Formulation Initiatives: The EU-Community Approach," in Electronic Participation, vol. 9249, E. Tambouris et al., Eds. (Lecture Notes in Computer Science, Cham: Springer, 2015, pp. 147-160.
- [17] V. Bekkers, A. R. Edwards, R. Moody, and H. Beunders, "Caught by surprise? Micro-mobilization, new media and the management of strategic surprises," Public Management Review, vol. 13, no. 7, pp. 1003-1021, 2011.
- [18] E. Loukis and Y. Charalabidis, "Active and passive crowdsourcing in government," in Policy Practice and Digital Science: Integrating Complex Systems, Social Simulation and Public Administration in Policy Research, M. Janssen, M. Wimmer, and A. Deljoo, Eds. (Public Administration and Information Technology Series, Cham: Springer Verlag, 2014, pp. 261-289.
- [19] G. Badaro et al., "A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications, and Visualizations," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 18, no. 3, pp. 27:1-52, 2019.
- [20] M. N. Al-Kabi, A. H. Gigieh, I. M. Alsmadi, and H. A. Wahsheh, "Opinion Mining and Analysis for Arabic Language," International Journal of Advanced Computer Science and Applications, vol. 5, no. 5, pp. 181-195, 2014.
- [21] A. Omar and M. Aldossari, "Lexical Ambiguity in Arabic Information Retrieval: The Case of Six Web-Based Search Engines," International Journal of English Linguistics, vol. 10 no. 3, pp. 219-228, 2020.
- [22] A. Omar, B. Alghayesh, and M. Kassem, "Authorship Attribution Revisited: The Problem of Flash Fiction A morphological-based Linguistic Stylometry Approach," Arab World English Journal, vol. 10, no. 3, pp. 318-329, 2019.
- [23] A. Omar and W. Hamouda, "The Effectiveness of Stemming in the Stylometric Authorship Attribution in Arabic," International Journal of Advanced Computer Science and Applications, vol. 11 no. 1, pp. 116-122, 2020.
- [24] A. Omar, I. Shaalan, and W. Hamouda, "Ambiguity Resolution in Arabic Localization.," Applied Linguistics Research Journal, vol. 5 no. 1, pp. 1-6, 2021.
- [25] K. Ryding, "Arabic inflectional morphology," in Arabic: A Linguistic Introduction Cambridge: Cambridge University Press, 2014, pp. 89-106.
- [26] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing: Challenges and Solutions," ACM Transactions on Asian Language Information Processing, vol. 8, no. 4, p. Article 14, 2009.
- [27] A. Omar, "An Evaluation of the Localization Quality of the Arabic Versions of Learning Management Systems," International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, pp. 443-449, 2021.
- [28] A. Omar, "Authorship attribution of Morsi Gameel Aziz's lyrics: A clustering-based stylometry approach," Journal of Language and Linguistic Studies, vol. 17, no. 1, pp. 542-557, 2021.
- [29] N. Boudad, R. Faizi, R. Oulad Haj Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," Ain Shams Engineering Journal, vol. 9, no. 4, pp. 2479-2490, 2018/12/01/ 2018.
- [30] A. Althagafi, G. Althobaiti, H. Alhakami, and T. Alsabait, "Arabic Tweets Sentiment Analysis about Online Learning during COVID-19 in Saudi Arabia" International Journal of Advanced Computer Science and Applications(IJACSA)," vol. 12, no. 3, pp. 620-625, 2021.
- [31] S. Kemp, "Digital 2021: Saudi Arabia," Data Reportal 10 February 2021 2021, Available: <https://datareportal.com/reports/digital-2021-saudi-arabia>.
- [32] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, A Practical Guide to Sentiment Analysis. New York: Springer International Publishing, 2017.
- [33] G. Gupta and G. S. Bhatthal, Sentiment Analysis of English Tweets Using Data Mining: Data Mining, Sentiment Analysis. Munich, Germany: BookRix, 2018.
- [34] H. Tankovska, "Countries with the most Twitter users 2021," Statista, New York, NY Feb 9, 2021 2021.
- [35] A. Omar, "Feature Selection in Text Clustering Applications of Literary Texts: A Hybrid of Term Weighting Methods," International Journal of Advanced Computer Science and Applications, vol. 11 no. 2, pp. 99-107, 2020.
- [36] Y. Shao, Topics in Dimension Reduction. Minneapolis, Minnesota: University of Minnesota, 2007.
- [37] P. D. Waggoner, Modern Dimension Reduction. Cambridge: Cambridge University Press, 2021.
- [38] H. Moisl, Cluster Analysis for Corpus Linguistics. Berlin: De Gruyter, 2015.
- [39] A. Omar, "Identifying themes in fiction: A centroid-based lexical clustering approach," Journal of Language and Linguistic Studies, vol. 17, no. Special Issue 1, 580-594 2021.
- [40] A. Omar, "The Detective and Sensation Fiction of Wilkie Collins: A Computational Lexical-Semantic Analysis," Arab World English Journal (AWEJ), vol. 11 no. 1, pp. 195-211, 2020.
- [41] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008.

AUTHORS' PROFILE

Fahad Alliheibi is a Full Professor of Arabic Language and Linguistics in the Department of Arabic Language, Faculty of Arts, King Abdulaziz University (KSA). Prof. Alliheibi received his PhD degree in Linguistics in 1999 from Durham University, UK. His research interests include Arabic Linguistics, Pragmatics, Text Linguistics, and Translation. ORCID: 0000-0002-2233-5165

Abdulfattah Omar is an Associate Professor of English Language and Linguistics in the Department of English, College of Science & Humanities, Prince Sattam Bin Abdulaziz University (KSA). Also, he is a standing lecturer of English Language and Linguistics in the Department of English, Faculty of Arts, Port Said University, Egypt. Dr. Omar received his PhD degree in computational linguistics in 2010 from Newcastle University, UK. His research interests include computational linguistics, digital humanities, discourse analysis, and translation studies. ORCID: 0000-0002-3618-1750

Nasser Al-Horais is a Full Professor of Arabic Language and Linguistics in the Department of Arabic Language and its Arts, College of Arabic Language & Social Studies, Qassim University (KSA). Prof. Al-Horais received his PhD degree in Linguistics in 2009 from Newcastle University, UK. His research interests include Arabic Syntax, Generative Linguistics (Minimalist Program), Discourse Analysis, and Comparative Syntax. ORCID: 0000-0002-2511-9791