



Automatically detecting and understanding the perception of COVID-19 vaccination: a middle east case study

Wajdi Aljedaani¹ · Ibrahem Abuhaimeed² · Furqan Rustam³ · Mohamed Wiem Mkaouer⁴ · Ali Ouni⁵ · Ilyes Jenhani⁶

Received: 8 February 2022 / Revised: 19 July 2022 / Accepted: 21 July 2022 / Published online: 4 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Introduction The development of COVID-19 vaccines has been a great relief in many countries that have been affected by the pandemic. As a result, many governments have made significant efforts to purchase and administer vaccines to their populations. However, accommodating such vaccines is typically confronted with people's reluctance and fear. Like any other important event, COVID-19 vaccines have attracted people's discussions on social media and impacted their opinions about vaccination.

Objective The goal of this study is twofold: First, it conducts a sentiment analysis around COVID-19 vaccines by automatically analyzing Arabic users' tweets. This analysis has been spread over time to better capture the changes in vaccine perceptions. This will provide us with some insights into the most popular and accepted vaccine(s) in the Arab countries, as well as the reasons behind people's reluctance to take the vaccine. Second, it develops models to detect any vaccine-related tweets, to help with gathering all information related to people's perception of the virus, and potentially detecting vaccine-related tweets that are not necessarily tagged with the virus's main hashtags.

Methods Arabic Tweets were collected by the authors, starting from January 1st, 2021, until April 20th, 2021. We deployed various Natural Language Processing (NLP) to distill our selected tweets. The curated dataset included in the analysis consisted of 1,098,376 unique tweets. To achieve the first goal, we designed state-of-the-art sentiment analysis techniques to extract knowledge related to the degree of acceptance of all existing vaccines and what are the main obstacles preventing the wide audience from accepting them. To achieve the second goal, we tackle the detection of vaccine-related tweets as a binary classification problem, where various Machine Learning (ML) models were designed to identify such tweets regardless of whether they use the vaccine hashtags or not.

Results Generally, we found that the highest positive sentiments were registered for Pfizer-BioNTech, followed by Sinopharm-BIBP and Oxford-AstraZeneca. In addition, we found that 38% of the overall tweets showed negative sentiment, and only 12% had a positive sentiment. It is important to note that the majority of the sentiments vary between neutral and negative, showing the lack of conviction of the importance of vaccination among the large majority of tweeters. This paper extracts the top concerns raised by the tweets and advocates for taking them into account when advertising for the vaccination. Regarding the identification of vaccine-related tweets, the Logistic Regression model scored the highest accuracy of 0.82. Our findings are concluded with implications for public health authorities and the scholarly community to take into account to improve the vaccine's acceptance.

Keywords Sentiment Analysis · COVID-19 pandemic · COVID-19 Vaccine · Vaccine uptake · Machine learning · Deep learning · NLP

1 Introduction

The COVID-19 pandemic has led to disruption of nearly every aspect of human life, leading to job losses, and deaths, among other unfortunate events. In that regard, the development of vaccines has been a great relief and offers hope for a better future. Many governments have invested a lot

Wajdi Aljedaani
wajdialjedaani@my.unt.edu

Extended author information available on the last page of the article

of money and efforts to ensure that their citizens get vaccinated in order to reach 'herd immunity' status. However, it is unfortunate that some people are not willing to take the vaccines. There are many reasons that could make people resent vaccination, which can be understood by evaluating their perceptions towards the vaccines.

Several previous studies have examined public sentiments on vaccines using Twitter Data Set. For example, Villavicencio et al. (2021) used English and Filipino tweets to get public opinions on vaccination in the Philippines, while Nurdeni, Budi, and Santoso conducted a similar study in Indonesia using English tweets. Another study by Batra et al. (2021) evaluated sentiments in India, the USA, Norway, and Pakistan using English tweets. However, there are no studies that have been done in Arabic countries, or using Arabic tweets, in order to highlight how people in such countries perceive the COVID-19 vaccines. Furthermore, there are no studies that have analyzed why people are reluctant to take the vaccines and what are the main concerns that people are raising on social media against them taking the vaccines. Yet, this is critical to know so that governments and organizations would be able to perform the proper measures to address these concerns either through advertising or targeting that audience with further explanations. Another alarming problem that organizations are facing is the spread of misleading information on social media in general and on Twitter and particular. Therefore it will be interesting to automatically detect all information related to vaccines before it is widespread so that proper information can be circulated instead. That is why there is a need to detect any vaccine-related to its regardless of whether they are hashtagged as a vaccine or not.

To address the above mentioned challenges, this study aims to conduct an Arabic sentiment analysis of COVID-19 Vaccines tweets. In particular, we utilize Twitter data to understand the perceptions of people towards the COVID-19 vaccine using machine and deep learning techniques. We specifically utilize over a million tweets that were made in Arabic. The outcomes of this research will be very useful to public health authorities in understanding public perceptions towards the COVID-19 vaccine and determining strategies that may be useful in making vaccination policies and campaigns. To the best of our knowledge, this is the first study to utilize Arabic tweets to conduct a sentimental analysis of vaccines.

Specifically, the research will be answering the following three questions:

RQ₁: What is the highest accepted vaccine in the Arab world?

This research question seeks to identify the most popular vaccine among Arabic people. We evaluate the most popular vaccine among the existing ones, including Moderna, Johnson & Johnson's Janssen, PfizerBioNTech, Novavax,

AstraZeneca, and Sputnik. In this study, we use public sentiments to establish whether people have positive, neutral, or negative sentiments towards vaccines.

RQ₂: What are the reasons that prevent people from taking the vaccine?

This research question seeks to understand why people are reluctant to take vaccines. Despite efforts to administer vaccines by most governments, some people are hesitating or reluctant to take the vaccines. We utilize the twitter data-set to generate reasons that could make people resent vaccines, fear them, or have a negative perception towards them.

RQ₃: To what extent can machine learning and deep learning detect vaccine-related tweets?

In this research question, we seek to know whether machine learning techniques can be used to identify tweets that have not been written explicitly as COVID-19 tweets. In addition, we want to know the extent of effectiveness of such techniques in determining COVID-19-related tweets in our Twitter data set. We selected a variety of common techniques that we could compare in terms of their precision, accuracy, and recall.

To summarize, the paper makes the following contributions:

- **RQ₁.** To find the highest accepted vaccine, we extracted a large corpus of 1,098,376 Arabic tweets about COVID-19. This corpus represents a dataset for the community to leverage for extension purposes¹.
- **RQ₂.** We manually analyzed a sample of 2,500 randomly selected tweets to extract what are the main reasons that push tweeters to advise against getting vaccinated.
- **RQ₃.** We used natural language processing, machine learning, and deep learning techniques to address our investigated research questions. We report the best performing models, along with providing their source code as part of the replication package².

Paper organization. The related works discuss in Sect. 2, and we describe our study methodology, such as experiment overall, data collection, data preprocessing, feature engineering, machine learning, deep learning, sentiment analysis, and evaluation parameters in Sect. 3. Then in Sect. 4, where we present and discuss the study results. Finally, Sect. 5 concludes the paper and highlights the direction of future work.

¹ <https://wajdialjedaani.github.io/ArTweetCovid19/>.

² <https://wajdialjedaani.github.io/ArTweetCovid19/>.

2 Related work

Vaccination apprehensions and behavioral tendencies among the masses have been observed extensively by researchers. A variety of obstacles to vaccine campaigns' effectiveness have been reported. For vaccine approval, motivational and information conversion methods have been used (World Health Organization et al 2020a). These tactics include public awareness programs to offer information and raise public awareness of the seriousness of the issue. The main goal of these public information programs is to learn about people's objections, fears, and existing public perceptions of the vaccine.

In recent years, social media has been the preferred method of interacting with the general public. According to researchers and experts in public relations (PR), social media plays a critical role in engaging people and framing successful campaigning tactics to reach out to individuals (Allagui and Breslow 2016). The researchers in Jones and Roy (2021) discovered that social media has powerful PR implications and a long-standing debate in the literature. According to the researchers, using social media in public relations will result in successful and widespread interaction and positively influence public behavior (Valentini 2015). Many countries across the world have begun, and some are preparing, to launch the COVID-19 vaccine in the coming months of 2021, and they must monitor the behavioral habits of their populations. The participants involved in the process and the policies framed by governments and the concerned agencies would significantly impact these behavioral trends in adopting the COVID-19 vaccine. The ability to collect real-time data and make dynamic decisions based on the data would significantly impact the effectiveness of vaccine campaigns. Engagement with the target community by listening to and reacting to their concerns, desires, and difficulties relevant to vaccination can steer the decision-making process toward an efficient and safe vaccination drive (World Health Organization et al. 2020b).

Sentiment analysis uses Natural Language Processing (NLP) methods to determine the beliefs, emotions, and perceptions of human society about someone or something (Keith Norambuena et al. 2019). It is not sufficient to identify and consider individual terms in the text during the sentiment analysis process. Therefore, it necessitates examining sentences in terms of their linguistic structures. The true expression of the emotion can be deduced by focusing on the vocabulary and linguistic properties of the language. Heuristic approaches are often used in linguistic construction research. Researchers have identified heuristics for sentiment analysis in a variety of domains. For example, in Pang et al. (2002), the analysis was conducted

on film critics based on the premise that negation scope is defined as the terms that appear between the original punctuation mark and the negator. Parts of Speech marking (POS) are used in another analysis by the authors (Taboada et al. 2011) to produce data for negation scope recognition. Sentiment research has the potential to be complicated. The method begins with discovering and selecting relevant content on the subject of concern. Text content processing is not an easy task; the enormous linguistic subtleties of natural languages present many difficulties. Using acceptable grouping, the emotion orientation must be calculated. Various methods and approaches can be used to address these issues (Bakshi et al. 2016). Sentiment analysis is a classification issue in which text fragments are categorized as Positive, Negative, or Neutral regarding the targeted subject (Bakshi et al. 2016).

Several researchers have conducted related studies that may be important to review in order to elucidate the gap that we will be addressing in our study. In research by Vilavicencio et al. (2021) in the Philippines, English and Filipino tweets were collected to explore attitudes towards the COVID-19 vaccine. The researchers utilized Natural Language Processing (NLP) techniques to understand the general sentiments, while further sentimental analysis was done using Naïve Bayes Model. The results gave an accuracy of 81.77%. In another study in Indonesia by Nurdini et al. (2021), the researchers collected English tweets between October and November 2020 to assess the opinions of social media users on Sinovac and Pfizer vaccines. The Support Vector Machine technique used in the study yielded accuracy of 85% and 78% with the Sinovac and Pfizer datasets, respectively. Another study by Cimorelli et al. (2021) collected English tweets in February 2021 to understand the public sentiments towards COVID-19 vaccination. The researchers used R statistical software and found that the general sentiment for the vaccine was 46%, compared to a national average of 54%. In research by Cotfas et al. (2021) in the UK, the researchers collected tweets to analyze public sentiments on COVID-19 vaccines in November 2020. They used the following analytical classification methods: Multinomial Naïve Bayes (MNB), Random Forest (RF), Support Vector Machine (SVM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Convolutional Neural Network (CNN). The paper achieved an accuracy of 78.94%. The paper achieved an accuracy of 78.94% in determining the correct sentiment.

A study by Chowdhury et al. (2021) in Bangladesh conducted a survey to determine public opinions on COVID-19 vaccines. Various analytical methods were used, but the BERT model and Naïve Bayes gave the highest accuracy of 84% and 81%, respectively. Another research by Batra et al. (2021) collected tweets from Norway, India, Pakistan, the USA, Sweden, and Canada to evaluate the cross-cultural

emotions towards the COVID-19 vaccine. The classification models that were used in the study were Deep Neural Network (DNN), Long Short Term Memory (LSTM) Network, and the Convolution Neural Network (CNN). The researchers found that people in neighboring countries showed similar attitudes towards COVID-19 vaccination, even if their actions and reactions towards the pandemic were different. In a study by Yousefinaghani et al. (2021) in Canada, the researchers analyzed tweets between January 2020 and January 2021 to get public opinions about COVID-19 vaccines. The researchers used the Valence Aware Dictionary and Sentiment Reasoner (VADER) method and found predominantly positive sentiments from the data set. Finally, research by Hussain et al. (2021) utilized Facebook and Twitter data to explore public attitudes towards COVID-19 vaccines in the USA and the United Kingdom. The researchers utilized deep learning-based techniques and natural language processing, which indicated higher positive sentiments in the UK compared to the USA. It is evident that most of the existing works analyzed the sentiments towards vaccination in general and not towards specific vaccines and were also in other languages apart from Arabic.

Our study differs considerably from the related studies that have been reviewed. To the best of our knowledge, this is the first study that has analyzed Arabic tweets and focused on the Arabic world in general. Table 1 presents a summary of the systematic analysis studies in the related work. In this study, we are conducting sentiment analysis for each specific vaccine. Hence, we are not only utilizing Arabic tweets in this study, but our study is also unique in that we want to analyze Arabic tweets for each vaccine. Using sentiment analysis, we will be able to determine whether people have a positive, neutral, or negative attitude towards the vaccines. We have used three methods of sentiment analyzers, such as Mazajak, TextBlob, and CAMeL, and various machine learning and deep learning techniques, which are not common in previous studies. In addition, our focus period is from 1st January 2021 to 20th April 2021. We have also used the largest dataset, which is provided in Table 3, compared to all the other studies. Therefore, this study will significantly contribute to the existing literature on public opinion toward COVID-19 vaccines.

Besides our work, there exist two large-scale Arabic tweets datasets (Haouari et al. 2020; Alqurashi et al. 2020). However, both of these datasets are related to Arabic tweets of COVID-19 in general. Our study provides the first dataset related to Arabic tweets about COVID-19 vaccines. Also, our dataset comes with the raw data and the corresponding sentiment of each tweet, which is not available in the other tweets datasets.

3 Materials and methods

This research uses numerous methods to overcome issues with classification (Fig. 1), as presented in Fig. 2. In the figure, we selected 34 keywords, which we used to get 2,046,073 tweets from the Twitter library. After collecting the tweets, we applied four phases, removing links, duplicates, and filtering, to develop a final dataset of 1,098,376 tweets. Then, we performed data pre-processing involving the removal of punctuation marks and usernames, among others, to prepare the data for the following sentiment analysis. In the second portion of the figure, sentiment analysis TextBlob, Mazajak (Farha and Magdy 2019), CAMeL (Obeid et al. 2020), and feature engineering techniques Term Frequency-Inverse Document Frequency(TF-IDF) were used to prepare the data for machine learning algorithms. We used five machine learning classifiers Random Forest (FR), Logistic Regression (LR), Gaussian Naive Bayes (GNB), Gradient Boosting Machine (GBM), Extra Tree Classifier (ETC), and three deep learning classifiers Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Multilayer Perceptron (MLP). To analyze the results, we used the evaluation metrics accuracy, recall, precision, and F1-score.

3.1 Data collection

We used Twitter, one of the most popular social media, as the data source for the study. In particular, based on Twitter posts, we are interested in the sentiment analysis for the COVID-19 vaccination. To collect a dataset that is related to vaccinations, we used keywords that are relative to our topic. The dataset includes all of the tweets linked to the Covid-19 vaccine. The Arabic language was chosen because it is the native language in the middle east. Relevant keywords, such as "Take vaccine" (i.e., "خذ اللقاح" in Arabic) was used to filter the messages/tweets. We collected only the related Arabic tweets based on which were determined by 34 keywords listed with the corresponding English words in Table 3. The Twint library³ was used to retrieve tweets from Twitter. Fields including username, date, and text were used to retrieve the tweets. Table 2 depicts a sample dataset. We collected our dataset from the period is 1st January 2021 to 20th April 2021. Figure 1 presents the distribution of the number of tweets collected each month. In the data collection process, there were four phases to finalize the studied dataset. The four sequential phases are as follows:

³ <https://github.com/twintproject/twint>.

Table 1 Summary of the systematic analysis studies in related work

Study	Year	Purpose	Sentiment analysis	Approach	Dataset Size	Duration	Availability	Language	Feature extraction	Techniques	Study location
									TF-IDF		
Adamu et al. (2021)	2020	Analysing public sentiment regarding Covid-19 vaccines	Manual annotation	ML	5,200 Tweets	N/A	No	English	TF-IDF	SVM, KNN	N/A
Elhishi et al. (2021)	2021	Understanding the public attitudes towards COVID19	Amazon Com-preend	ML	1M Tweets	Mar20–Jan21	No	English	TF-IDF	K-means clus-tering	N/A
Kumares (yyy)	2021	Analysing public sentiment regarding Covid-19 vaccines	VADER	ML	1200 Tweets	N/A	No	English	TF-IDF	Naïve Bayes, LR	India
Liu et al. (2021)	2021	Understand-ing public perceptions of COVID-19 vaccines	Manual annota-tion	ML/DL	5000 Tweets	Nov20–Jan21	No	English	TF-IDF	BiER, LR, RF, SVM	N/A
Aygiin et al. (2021)	2021	Analyzing public sentiments related to COVID-19 vaccines	Aspect-Based	DL	928,402 Tweets	Nov20–Mar21	No	English, Turk-ish	TF-IDF, Word-2Vec	BERT	8 Countries
Yang and Sorn-teriamvanich (2021)	2021	Investigat-ing Public Perception of COVID-19 Vaccine	TextBlob	ML	190,000 Tweets	Dec20–Jun21	No	English, Japa-nese	BoW	Naïve Bayes	Japan, USA, UK
Jayasurya et al. (2021)	2021	Analyzing of Public Sentiment on COVID-19 Vaccination	VADER	ML	431,986 Tweets	Feb20–Apr21	Yes	English	TF-IDF	14 Models	N/A
Cotfas et al. (2021)	2021	Analyzing the opinions regarding COVID-19 vaccination	Manual annota-tion	ML	752,951 Tweets	Nov20–Dec20	No	English	BoW	MNB, RF, SVM, Bi-LSTM, CNN, BERT	UK

Table 1 (continued)

Study	Year	Purpose	Sentiment analysis	Approach	Dataset		Feature extraction	Techniques	Study location
					Size	Duration			
Rahul et al. (2021)	2021	Analyzing Public Sentiments Regarding COVID-19 Vaccines	TextBlob	ML	13,109 Tweets	Jan20–Jun21	No	English	TF-IDF
Reshi et al. (2022)	2022	Analyzing the global perceptions about COVID-19 vaccination	TextBlob, VADER, AFINN	ML, DL	20,000 Tweets	Mar20–Apr20	No	English	TF-IDF
Mishra et al. (2022)	2022	Compared the sentiments of before and after the second wave	Manual annotation	ML	48,913 tweets	Feb21–Apr21	No	English	TF-IDF, Bow
Nezhad and Dehimi (2022)	2022	Investigating Iranian views toward COVID-19	Persian Analyser	ML	80,3278 Tweets	Apr21–Sep21	No	Persian	Word2vec
Zulfiker et al. (2022)	2022	Analyzing the views and opinions about the vaccines	Manual annotation	ML, DL	1,075 Tweets	June20–July21	No	English	word2vec
Nuser et al. (2022)	2022	Analyzing user sentiment towards the COVID-19 vaccine	Manual annotation	DL	13,190 Tweets	Jan21–Feb21	No	English	N/A
Paliwal et al. (2022)	2022	Analysing the sentiments of the people towards the emergency	VADER, TextBlob	ML	7,640 tweets	Jan21	No	English	TF-IDF
Akpatsa et al. (2022)	2022	Examining Covid-19 vaccine-related Twitter-emerged sentiments	Manual annotation	ML	15,239 Tweets	Jan21	No	English	TF-IDF
									LR, RF, SVM, NB

Table 1 (continued)

Study	Year	Purpose	Sentiment analysis	Approach	Dataset Size	Duration	Availability	Language	Feature extraction	Techniques	Study location
This work	2022	Analyzing the sentiment of Arabic tweets about COVID-19 vaccine	TextBlob, Mazajak, CA-MeL	ML, DL	1,098,376 Tweets	Jan21–Apr21	Yes	Arabic	TF-IDF	LR, RF, ETC, GNB, GBM, LSTM, CNN, MLP	Global

- **Phase (1):** The first phase starts with collecting all the tweets relevant to the purpose of our study. To do so, We manually analyzed Arabic keywords and hashtags associated with COVID-19 vaccination on Twitter. Then, we looked into previous studies on different languages, such as English (the list provided in Table 1). After gathering all the potential keywords related to our study, we snowballed to avoid missing any crucial keywords, which helped us choose the proper keywords and cover all the relevant keywords. As a result, we found that 34 keywords are relevant to our study, and the total tweets yielded 2,046,073.
- **Phase (2):** The second phase performed with removing links: 111 tweets were removed which were not helpful to the study purposes, and 2,045,962 tweets made it to the next stage.
- **Phase (3):** In the third phase, we removed the duplicated tweets for the second stage. For instance, a tweet used two or more keywords that were selected from the 34 keywords. After performed the removal: 895,394 tweets were discarded, and 1,150,568 made it to the next stage.
- **Phase (4):** In this phase, we aim to select the relevant tweets for our study. In fact, we need to filter out unhelpful tweets, such as advertisements or spam. To do so, we manually analyzed the top 2K frequent words and found that only the top 1K frequent words were relevant to our study. After selecting the 1K frequent words, 52,104 tweets were removed, and 1,098,376 made it to the final phase.
- The final dataset that used in our study is yielding a total of 1,098,376 tweets.

3.2 Data preprocessing

The tweets were retrieved using unique hashtags and search keywords as a first step in the study's development (see Table 3). The basic hashtags and keywords used in the tweet extraction have been defined in Sect. 3.2. Following the tweet extraction procedure, the tweet dataset was cleaned by preprocessing. The data elements in the tweets that were not appropriate for the sentiment analysis process were removed during the cleaning process. Usernames, # symbols, hyperlinks, punctuation marks, and stop words are examples of data items. The Natural Language Toolkit Library (NLTK)⁴ was used to do preprocessing on tweets. NLTK (Bird 2005; Loper and Bird 2002) includes over 50 corpora, lexical analysis resources,

⁴ <https://www.nltk.org/>.

Fig. 1 Distribution of the number of tweets collected each month

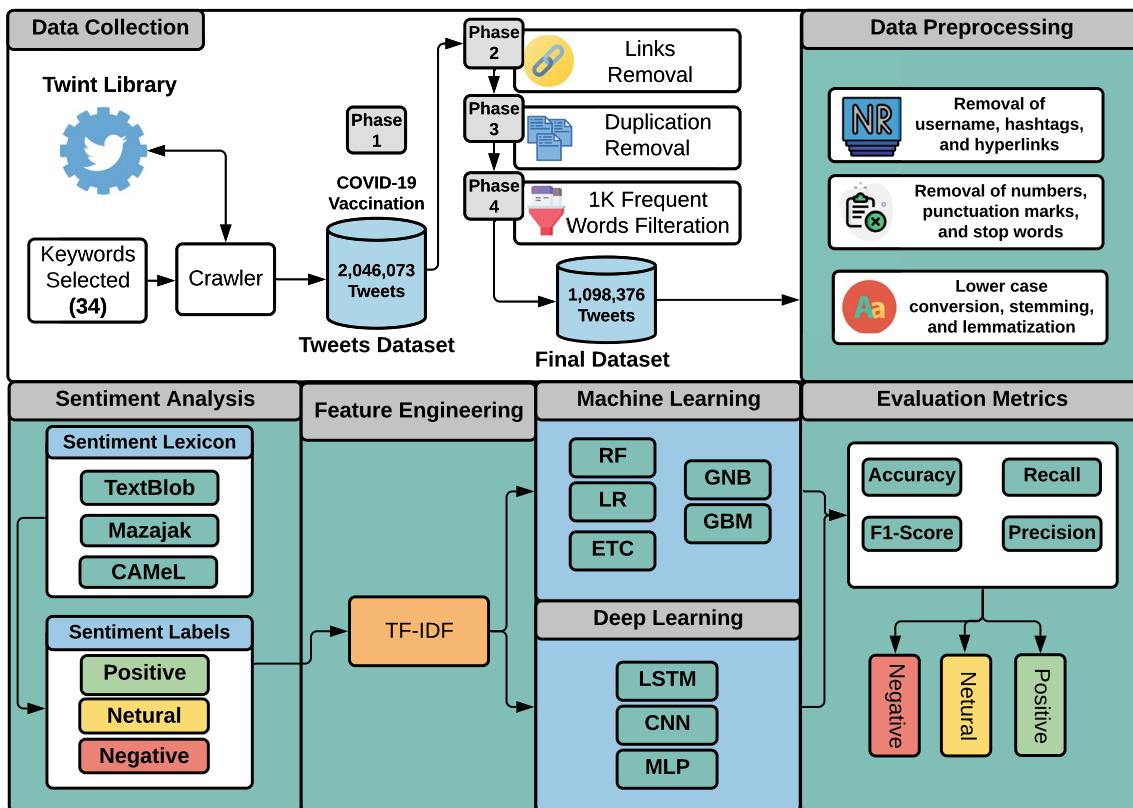
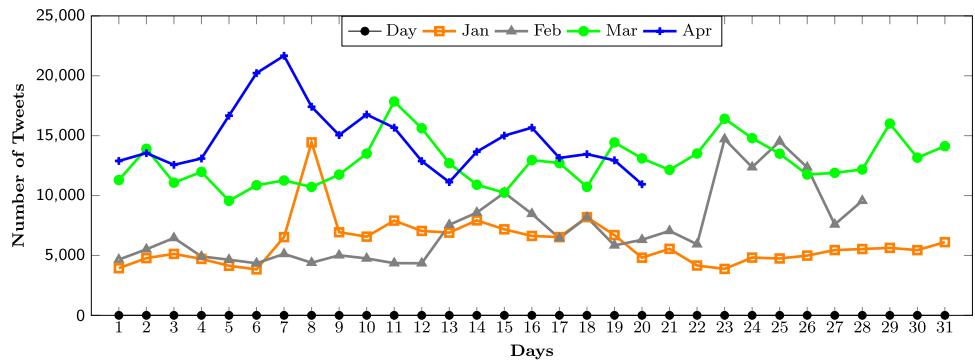


Fig. 2 Overview approach of the study

Table 2 Sample of tweets dataset

Arabic tweet	English translation
اللناح خطوه في طريق العوده للحياة الطبيعية	The vaccine is a step on the way back to normal life
ربما يكون اللناح الصيني هو الامن الى الان	Perhaps the Chinese vaccine is the safest for now
توقع حتى لو اخذت اللناح راح تصاب	I expect that even if you take the vaccine, you will get infected

and a range of text processing libraries. The essential and foundational NLP functions for tagging, parsing, tokenization, and semantic reasoning are all included in these text

processing libraries. These libraries are suitable for use in industrial applications (Kulkarni and Shivananda 2021).

Table 3 Statistics of vaccination types per tweets

No.	Keywords	English Translation	Phase (1)	Phase (2)	Phase (3)	Phase (4)	# of Final dataset
			Tweets	# After removing Links	# After Remove Duplication	# After Filtering based on 1K frequent words	
01	اللّاح	Vaccination	594,739	594,739	506,487	506,327	506,327
02	ارفض _التطعيم	Vaccination_refusal	250	250	161	160	160
03	لّاح	Vaccine	443,870	443,795	115,628	114,529	114,529
04	ارفض _التطعيم	Vaccine_reject	604	604	273	273	273
05	ارفض _اللّاح	Vaccination_conspiracy	2,937	2,937	2,698	2,698	2,698
06	ارفض _التطعيم	Vaccination_fool	833	833	761	760	760
07	اللّاح_ كذبة	Sputnik	45,101	45,101	36,805	36,759	36,759
08	فائزز	Pfizer	193,926	193,926	140,047	90,770	90,770
09	اللّاح_ مأجومي	Vaccination_not_safe	180	180	180	180	180
10	اللّاح_ الروسي	Russian_vaccine	13,027	13,027	242	242	242
11	اللّاح_ الصيني	Chinese_vaccine	8,355	8,355	114	114	114
12	اللّاحات	Vaccinations	85,662	85,662	62,194	62,183	62,183
13	تطعيم _كورونا	Corona_doses	21,493	21,493	16,042	16,037	16,037
14	تطعيم _كوفيد	COVID_Vaccinations	3,833	3,833	1,941	1,941	1,941
15	جرعة _كورونا	Corona_dose	28,292	28,292	15,054	15,054	15,054
16	سينوفارم	Sinopharm	10,297	10,297	6,860	5,341	5,341
17	لا _للتّطعيم	NO_vaccinations	1,776	1,776	1,173	1,172	1,172
18	لا _لّاح	No_vaccines	352	352	188	188	188
19	لّاح_ كرونا	Corona_vaccine	7,233	7,233	4,855	4,839	4,839
20	لّاح_ كورونا	Corona_vaccines	207,313	207,313	98,071	98,032	98,032
21	لّاح_ كوفيد	COVID_vaccines	27,841	27,841	8,933	8,930	8,930
22	لّاح_ أكسفورد	Oxford_vaccine	13,905	13,905	6	6	6
23	لّاح_ جونسون_آند_جونسون	Johnson & Johnson_vaccine	4,057	4,057	9	9	9
24	لّاح_ فائزز	Pfizer_vaccine	37,259	37,259	57	57	57
25	لّاح_ كوفيد_١٩	COVID_19_vaccine	25,608	25,608	0	0	0
26	لّاح_ موديرنا	Moderna_vaccine	2,017	2,017	7	7	7
27	لّاحات_ كورونا	Corona_vaccinations	34,196	34,196	20,070	20,069	20,069
28	ماراح_ اخذ_ اللّاح	Not_taking_vaccines	683	683	27	27	27
29	موديرنا	Moderna_vaccine	7,835	7,835	1,945	1,943	1,943
30	نعم _لّاح	Vaccine_yes	141	142	55	55	55
31	خذ_ الخطوة_ خذ_ اللّاح	Take_step_take_vaccine	15,262	15,262	190	190	190
32	خذ_ اللّاح	Take_vaccine	22,467	22,456	156	156	156
33	الجرعة_ الأولى	First_dose	118,844	118,828	64,712	64,705	64,705
34	الجرعة_ الثانية	Second_dose	65,885	65,875	44,627	44,623	44,623
Total			2,046,073	2,045,962	1,150,568	1,098,376	1,098,376
Removed tweets			N/A	111	895,394	52,104	N/A

Table 4 Sample of tweets dataset includes the corresponding tags

Arabic tweet	English translation	Tag
اللقاء خطوه في طريق العوده للحياة الطبيعية	The vaccine is a step on the way back to normal life	Positive
ربما يكون اللقاح الصيني هو الامن الي الان	Perhaps the Chinese vaccine is the safest for now	Neutral
اتوقع حتى لو اخذت اللقاح راح تصاب	I expect that even if you take the vaccine, you will get infected	Negative

In the following subsections, the pre-processing steps have been described.

3.2.1 Removal of username, hashtags, and hyperlinks

People commonly use @username on Twitter to refer to or tag their friends and similar people, as well as hashtags and hyperlinks in their messages. Since usernames, hashtags, and hyperlinks in tweets are not useful for sentiment analysis, they have been omitted from the tweets. We also removed any words that non-Arabic language.

3.2.2 Removal of numbers, punctuation marks, and stop words

It is also irrelevant to use numbers, punctuation marks, or stop words to locate the sentiment in the letter. As a result, all non-alphabetic characters have been deleted, including numbers and punctuation marks. Additionally, using the NLTK library features, all stop terms have been deleted. Using a sampling of tweets, An example can be stop words such as (في، إلى، على) (from, to, in). We removed Arabic diacritics and punctuation marks (؟, !, ., ؟) as well as special characters such as (#, %, &).

3.2.3 Lower case conversion, stemming, and lemmatization

All the resulting tweet text has been translated to lowercase letters to minimize redundancy and simplify the results. The conversion is necessary because computers handle lowercase letters and words differently than uppercase letters and words. For example, 'a' is not the same as 'A,' and "GO," "Go," and "go" are viewed as three distinct words. Humans, on the other hand, use these words in various ways in different contexts in their written natural language to convey the same meaning. This conversion is performed to minimise the difficulty in NLP processes by bridging the distance between the accessibility of letter case use among humans and computers. Additionally, two text normalising NLP methods, such as stemming and lemmatization, were used to optimise the text data for sentiment analysis. This text normalisation methods were used to change the text in order

to make labelling easier. Since a word in a written human language text may have several meanings depending on the context in which it is used. For example, given their root word “go,” words like “goes,” “gone,” and “going” all have the same meaning. So, when looking for the root word “go” and “going” or “gone,” the words “goes,” “going,” or “gone” serve no different purpose than when searching for the root word “go.” Inflection refers to the distinction made between the different types of a single root word. The inflection of these terms has been omitted from the tweet texts, allowing Stemming and Lemmatization to produce root words from the various inflected words. To obtain the desired results, these two text processing analysis techniques function in various ways. Inflections have been omitted by using these two protocols in the current data pre-processing task to minimize data complexity. The consequence of using these protocols was that various inflected words were reverted to their root words. For example, all instances of the words “goes,” “gone,” and “going” were replaced with their root word “go”. Below, we provide an example for the Arabic language:

3.3 Sentiment analysis

This subsection presents sentiment, which involves labeling data into either positive, negative, or neutral. We are using three sentiment analysis techniques: TextBlob, Mazajak, and CAMeL. This subsection discusses the following topics: data labeling, TextBlob, Mazajak, and CAMeL.

Table 5 TextBlob sentiment score range

Negative	Polarity score < 0
Neutral	Polarity score = 0
Positive	Polarity score > 0

3.3.1 Data labeling

Tweets were classified into 3 categories: positive, negative, and neutral as follows:

- **Positive:** A positive sentiment is associated with positive words or phrases. For example, if a text expresses excitement, kindness, protection, joy, etc., it is usually classified as positive.
- **Negative:** A negative sentiment is expressed by a negative sentence associated with it. For example, any text that expresses sadness, anger, intimidation, rage, or oppression is classified as negative (Hu et al. 2012).
- **Neutral:** When the creator of the tweet does not express a personal opinion and is simply sharing information.

Table 4 presents an example of the tweets with tags that assigned to them.

TextBlob. We use TextBlob, a lexicon-based tool for natural language processing (NLP) tasks on raw text (Loria 2018). TextBlob is a Python library that provides a programming interface for processing text data using textblob algorithm implementations. A textblob, for example, can be used to analyse text sentiments, extract noun phrases, construct POS tags, translate, classify, and more (Vijayarani and Janani 2016). In a nutshell, the TextBlob library has a number of built-in features that help with language processing. It can function in a variety of languages, including Spanish, English, and Arabic. TextBlob is used to analyse emotions in tweet data of negative, positive, or neutral polarity, according to study (Munjal et al. 2018). This library is built on top of the NLTK (Natural Language Toolkit), and its sentiment analysis algorithm is compatible with both NLTK and pattern processing (Laksono et al. 2019). Its dictionary contains about 2918 lexicons. In TextBlob, the polarity measure is dependent on either subjectivity (i.e., personal opinions) or objectivity (i.e., facts). (polarity ranking, subjectivity score) (Sohangir et al. 2018) is how the sentiment analyser returns sentiment ratings.

The sentiment score range for TextBlob is seen in the Table 5. The scores below 0 mean that the sentiments are negative polarised, while the scores above +1.0 indicate that the sentiments are positive polarised. In terms of subjectivity, 230 ratings below 0.0 mean that the emotions are founded on reality, while scores above 1.0 suggest that they are based on personal opinions. There is one more setting in

TextBlob, namely the ‘intensity’ which is used to measure subjectivity. The intensity of a phrase decides whether it modifies the next word. Adverbs are used as modifiers in English (‘very good’). However, for the phrase “This was a useful case, but I’d like another,” It gives a 0.0 for both subjectivity and polarity, which is not the best result we should hope for. If the library returns 0.0, it’s because the sentence did not have any terms in the NLTK training collection that had polarity, or because TextBlob uses a weighted average sentiment value over all the words in each study.

Algorithm 1: TextBlob algorithm for sentiment analysis

Input: Worldwide Covid19 Vaccination Tweets
Result: Polarity Score $> 0 \rightarrow$ (Positive)
 Polarity Score $= 0 \rightarrow$ (Neutral)
 Polarity Score $< 0 \rightarrow$ (Negative)
 initialization;
loop (tweet in tweets)
 Compute Polarity Score TextBlob
 (tweet)
condition:
if (Polarity Score > 0) **then**
 Tweet Sentiment = Positive;
elseif (Polarity Score $= 0$) **then**
 Tweet Sentiment = Neutral;
else
 Tweet Sentiment = Negative;
condition end
loop end

Mazajak. It⁵ is a deep learning and massive Arabic word embedding-based online Arabic sentiment analysis framework. This is the first move, which helps to eliminate inconsistencies and normalize the data into a logical format that can be conveniently managed. The measures are primarily dependent on the work done in the previous phase (El-Beltagy et al. 2017). The measures below were included in our implementation:

- **Letter normalisation:** The process of bringing together letters that occur in various ways (Darwish and Magdy 2014).
- **Elongation removal:** deleting letters that occur repeatedly, particularly in social media data (Darwish et al. 2012).
- **Cleaning:** Unknown characters, diacritics, punctuation, URLs, and other items should be removed.

⁵ <http://mazajak.inf.ed.ac.uk:8000/>.

Each row represents a letter, and its corresponding embedding represents each word, in a two-dimensional matrix. 300 was chosen as the embedding D scale. In sentence construction, the short sentences are arranged to form the next to the largest in the set. The skip-gram architecture was used to build word embedding using the word2vec (Mikolov et al. 2013) program encoding. Unlike the currently available AraVec (Soliman et al. 2017), which was developed with a corpus of 67 million tweets, the embedding were created with a corpus of 250 million unique Arabic tweets, rendering it the largest Arabic word embedding set. The tweets were collected over many years, from 2013 to 2016, to ensure that a wide range of topics were addressed. Due to the vast and varied corpus, multiple dialects are covered, which helps to reduce the impact of dialectal heterogeneity. The embedding were created using the same pre-processing steps as were used in the sentiment analysis scheme.

Mazajak uses CNN to train the algorithm, and the LSTM that is made by this model is followed. The CNN performs the role of a feature extractor, learning local patterns within the sentence and providing representative features. The LSTM uses the derived features while still considering the context and word order. A max pooling layer receives the embedding, applies CNN pre-processing to them, and then moves them on to the following layers. The use of full pooling (weighing the importance of data) is employed to do this, which implies that sentiment would typically be conveyed in plain language. The characteristics were applied to an LSTM, which expanded, and then a softmax sheet was created and used to generate a distribution of probabilities over the output groups.

CAMEL. CAMEL⁶ contains Python APIs as well as a series of command-line utilities that expose functionality across the thin layers of an open-source package. CAMEL Tools are all used for pre-processing, morphological analysis, disambiguation, Dialect Identification (DID) (Salameh et al. 2018), Named Entity Recognition (NERs) (Shaaalan and Raza 2009) (NER), and Sentiment Analysis (SA) (Abdul-Mageed and Diab 2014). The majority of the pre-processing utilities that are commonly used in Arabic NLP have been reimplemented in CAMEL Tools. Different software and programmes perform subtly different pre-processing procedures that aren't really well documented or accessible for standalone use. We hope to save the time it takes to write Arabian NLP applications by including these utilities in the package and to ensure pre-processing is consistent from project to project. An Arabic sentence is classified as positive, negative, or neutral by our analyser. The tool fine-tuned multilingual BERT (mBERT) (Devlin et al. 2018) and AraBERT (Baly et al. 2020) on the role of Arabic SA using

HuggingFace's Transformers (Wolf et al. 2019). It was fine-tuned by introducing a completely connected linear layer with a softmax activation feature.

Since Python has become both an open-source and designed with Morphological Analysis in mind, there is a general Arabic toolkit (CAMEL Tools) for building pre-analysis for open-source projects as well as pre-identified language and called entity recognition for morphological and sentiment analysis. CAMEL tool provides command-line interfaces(CLIs) and application programming interfaces are used with the tools (APIs).

Morphological modelling, parsing, divisions, segmentation, stemming, and object identification Named Entity Recognition (NERs) are functions that can be applied to construct higher-level applications, such as localization, correction of spelling, and questions. Other activities including Dialect Identification (DID) and Sentiment Analysis (SA) are designed for the analysis of extra data to calculate different input text features. We take a quick look at the various attempts for a number of Arabic NLP activities:

Morphological Modelling The depth of the information analyzer morphology provides very superficial analysers, on the one hand (Vanni and Zajac 1996), while on the other, analyzers have form-based, functional, lexical and morphemic features (Buckwalter 2002; Smrz 2007; Boudlal et al. 2010).

Named Entity Recognition ANERsys, one of the first Arabic NER systems, was developed by Benajiba and Rosso (2007). They also developed their own language instruments - ANERcorp (an annotated corpus of personal, place and organisational names) and ANERgazet - popular in Arabic NER literature (i.e., Person, Location and Organization gazetteers). Our scheme is equalised with the functions of Benajiba and Rosso, and we use all publicly available training data in the Tools (Benajiba and Rosso 2008).

Dialect Identification A fine grain DID system was developed by Al-Badrashiny et al. (2014), which covers dialects from 25 cities in different Arab countries. Elfardy and Diab (2012) also suggested token dialect detection instructions. In order to train a Naive Bayes classifier, a token-level DID process is combined with other functionalities (Elfardy and Diab 2012). In the sense of social media, Sadat et al. (2014) used a bigram-level model to characterise the dialect of phrases across 18 Arab dialects.

3.3.2 Final voting

This subsection represents the final voting process for our dataset. As mentioned above, we used three sentiment analyzers: CAMEL, TextBlob, and Mazagk. Each analyzer can label the tweets based on their pre-trained data, which means it can produce different sentiments for the same tweet. However, we want to ensure that each tweet sentiment in our

⁶ https://github.com/CAMEL-Lab/camel_tools.

dataset is reliable and accurate. Therefore, we used three sentiment analyzers, and the process we pursued was as follows:

- If the tweet is labeled the same by **all** the three sentiment analyzers, for example, all analyzers labeled the sentiment positive. Then, we confirmed that the tweet was positive.
- If the tweet is labeled the same by two analyzers, for example, **two** analyzers labeled the sentiment positive. Then, we verified that the tweet was positive.
- If the tweet is labeled differently between all the three analyzers, for example, negative, positive, and neutral. Then, we have to introduce a fourth sentiment analyzer or look at the tweet manually to verify which sentiment label between the three analyzers is more accurate. In our final voting, we did not encounter a case with three different labels. Therefore, all three sentiment analyses were sufficient for us to ensure all tweets in the final dataset were labeled correctly.

3.4 Feature engineering

Feature engineering involves identifying important features of data in order to learn more about the main features or train machine learning algorithms (Bocca and Rodrigues 2016). Hence, feature engineering improves the benefit of machine learning algorithms. Furthermore, it helps in locating important aspects in the data to improve the work of learning algorithms (Heaton 2016). We divided our corpus into ‘testing subset’ and ‘training subset’, allocating a ratio of 1:3, respectively. We utilized feature extraction technique to train the subset. Testing and training data were done an attribute extraction methods. The feature engineering approach we used in this paper was Term Frequency-Inverse Document Frequency (TF-IDF).

3.4.1 TF-IDF features

Another method for extracting features is to use term frequency-inverted document frequency (TF-IDF). This method is widely used in the correction of musical data and text interpretation (Bei 2008). It gives terms in a language weight based on the reciprocal frequency of the document and the frequency of the terms (Rustam et al. 2019; Robertson 2004). The importance of words with a higher weighted score is considered higher (Zhang et al. 2011). The TF-IDF can be expressed mathematically as follows:

$$tfidf = tf_{t,d} * \log \frac{N}{D_{i,t}}, \quad (1)$$

where $tf_{t,d}$ represents the frequency of term t in document d , N represents the number of documents, and $D_{i,t}$, t represents the number of documents that comprise the term t .

We chose TF-IDF because it is the most common and heavily used technique for extracting features in similar studies, as shown in Table 1. A recent study on the English COVID-19 vaccination dataset compared TF-IDF and BoW and found that TF-IDF performed better than BoW (Mishra et al. 2022). Therefore, we chose TF-IDF for extracting the feature in our study.

3.5 Machine learning techniques

Various machine learning algorithms use feature vectors for training based on datasets, then the polarity of the content is determined by semantic analysis, which provides a variety of similarities and alternative words. The algorithms of Machine Learning are discussed in detail in this section, as well as the details of how to set their hyperparameters and how we can implement Machine Learning models. The machine learning algorithms were implemented using the Sci-kit Learn library (Loper and Bird 2002). The deployment of each of the five supervised machine learning models was done using the sci-kit module. Supervised machine learning models, like the LR, RF, SVC, GNB, and ETC, are widely used for finding solution of problems related to regression and classification (Kotsiantis et al. 2007). The specifications of the implementations of the algorithms are listed in Table 6, together with their hyperparameters.

3.5.1 Random forest (RF)

By composing the outcomes of sub-trees, a random forest (or ensemble model) is used to construct high-precision predictions. Random forest uses the technique of bagging to train multiple decision trees using bootstrap samples (Biau and Scornet 2016). After training, the Bootstrap samples conduct subsampling and overwrite the dataset (Breiman 1996). To help the prediction process, the RF approach employs decision trees. Attribute collection (Liaw and Wiener 2002) is the name for this mechanism. After the models have been trained, the results are combined through voting in this form of ensemble classification. Several previous studies (Rupapara et al. 2022; AIOMar et al. 2021) used ensemble approaches and procedures. Boosting (Schapire Robert 1999) and bagging (Breiman 1996; Freund and Schapire 1997) are two of the most well-known ensemble approaches. Bootstrap aggregation, also known as bagging, is a method of training several models on bootstrapped samples. The following Eqs. 2 and 3 can be used to classify a random forest:

Table 6 Summary of the hyperparameter in machine learning algorithm

Algorithm	Hyperparameter	Value	Description
LR	multi_class	Multinomial	Best to solve the multi-class classification problem
	random_state	1000	mm
	C	3.0	Inverse of regularization strength
	solver	newton-cg	
RF	n_estimators	500	The number of decision trees
	random_state	5	The random state sample was taken these random decisions to be managed
	max_depth	200	The maximum depth between each tree
ETC	n_estimators	500	The number of decision trees
	random_state	5	The bootstrapping of the samples used when building trees
	max_depth	200	Maximum depth of both the estimated regression estimation techniques
SVC	kernel	linear	It maps the observations into some feature space
	C	1.0	The penalty parameter of the error term
	random_state	500	The opposite of the power of regularization; it must have been a positive float
GNB	N/A	N/A	Default setting
LSTM	LSTM Units	1000	LSTM layers units
	Activation Function	ReLU	Activation function used in model
	Epochs	100	number of epoches
	Loss Function	Adam	Loss function used in model
	Optimizer	categorical_crossentropy	Optimizer used in model
	1DConv	64 filters	1 dimensional convolutional neural network with 64 filter
CNN	Kernel	3×3	kernel size
	Activation Function	ReLU	Activation function used in model
	Epochs	100	number of epoches
	Loss Function	Adam	Loss function used in model
	Optimizer	categorical_crossentropy	Optimizer used in model
	Neurons	64	Neurons in input layer and hidden layer
MLP	Hidden layer	2	number of hidden layer and hidden layer
	Activation Function	ReLU	Activation function used in model
	Epochs	100	number of epoches
	Loss Function	Adam	Loss function used in model
	Optimizer	categorical_crossentropy	Optimizer used in model

$$rf = mode\{tr_1, tr_2, tr_3, \dots, tr_n\} \quad (2)$$

$$rf = mode\left\{ \sum_{i=1}^n tr_i \right\} \quad (3)$$

where, $tr_1, tr_2, tr_3, \dots, tr_n$ are decision trees in random forest and n is the number of trees.

Random forest was used with poor learners up to 300 to achieve high-level precision, and the n estimator value was set to 300. The n estimator parameter specifies the number of trees that will be added to the prediction process. Random forest used bootstrap samples to train three hundred decision trees in the experiment segment. The final predictions of decision trees were achieved through a voting procedure (Freund and Schapire 1997). The random forest

parameter max depth has 60 possible values. The depth level or maximum depth of decision trees has been calculated using the parameter “max depth.” By consolidating the degree of depth, it reduces the difficulty of decision trees. It also decreases the likelihood of over-fitting in decision trees (Breiman 1996). Another parameter in the RF model was “random state.” It is used to ensure that samples are random while the classifier is being trained. The best results were achieved in this analysis using these hyper-parameters and a random forest model.

3.5.2 Logistic regression (LR)

Another supervised ML model is LR serves, which is used for classification based on the principle of likelihood (Safdari et al. 2019). The logistic regression method (LR)

Table 7 Results of the experiment of all models with TF-IDF features

Models	Accuracy	Precision	Recall	F1-score
LR	0.82	0.82	0.79	0.80
RF	0.76	0.77	0.70	0.72
SVC	0.82	0.82	0.78	0.80
GNB	0.57	0.61	0.64	0.56
ETC	0.76	0.80	0.68	0.72
LSTM	0.82	0.82	0.82	0.82
CNN	0.81	0.81	0.80	0.81
MLP	0.82	0.82	0.81	0.81

is a mathematical method based on a logistic equation. Linear regression is like LR. LR attempts to determine whether something is true or false. It can handle both discrete and continuous data, such as weight and age. The LR relationship is established between an absolute dependent variable and one or more independent variables (where the dependent variable is commonly referred to as the target class) by assuming probabilities and employing a “S”-shaped logistic function. We will demonstrate it as follows:

$$g(x) = \frac{L}{1 + e^{-k(v - v_0)}} \quad (4)$$

The logistic function’s vector v and S-shaped curve have values ranging from $-\infty$ to $+\infty$ of real numbers. Since certain target classes, such as Negative, have a limited corpus, this analysis uses the “liblinear” hyperparameter to improve. The next parameter, multi-class, is used in combination with the value of “multinomial” in binary-classifiers since it is more appropriate. Here, “C” denotes the third statement, which regularises inversely, and it retains the regularising stability change by specifying the inversely organised to Lambda and minimises the probability of the model over-fitting problem (Lee et al. 2022). The LR classifier was chosen because it is better suited for binary classification (Sebastiani 2002).

3.5.3 Gaussian naive bayes (GNB)

The Naive Bayes algorithm, a probabilistic algorithm used for a variety of classification functions, is based on Bayes theorem. The extension of Naive Bayes is Gaussian Naive Bayes. While other functions can be used to estimate data distributions, the Gaussian or normal distribution is the easiest to implement because the mean and standard deviation must be calculated for the training data. The Naive Bayes algorithm is a probabilistic classification algorithm that can be used to solve a variety of problems. It is based on a probabilistic model that is easy to code and makes predictions in real time. This algorithm is commonly used to

solve real-world problems because it can be tuned to respond quickly to user queries.

3.5.4 Gradient boosting machine (GBC)

For classification tasks, GBC employs a boosting process. GBC matches a set of slow learners (decision trees), which are sequentially educated using the predecessors’ classifier errors. The first decision tree classification is supported by data collection and then the second decision tree is educated using mistakes from the first classification and added to the first classification and so forth. This sequential combination of classifiers can minimise errors and enhance the precision of classification. GBCs have been used in astronomical research in many previous studies including the distinction of pulsar noise signals (Bethapudi and Desai 2018), photometric astrophysics over supernovae detection and automatic galaxy discovery and classification using Galaxy Zoo (Tramacere et al. 2016) (for more information on GBC, see Friedman (2001)). GBC defines the mean quadrather error (MSE) which is reduced to a minimum:

The learning rate r (Bethapudi and Desai 2018) determines how strongly a classifier should be fused with its predecessor, and the number of residual values equals n . Consequently, GBC basically changes estimates where the number of residuals is equal to 0 or the minimum and the predicted values are similar enough to the actual values. The GBC algorithm, which we implemented using the “scikit-learn” program, has four parameters, as seen in the Table 7.

3.5.5 Extra tree classifier (ETC)

The Extra-Tree Classifier (ETC) is a highly randomized tree classifier. ETC is a decision tree combination classifier that is constructed differently from traditional decision trees (Amaar et al. 2022). That is, to determine the optimal split for dividing the samples of a node into two classes, random splits are selected from the individual features chosen randomly, including the maximum features from which the optimal split is selected. Therefore, ETC seems to fit different dataset sub-samples through the process of averaging the differential problems inherent in the singular decision tree method, which helps to prevent overfitting and increase the predictive accuracy.

3.6 Deep learning

This subsection addresses the deep learning algorithms used in this research. We discuss the three deep learning algorithms selected in the study to solve classification problem. The algorithms used are Long Short-Term Memory (LSTM),

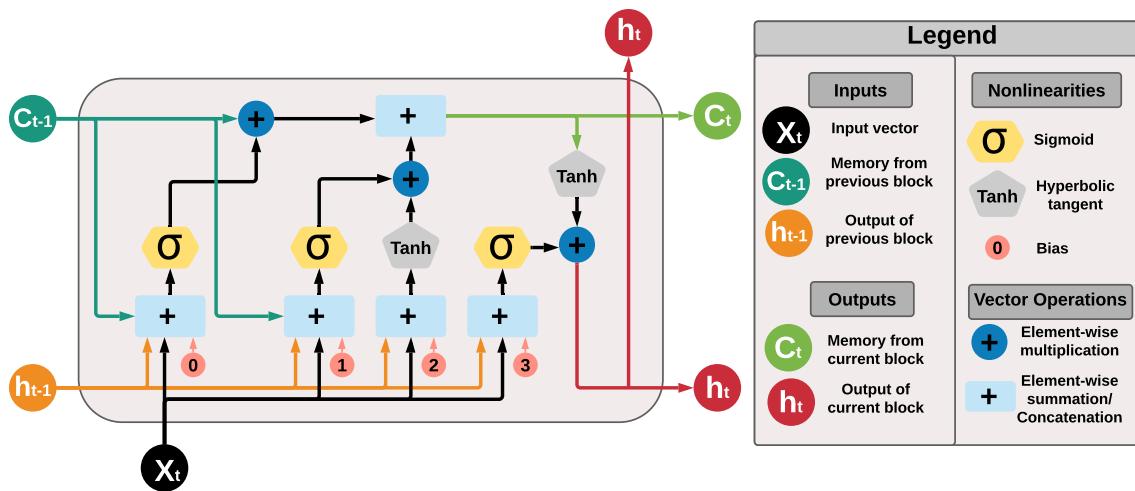


Fig. 3 The structure of the Long Short-Term Memory (LSTM) neural network. Reproduced from Yan Yan (2016)

Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP).

3.6.1 Long short-term memory (LSTM)

LSTM was first introduced in 1997 by Schmidhuber and Hochreiter (1997) and reemerged as a successful architecture in 2014 after gaining remarkable performance in statistical machine translation (Zhou et al. 2015). In deep learning, LSTM is mainly used to process the sequence data [?]. Before the introduction of LSTM, Recurrent Neural Network (RNN) was widely used in the sequence analysis of various lengths. However, as the standard RNN is equivalent to multi-layer feed-forward neural networks, the large number of historical data brought in by long sequences is expected to lead to gradient loss and data loss (Zhang et al. 2018). Therefore, LSTM has been proposed to address the problem of vanishing gradient by replacing the self-connected hidden units with memory blocks. The LSTM block uses a purpose-built memory cell to store data and is best for finding and exploiting long-distance content. Memory units enable the network to know when to read new information and when to forget old information. The memory block in LSTM uses purpose-built memory cells to store information, and it is better at finding and exploiting long-range context. The memory units enable the network to be aware of when to learn new information and when to forget old information (Ye et al. 2021). The basic LSTM structure is shown in Fig. 3.

The structure adds a memory cell to store historical information, and the renewal, deletion, and extraction of historical information are controlled by three consecutive gates, the input gate, the forget gate, and the output gate. The input gate is used to detect how incoming vectors change the state

of a memory cell. Exit gate allows the memory cell to influence the output. Finally, the output gate allows the memory cell to remember or forget the data it has previously stored.

3.6.2 Convolutional neural network (CNN)

Convolutional Neural Network (CNN), one of the architectures of the deep neural network, has been widely used for solving various artificial intelligence tasks in recent years (Hoang et al. 2019). CNN is a type of in-depth learning model that processes data with a grid pattern, inspired by the animal cortex of vision's organization (Hubel and Wiesel 1968; Fukushima and Miyake 1982). It is designed to learn automatically and adapt to the characteristics of features, from bottom to top-level patterns. In CNN, text-classification criteria are similar to image classification, the only difference being that instead of pixel values the word vectors are used (Amin and Nadeem 2018). CNN is a mathematical formulation that is often composed of three types of layers (or building blocks): convolution, integration, and fully integrated layers. The first two layers, the layers of convolution and the pooling, make the element extraction, and the third, the fully connected layer, sets the maps of the extracted elements for the final output, such as isolation. The convolution layer plays a major role in CNN, which is made up of several mathematical operations, such as convolution, a special type of direct operation. Since one layer feeds the output to the next layer, the extracted features can become extremely complex. The performance of the model under certain kernels and weights is calculated by the function of the loss through forward propagation on the training data set. The learnable parameters, i.e., weights and kernels, are revised according to the loss rate by backpropagation via a gradient descent optimization algorithm (Yamashita et al.

2018). The key reasons behind the consideration of CNN as a classical model are as follows. First, the main interest in using CNN lies in the idea of using a weight-sharing concept, because the number of parameters that need training is significantly reduced, which has led to general improvement (Arel et al. 2010). Subsequently, because of the small number of parameters, CNN can be trained smoothly and does not suffer over-fitting (Smirnov et al. 2014). Second, the classification phase is integrated with the feature extraction stage (Lawrence et al. 1997), both of which use the learning process. Third, it is much easier to use large networks using CNN as compared to general models of Artificial Neural Networks (ANN) (Rustam et al. 2022).

3.6.3 Multilayer perceptron (MLP)

Modern times have seen methods of deep learning being used in various fields for example medical image processing, computer vision, natural language processing commonly referred to as NLP, automatic speech recognition and several others (Wang et al. 2019). This research uses MLP, which is a deep learning model and from the class of feed-forward ANNs (Driss et al. 2017). There are three layers of nodes in the MLP, the first is the input layer, next is a hidden layer and the third is the output layer. The flow of information or data in the MLP is from the input layer to the output one, that is, in one direction only. Apart from the nodes of the input layer, the rest of the nodes are neurons that learn from the data through a nonlinear activation function. What makes MLP different from a linear perceptron is that it has a nonlinear activation function and several layers. Further, it has the ability to discriminate between data that cannot be separated linearly. As previously stated, the MLP is a feed-forward multilayer network, which is a form of ANN. ANN of this type works in the way that input layer nodes store the network's input features whereas the layers that are hidden collect the input layer's weighted inputs and transmit this output to the next layer. As a result, the classification results of the input data go to the output layer (Tarkhaneh and Shen 2019). The steps to follow in an MLP can be learned through various algorithms, of which a common method of supervised learning is back-propagation (Manik et al. 2019). There are four phases in the back-propagation method, that are: initialization of weights, feed-forward, error back-propagation and updating weights.

3.7 Evaluation metrics

The metrics used to assess how machine learning models perform are precision, accuracy, their F1-score, and recall.

3.7.1 Accuracy

The accuracy of the model is measured in terms of the ratio of correct predictions to the total number of predictions from the classifiers on test data (Aljedaani et al. 2022). The score for accuracy can be a maximum of 1, depicting that each of the classifier's predictions is correct. Whereas, the least score for accuracy is 0. To calculate accuracy, the following formula can be used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP and TN are true positive and true negative respectively and FP and FN are false positive and false negative respectively.

3.7.2 Precision

Precision (Aljedaani et al. 2022), defined as a positive predictive value, is the proportion of correctly classified instances among all true classified instances. A score of 1 in precision indicates that every instance of positive data is positive. However, what needs to be noted is that it has no impact on the number of positive instances which were predicted to be positive but are labeled negative. Precision can be calculated as

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3.7.3 Recall

Recall (Aljedaani et al. 2021) also known as sensitivity, depicts the relative number of instances classified as positive from all of the positive instances. We can measure recall as following

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

3.7.4 F1-Score

Recall and precision cannot be considered to be accurate indicators of a classifier's performance. For this, F1-score (Fang et al. 2021) is regarded as more essential since it includes both recall and precision and provides a value ranging between 0 and 1. It is measured using the harmonic mean of recall and precision as

$$F_1 Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$



Fig. 4 WordCloud for the tweets dataset showing the most used words. The shape of the wordcloud matches the Twitter's logo as a reference to the source of the data (Wordclouds 2022)

3.8 Thematic analysis

While **RQ1** is being addressed quantitatively, **RQ2** is intended to deliver qualitative insights into people's perception of the vaccines as they engage in discussing it on Twitter. Therefore **RQ2** dives deeper into the content of the tweets to extract the reasons behind the reluctance to adopt the vaccine. This represents valuable information, as it helps detect any false information circulating on social media and provides concrete steps towards responding to them through

media campaigns. In this subsection, detail the sampling of tweets, along with their corresponding thematic analysis.

Tweets Selection. To sample the tweets, we adopted the guidelines outlined by Ruggeri and Samoggia (2018):

- **Accounts.** Our tweets should be selected from *credible* Twitter accounts, having a minimum of 100 tweets, retweets, or replies. This allows the selection of active accounts.
 - **Tweet length.** Extracting information from a tweet goes through reading and comprehending its text. Therefore, short tweets may not provide sufficient information, and therefore avoid the selection of tweets with less than 100 characters.

This process resulted in sampling 2500 tweets for the next phase.

Thematic Analysis. We follow the guidelines of Braun and Clarke (2006) in order to identify *themes* in each tweet, along with its *confidence level*. We adopted an inductive bottom-up approach: the identified themes (i.e., reasons to avoid taking the vaccines) are explosively extracted from various tweets, belonging to multiple individuals (Patton 1990). We also try to create as many themes as possible, without forcing them into pre-existing coding frames (Braun and Clarke 2006; Aljedaani et al. 2021). Tweets were primarily analyzed by the first author, because of his familiarity with the dialect of that region. Tweets were analyzed in batches of 500. After each batch, a meeting between authors is scheduled to review the manual analysis. These review meetings result in gathering preliminary themes into more comprehensive categories.

The review initially started with up to 9 themes, and by the end of the 10th review meeting, there were 6 main categories of themes that were finalized. These categories represent the main motivating factors advocating against

Fig. 5 Percentage of the tweets about all the six vaccinations types in the collected dataset

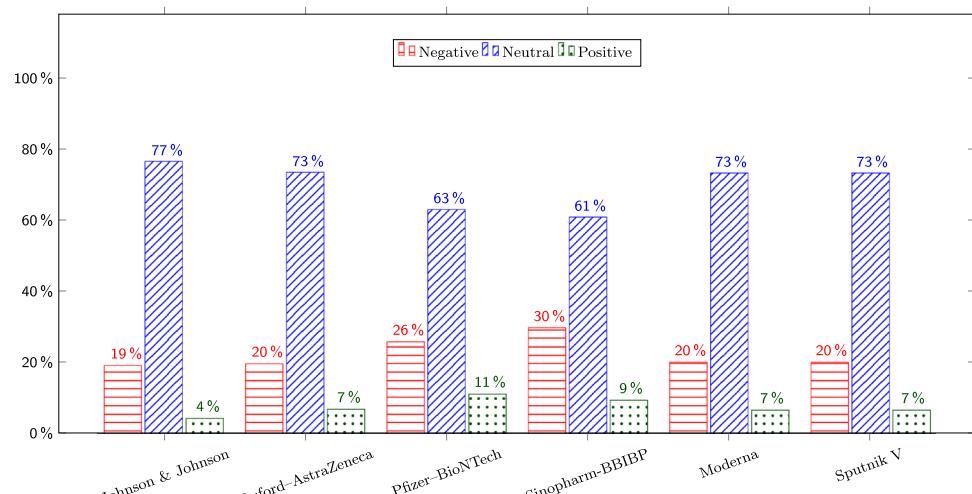


Fig. 6 Percentage of the tweets in each month along with their sentiment and total number of tweets in the collected dataset

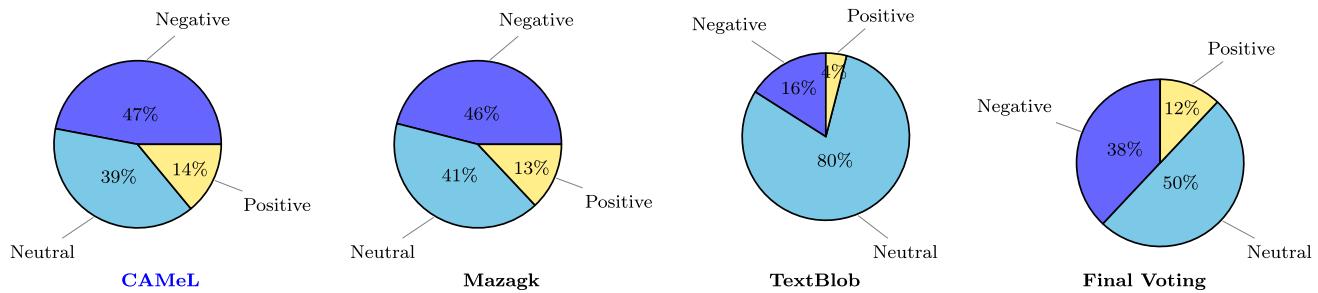
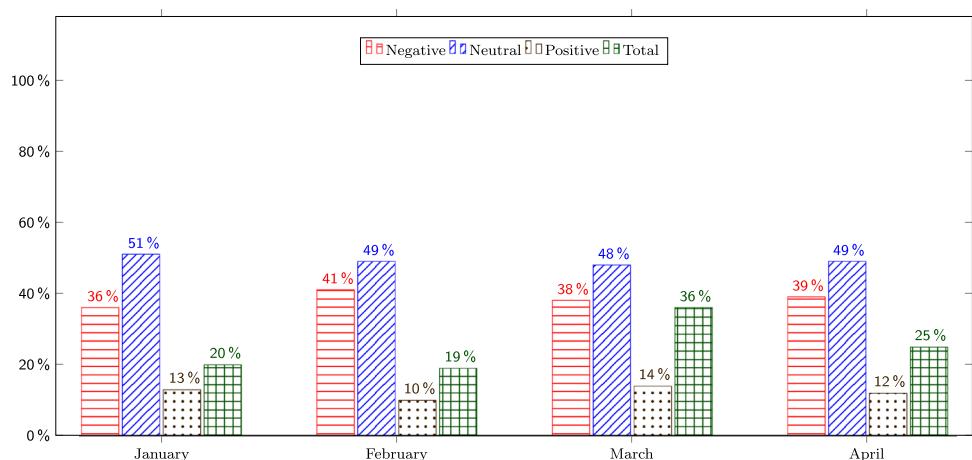


Fig. 7 Percentage of the voting in the sentiment analyzers

taking vaccines. The result of this manual analysis answers our second investigation.

4 Study results

In this section, we present the results of our study.

RQ₁: What is the highest accepted vaccine in the Arab world?

Motivation. In this research question, we want to identify the vaccine that has the highest popularity in the Arab world. Currently, there are various vaccines with brand names such as Moderna, Johnson & Johnson's Janssen, PfizerBioNTech, Novavax, AstraZeneca, Sputnik, and others. Although the public sentiments towards such vaccines had been identified in previous studies using English tweets, we wanted to determine their popularity using Arabic tweets and a different dataset.

Approach. In this study, we use sentimental analysis to evaluate whether tweets pertaining to a given vaccine were positive, neutral, or negative. To perform the sentimental analysis, we utilized TextBlob, Mazajak, and CAMEL. We applied each of the methods to our dataset and then interpreted the results.

Vaccination Types. Our study focused on public sentiments towards a wide range of available vaccines, as follows: Johnson & Johnson, Oxford-AstraZeneca, Pfizer-BioNTech, Sinopharm-BBIP, Moderna, and Sputnik V. In Fig. 5, we provide vaccination types per tweets, which are derived from Arabic keywords. A word cloud is provided in Fig. 4 showing the most used words in the tweets in the Arabic language. From our results in Fig. 5, Pfizer-BioNTech attracted the highest percentage of positive sentiments (11%), followed by Sinopharm (9%), Oxford-AstraZeneca (7%), and Moderna (7%). The largest percentage of neutral tweets were from Johnson & Johnson (77%), followed by Oxford-AstraZeneca (73%) and Moderna (73%). The highest percentage of negative tweets were for Sinopharm-BBIBP (30%), followed by Sputnik V (29%) and Pfizer-BioNTech (26%). From the results presented in Fig. 5, we observe that, while most of the Arabic people had positive sentiments towards Pfizer-BioNTech, they had neutral emotions towards Johnson & Johnson and negative emotions towards Sinopharm-BBIBP.

Since there are no other studies that looked at vaccines acceptance over others in social media. To extract some potential answers to why Pfizer-BioNTech is experiencing more positive sentiments than others, we looked at the

actual tweets of our dataset, and we collected the following heuristics:

1. Availability: Pfizer-BioNTech vaccine was the first vaccine available among other vaccines. Since it appeared when the population was desperate for a solution, it was widely accepted and adopted, making it be the “savior” according to several tweets.
2. Safety: Pfizer-BioNTech has been approved to be used for younger people aged 12 and older compared to other vaccines such as Moderna and Janson & Janson for 18 years and above.
3. Storage temperature: Pfizer-BioNTech needs an ultra freezer “Sub-Zero”, while Moderna and others can be stored in a regular freezer. This makes people think that Pfizer-BioNTech is more important than other vaccines.
4. Side effect: People who were vaccinated with Pfizer-BioNTech shared their positive experiences afterward of having no side effects or minimal after taking the vaccine shot compared to other vaccines. For example, Johnson & Johnson vaccine was reported to cause, in some cases, blood clots and causes to be paused for further investigation.

Timeline of Vaccination Types. We also sought to explore how sentiment toward Covid-19 vaccines changed over time and expressed our results in Fig. 6. It was evident that February and April had the highest percentage of positive sentiments, while January had the highest number of neutral sentiments. March had the highest percentage of negative emotions towards the vaccines. The overall percentages in January, February, March, and April were 20%, 19%, 36%, and 25%, respectively. It is evident that people are developing more favorable opinions of the Covid-19 vaccines over time.

Sentiment analyzer. Sentiment analysis involves the classification of emotions into either positive, negative or neutral, which is done using various techniques. The sentiment score range is either negative, positive, or equal to zero as given in Table 5. The sentiment analyzers we used were as follows: Camel, Mazagk, TextBlob, and Final Voting. A summary of the statistics we obtained using the sentiment analyzers is given in Fig. 7. The highest percentage of positive sentiments were recorded using Camel, followed by Mazagk. For the neutral sentiments, the highest percentage was realized using TextBlob followed. Finally, on the negative sentiments, the highest percentage was recorded using Camel, which was closely followed by Mazagk. Figure 7 provides pie charts to show the percentages of positive, negative, and neutral tweets in each of the four sentiment analyzers. Overall sentiment analyzers (final voting) showed that 50% of the tweets collected voted as neutral

while the 38% tagged as negative, and only 12% voted as positive tweets.

Summary RQ₁

We found that these vaccines *Pfizer-BioNTech* and *Oxford-AstraZeneca* are the most popular. *Pfizer-BioNTech* vaccine is the highest positive rate, and *Sinopharm-BBIBP* is the highest negative rate.

RQ₂: What are the reasons that prevent people from taking the vaccine?

Motivation. Our second research question focuses on the reasons behind people not taking vaccines. The aim was to understand why the public was not quick to get vaccinated, despite most governments making every effort to purchase and distribute the vaccines. Identifying the reasons that are making people not to take vaccines is important, especially for public health authorities, in devising strategies on how to increase vaccination rates.

Approach. To answer this research question, we performed a manual analysis of a sample of tweets. The sample was randomly selected from the final dataset. We evaluated a sample size of 2500 tweets out of our final 1,098,376 tweets (described in detail in Sect. 3.1), a statistically significant sample with a confidence level of 99% and a confidence interval of 2.58%.

Reason 1: Lack of trust in vaccines. One of the takeaways from our study is that people are avoiding taking vaccines because of a lack of trust. We found that people do not have a lot of trust in vaccines, which has been a problem for many years. For the COVID-19 vaccines, the major reason for lack of trust is inadequate information and ignorance, which could be improved through public education initiatives. In some cases, people are failing to trust public health institutions, or even the healthcare systems, making them resent the vaccinations. Trust is a very important aspect of vaccination uptake and should be given attention to ensure that people get vaccinated. The lack of trust and skepticism towards COVID-19 vaccines were identified in a previous study by Chowdhury et al. (2021) using English tweets, and our study has established the same using Arabic tweets.

Reason 2: Vaccine timeline. We have also realized that the timeline within which the COVID-19 vaccines have been created has led to suspicion over their authenticity. Most previous vaccines were developed and tested over many years. With the current pandemic, the processes between the development and administration of vaccines have taken less than a year, and some people think that such a rush may not be a cause of concern. It would be important to note that the vaccine development methodology used for COVID-19 has been employed for nearly two decades, and therefore, the development of the vaccine was fast enough. The effect

of time on public perception towards COVID-19 was identified in a previous study by Cotfas et al. (2021) using English tweets from the UK, and we have made a similar observation using Arabic tweets.

Reason 3: Perceptions on vaccine effectiveness. A section of the population is also concerned about the effectiveness of the vaccines. Our sentimental analysis indicated that most people are skeptical about the duration of effectiveness of the vaccines. For example, some vaccines require more than one dose, which raises suspicion on how long the vaccines are protecting them. Although trials and studies were done with human subjects, there is worry that the long-term effectiveness is not well established because the trials were done for a very short duration. In some cases, people have questioned whether a vaccinated person can still get or spread the virus. The need to counter fake news regarding vaccine effectiveness was identified in a previous study by Cotfas et al. (2021), and we made the same observation using Arabic tweets.

Reason 4: Side effects of the vaccines. There is a perception that COVID-19 vaccines have negative side effects, which we identified in our study. Some people are also worried about the long-term effects of the vaccines, which have not been established since the pandemic has only existed for less than two years. In some cases, there have been reported side effects such as fever, aches, blood clots, and others. Unfortunately, many people are not aware that such effects are mild, short-term, and are only manifested among very few people that take the vaccines. We are convinced that the perception of side effects is one of the leading causes of the negative sentiments that we found in our study, which is an important aspect for health authorities. Previous studies by Cimorelli et al. (2021), and Villavicencio et al. (2021) found that people were having negative perspectives toward COVID-19 vaccines due to their side effects, and we have established similar results using Arabic tweets.

Reason 5: Lack of concern about the virus. We have also noted that some people do not realize the seriousness of the COVID-19 pandemic and see no need for taking the vaccine. For example, the younger population is more resistant to taking the vaccines because they think only the older ones are more vulnerable to it. A previous study by Cimorelli et al. (2021) had also indicated that older people are more likely to get infected by COVID-19, and we have established using Arabic tweets that such a perception is making the younger ones not to get vaccinated. Other people believe that, after getting infected with the virus for the first time, their bodies develop natural immunity, and they do not need the vaccine. The inability to see the COVID-19 pandemic as a serious condition has even made people to believe that lockdowns,

vaccinations, and other public health measures are unnecessary. It is important that such opinions are overcome so that more people embrace vaccination.

Reason 6: The emergence of conspiracy theories. We noticed that the existence of conspiracy theories is affecting public perception towards COVID-19 vaccines, making them not to get the vaccinations. Such misinformation and misconceptions were evident from our dataset and the sentiments that we got in this study. Examples of such theories include the association of the vaccines with infertility, autism, and long-term complications, among others. The conspiracy theories need to be countered with information, which will lead to favorable attitudes towards vaccines. Using English tweets, the existence of misinformation was found by Yousefinaghani et al. (2021) as a disincentive towards vaccination, and our study has substantiated the finding using English tweets.

Summary RQ₂

We found that people are not taking the vaccine for mainly six reasons. The reasons are (1) lack of concern about the virus, (2) vaccine timeline, (3) perceptions on vaccine effectiveness, (4) side effects of the vaccines, (5) lack of concern about the virus, and (6) the emergence of conspiracy theories.

RQ₃: To what extent can machine learning and deep learning detect vaccine-related tweets?

Motivation. In this research question, we want to see how machine learning and deep learning perform to detect tweets that are not tagged as COVID-19 tweets. Some people tweet about COVID-19 related without tagging it, and we want to see whether we could detect such tweets using various techniques. This might be important because some people do not like to have miscommunication threats. Therefore, we used these techniques to detect any misinformation on this topic.

Approach. We use various machine learning techniques, including Random Forest (RF), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Gradient Boosting Machine (GBM), and Extra Tree Classifier (ETC). We also used various deep learning techniques, including Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP) (Fig. 8).

The results of machine learning and deep learning are shown in Table 8 and Fig. 9. According to the results, the performance of LR, SVC, LSTM, and MLP showed better than other models with 0.82 accuracy, precision, recall, and F1 score, respectively. We can see that from Fig. 9 that GNB provided performs the worst as compared to all other models.

Comparing machine learning with deep learning, all three deep learning models show a close accuracy range between 0.81 and 0.82. However, machine learning models show that

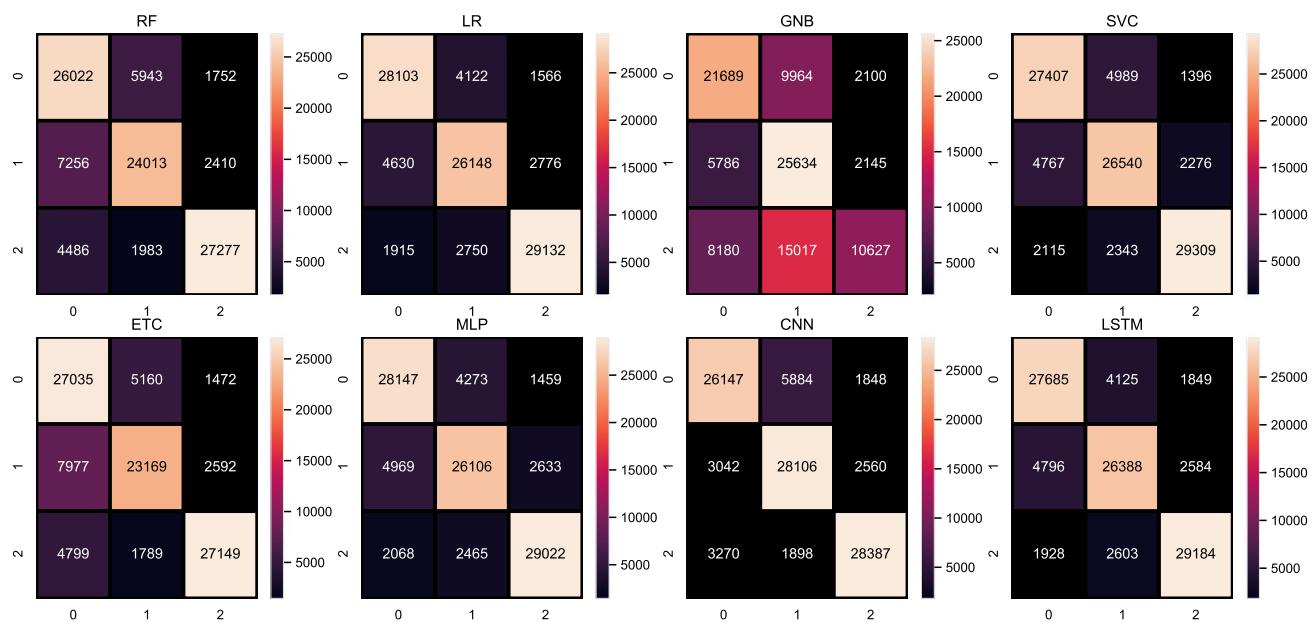


Fig. 8 Confusion matrices of all models with TF-IDF features

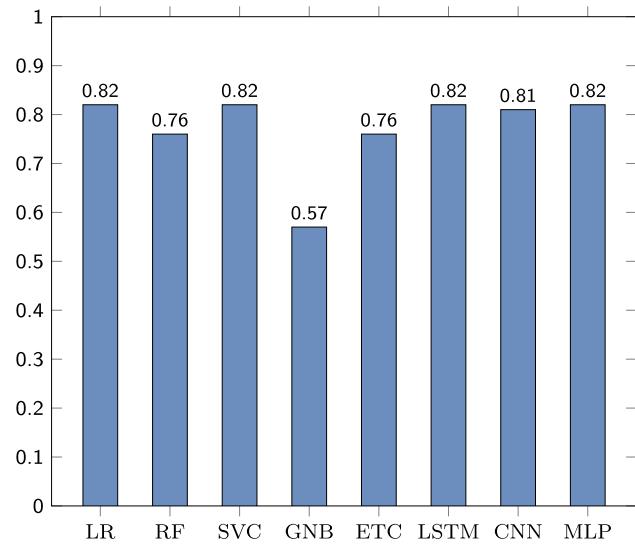


Fig. 9 Comparison of accuracy of all models with TF-IDF features

there are two models out-performs from all four models: LR and SVC models.

Figure 8 shows the confusion matrix for each models. In the confusion matrix, Negative, Neutral, and Positive are represented by 0, 1, and 2, respectively. These confusion matrices show that the models are equally good for each target class. The accuracy score of LR, SVC, LSTM, and MLP is somehow equal after the roundoff of values, but

Table 8 Per-class precision, recall and f1-score of all models with TF-IDF features

Models	Class	Precision	Recall	F1-score
LR	Negative	0.83	0.83	0.83
	Neutral	0.83	0.86	0.84
	Positive	0.80	0.68	0.73
RF	Negative	0.73	0.77	0.75
	Neutral	0.77	0.81	0.79
	Positive	0.82	0.51	0.63
SVC	Negative	0.82	0.83	0.82
	Neutral	0.83	0.86	0.84
	Positive	0.81	0.66	0.73
GNB	Negative	0.76	0.60	0.67
	Neutral	0.81	0.49	0.61
	Positive	0.25	0.83	0.39
ETC	Negative	0.76	0.74	0.75
	Neutral	0.75	0.85	0.80
	Positive	0.88	0.46	0.60
LSTM	Negative	0.82	0.83	0.83
	Neutral	0.81	0.85	0.84
	Positive	0.82	0.78	0.79
CNN	Negative	0.82	0.83	0.82
	Neutral	0.81	0.84	0.81
	Positive	0.80	0.68	0.75
MLP	Negative	0.82	0.83	0.82
	Neutral	0.81	0.85	0.82
	Positive	0.82	0.68	0.75

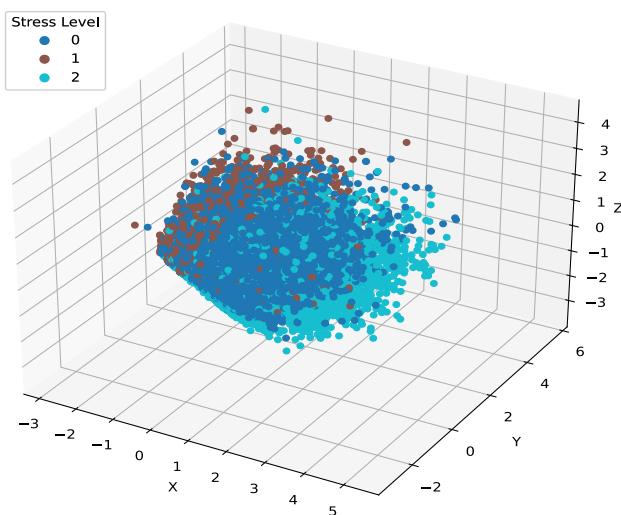


Fig. 10 Feature space for each target class

there is a little bit of difference in the number of correct predictions. LR gives 83,383 correct predictions out of 101142. Similarly, SVC gives 83,256, LSTM gives 83,257 and MLP gives 83,275 correct predictions. These results show that LR is significant compared to all other used models with the highest correct prediction ratio.

We also show the per-class accuracy of models in Table 8 and Fig. 10, which shows that all models are equally good for all target classes. These all results show that models are not over-fitted on any single class.

To further discuss the analysis on models performance, we also illustrate the feature space. Figure 10 shows the features of with respect to the target classes. We can see that features are somehow linearly separable which help to achieve the significant 0.82 accuracy score but their is little bit overlapping in samples. This little overlapping of samples is a limitation of this study. We will consider a more linearly separable feature space in our future work and will work on feature engineering techniques for this.

Summary RQ₃

We were able to automatically detect tweets that take about vaccines without using the hashtag the vaccine with an accuracy of 0.82 on average.

5 Conclusion

In this study, we explored public sentiments toward COVID-19 vaccines with an aim to establish the reasons that people were not taking the vaccines. In addition, we hoped to contribute to the existing literature by providing the first study that utilized Arabic tweets to evaluate public opinions of COVID-19 vaccines. We established that, among the

various vaccines, Pfizer-BioNTech had the highest positive sentiment. However, in all the vaccines, the positive sentiments fluctuated between January and April 2021, ranging between 10% and 14%. Some of the reasons for people not taking vaccines that we found in our study include lack of trust, vaccine timeline, perception of vaccine effectiveness, side effects of the vaccines, lack of public concern about the vaccines, and the emergence of conspiracy theories. The findings of our study would benefit public health authorities that are trying to convince the public to take the vaccines.

In the future, we recommend that researchers get public sentiments on COVID-19 vaccines in other languages, such as French, Spanish, Greek, and German, among others. Other studies may be done to compare the sentiments in different countries, regions, or languages. We also recommend expanding the dataset by collecting datasets from other social media sites such as Facebook and evaluating public sentiments towards the vaccines. Other researchers may also replicate the study using different sentiment analyzers and machine learning techniques. Finally, researchers may also explore ways to deal with the factors hindering people from getting vaccines.

Availability of code and data The materials used in this study are available at (Replication package 2022).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abdul-Mageed M, Diab M (2014) Sana: a large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis. In: Proceedings of the 9th international conference on language resources and evaluation (LREC'14)
- Adamu Hassan, Jiran MJB M, KH Gan, Samsudin N-H (2021) Text analytics on twitter text-based public sentiment for COVID-19 vaccine: a machine learning approach. In: 2021 IEEE international conference on artificial intelligence in engineering and technology (IICAIET), pp 1–6. IEEE
- Akpatsa SK, Li X, Lei H, Obeng V-HKS (2022) Evaluating public sentiment of COVID-19 vaccine tweets using machine learning techniques. *Informatica* 46(1), pages 69–75
- Al-Badrashiny M, Eskander R, Habash N, Rambow O (2014) Automatic transliteration of romanized dialectal arabic. In: Proceedings of the 18th conference on computational natural language learning, pp 30–38
- Aljedaani W, Aljedaani M, AlOmar EA, Mkaouer MW, Ludi S, Khalaf YB (2021) I cannot see you—the perspectives of deaf students to

- online learning during COVID-19 pandemic: Saudi Arabia case study. *Edu Sci* 11(11):712
- Aljedaani W, Mkaouer MW, Ludi S, Javed Y (2022) Automatic classification of accessibility user reviews in android apps. In: 2022 7th international conference on data science and machine learning applications (CDMA), pp 133–138. IEEE
- Aljedaani W, Mkaouer MW, Ludi S, Ouni A, Jenhani I (2022) On the identification of accessibility bug reports in open source systems. In: Proceedings of the 19th international web for all conference, pp 1–11
- Aljedaani W, Rustam F, Ludi S, Ouni Ali, Mkaouer MW (2021) Learning sentiment analysis for accessibility user reviews. In: 2021 36th IEEE/ACM international conference on automated software engineering workshops (ASEW), pp 239–246. IEEE
- Allagui I, Breslow H (2016) Social media for public relations: lessons from four effective cases. *Public Relat Rev* 42(1):20–30
- AlOmar EA, Aljedaani W, Tamjeed M, Mkaouer MW, El-Glaly YN (2021) Finding the needle in a haystack: on the automatic identification of accessibility user reviews. In: Proceedings of the 2021 CHI conference on human factors in computing systems, pp 1–15
- Alqurashi S, Alhindi A, Alanazi E (2020) Large Arabic twitter dataset on COVID-19. [arXiv:2004.04315](https://arxiv.org/abs/2004.04315)
- Amaar A, Aljedaani W, Rustam F, Ullah S, Rupapara V, Ludi S (2022) Detection of fake job postings by utilizing machine learning and natural language processing approaches. *Neural Process Lett* 54(3):2219–2247
- Amin MZ, Nadeem N (2018) Convolutional neural network: text classification model for open domain question answering system. [arXiv:1809.02479](https://arxiv.org/abs/1809.02479)
- Arel I, Rose DC, Karnowski TP (2010) Deep machine learning—a new frontier in artificial intelligence research [research frontier]. *IEEE Comput Intell Mag* 5(4):13–18
- Aygün İ, Kaya B, Kaya M (2021) Aspect based twitter sentiment analysis on vaccination and vaccine types in COVID-19 pandemic with deep learning. *IEEE J Biomed Health Inf* 26(5):2360–2369
- Bakshi RK, Kaur N, Kaur R, Kaur G (2016) Opinion mining and sentiment analysis. In: 2016 3rd international conference on computing for sustainable global development (INDIACOM), pp 452–455. IEEE
- Baly F, Hajj H, et al (2020) Arabert: transformer-based model for arabic language understanding. In: Proceedings of the 4th workshop on open-source arabic corpora and processing tools, with a shared task on offensive language detection, pp 9–15
- Batra R, Imran AS, Kastrati Z, Ghafoor A, Daudpota SM, Shaikh S (2021) Evaluating polarity trend amidst the coronavirus crisis in peoples' attitudes toward the vaccination drive. *Sustainability* 13(10):5344
- Bei Yu (2008) An evaluation of text classification methods for literary study. *Lit Linguist Comput* 23(3):327–343
- Benajiba Y, Rosso P (2007) Anersys 2.0: Conquering the ner task for the arabic language by combining the maximum entropy with pos-tag information. In: IICAI, pp 1814–1823
- Benajiba Y, Rosso P (2008) Arabic named entity recognition using conditional random fields. In: Proceedings of the workshop on HLT & NLP within the Arabic World, LREC, vol 8, pp 143–153. Citeseer
- Bethapudi S, Desai S (2018) Separation of pulsar signals from noise using supervised machine learning algorithms. *Astron Comput* 23:15–26
- Biau G, Scornet E (2016) A random forest guided tour. *Test* 25(2):197–227
- Bird SG (2005) Nltk-lite: efficient scripting for natural language processing. [arXiv:cs/0205028](https://arxiv.org/abs/cs/0205028)
- Bocca FF, Rodrigues LHA (2016) The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Comput Electron Agric* 128:67–76
- Boudlal A, Lakhouaja A, Mazroui A, Meziane A, Bebah MOAO, Shoul M (2010) Alkhalil morpho sys1: a morphosyntactic analysis system for arabic texts. In: International Arab conference on information technology, pages 1–6. Elsevier Science Inc New York, NY
- Braun V, Clarke V (2006) Using thematic analysis in psychology. *Qual Res Psychol* 3(2):77–101
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Buckwalter T (2002) Buckwalter arabic morphological analyzer version 1.0. Linguistic Data Consortium, University of Pennsylvania
- Chowdhury AA, Das A, Saha SK, Rahman M, Hasan KT (2021) Sentiment analysis of COVID-19 vaccination from survey responses in bangladesh. *BMC Public Health*
- Cimorelli A, Bollinger M, Newman K, Wood B (2021) Social media sentiments towards vaccines. Available at SSRN 3833625
- Cotfas L-A, Delcea C, Roxin I, Ioanăs C, Gherai DS, Tajariol F (2021) The longest month: analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access* 9:33203–33223
- Darwish K, Magdy W (2014) Arabic information retrieval. *Found Trends Inf Retr* 7(4):239–342
- Darwish K, Magdy W, Mourad A (2012) Language processing for Arabic microblog retrieval. In: Proceedings of the 21st ACM international conference on information and knowledge management, pp 2427–2430
- Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- Driss SB, Soua M, Kachouri R, Akil M (2017) A comparison study between mlp and convolutional neural network models for character recognition. In: Real-time image and video processing 2017, vol 10223, p 1022306. International Society for Optics and Photonics
- El-Beltagy SR, Kalamawy MEL, Soliman ABakr (2017) Niletmrg at semeval-2017 task 4: arabic sentiment analysis. [arXiv:1710.08458](https://arxiv.org/abs/1710.08458)
- Elfardy H, Diab MT (2012) Simplified guidelines for the creation of large scale dialectal arabic annotations. In: LREC, pp 371–378. Citeseer
- Elhishi S, El-Deeb R, El-Gamal FE-ZA, Sakr NA, El-Metwally S (2021) Analyzing public perceptions toward COVID-19 vaccination process using social media and machine learning. In: The 7th annual international conference on arab women in computing in conjunction with the 2nd forum of women in research, pp 1–4
- Fang F, Wu J, Li Y, Ye X, Aljedaani W, Mkaouer MW (2021) On the classification of bug reports to improve bug localization. *Soft Comput* 25(11):7307–7323
- Farha IA, Magdy W (2019) Mazajak: an online Arabic sentiment analyser. In: Proceedings of the fourth arabic natural language processing workshop, pp 192–198
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat*, pp 1189–1232
- Fukushima K, Miyake S (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets, pp 267–285. Springer
- Haouari F, Hasanain M, Suwaileh R, Elsayed T (2020) Arcov-19: The first arabic covid-19 twitter dataset with propagation networks. [arXiv:2004.05861](https://arxiv.org/abs/2004.05861)
- Heaton J (2016) An empirical analysis of feature engineering for predictive modeling. In: SoutheastCon 2016, pp 1–6. IEEE
- Hoang V-T, Jo K-H (2019) Pydnet: an efficient CNN architecture with pyramid depthwise convolution kernels. In: 2019 international

- conference on system science and engineering (ICSSE), pp 154–158. IEEE
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243
- Hu Y, John A, Wang F, Kambhampati S (2012) Et-Ida: joint topic modeling for aligning events and their twitter feedback. In: Proceedings of the AAAI conference on artificial intelligence, vol 26
- Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, 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. *J Med Internet Res* 23(4):e26627
- Jayasurya GG, Kumar S, Singh BK, Kumar V (2021) Analysis of public sentiment on COVID-19 vaccination using twitter. *IEEE Trans Comput Soc Syst*, vol. 9, page 1101–1111
- Jones I, Roy P (2021) Sputnik V COVID-19 vaccine candidate appears safe and effective. *The Lancet* 397(10275):642–643, vol. 397
- Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. *Emerg Artif Intell Appl Comput Eng* 160(1):3–24
- Kulkarni A, Shivananda A (2021) Implementing industry applications. In: Natural language processing recipes, pp 135–211. Springer
- Kumaresh S (2021) Sentiment analysis of covid-19 vaccine in a social media platform using machine learning techniques. *syndicate-The J Manage*, 39, vol.21 pages 39–51
- Laksono RAdi, Sungkono KR, Sarno R, Wahyuni CS (2019) Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In: 2019 12th international conference on information & communication technology and system (ICTS), p 49–54. IEEE
- Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. *IEEE Trans Neural Netw* 8(1):98–113
- Lee E, Rustam F, Washington PB, Barakaz FE, Aljedaani W, Ashraf I (2022) Racism detection by analyzing differential opinions through sentiment analysis of tweets using stacked ensemble gcr-nn model. *IEEE Access* 10:9717–9728
- Liaw A, Wiener M et al (2002) Classification and regression by randomforest. *R news* 2(3):18–22
- Liu S, Li J, Liu J et al (2021) Leveraging transfer learning to analyze opinions, attitudes, and behavioral intentions toward COVID-19 vaccines: social media content and temporal analysis. *J Med Internet Res* 23(8):e30251
- Loper E, Bird S (2002) Nltk: the natural language toolkit. [arXiv:cs/0205028](https://arxiv.org/abs/cs/0205028)
- Loria S (2018) textblob documentation. Release 015:2
- Manik A, Adiwijaya A, Utama DQ (2019) Classification of electrocardiogram signals using principal component analysis and levenberg marquardt backpropagation for detection ventricular tachyarrhythmia. *J Data Sci Appl* 2(1):29–37
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. [arXiv:1301.3781](https://arxiv.org/abs/1301.3781)
- Mishra S, Verma A, Meena K, Kaushal R (2022) Public reactions towards COVID-19 vaccination through twitter before and after second wave in India. *Soc Netw Anal Min* 12(1):1–16
- Munjal P, Narula M, Kumar S, Banati H (2018) Twitter sentiments based suggestive framework to predict trends. *J Stat Manage Syst* 21(4):685–693
- Nezhad ZB, Deihimi MA (2022) Twitter sentiment analysis from Iran abou COVID 19 vaccine. *Diabetes Metab Syndr Clin Res Rev* 16(1):102367
- Norambuena BK, Lettura EF, Villegas CM (2019) Sentiment analysis and opinion mining applied to scientific paper reviews. *Intell Data Anal* 23(1):191–214
- Nurdeni DA, Budi I, Santoso AB (2021) Sentiment analysis on covid19 vaccines in Indonesia: from the perspective of sinovac and pfizer. In: 2021 3rd East Indonesia conference on computer and information technology (EICoCIT), pp 122–127. IEEE
- Nuser M, Alsukhni E, Saifan A, Khasawneh R, Ukkaz D (2022) Sentiment analysis of COVID-19 vaccine with deep learning. *J Theor Appl Inf Technol* 100(12), pages 4513–4521
- Obeid O, Zalmout N, Khalifa S, Taji D, Oudah M, Alhafni B, Inoue G, Eryani F, Erdmann A, Habash N (2020) Camel tools: an open source python toolkit for arabic natural language processing. In: Proceedings of the 12th language resources and evaluation conference, pp 7022–7032
- Paliwal S, Parveen S, Alam MA, Ahmed J (2022) Sentiment analysis of COVID-19 vaccine rollout in India. In: ICT systems and sustainability, pp 21–33. Springer
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. [arXiv:cs/0205070](https://arxiv.org/abs/cs/0205070)
- Patton MQ (1990) Qualitative evaluation and research methods. SAGE Publications, Inc, Thousand Oaks
- Rahul K, Jindal BR, Singh K, Meel P (2021) Analysing public sentiments regarding COVID-19 vaccine on twitter. In: 2021 7th international conference on advanced computing and communication systems (ICACCS), vol 1, pp 488–493. IEEE
- Replication package (2022) https://wajdialjedaani.github.io/ArTwe_etCovid19/
- Reshi AA, Rustam F, Aljedaani W, Shafi S, Alhossan A, Alrabiah Z, Ahmad A, Alsuwailem H, Almangour TA, Alshammari MA, et al (2022) Covid-19 vaccination-related sentiments analysis: a case study using worldwide twitter dataset. In: Healthcare, vol 10, p 411. MDPI
- Robertson S (2004) Understanding inverse document frequency: on theoretical arguments for IDF. *J doc*, vol. 60, pages 503–520
- Ruggeri A, Samoggia A (2018) Twitter communication of agri-food chain actors on palm oil environmental, socio-economic, and health sustainability. *J Consum Behav* 17(1):75–93
- Rupapara V, Rustam F, Aljedaani W, Shahzad HF, Lee E, Ashraf I (2022) Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Sci Rep* 12(1):1–15
- Rustam F, Ashraf I, Mehmood A, Ullah S, Choi GS (2019) Tweets classification on the base of sentiments for 37 us airline companies. *Entropy* 21(11):1078
- Rustam F, Reshi AA, Aljedaani W, Alhossan A, Ishaq A, Shafi S, Lee E, Alrabiah Z, Alsuwailem H, Ahmad A et al (2022) Vector mosquito image classification using novel RIFS feature selection and machine learning models for disease epidemiology. *Saudi J Biol Sci* 29(1):583–594
- Sadat F, Kazemi F, Farzindar A (2014) Automatic identification of arabic dialects in social media. In: Proceedings of the first international workshop on Social media retrieval and analysis, pp 35–40
- Safdar N, Alrubaye H, Aljedaani W, Baez BB, DiStasi A, Mkaouer MW (2019) Learning to rank faulty source files for dependent bug reports. In: Big data: learning, analytics, and applications, vol 10989, pp 60–78. SPIE
- Salameh M, Bouamor H, Habash N (2018) Fine-grained arabic dialect identification. In: Proceedings of the 27th international conference on computational linguistics, pp 1332–1344
- Schapire RE (1999) A brief introduction to boosting. In: Ijcai vol 99, pp 1401–1406. Citeseer
- Schmidhuber J, Hochreiter S (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Sebastiani F (2002) Machine learning in automated text categorization. *ACM Comput Surv (CSUR)* 34(1):1–47
- Shaaalan K, Raza H (2009) Nera: Named entity recognition for arabic. *J Am Soc Inf Sci Technol* 60(8):1652–1663
- Smirnov EA, Timoshenko DM, Andrianov SN (2014) Comparison of regularization methods for imangenet classification with deep convolutional neural networks. *Aasri Procedia* 6:89–94

- Smrz O (2007) Elixirfm—implementation of functional arabic morphology. In: Proceedings of the 2007 workshop on computational approaches to Semitic languages: common issues and resources, pp 1–8
- Sohangir S, Petty N, Wang D (2018) Financial sentiment lexicon analysis. In: 2018 IEEE 12th international conference on semantic computing (ICSC), pp 286–289. IEEE
- Soliman AB, Eissa K, El-Beltagy SR (2017) Aravec: a set of arabic word embedding models for use in arabic nlp. Procedia Comput Sci 117:256–265
- Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Comput Linguist 37(2):267–307
- Tarkhaneh O, Shen H (2019) Training of feedforward neural networks for data classification using hybrid particle swarm optimization, mantegna lévy flight and neighborhood search. Heliyon 5(4):e01275
- Textblob: Simplified text processing. <https://textblob.readthedocs.io/en/dev/>. Accessed: 2021-07-03
- Tramacere A, Paraficz D, Dubath P, Kneib J-P, Courbin F (2016) Asterism: application of topometric clustering algorithms in automatic galaxy detection and classification. Mon Not R Astron Soc 463(3):2939–2957
- Valentini C (2015) Is using social media “good” for the public relations profession? A critical reflection. Public Relat Rev 41(2):170–177
- Vanni M, Zajac R (1996) The temple translator’s workstation project. In: TIPSTER TEXT PROGRAM PHASE II: proceedings of a Workshop held at Vienna, Virginia, May 6–8, 1996, pp 101–106
- Vijayarani S, Janani R et al (2016) Text mining: open source tokenization tools—an analysis. Adv Comput Intell Int J (ACII) 3(1):37–47
- Villavicencio C, Macrohon JJ, Inbaraj XA, Jeng J-H, Hsieh J-G (2021) Twitter sentiment analysis towards COVID-19 vaccines in the philippines using naïve bayes. Information 12(5):204
- Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: A survey. Pattern Recognit Lett 119:3–11
- Wolf T, Debut L, Sanh Victor, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, et al (2019) Huggingface’s transformers: state-of-the-art natural language processing. [arXiv:1910.03771](https://arxiv.org/abs/1910.03771)
- Wordclouds (2022) <https://www.wordclouds.com/>
- World Health Organization et al (2020a) Behavioural considerations for acceptance and uptake of COVID-19 vaccines: who technical advisory group on behavioural insights and sciences for health, meeting report, 15 october 2020. WHO
- World Health Organization et al. (2020b) Guidance on developing a national deployment and vaccination plan for COVID-19 vaccines: interim guidance, 16 november 2020. Technical report, World Health Organization
- Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9(4):611–629
- Yan S (2016) Understanding lstm and its diagrams. <https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714>, March. Accessed: 04/06/2021
- Yang X, Sornlertlamvanich V (2021) Public perception of COVID-19 vaccine by tweet sentiment analysis. In: 2021 international electronics symposium (IES), pp 151–155. IEEE
- Ye X, Zheng Y, Aljedaani W, Mkaouer MW (2021) Recommending pull request reviewers based on code changes. Soft Comput 25(7):5619–5632
- Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S (2021) An analysis of COVID-19 vaccine sentiments and opinions on twitter. Int J Infect Dis, vol. 108, pages 256–262
- Zhang W, Yoshida T, Tang X (2011) A comparative study of tf* idf, lsi and multi-words for text classification. Expert Syst Appl 38(3):2758–2765
- Zhang J, Li Y, Tian J, Li T (2018) Lstm-cnn hybrid model for text classification. In: 2018 IEEE 3rd Advanced information technology, electronic and automation control conference (IAEAC), pp 1675–1680. IEEE
- Zhou C, Sun C, Liu Z, Lau F (2015) A c-lstm neural network for text classification. [arXiv:1511.08630](https://arxiv.org/abs/1511.08630)
- Zulfiker MS, Kabir N, Biswas AA, Zulfiker S, Uddin MS (2022) Analyzing the public sentiment on COVID-19 vaccination in social media: Bangladesh context. Array 15:100204

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Wajdi Aljedaani¹  · Ibrahem Abuhamed² · Furqan Rustam³ · Mohamed Wiem Mkaouer⁴ · Ali Ouni⁵ · Ilyes Jenhani⁶

Ibrahem Abuhamed
abuhoimed@psu.edu.sa

Furqan Rustam
furqan.rustam1@gmail.com

Mohamed Wiem Mkaouer
mwmvse@rit.edu

Ali Ouni
ali.ouni@etsmtl.ca

Ilyes Jenhani
ijenhani@pmu.edu.sa

² Prince Sultan University, Riyadh, Saudi Arabia

³ University of Management and Technology, Lahore, Pakistan

⁴ Rochester Institute of Technology, Rochester, USA

⁵ ETS Montreal, University of Quebec, Quebec, Canada

⁶ Prince Mohammad Bin Fahd University, Dhahran, Saudi Arabia

¹ University of North Texas, Denton, USA