Public Sentiment Analysis in Social Media on the SARS-CoV-2 Vaccination Using VADER Lexicon Polarity(*)

Dr. Jeehaan Algaraady
Department of English, Taiz University
Taiz, Yemen

Dr. Mohammad Mahyoob
Department of Languages and Translations,
Faculty of Science and Arts, Alula, Taibah
University, Madinah, Saudi Arabia

تحليل المشاعر العامة في وسائل التواصل الاجتماعي حول التطعيم ضد فيروس كورونا-19 باستخدام قطبية معجم VADER

د/ جيهان الجرادي
قسم اللغة الإنجليزية ـ جامعة تعز

د/ محمد مهيوب
قسم اللغات والترجمة
كلية العلوم والآداب ـ العلا ـ جامعة طيبةـ المملكة العربية السعودية

الملخص

شكلت معالجة اللغة الطبيعية مؤخرًا مجالًا مهمًا في اللغويات الحاسوبية والذكاء الاصطناعي، حيث أصبح العالم الافتراضي والرقمي جانبًا أساسيًا في حياتنا اليومية. ويعد تحليل المشاعر والتنقيب في البيانات من أهم فروع معالجة اللغة الطبيعية، والذي لفت انتباه الباحثين إلى تناول العديد من القضايا المطروحة على منصات التواصل الاجتماعي والتنقيب عن أراء ومشاعر مستخدمي هذه المنصات. تستكشف هذه الدراسة مشاعر وآراء الجمهور تجاه تلقي جرعات تطعيم فيروس كورونا-19 في المملكة العربية السعودية. حيث تبحث هذه الدراسة عن الدوافع والموانع في أخذ جرعات اللقاح الأولى والثانية وكيف اختلف وعي الجمهور ومواقفه في الجرعتين. تناولت هذه الدراسة 6,232 تغريدة وتعليق تم جمعها من منصات التواصل الاجتماعي الرسمية (Twitter وYouTube) بين 19 ديسمبر 2020 و10 ديسمبر 2021. حيث تم استخدام قطبية معجم NLTK VADER في عملية تحليل المشاعر العامة لدى الجمهور، بينما تم تصنيف الآراء بناءً على منهجية القياس متعدد الأبعاد. في حالة جرعة اللقاح الأولى، اظهرت النتائج أنه من بين 2,989 رأيًا مسجلًا، كان 61.5٪ من نسبة الجمهور على استعداد لتلقي التطعيم ضد فيروس كورونا -19 – ويثق الغالبية منهم في سلامة اللقاح وبإجراءات وقرارات وزارة الصحة. في حين أن 21.1٪ أبدو مشاعراً معاكسة تجاه التطعيم، حيث أنهم لا يثقون باللقاح ويخشون من أعراضه الجانبية. أما في حالة جرعة اللقاح الثانية، من بين 3,243 تغريدة وتعليق، 63.2٪ أبدو مواقف إيجابية تجاه أخذ اللقاح. وكانت الثقة في سلامة اللقاح وعدم المنع من العمل والسفر والأنشطة الأخرى هي الدوافع الأساسية لتلقي اللقاح في هذه المرحلة. بينما سجلت المشاعر السلبية 30.3٪ من نسبة التغريدات والتعليقات، وكانت الموانع من أخذ التطعيم هي الشك في سلامة اللقاح واعراضه وتضارب البروتوكولات المتبعة دولياً. إن من شأن تحديد المشاعر والمواقف العامة تجاه التطعيم التعرف على الأسباب الكامنة وراء رفض اللقاح أو قبوله من قبل الناس والذي من شأنه أن يساعد صانعي القرارات الصحية في تخطيط وتنفيذ استراتيجيات موائمة للتوعية بأهمية التطعيم وخلق توجهات إيجابية نحوه.

الكلمات المفتاحية: تحليل المشاعر، تحليل المحتوى، التنقيب في البيانات، تصنيف الآراء، كوفيد-19، التطعيم، القياس متعدد الأبعاد.

Public Sentiment Analysis in Social Media on the SARS-CoV-2 Vaccination Using VADER Lexicon Polarity

Jeehaan Algaraady
Department of English, Taiz University, Taiz, Yemen

Mohammad Mahyoob
Department of Languages and Translations, Faculty of Science and Arts, Alula, Taibah University, Madinah, Saudi Arabia

Abstract

Recently Natural Language Processing (NLP) constituted an important area of computational linguistics and artificial intelligence, where the virtual and digital world has become an essential aspect of our daily lives. Sentiment analysis and data mining are sub-fields of NLP, which draw the attention of researchers to search and mine various issues on social media. This study explores the public's sentiments and opinions towards the SARS-CoV-2 vaccination doses in Saudi Arabia. It tries to provide insights on the motivations and barriers in taking the first and second vaccine doses and how the public's awareness and attitudes differ in the two doses. The research objects are 6.232 public tweets and comments that have been harvested from official social media platforms (Twitter and YouTube) between December 19, 2020, and December 10, 2021. The sentiment analysis measured polarity using the NLTK VADER analyzer, and the opinions were identified and classified based on the multidimensional scaling method. The results show that in the case of the first vaccine dose of the 2989 opinions enrolled, 61.5% of the public were willing to take the COVID-19 vaccination—the majority trust the vaccine safety and the Ministry of Health measures and decisions. While 21.1% show negative attitudes towards the vaccination, most of them untrust the vaccine and are worried about its syndromes. In the case of the second vaccine dose of the 3,243 opinions enrolled, 63.2% also show positive attitudes towards taking the vaccine. Trusting the vaccine safety and not being prevented from work, travel, and other activities are the primary motivations to receive the vaccine in this phase. While negative sentiments scored 30.3%, the most frequent determinant is the suspicion of the vaccine safety, symptoms, and decision discrepancies. Identifying public sentiments and attitudes toward COVID-19 vaccination would provide a better understanding of the reasons behind vaccine rejection or acceptance would help the health policymakers better develop and implement vaccine awareness strategies and appropriate communication to enhance vaccine taking.

Keywords: Sentiment Analysis, Content Analysis, Data Mining, Opinions Classification, COVID-19, Vaccination, Multidimensional Scaling

Introduction

Sentiment analysis and data mining are sub-fields of Natural Language Processing, drawing the attention of researchers to investigate various issues on social media. The rapid and devastating spread of the COVID-19 pandemic worldwide caused an immediate response from public health to contain the pandemic. Some COVID-19 vaccines were developed, approved, and rollout in less than a year. It is expected that over 70% of the population needs to be vaccinated to reach the level of herd community (Orenstein & Ahmed, 2017; Aguas et al., 2021). Measuring the public's sentiments and opinions towards vaccination is extremely important to identify the public's support for vaccination. And that helps achieve the public health goal to reach the level of herd community, stop outbreaks of vaccine-preventable illnesses, and ensure adoption of novel vaccines (Callender, 2016). Identifying public attitudes toward COVID-19 would provide a better understanding of the reasons behind vaccine hesitancy and how to develop better and implement vaccine awareness strategies (Eibensteiner, 2021). In addition, public health policymakers could design appropriate and effective communication to reach out to the public (Mitra et al., 2016; Salathé & Khandelwal, 2011). Recently, social media has been increasingly used for expressing and sharing individuals' opinions on various topics. Currently, the COVID-19 pandemic increased the use of social media by users besides public health professionals discussing many issues, including a vaccine. Social media network is a platform for surveillance and a helpful communication tool for worldwide health actors (Deiner et al., 2019). The individual's sentiments and views on social media are crucial in figuring out the public mode on different topics. Various government sectors have tried to preserve and mitigate the consequences of the pandemic. They have taken many precautionary measures, such as switching to online education (Mahyoob, 2021).

Consequently, the public's sentiments are influenced positively or negatively by individuals' opinions on vaccines. The information disseminated on social media could affect people's decision to accept, delay or refuse vaccination (Rosselli et al., 2016; Broniatowski et al., 2018). To detect these opinions and thoughts, sentiment analysis which is an application of opinion mining is, emerged as a technique to computationally analyze a piece of text applying natural language processing (NPL) (Padmaja & Fatima, 2013).

People's discussion on taking the COVID-19 vaccination started after the rollout of vaccines in the USA, UK, China, and Russia on social media, which highly increased in January 2020. A global survey on public

acceptance of COVID-19 vaccines revealed wide-ranging acceptance rates of below 55% to a high of about 90% (Lazarus et al., 2021). In KSA, people started discussing taking the first vaccine dose in December 2020, as the vaccine reached KSA on December 16, 2020. It was first limited to older people above 65 years old, those suffering from chronic diseases, and health workers (the first defense line), and later it was available for all adults. The health centers started giving the second dose on July 5, 2021, for people aged 40 years old or above, and later it was available for all adults. This discussion comprises a rich research object that needs to be analyzed to identify the public's support for vaccination.

Indeed, many researchers have conducted a content analysis of the public's tweeters about vaccination to explore and assess their sentiments and attitudes towards vaccination (Nuzhath et al., 2020; Piedrahita-Valdés et al., 2021; DeVerna et al., 2021; Kwok et al., 2021; Ritonga et al., 2021). Other studies discussed the role of Twitter opinions conversation on social media on vaccine hesitancy and oppositions (Bonnevie et al., 2021; Cossard et al.,2020; Puri et al., 2020). Indeed, this is the first study exploring public sentiments and reactions towards the COVID-19 vaccination in KSA. The study aims to assess the public emotions or attitudes towards the first and second COVID-19 vaccine doses between December 2020 and December 2021 on Twitter and Facebook in Saudi Arabia and the motives and barriers behind these opinions for and against the COVID-19 vaccination. It studies how their awareness and attitudes differ in the two doses. The sentiment metrics are assigned to the NLTK (VADER) dataset analyzer. Then dataset contents are categorized into motives for accepting the vaccine, including "trust the vaccine and their worries about their woks, study, travel, etc." and barriers or hesitancy for refusing the vaccine, including "untrust the vaccine, vaccine side effects, and vaccine protocols discrepancies". These categories are clustered based on the classic multidimensional scaling method representing the positive and the negative public attitudes. The authors employed the YouTube platform and the Twitter platform as a popular outlet and source expressing discussions and the actual and real-time opinions related to health information (Love et al., 2013). Indeed, providing the policymakers with an overall public attitude towards the vaccine enables them to enhance vaccine confidence among the population.

The rest of the paper is structured as follows: Section 2 introduces a brief description of the related studies in the literature. In section 3, the authors present the proposed method and data collection and describe the tool used in this study. Section 4 discusses the results. Section 5 concludes.

Literature review

Social networks emerged as an essential source for opinion expression and information mining. Many studies have been conducted to detect people's sentiments and opinions on the COVID-9 vaccine. Though many studies detected people's sentiments and analyzed datasets related to different topics about COVID-19, few studies dealt with the COVID-19 vaccine. Previous studies have performed sentiment and attitude analysis concerning the COVID-19. Villavicencio et al. (2021) proposed a Naïve Bayes model to perform a sentiment analysis of English and Filipino tweets (993) and classify them into positive, negative, and neutral using the RapidMiner data science software and scored 81.77% accuracy. Kwok et al. (2021) explored Australian Twitter users' sentiments about the COVID-19 vaccine between January and October 2020. They utilized the R library package *syuzhet* to assign each tweet with its sentiments (positive, negative) and eight emotions (anticipation, fear, disgust, trust, surprise, sadness, joy, and anger). Last, they identified three topics in the tweets: attitudes toward COVID-19 and the vaccination, misconceptions and complaints about COVID-19 control, and advocacy of infection control measures against COVID-19. Bonnevie et al. (2021) quantified the increase in Twitter conversations around vaccine opposition during the COVID-19 pandemic in the United States. They started collecting tweets, categorized them into topics, and then traced them. After four months of observation, they noticed an apparent increase in vaccine rejection on Twitter. Exposure to these increased amounts of vaccine opposition may mislead people to oppose vaccines, which could drastically impact the health of populations for decades to come. Therefore, to ensure the most comprehensive support for a COVID-19 vaccine, it is crucial to identify and investigate the messages used by vaccine opponents.

Thelwall et al. (2021) investigated the types of vaccine hesitancy information shared on Twitter to address the public's misleading attitudes. They discussed vaccine safety, conspiracies, and vaccine development speed as the main themes in the tweets. Their findings revealed that 79% of those who showed negative attitudes towards vaccines expressed right-wing views, conspiracy theories, or fear of the deep state. Lyu et al. (2021) utilized Kwok et al. (2021)'s method to explore public perceptions, concerns, emotions, and topics in general discussions related to the COVID-19 vaccine on social media and how they influence the achievement of herd immunity goals. Monselise et al. (2021) investigated the public sentiment and topics related to COVID-19 vaccines in their discussions about the vaccines on social media for 60 days starting from December 16, 2020,

when the vaccines were begun in the United States. The sentiments were identified using the sEntiment Reasoner sentiment analysis library and Valence Aware Dictionary, using sentence bidirectional encoder representations from transformer embeddings. The discussion topics were identified by nonnegative matrix factorization. Their results revealed that fear was the leading emotion in tweets, then joy, and the primary public concern was about the administration and access to vaccines. Hussain et al. (2021) utilized natural language processing and deep learning-based techniques to predict average sentiments, sentiment trends, and discussion topics on social media in the United Kingdom and the United States from March 1 to November 22, 2020. These aspects were analyzed longitudinally and geospatially, and manually reading randomly selected posts on points of interest helped recognize underlying themes and validated insights from the analysis. The finding revealed that public optimism over the vaccine trial, development, and effectiveness were identified besides corporation control, concerns over their safety, and economic viability.

Methodology

This study comprises four phases. Phase one concerns the data collection. The data is initially preprocessed in the second phase by cleaning and removing irrelevant information. The third phase introduces the use of NLTK'S VADER analyzer for analyzing and assigning the sentiments metrics to the data. The fourth phase involves classifying the data to introduce the main themes (motives and barriers for taking the vaccine) using the classic multidimensional scaling method.

a.Data collection

This section introduces the construction of the study harvested data. A total of 6,232 tweets, retweets, and comments have been collected from the Saudi Ministry of Health (MOH) official accounts on Twitter[1] and YouTube[2] concerning COVID-19 vaccination two doses. The data represents (n = 2,989) public's tweets and comments on receiving the COVID-19 vaccination after the declaration of the first dose registration in Saudi Arabia within the timeline of December 15, 2020, to July 1, 2021, and (n = 3,243) tweets and comments on the second dose register declaration from July 5, 2021, to December 20, 2021. Table one below shows a detailed description of the harvested data where type-token-ratio (TTR) measures the language elaboration. In collecting tweets, the irrelevant tweets are not considered for this analysis because they have no value for achieving the main goals of this analysis, i.e., they will not

---

(1) https://twitter.com/SaudiMOH?ref_src=twsrc%5Egoogle%7Ctwcamp%%7Ctwgr%5Eauthor
(2) https://www.youtube.com/channel/UCP_wyoQnTjZbFJt7DLGlE4A

provide any reasons (motivations and barriers) for accepting or refusing vaccination.

Table 1. The absolute number of Tweets and YouTube comments

	Tweets & Comments number	Tokens	Types	Type-token-ratio (TTR)
1st dose registration (December 19 202-July 1, 2021)	**2,989**	**39505**	**3590**	**0.0909**
2nd dose registration (July 5, 2021 – December 10, 2021)	**3,243**	**33180**	**3272**	**0.0986**

b. Preprocessing the data

After collecting the data and for results accuracy, the authors need to prune the text data, so the series of preprocessing steps were conducted to remove irrelevant information from the dataset such as stop words, users' names, posters, digits, abbreviations, profiles, timing, and other special characters using a regular expression (Regex) in Python. Then, tokenize the tweets into individual words, tokens, and stems using the NLTK function (Porter Stemmer). After the data preprocessing phase, the word types (Arabic without sentiment weights) are converted into English versions to facilitate sentiment analysis. Now, the data are ready for sentiment classification. In the third phase, the data sentiments were classified. Each word is given its appropriate weight between (1 and -1) for positive, negative, and neutral with the support of Valence Aware Dictionary and sEntiment Reasoner (VADER), a python lexicon, and a rule-based sentiment analysis tool. It calculates input sentiment scores and expresses sentiments in social media (Hutto & Gilbert, 2014). VADER is designed to determine the sentiments of social media posts based on individual words and sentences (Elbagir & Yang, 2019). First, we applied a sentiment intensity analyzer to classify the preprocessed data, and the outcome metric has four parts: positive, negative, neutral, or compound score, as shown in table two. Then the polarity scores method was applied to define the sentiment. The compound score is the sum of the lexicon ratings, standardized values between -1 and 1, and was used as a classifier, i.e., comment or tweet with a compound score more than or equal to 0.05 is classified as a positive sentiment. While a comment or tweet with a compound score less than or equal to -0.05 is identified as a negative sentiment**,** and any score between those values is considered a neutral sentiment.

Table 2. Sentiment scores as positive, negative, neutral, and compound assigned by NLTK VADER Sentiment Analyzer

SA ('The vaccine has side effects that are produced after the vaccine')
(اللقاح له عوارض جانبية تنتج بعد اللقاح)
{'neg': 0.104,
'neu': 0.896,
'pos': 0.0,
'compound': -0.4874}
SA('Take the vaccine, it is effective and excellent, and the government will not bring something harmful to people')
(خذو اللقاح ترا فعال وممتاز ومستحيل الحكومة تجيب شي مضر للناس)
{'neg': 0.0,
'neu': 0.779,
'pos': 0.221,
'compound': 0.7783}

c. Public's opinions categorization

The public's opinions on the vaccine have been classified into five categories; trusting vaccination, keeping life activities (work, study, travel, etc.) as public positive orientation or motivations towards taking the vaccination; and untrusting vaccination, vaccination syndromes, and vaccine protocols discrepancies as public negative orientation and barriers to taking the vaccine. The first opinion category conveys positive public beliefs such as vaccination is a cause of increasing immunity, decreasing the affected cases, and lightening the effects of COVID-19. They trust their government and its measures and recommendations. The second category displays that people are willing to receive a vaccination that enables them to practice their life activities smoothly, such as traveling, studying, working, entering public and private institutions, etc. The third category reflects the people's convictions about the speed of development of the COVID-19 vaccine, and they think that it is still under experiment and harmful. The fourth category reveals people's resistance towards vaccinations due to the vaccine's side effects. The last category indicates that the public is against vaccination, and they are upset by the inconsistencies in the proposed protocols related to taking the vaccine. These categories are labeled by enquiring specific phrases as representatives of the proposed categories from the data with Boolean operators. These phrases are selected manually by scanning the data and assessing (n= 1000 tweets). Then the opinion categories are labeled by enquiring these phrases from data with Boolean operators and revised.

These categories have been used as codes in the QDA software (Qualitative Data Analysis software) for facilitating clustering the same data and thus classifying the data based on the classic multidimensional scaling. QDA software provides a data annotation and a lexicon approach with assessing metrics for opinion mining and allows a systematic qualitative and quantitative content analysis (Mahyoob et al., 2020). It is worth mentioning that there is an overlapping between the categories as they are in some cases labeled to the same data, which is illustrated in the results section. After annotating the data with the five categories, the public's opinions are compared based on these categories.

Results and discussion

The results of the top ten positive and negative words, the comments and tweets sentiment analysis using VADER analyzer, and public opinions categories on taking the COVID-19 vaccine are discussed in this section. Table three displays the top ten positive words related to the keyword "vaccine" for the first and second doses declarations with their frequencies in the second column, sentiment polarity scores in the third column, and their weights in VADER in the fourth column. For both doses, the word Allah, الله "God" is the highest positive word with a score of 0.2732 and weighs 1.1 in VADER (word occurrence 685 times for the first dose and 776 times for the second dose). The following positive word is - accept, "وافق" with a positive score of 0.3818 and weighs 1.6 according to VADER. While the words solution, "حل" and natural, "طبيعي" are the lowest positive with a score of 0.3182 and 0.3612, respectively.

Table 3. The top ten positive words in the context of COVID-19 for the first and second doses declarations

Word		Frequency		Vader weights	
English	**Arabic**	**First dose**	**Second dose**	**Word score**	**Word polarity**
Allah	الله	**685**	**776**	**0.2732**	**1.1**
accept	وافق	**690**	**884**	**0.3818**	**1.6**
thanks	شكرا	**455**	**340**	**0.4404**	**1.9**
good	خير	**265**	**149**	**0.4404**	**1.9**
effective	فعّال	**290**	**135**	**0.4767**	**2.1**
wellness	عافية	**200**	**165**	**0.4404**	**1.9**
free	مجان	**40**	**90**	**0.5106**	**2.3**
want	يريد	**75**	**109**	**0.0772**	**0.3**
solution	حل	**50**	**60**	**0.3182**	**1.3**
natural	طبيعي	**35**	**45**	**0.3612**	**1.5**

Similarly, table four displays the top ten negative words related to the keyword "vaccine" for the first and second doses declarations. For both

doses, the word NO, "لا" is the highest positive word with a score of -0.296 and weighs -1.2 in VADER (word occurrence 1210 times for the first dose and 924 times for the second dose). The table below illustrates that the terms "stop, dangerous, and forced scored high frequency and weights". While the words poison, "سم" and Conspiracy, "مؤامرة" are the lowest frequent negative with a score of -0.5267and -0.5423, respectively.

Table 4. The top ten negative words in the context of COVID-19 for the first and second doses

Word		Frequency		VADER weights	
English	Arabic	First dose	Second dose	Word score	Words polarity
no	لا	1210	924	-0.296	-1.2
stop	وقف	280	140	-0.296	-1.2
dangerous	خطير	265	180	-0.4767	-2.1
forced	مجبر	223	422	-0.4588	-2.0
affected	مصاب	190	106	-0.1531	-0.6
complain	حسبنا	132	129	-0.3612	-1.5
refuse	رفض	102	82	-0.296	-1.2
died	مميت	75	35	-0.5574	-2.6
poison	سم	65	35	-0.5423	-2.5
conspiracy	مؤامرة	55	38	-0.5267	-2.4

Figure one summarizes the overall frequency of the positive and negative tweets and comments in the keyword context (Corona). It is noted that the sentiments related to the COVID-19 vaccine changed from phase to phase. Generally, the positive sentiment was dominant for the first and second doses in KSA. Positive opinions showed more significant engagement metrics than negative opinions, which reached 61.5% and 63.2%, respectively, forming approximately two-thirds of sentiments. It is similar to the results of recent studies on the public's sentiments towards the COVID-19 vaccine by (Kwok et al., 2021; Hussain et al., 2021; Piedrahita-Valdés et al., 2020).

In contrast, the negative sentiment scored 21.1% in the first dose. It increased to 30.3% in the second dose, which displays the publics' negative experience with the vaccine and its influence by the negative disseminated news about the vaccine on social media. The neutral sentiment scored 15.7% in the first dose and decreased to 8.2% in the second dose, revealing that public attention to the vaccine increased.
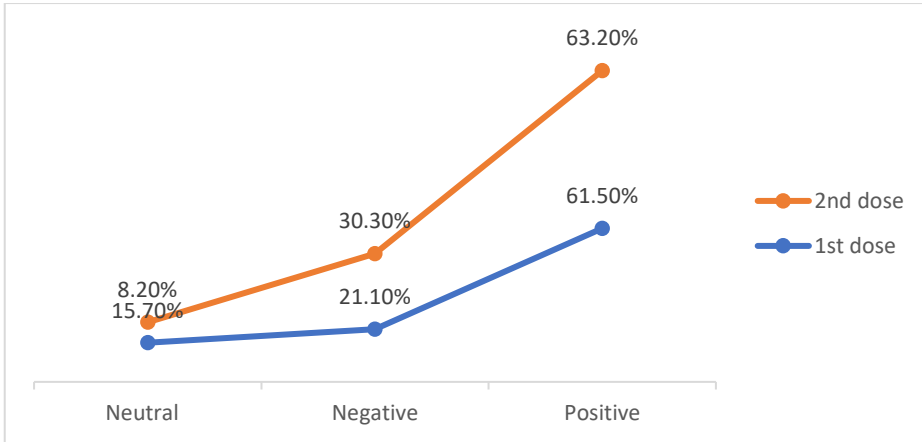
Fig. 1. Overall sentiment polarity of Tweets and YouTube comments

Opinions' categories frequencies and relations in data

a. **Phase one: The public's opinions frequency and relations for the first dose declaration of the COVID-19 vaccination**

Figure two below illustrates the public's opinions categories frequency and similarity for the first dose from December 15, 2020, to July 1, 2021. It introduces the main motivations and barriers to receiving the vaccine. Each circle in the map represents an opinion category that has been assigned in the data. The similarity between the opinions categories based on the classic multidimensional scaling method is defined by the distance between the circles. The larger the circle in the map, the more frequent the category it represents. The connecting lines between the categories' circles displayed coincidences between every two categories. It is worth mentioning that the more two opinions intersect, the more comparably they are utilized in the data, the closer they are positioned or clustered together on the map. Of the 2,989 opinions enrolled, the "trusting the vaccine" category scored the highest positive frequency (52.9%) and was assigned to 1627 opinions, while the second category, "work, study, and other life activities", scored 2.5% and was given to 73 opinions.

In contrast, the "untrusting the vaccine" category scored the highest negative frequency, with 35.4% allocated to 1022 opinions because of the safety of rapidly developed COVID-19 vaccines. It was monitored as the main reason for untrusting the vaccine, which aligns with Elbagir & Yang (2019); Mahyoob et al. (2020), who stressed the assurance of vaccine safety as the main reason for accepting the vaccination. The second frequent negative category is "vaccine's side effects", which reached a 7.6% score and introduced the attitudes of 220 subjects. The "protocols discrepancies" category scored the lowest negative value with 1.6% and represented only 47 opinions of the obtained data.
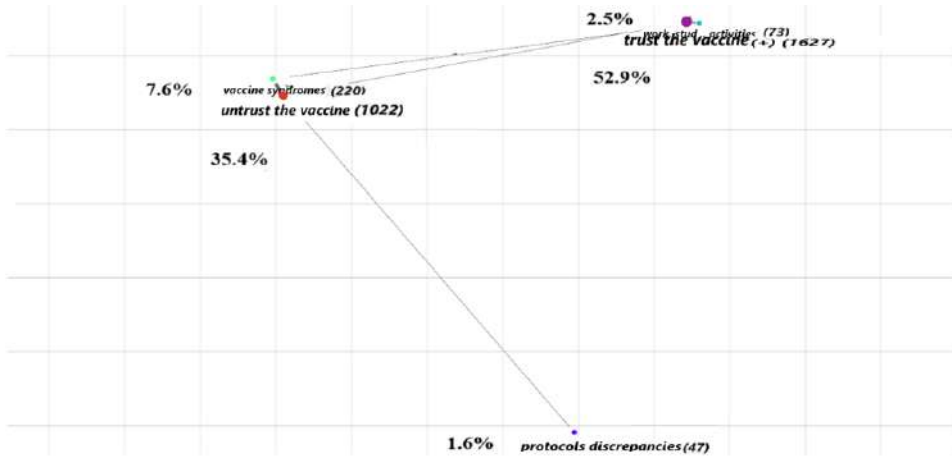
Fig. 2. public's opinions frequency and relations for the vaccine first dose

The low score of the category "work, study, travel, life activities" in this phase displays that the public does not show considerable worries about their work, travel, and other life activities. Because at this phase, there were no official obligations to be immunized to travel, work, study, or enter any public or private sectors. In the same vein, the map shows that unmentioned opinions with a 1.6% score were upset with the protocols discrepancy because the protocols were still unknown and consequently unannounced, so no considerable reactions were monitored from the public's side.

The closure distance between the two categories: "trust the vaccine and work, study, and other life activities" means they share the positive tone towards receiving the vaccination, while the closure distance between "untrusting the vaccine" and "vaccine's side effects" categories indicates their negativity along with "protocols discrepancy". The category: "untrusting the vaccine" is connected and closer to the "vaccine syndromes" category rather than the "protocols discrepancies" category. Furthermore, the distance between "vaccine syndromes" and "protocols discrepancies" is further with no connection. It means that those who untrust the vaccine are at most worried about the vaccine effectiveness and side effects of the rushed speed developed the vaccine that is similarly found by (Nuzhath et al. 2020). In contrast, the positive "work, study, travel, life activities" category is far from negative. It connects only with the two negative categories, "untrusting the vaccine" and "vaccine's syndromes", which means they are untrusting the vaccine but willing to receive the vaccine because they want to keep their activities on safely.

b. Phase two: The public's opinions frequency and relations for the ^{second} dose declaration of the COVID-19 vaccination

Figure three shows the opinions categories relation for the second dose of the COVID-19 vaccination, and the positive categories are clustered together on the left side. In contrast, the negative categories are clustered on the right side. Compared to the first phase, the number of those who trusted the vaccine decreased after the declaration of the second dose of the COVID-19 vaccination from 52.9 in the first phase to 43.0% score in the second phase. The number of people who accepted to receive the vaccine to keep their work, study, travel, shopping, meetings, etc. notably increased from 2.5% in the first phase to 14.5 scores in the second phase as the vaccine becomes officially obligated for practicing all these activities on October 10, 2021. Similarly, the number of people who untrusted the vaccine also decreased to 22.2%, but their worry about the vaccine side effects and upset with the protocol's discrepancies increased more to reach 14.3% and 5.9%, respectively. That indicates some suffered from the vaccine's side effects and suspect the vaccine due to the changing proposed protocols regarding the age, time of immunity offered by the vaccine, and the number of doses from time to time.

As displayed in the map, the two positive codes: "trusting the vaccine" and "work, study, and life activities", are more coincidences and closer to each other that indicating they share some of the labeled data, i.e., some of the public accept to receive the vaccine to keep their activities going on also trust the vaccine. However, it implies that some people are willing to receive the vaccine for keeping life activities and do not necessarily trust the vaccine. In contrast, the distance between the negative opinion categories "vaccine syndromes" and the positive opinion categories: "work, study, life activities" is too much with a thinner connecting line, which indicates some accept the vaccination for keeping their work, study, etc.; but they were worried about the vaccine syndrome. In contrast, there is no connecting line between the positive opinion "trusting the vaccine" and any other negative opinions, which indicates that those who trusted the vaccine did not have any negative beliefs on the vaccine.

In contrast, there are connecting lines between all the three negative opinions categories which means they are assigned similarly to the segments in different degrees as displayed by the size of the lines. The category: "untrusting the vaccine" is closer to the "vaccine syndromes" category rather than the "protocols discrepancies", and the distance between "vaccine syndromes" and "protocols discrepancies" categories is further. It means that among those who untrust the vaccine, some express concerns about the

potential vaccine syndromes of the COVID-19 vaccine rather than showing upset of the discrepancy in the proposed protocols regarding taking the vaccine, which is in line with (Nuzhath et al. 2020, Kwok et al., 2021). Moreover, very few of those worried about the vaccine syndromes show negative attitudes to the vaccine because of the inconsistency in the proposed vaccination protocols.
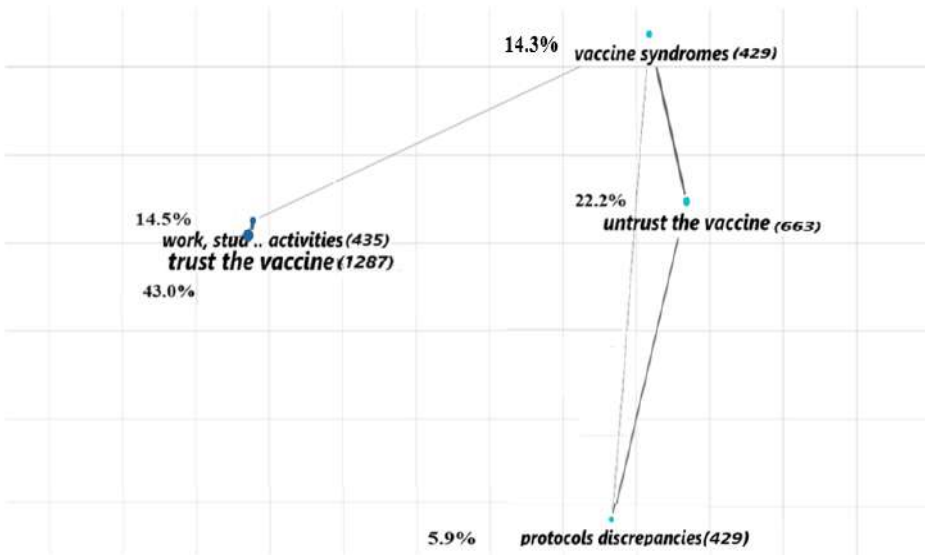

Fig. 3. Public's opinions frequency and relations for the vaccine ^{second} dose

Moreover, only one negative category, "vaccine syndromes", is connected to the positive category: "work, study, life activities" with a thinner connecting line, which reveals some of those who are obligated to receive the vaccine for keeping their life activities show negative tone and are worried of the side effects of the vaccine.

Conclusion

The rapid development and rollout of the SARS-CoV-2 raise various public sentiments that demand understanding, affecting vaccination taking. In this study, the authors tapped into 6,232 tweets and comment harvested from official social media platforms (Twitter and YouTube) between December 19, 2020, and December 10, 2021, not only to assess the public sentiments towards the COVID-19 vaccination doses in KSA but also to gain insights into their motives and barriers behind these sentiments to accept or reject the vaccine and how their awareness and attitudes differ for the two doses. The metrics were assigned and classified using NLTK'S Valence Aware Dictionary and sEntiment Reasoner (VADER) analyzer**.** Through sentiment mining and analysis, the results revealed that the positive

sentiment about the COVID-19 vaccination two doses was the dominant polarity and having higher engagements towards both doses. These results figurate the main reasons behind public positive and negative attitudes towards the vaccine as discussed on social media. The main motive behind this positivity is the public's trust in vaccine safety and the Ministry of Health's measures and decisions. Then, their concerns on their work, study, travel, and other activities, while the negative sentiment increased after the declaration of the second dose as people were unwilling to take the vaccine and show their worries on the vaccine safety, syndromes, and decisions and protocols discrepancy. Understanding public sentiments and the reasons behind the vaccine rejection or acceptance of COVID-19 vaccination doses would help the health policymakers design better and implement vaccine awareness strategies and appropriate, effective communication to boost vaccine taking.

Declarations of Competing Interest

The authors declare no conflict of interest.

References

Aguas, R., King, J. G., Gonçalves, G., Ferreira, M. U., & Gomes, M. G. M. (2021). Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. *MedRxiv*, 2020-07.

Bonnevie, E., Gallegos-Jeffrey, A., Goldbarg, J., Byrd, B., & Smyser, J. (2021). Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *Journal of communication in healthcare*, *14*(1), 12-19.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., ... & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health*, *108*(10), 1378-1384.

Callender, D. (2016). Vaccine hesitancy: more than a movement. *Human Vaccines & Immunotherapeutics*, *12*(9), 2464-2468.

Cossard, A., Morales, G. D. F., Kalimeri, K., Mejova, Y., Paolotti, D., & Starnini, M. (2020, May). Falling into the echo chamber: the Italian vaccination debate on Twitter. In *Proceedings of the International AAAI conference on web and social media* (Vol. 14, pp.130-140).

Deiner, M. S., Fathy, C., Kim, J., Niemeyer, K., Ramirez, D., Ackley, S. F., ... & Porco, T. C. (2019). Facebook and Twitter vaccine sentiment in response to measles outbreaks. *Health informatics journal*, *25*(3), 1116-1132.

DeVerna, M., Pierri, F., Truong, B., Bollenbacher, J., Axelrod, D., Loynes, N., ... & Bryden, J. (2021). CoVaxxy: A global collection of English Twitter posts about COVID-19 vaccines. *arXiv e-prints*, arXiv-2101.

Eibensteiner, F., Ritschl, V., Nawaz, F. A., Fazel, S. S., Tsagkaris, C., Kulnik, S. T., ... & Atanasov, A. G. (2021). People's willingness to vaccinate against COVID-19 despite their safety concerns: Twitter poll analysis. *Journal of medical Internet research*, *23*(4), e28973.

Elbagir, S., & Yang, J. (2019, March). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 122, p. 16).

Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., ... & Sheikh, A. (2021). Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. *Journal of medical Internet research*, *23*(4), e26627.

Hutto, C., & Gilbert, E. (2014, May). 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* (Vol. 8, No. 1).

Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Twitter Speaks: An Analysis of Australian Twitter Users' Topics and Sentiments About COVID-19 Vaccination Using Machine Learning. *Journal of Medical Internet Research*.

Lappeman, J., Munyai, K., & Kagina, B. M. (2021). Negative sentiment towards COVID-19 vaccines: A comparative study of USA and UK social media posts before vaccination rollout. *F1000Research*, *10*(472), 472.

Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., ... & El-Mohandes, A. (2021). A global survey of potential acceptance of a COVID-19 vaccine. *Nature medicine*, *27*(2), 225-228.

Love, B., Himelboim, I., Holton, A., & Stewart, K. (2013). Twitter as a source of vaccination information: content drivers and what they are saying. *American journal of infection control*, *41*(6), 568-570.

Lyu, J. C., Le Han, E., & Luli, G. K. (2021). COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. *Journal of medical Internet research*, *23*(6), e24435.

Mahyoob, M., Al-Garaady, J., & Alrahaili, M. (2020). Linguistic-based detection of fake news in social media. *Forthcoming, International Journal of English Linguistics*, *11*(1).

Mahyoob, M. (2021). Online Learning Effectiveness During the COVID-19 Pandemic: A Case Study of Saudi Universities. *International Journal of Information and Communication Technology Education (IJICTE)*, *17*(4), 1-14.

Mitra, T., Counts, S., & Pennebaker, J. W. (2016, March). Understanding anti-vaccination attitudes in social media. In *Tenth International AAAI Conference on Web and Social Media*.

Monselise, M., Chang, C. H., Ferreira, G., Yang, R., & Yang, C. C. (2021). Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. *Journal of Medical Internet Research*, *23*(10), e30765.

Nuzhath, T., Tasnim, S., Sanjwal, R. K., Trisha, N. F., Rahman, M., Mahmud, S. F., ... & Hossain, M. M. (2020). COVID-19 vaccination hesitancy, misinformation and conspiracy theories on social media: A content analysis of Twitter data.

Orenstein, W. A., & Ahmed, R. (2017). Simply put: Vaccination saves lives. *Proceedings of the National Academy of Sciences*, *114*(16), 4031-4033.

Ritonga, M., Al Ihsan, M. A., Anjar, A., & Rambe, F. H. (2021, February). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012045). IOP Publishing.

Rosselli, R., Martini, M., & Bragazzi, N. L. (2016). The old and the new: vaccine hesitancy in the era of the Web 2.0. Challenges and opportunities. *Journal of preventive medicine and hygiene*, *57*(1), E47.

Padmaja, S., & Fatima, S. S. (2013). Opinion mining and sentiment analysis-an assessment of peoples' belief: A survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing*, *4*(1), 21.

Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J.

Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human vaccines &immunotherapeutics*, *16*(11), 2586-2593.

R., Guillem-Saiz, J., ... & Machío-Regidor, F. (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines*, *9*(1), 28.

Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS computational biology*, *7*(10), e1002199.

Thelwall, M., Kousha, K., & Thelwall, S. (2021). Covid-19 vaccine hesitancy on English-language Twitter. *Profesional de la información (EPI)*, *30*(2).

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information*, *12*(5), 204. [1] Orenstein WA, Ahmed R. Simply put: vaccination saves lives. National Acad Sciences; 2017.