Sentiment Analysis of COVID-19 Vaccination in Saudi Arabia

Sawsan Alowa, Lama Alzahrani, Noura Alhakbani, and Hend Alrasheed*
Department of Information Technology, King Saud University, Riyadh 11451, Saudi Arabia;
442202912@student.ksu.edu.sa; 442203072@student.ksu.edu.sa; nhakbani@ksu.edu.sa
*Correspondence: halrasheed@ksu.edu.sa

Abstract
Since the COVID-19 vaccine became available, people have been sharing their opinions on social media about getting vaccinated, causing discussions of the vaccine to trend on Twitter alongside certain events, making the website a rich data source. This paper explores people's perceptions regarding the COVID-19 vaccine during certain events and how these events influenced public opinion about the vaccine. The data consisted of tweets sent during seven important events that were gathered within 14 days of the first announcement of each event. These data represent people's reactions to these events without including irrelevant tweets. The study targeted tweets sent in Arabic from users located in Saudi Arabia. The data were classified as positive, negative, or neutral in tone. Four classifiers were used —support vector machine (SVM), naïve Bayes (NB), logistic regression (LOGR), and random forest (RF)—in addition to a deep learning model using BiLSTM. The results showed that the SVM achieved the highest accuracy, at 91%. Overall perceptions about the COVID-19 vaccine were 54% negative, 36% neutral, and 10% positive.

Keywords:
COVID-19 vaccine; sentiment analysis; BiLSTM; polarity; machine learning; SMOTE; oversampling

1. Introduction

COVID-19 is upending people's lives worldwide and is adversely impacting many aspects of society as well as the global economy. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1,2]. Clinical manifestations range from mild or asymptomatic to more severe illnesses that can lead to respiratory arrest and death [3]. One of the earliest precautionary measures governments took worldwide was applying community containment. Community containment is a strategy intended to slow the transmission rate of new viruses without an available vaccine or a highly effective antiviral therapy. Community mitigation measures against COVID-19 included physical/social distancing, the closure of schools, restaurants, and movie theaters, and encouraging organizations to allow employees to work from home. Also, large gatherings such as festivals, graduations, and sporting events were discouraged or prohibited [4]. The economic impact of community mitigation has destroyed many businesses. For example, in the United States alone, more than 40 million people filed initial claims for unemployment benefits [5,6]

Multiple vaccines have been developed to contain the COVID-19 pandemic, including vaccines from AstraZeneca, Johnson & Johnson, Moderna, and Pfizer-BioNTech. Given their high degree of development and otherwise lengthy vaccine approval timeframe, the COVID-19 vaccine received temporary emergency approval. Still, both manufacturers and researchers emphasized the vaccine's safety and efficacy in reducing the risk of serious illness associated with COVID-19 infection. They guaranteed that the new vaccine had undergone rigorous testing to verify its safety before being recommended for widespread use [7].
Upon the availability of the COVID-19 vaccine, world governments announced plans to provide COVID-19 vaccinations to end the global pandemic [8–11]. For instance, in Saudi Arabia, the Ministry of Health has urged citizens and residents to get vaccinated as an essential step to return to normalcy [12]. However, many individuals have responded to these efforts with skepticism and have raised various concerns about the vaccine. Additionally, many activists have fueled concerns about vaccine safety on social media platforms, particularly Twitter [13]. In response, officials and experts have emphasized the vaccine's safety and have urged citizens and residents to receive the vaccine for the well-being and safety of all [14].

Social media is vital to this issue as it connects people from different geographical locations and allows them to exchange their opinions and ideas. The new measures imposed by authorities regarding vaccination enforcement have prompted many to express their thoughts on social media. Social networks have been used to discuss controversial topics that attract people's attention. Some people use information posted on social network platforms and consider the websites a reliable source. Thus, social media is considered one of the largest information systems, one that can yield rich data for analytical purposes [15].
Therefore, policymakers and government officials must understand people's attitudes toward specific issues as portrayed on social media. In this respect, the potential of sentiment analysis can be exploited. Sentiment analysis is a commonly used task in natural language processing (NLP). It comprises subtasks including polarity analysis, which detects the tone of the text (positive, negative, or neutral) [16]; emotion recognition, which recognizes feelings, such as

Manuscript received February 5, 2023
Manuscript revised February 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.2.2

happiness, sadness, or fear, behind the given text; and subjectivity identification, which classifies the text as subjective or objective [17,18]. The present study used sentiment analysis to analyze public opinions about COVID-19 vaccination in Saudi Arabia as posted on Twitter. Our results demonstrated that negative sentiment was the most dominant tone among the analyzed posts. We also gained insight into changes in public opinion alongside certain events over time.

The Twitter platform has become an important channel for governments and health sectors to disseminate crucial information and to communicate and understand public concerns and attitudes. The massive amount of available data generated every second on Twitter has garnered broad interest from researchers. The motivation of this research was to assist the Saudi Arabian health sector and government in understanding people's opinions (positive, negative, or neutral) regarding the COVID-19 vaccine during selected events.

We obtained insights into public opinions concerning the COVID-19 vaccine in the Saudi community by performing a sentiment analysis of "tweets" related to the research topic. This study answered the following research questions:

1. What were public perceptions in Saudi Arabia about COVID-19 vaccination during specific events?
2. Which event had the most significant impact on public opinions about the COVID-19 vaccine based on the number of tweets per event? What was the overall sentiment regarding that event?

The main contributions of this paper are as follows:

- The detection of social polarity, via sentiment analysis, concerning COVID-19 vaccination during specific events in Saudi Arabia,
- The estimation of the overall sentiments of Saudi citizens regarding the most prominent events,
- The exploration of the effect of applying the oversampling technique with classical machine learning algorithms.

This paper is structured as follows. Section 2 presents related work that employed sentiment analysis on tweets. Section 3 describes the materials and methods used in the current study, while section 4 presents the research results. Section 5 presents a discussion of these results, while section 6 provides conclusions based on this discussion.

2. Related Work

Many studies have applied sentiment analysis to study public opinions on social media. For instance, Hung et al. [19] used machine learning methods to analyze COVID-19 discussions on Twitter over a one-month period from users located in the United States. The main goal was to understand sentiments toward COVID-19 in addition to predominant associated topics. The tweets under study were classified as positive, negative, or neutral sentiments, and the geographic areas in which users were located were considered. Latent Dirichlet allocation (LDA) was used for topic modeling, along with the VADER tool for sentiment analysis. The results indicated a range of positive and negative sentiments, and five major topics were identified (emotional support, social change, healthcare environment, business economy, and psychological stress). The study concluded that people who expressed the most negative and most positive sentiments via tweets lived in states with the highest and lowest infection rates, respectively. The authors ensured that their results from both topic modeling and sentiment analysis were validated. First, the topic modeling results were validated using centrality measurements and by plotting the social network graphs associated with each dominant topic. Second, the sentiment analysis obtained from VADER was validated by manually annotating a random sample of the dataset and then measuring its sensitivity and specificity to verify the quality of the sentiments it contained in comparison to manual human sentiments.

Another study, conducted by Manguri et al. [16], mined people's opinions on COVID-19. The data were collected from Twitter for seven consecutive days using the Tweepy library, and the search mainly utilized two hashtags: "#COVID-19" and "#coronavirus." Then, the data were analyzed using the TextBlob library, which employs the naïve Bayes (NB) classifier. The results showed that most tweets expressed neutral sentiments about COVID-19. The authors also found that people who used the COVID-19 hashtag were more optimistic than those who used the coronavirus hashtag. However, despite the fact that the authors mentioned evaluating their sentiment analysis, no such evaluation was covered in their paper.

An interesting observation, made by Samuel et al. [20], was that tweet length directly affects the performance of machine learning algorithms. The authors examined Twitter textual analytics during COVID-19 to assess trends in public fear and negative sentiment over time as COVID-19 peaked in the United States. Using NB and logistic regression (LOGR), the study conducted an exploratory sentiment analysis to model two machine learning methods for the purpose of classifying positive and negative sentiments from different tweet lengths. The research found that the length of tweets influences classification accuracy, with both methods showing weak performance with long tweets. Meanwhile, the performance of the NB classifier was better than that of LOGR when analyzing short tweets.

Unlike other studies that used supervised machine learning algorithms, Alanezi and Hewahi [21] performed sentiment analysis using clustering methods. Their research was aimed at examining the effects of social distancing on people during COVID-19. The researchers compared k-means clustering and mini-batch k-means clustering for two datasets collected from Twitter in both Arabic and English. They found that better results were achieved with the English-language dataset and by using k-means clustering. According to the authors, the reason behind the variation in the performance of the two datasets was that some of the Python libraries used did not support the Arabic language.

Du et al. [22] proposed hierarchical machine learning based on a three-level support vector machine (SVM) classification to explore sentiment-related refusal of the human papillomavirus (HPV) vaccine as expressed on Twitter. The study classified tweet polarity into three categories (positive, negative, and neutral) and examined the causes of negative sentiments, such as safety, efficacy, emotional resistance, and cost. The system's overall performance in automatically predicting the labels of unannotated tweets was promising. The system predicted the labels of 184,214 tweets in the collected unannotated corpus. The results showed that the system achieved an accuracy rate of around 70% for both micro-averaging and macro-averaging. Moreover, the authors added an additional step in which they performed a time series analysis on the predicted sentiment. This step allowed them to find trends in different sentiment categories.

Raghupathi et al. [23] studied public perceptions regarding vaccination during a measles outbreak in the United States. Tweets related to vaccines were collected from January 1, 2019, to April 5, 2019. After applying sentiment analysis, the data were clustered using k-means into topics using the term frequency inverse document frequency (TF-IDF) technique. The results showed that most of the tweets were concerned with finding new vaccines for these diseases (i.e., the Ebola virus, influenza, and HPV), while the remaining tweets were sent between people who were concerned about the measles outbreak as well as by supporters and opponents debating the efficacy of measles vaccination. The authors depended on VADER to extract sentiments from the tweets. However, an evaluation of the machine-extracted analysis was not available.

Salathé and Khandelwal [24] used Twitter data to study sentiments toward the influenza (H1N1) vaccine during the fall wave of the H1N1 pandemic in the United States. Different machine-learning classification algorithms were tested to classify the data as positive, negative, neutral, or irrelevant. The sentiment polarity of the training set was assigned manually by undergraduate students from Pennsylvania State University. To validate the sentiment analysis results, the findings were compared to vaccination rates provided by the Centers for Disease Control and Prevention (CDC). A correlation was found between opinions expressed on social media and the vaccination rate. Concerted efforts were made by the authors to develop an algorithm to create a network of Twitter users to track the flow of opinions regarding the H1N1 vaccine on the state level. However, a graphical representation of the network was not included.

In [25], the authors applied SVM and NB models to detect text sentiments in the context of COVID-19 vaccination. The study compared classical machine learning and deep learning approaches to select the best-performing algorithm. The authors reported that NB outperformed SVM and deep learning approaches with a 76% accuracy score.

An essential point was raised by Praveen et al. [26] in their study of the unbalanced data issue; they had some concerns that not handling this issue could affect the overall performance of the sentiment analysis. Therefore, their solution included selecting the same number of tweets per month. The authors studied Indian citizens' opinions toward the COVID-19 vaccine. The Twint library was used to collect the data from social media posts that included the phrase "COVID vaccine." The study consisted of two parts. First, sentiment analysis was performed using the TextBlob Python library to find overall perceptions about the COVID-19 vaccine. Second, topic modeling was applied using LDA on negative-sentiment posts to identify the issues that concerned citizens regarding the COVID-19 vaccine. The results showed that most of these discussions on social media had a neutral tone, followed by a positive tone and then a negative tone. The study also found that the main issues that concerned Indian citizens regarding the COVID-19 vaccine were allergic reactions and fear of side effects.

Nurdeni et al. [27] conducted sentiment analysis on citizen opinions about two types of vaccines used in Indonesia: Sinovac and Pfizer. The data were collected from Twitter using Twitter's API from October 2020 to November 2020. The dataset was split in two based on the vaccine type. Also, the data were labeled manually into three classes: positive, negative, and neutral. Three algorithms were applied for both datasets. Accuracy, precision, recall, and the F1-score were measured using NB, SVM, and random forest (RF). Both vaccines received positive reviews; however, Pfizer earned more positive perceptions than Sinovac. The best accuracy was achieved by SVM in both datasets.

Shahriar et al. [28] followed a different and more complicated approach compared to other studies, one that incorporated the four types of data analytics—descriptive, diagnostic, predictive, and prescriptive—to develop a data analytics-based framework to analyze people's perceptions of COVID-19 vaccination brands from a gender perspective.

The predictive analytics part was mainly based on deep learning models (RNN, CNN, GRU, BiLSTM, and LSTM). LSTM outperformed other deep learning models with 86% in precision, recall, and F1-score.

Another model was proposed by Alabrah et al. [29] to improve the accuracy of sentiment classification in the context of COVID-19 vaccine hesitancy among residents of gulf countries. The authors proposed feeding the sentiments extracted by three different methods (VADER, TextBlob, and Ratio) into an LSTM and four other machine learning algorithms (SVM, Fine-KNN, ensemble Boost, and Total Boost) separately. They also proposed passing the normalized features from the LSTM to the four machine learning algorithms. Compared to [28], the model proposed by [29] achieved a classification accuracy of 94.01% by both Fine-KNN and Ensemble Boost through the method incorporated using VADER and LSTM. On the other hand, a simpler architecture of deep learning implementation of LSTM with word embeddings was proposed by Khan et al. [30]. Their model, which was aimed at exploring people's perceptions regarding COVID-19 in Saudi Arabia, achieved a higher accuracy rate (95%) compared to [29]. A summary of the related work discussed above is provided in Table 1.

Table 1. Summary of Related Work

Reference	Technique	Accuracy	Language	Feature Extraction	No. Tweets	Virus
Salathé and Khandelwal (2011)	NB and Maximum Entropy	84.29%	English	N-gram	477,768	H1N1 vaccine
Du et al. (2017)	Hierarchical Classification (SVM)	F score: 78.6%	English	N-gram, POS, word clusters	184,214	HPV
Raghupathi et al. (2020)	VADER	-	English	TF-IDF Vectorizer	9,581	measles vaccination
Manguri (2020)	TextBlob	-	English	-	530,232	COVID-19
Hung et al. (2020)	LDA, VADER	Sen: 89.3% Spe: 77.3%	English	-	902,138	COVID-19
Alanezi and Hewahi (2020)	k-means, mini-batch k-means	-	English, Arabic	PCA	E: 23,490 A: 13,088	COVID-19
Samuel et al. (2020)	NB, LOGR	91%	English	N-gram	> 900	COVID- 19
Nurdeni et al. (2021)	NB, SVM, RF	85%	English	POS	3,242	COVID-19 vaccine
Praveen et al. (2021)	TextBlob, LDA	-	English	-	73,760	COVID-19 vaccine
Cotfas (2021)	MNB, RF, SVM, BiLSTM, CNN, BERT	78.94%	English	Bag-of-Words, word embeddings	2,349,659	COVID-19 vaccine
Alabrah et al. (2022)	LSTM, SVM, Fine-KNN, Ensemble Boost	94.61%	Arabic	Deep features	685	COVID-19 vaccine
Shahriar et al. (2022)	LSTM, RNN, CNN, BiLSTM, GRU	86%	English	Word2vec (Skip-gram)	59,750	COVID-19 vaccine
Khan et al. (2022)	SVM, NB, LOGR, LSTM	95%	Arabic	N-gram, TF-IDF, word embeddings (Fast Text Word2vec)	3,000	COVID-19 vaccine
Current study	SVM, NB, LOGR, RF, BiLSTM	91%	Arabic	TF-IDF Vectorizer, word embeddings	56,665	COVID-19 vaccine

3. Materials and Methods

We investigated public perceptions about the COVID-19 vaccine in Saudi Arabia during specific events by analyzing sentiments gleaned from Twitter data. Our method is described in the following steps (Figure 1).

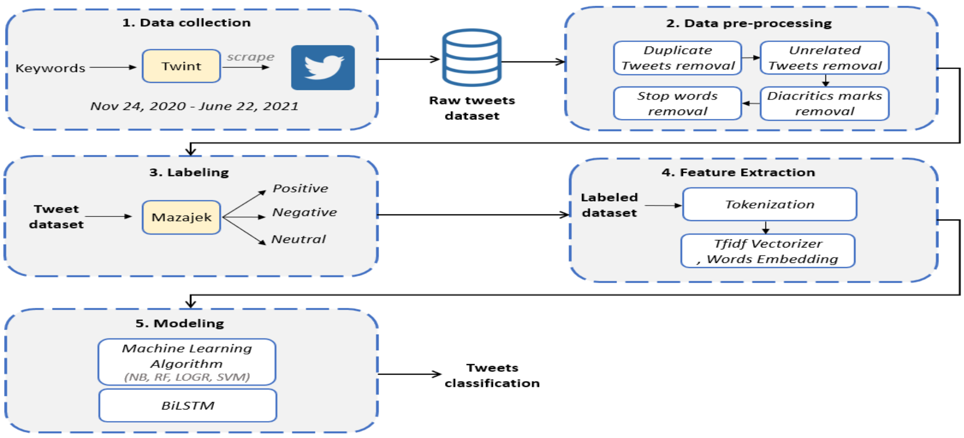

Figure 1. Method pipeline

3.1. Data Collection

The Python libraries Twint [31] and NLTK [32] were used to extract data from Twitter and to perform sentiment analysis, respectively. The Twitter API was not used since it has some limitations regarding the period and number of extracted tweets [33]. Data collection focused on collecting tweets generated during and around specific events between November 24, 2020, and June 22, 2021. The list of events included:

- lifting the international travel ban for Saudi citizens,
- resuming in-class school attendance starting September 2021,
- mandating vaccination for entry to any public or private establishment starting August 2021,
- Pfizer vaccine approval,
- AstraZeneca vaccine approval,
- appearance of India's COVID-19 variant, and
- appearance of Vietnam's COVID-19 variant.

The collected tweets needed to satisfy three criteria with respect to location, language, and date. The location was set to Saudi Arabia in the following format: (23.885942, 45.079162, 400 km). This represents Saudi Arabia's coordinates with a 400-km radius. The language was set to Arabic. The tweet dates needed to be within 14 days of the first day of the event (some events had overlapping dates). To extract all possible tweets related to an event, several keywords were used. Each search included a combination of keywords related to both vaccines, such as (لقاح، اللقاح، تطعيم), and to the event itself, such as (، تحصين، حضورياً، السفر اشتراط). Table 2 lists the search keywords, the collection period, and the number of extracted tweets. The search resulted in 92,998 tweets and 56,665 filtered tweets.

Table 2. Summary of Keywords and Collected Tweets per Event

Event	Period	Search keywords	Collected tweets	Filtered tweets
Pfizer vaccine approval	Dec. 10, 2020 – Dec. 23, 2020	لقاح / تطعيم / فايزر / جرعة	1,950	754

AstraZeneca vaccine approval	Jan. 18, 2021 – Jan. 31, 2021	لقاح / تطعيم / استرازينيكا / أكسفورد / لقاح + استرازينيكا / اعتماد + لقاح /اعتماد + استرازينيكا / السعودية + استرازينيكا	533	441
Resuming in-class attendance	May 10, 2021 – May 24, 2021	حضوريا + لقاح / لقاح + حضوريا / حضوري + تطعيم / حضورياً + تطعيم / لقاح + الدراسة / لقاح + كورونا + دراسة / لقاح + كورونا / السعودية+ عودة+ كورنا / لقاح + عودة + الطلاب / لقاح + عودة + الدراسة / مدرسين + تطعيم / لقاح كورونا/ لقاح + الدراسة	2,313	520
Lifting the ban on international flights	May 17, 2021 – May 31, 2021	جرعة + السفر/ السفر+ لقاح / لقاح + السماح /كورونا + السفر+ لقاح/ لقاح + عمرة +العمرة / السعودية + العمرة	1,076	742
Mandating the vaccine to enter any public or private establishment	May 18, 2021 – June 1, 2021	التحصين + شرط / اشتراط + تحصين / لقاح + تحصين /لقاح + شرط	1,278	269
Appearance of India's COVID-19 variant	March 25, 2021 – April 8, 2021	"المتحور+ الهندي + لقاح / المتحور + الهندي + تطعيم/ المتحور + الهندي + جرعة / المتحور + لقاح / لقاح	521	300
Appearance of Vietnam's COVID-19 variant	June 4, 2021 – June 18, 2021	المتحور + الفيتنامي + لقاح / المتحور + الفيتنامي + جرعة / المتحور + الفيتنامي + تطعيم / المتحور+ لقاح / الفيتنامي + جرعة + المتحور / لقاح	1,170	344
General search keywords	May 10, 2021 – June 1, 2021	توكلنا / تطعيم / لقاح / تطعيم + السعودية	84,157	53,295
	Total		92,998	56,665

Although many authors have conducted studies and analyses of public sentiments regarding COVID-19 vaccine hesitancy and anti-vaccine attitudes on social media, especially on Twitter, the results cannot be generalized across countries. This is attributable to some potential factors observed in previous research, such as cultural differences, political issues, trust relationships, and detailed information about the vaccine reported by the government and health sector [22]. This study aimed to classify polarity on Twitter regarding COVID-19 vaccines as positive, negative, or neutral. This approach has previously achieved promising results and is commonly used to classify sentiments on social media [16,19,22,23,25].

3.2. Data Preprocessing

Arabic citizens exhibit cultural and linguistic differences across regions. Thus, data pre-processing steps are essential when dealing with the Arabic language because Twitter users use informal written language, known as a dialect, that must be normalized. Furthermore, it is crucial to handle Twitter features, such as tweet size and content diversity, that may consist of texts, symbols, emojis, URLs, images, and videos. Thus, we applied the following pre-processing steps to the tweets before feeding them into the classifier.

First, duplicate tweets were removed from the dataset. Second, unrelated tweets, such as news tweets and tweets that did not contain an opinion, were identified and filtered out. Two features were inspected to determine unrelated tweets: tweet handle and tweet content.

- All tweets posted by a Twitter news handle were removed. The following keywords were used to identify Twitter news handles: منوعات, تعليم, مبادرة, وظيفة, مجمع, الأن, واتس, مشاهد, موجز, تحديث, وسم, الكرة, سناب, محافظة, إمارة, شركة, عكاظ, الاهرام, هاشتاق, اليوم, الجزائرية, إجازة, اجازة, المغربية, معهد, إدارة, رتويت, عناوين, ,جامعة, مدرسة, حقائق, دليل, الوطن, اون لاين, محتوى, خطابة, اليمنية, السورية, الأردنية, الكويتية, الامارتية, العربية, الدولية, الخليجية, شبكة, اونلاين, نيوز, صحيفة, جريدة, قناة, مجلة, جوال, القاهرة, الالكترونية, مواعيد, اخبار, أخبار, حجز, حجوزات. This set of keywords included known Twitter news accounts. In addition, it included keywords such as "news," "channel," and "online," which are known to be used by Twitter accounts that broadcast news (Figure 2).

- Tweets posted by non-Saudi users or residents were removed from the dataset. Such tweets were identified by the information Twitter users attached to their tweets, such as a flag emoji or an abbreviation (e.g., q8, Q8, QA, KW, Kuwait, Kuwaiti, uae, AE, leban, syri, qata, qtri, lib, yem, oman, egy, Egyp, EG, BH, Iraq, PS), in the Twitter handle that showed the country with which they identified (Figure 3). However, before such tweets were deleted, we checked that they did not discuss COVID-19 vaccines in Saudi Arabia. To examine this, we used the following set of keywords: ksa, Saudi, Tawakkalna, and المملكة, السعودية, توكلنا. Tweets with any of these words were not removed from the dataset, as shown in Figure 4.

Second, all punctuation and emojis were removed from the tweets using the repressor library in Python. Because this study aimed to extract only people's opinions from the tweets, we had to ensure that the tweets were free from

Figure 2. Example of tweets posted by Twitter news handles.

URLs and non-text elements, such as images and videos. Since the study was concerned with the Arabic language, several challenges related to its rules and grammar affected the interpretation of the word meanings. We normalized the Arabic text by replacing characters, such as replacing إ,أ,آ with ا, ي with ى, ة with ه, and ؤ with و.

Third, to avoid errors that may have affected classification accuracy, diacritical marks (Tashkeel) were removed from all collected tweets. Moreover, Tatweel, which extends the length of the line that connects letter pairs in a word, was removed. For instance, the extension (ـــــ) was removed from the word (لقـــــاح) to return it to its original form (لقاح). This was important for avoiding the creation of different forms of the same word.

Figure 3. Example of tweets by non-Saudi users that were removed.

Finally, sentence tokenization was applied to break each sentence into multiple words, making it easier to analyze.

Figure 4. Example of tweets by non-Saudi users with @SaudiNews50 tagged.

3.3. Labeling

Model training required a labeled dataset (with respect to sentiment). Due to the large size of the dataset, annotation was performed using Mazajak [34], which is a deep learning sentiment analysis model for Arabic text. Mazajak analyzed each tweet and assigned one of three sentiments (positive, negative, or neutral) as follows:

- Positive tweets contain positive opinions regarding the vaccine—for example, sharing news of becoming vaccinated or encouraging others to do so. The following tweet was assigned a positive sentiment: " تم بحمدالله وفضله ثم فضل هذا البلد العظيم اخذ لقاح كورونا" [Thanks to God then to this great country, I took the corona vaccine].
- Negative tweets contain negative opinions regarding the vaccine—for example, by explicitly expressing that the user does not want to be vaccinated, asking other people to remain unvaccinated, or sharing rumors to spread fear about the vaccine. The following tweet was

assigned a negative sentiment: “لقاح انا ما احتاجه و مقتنعه اني ما احتاجه فهذي ابسط حقوقي ك انسان اللي مقتنع باللقاح ويبغي ياخذه هذا حقه و اللي غير مقتنع ويرفض اللقاح برضو هذا حقه” [A vaccine that I do not need, and I am convinced that I do not need it. This is the simplest of my rights as a human being. The person who is convinced of the vaccine and wants to take it, this is his right, and the one who is not convinced and refuses the vaccine also, that is also his right].

- Neutral tweets lack a positive or negative opinion toward the vaccine. These tweets usually are in the form of a question. The following tweet was assigned a neutral sentiment: “هل شرط حضور المباراه التحصين ام يكتفي بالجرعه الواحده” [Do I need to be immunized to attend the game or is one dose enough?].

3.4. Modeling

This study experimented with both classical machine learning models and a deep learning model. Sections 1 and 2 provide an overview of the selected algorithms.

3.4.1. Applied Machine Learning Algorithms

Four supervised machine learning models were used and compared in the analysis: SVM, NB, LOGR, and RF. Both the SVM and NB models have been demonstrated to achieve superior results in sentiment analysis in different languages [35–37]

3.4.1.1. Support Vector Machine (SVM)

The SVM classifier has been widely used to classify text because of its simplicity and accuracy [38]. The SVM model requires a trained, labeled dataset based on a specific set of features extracted from the text.

3.4.1.2. Naïve Bayes (NB)

The NB model is a popular machine learning model in text content classification. NB has recently been applied in different domains, including marketing and entertainment [20,39–43]. Two types of NB algorithms are provided by the Scikit-learn library [44]: binary NB and multinomial NB. The binary, or Bernoulli, NB is used to assign the text to one of two classes of sentiments. The multinominal NB is used to iteratively count the number of times a word occurs in a document [45].

3.4.1.3. Logistic Regression (LOGR)

Logistic regression (LOGR) determines class probabilities by generating variables that contain natural logs of the odds of classes occurring. Then, it applies a maximum likelihood estimation algorithm to estimate the probabilities. [46].

3.4.1.4. Random Forest (RF)

Random Forest (RF) is an ensemble method made up of a combination of decision trees, called estimators. Each tree votes for its preferred class, after which the final prediction is given based on the most voted class [46].

3.4.2. Applied Deep Learning Model

One of the widely used methods in NLP is recurrent neural networks (RNNs). RNNs are employed in tasks such as text classification and document-level sentiment classification, for which they must handle variable sequential input length [47]. The limitation of RNNs is that they use either “sigma” cells or “tanh” cells, which fail to learn relevant information when the input gap is large. An enhanced type of RNN was introduced to overcome this problem: long short-term memory (LSTM). The structure of LSTM is aimed at extending the memory of RNNs and handling the issue of long-term dependencies by adding three gates: the input gate, which allows or blocks the update on input; the forget gate, which disables neurons found to be unimportant based on the learned weight by algorithm; and the output gate, which controls the neuron sate on the output [48,49]. LSTM has become one of the state-of-the-art models for handling wide machine learning problems [50].

A more advanced variant of LSTM is bidirectional LSTM (BiLSTM). BiLSTM consists of two LSTM cells, forward and backward LSTM. BiLSTM considers all of the previous and upcoming words to extract the sentiment—unlike LSTM, which has a one-directional structure that only considers past words to extract sentiments from the text [51]. In our study, we chose to use a deep learning model based on BiLSTM since it is the most advanced variant of LSTM and compared the results with our four selected machine learning algorithms.

3.5. Synthetic Minority Oversampling Technique

The synthetic minority oversampling technique (SMOTE) [52] was designed to handle imbalanced datasets by oversampling the minority classes in the training set. SMOTE generates synthetic samples based on the “feature space” similarity between existing minority samples without changing the size of the majority classes [53]. SMOTE considers the “feature space” instead of the “data space” to introduce synthetic instances by randomly selecting a minority class and then interpolating its k-nearest neighbors (KNN). New instances are generated by the KNN within the dataset by considering other instances near them. Imbalanced data can affect the training process of the machine learning models; as a result, imbalanced data introduces biases in the class prediction. According to [54], it is essential to handle the issue of imbalanced data as a preliminary step before the classification. In this study, we used SMOTE because the classes in our dataset were imbalanced. The class distribution was 10% positive, 36% neutral, and 54% negative. The implementation of SMOTE is provided as a class by an open-source "Imbalanced-learn" library, which is an MIT-licensed library relying on Scikit-learn [55].

3.6. Experiment Setup and Environment

The full implementation of both machine learning and deep learning models was accomplished using the Python language in Google Colaboratory (Colab). Colab is a cloud service based on the Jupyter notebook environment. The GPU was utilized to accelerate the training process [56]. First, the preprocessed data were split into 80:20 training and testing, respectively. Then, feature extraction was applied using the TF-IDF Vectorizer from the Scikit-learn (Sklearn) library for the machine learning part. The TF-IDF Vectorizer is equivalent to applying a count vectorizer followed by a TF-IDF transformer. Its primary function is to convert text into a matrix of TF-IDF features. TF refers to "term frequency," while "IDF" refers to inverse document frequency [57]. TF provides the word frequency in the corpus and represents the ratio of word occurrences in the document compared to the total number of words in the document. IDF measures the weight of the rare words across the corpus. Words that infrequently occur in the corpus have higher IDF scores. The product of TF and IDF gives the TF-IDF score of the words in the corpus [23]. Then, the performance of all machine learning classifiers can be validated using the Stratified KFold method with K = 10. The Stratified KFold method, provided by the Sklearn library, was selected among different cross-validation methods. The main concept of KFold cross-validation is to split the training set into KFolds, with the training performed on the K-1 fold, and the evaluation completed on the remaining 1-fold [58]. Stratified KFold was chosen because it guarantees that each training and testing set contains a representative ratio of each class (positive, negative, neutral) to avoid any possible estimation bias [59].

Before building the deep learning model, a set of hyperparameters was defined. We set the out-of-vocab token (oov _token) = '<oov> ', an argument provided by Keras API to substitute any unknown word in the vocabulary with a word of our choice. The main benefit of the out-of-vocab token is that it tells the model that information is present in the place of '<oov>'. Also, the maximum number of words = 5,000, representing the most common 5,000 words from our tokenized corpus, in addition to pad type = 'post' and truncating type = 'post'. The preprocessed data were tokenized using Keras API's "Tokenizer" class. The "Tokenizer" vectorizes the text into a sequence of integers, after which the sequence is split into a list of tokens. Using the "fit_on_texts" method, which updates the vocabulary based on the provided set, we created a training sequence that contained tokens from the training set and a test sequence that contained tokens from the testing set. Then, with the "texts_to_sequences" method, the text was converted into a list of integers, after which we applied the "pad_ sequences" function, which transformed the list of sequences into a 2D Numpy array. The function assumed several parameters: "maxlen," "padding," and "truncating." The maximum length (maxlen) = 100 limits, the maximum length of all sequences, while padding = "post" defines where to add the pad before or after the sequence in case the sequence is less than the defined length. Truncating = "post" is useful when a text is longer than the defined maximum length; it removes the extras from the beginning or end based on the selected truncating type.

The BiLSTM model is a sequential model built using Keras API. The model consists of seven layers: an embedding layer, a BiLSTM layer, two dense layers, and three dropout layers between every two layers (Figure 5). The dropout layer value is set to 0.3, where dropout is used as a regularization method to avoid overfitting [60]. The model layers employed were as follows: the first layer was the embedding layer, which took two parameters, vocab_size = 5,000 and embed_dim = 50, followed by the dropout layer. Then, a BiLSTM layer was followed by a dropout layer. Next, a dense layer with a "Relu" activation function was followed by a dropout layer. Finally, the output layer contained three neurons representing the number of classes in our dataset (positive, negative, and neutral) with a "softmax" activation function. The "softmax" activation function, used for multiclass classification problems, returned the probability for each data point of all target classes [61]. Finally, the model was compiled with a "categorical cross-entropy" loss function, in addition to using an "Adam" optimizer and an "accuracy" metric. The training was set for 10 epochs, a batch size of 32, and a validation split of 0.1 to monitor the model's performance during the training.

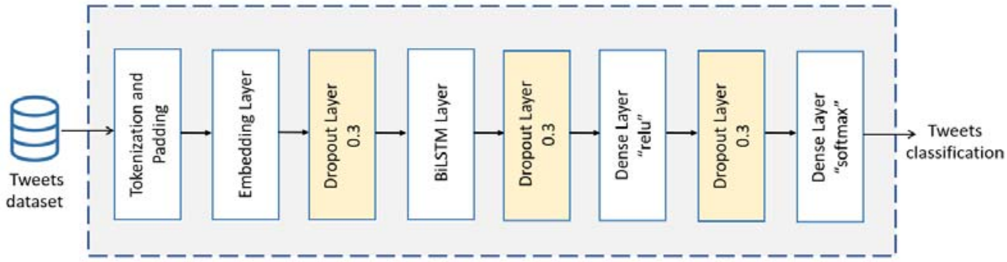

Figure 5. BiLSTM architecture.

3.7. Model Validation and Evaluation

We validated the performance of each of the machine learning models used in this study using a set of well-known performance measures and a confusion matrix. Our goal was to determine the model that performed better in classifying the sentiments contained in Arabic tweets.

3.7.1. Performance Measurements

It is essential to report how well the predictors can make correct predictions because doing so means that the results can be trusted. A set of performance measures is used to evaluate machine learning models [62]. To assess the performance of each classifier of a tweet sentiment, four metrics were used: precision, recall, F1-score, and accuracy. The first three measures were based on four prediction probabilities: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Table 3). TP represents the number of positive tweets correctly classified as positives. TN counts the number of tweets that were classified correctly as negative. FP refers to the number of negative tweets that were incorrectly classified as positive, while FN represents the number of positive tweets that were incorrectly classified as negative.

Table 3. Confusion Matrix Parameters

		Predicted	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Precision is the number of correctly identified positives across all classes compared to the total number of positive predictions for all classes. It is calculated as shown in (1):

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (1)$$

Recall shows the frequency of correctly predicted positives from actual positive samples, which measures the ability of a predictor to identify positive samples. It is calculated as shown in (2):

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (2)$$

The F1-score combines precision and recall into a single measure as a weighted average. It is calculated as shown in (3):

$$\text{F1} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (3)$$

The accuracy metric measures the ability of a classifier to correctly identify all samples. It is calculated as shown in (4):

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (4)$$

Note that accuracy does not consider the possible error types of classification that may occur in the models [63]. The performance of the "Mazajek" sentiment analysis was validated by taking a random sample of 300 tweets and annotating them manually by the authors as positive, negative, and neutral to check against the annotation of Mazajek. Then, the precision, recall, and F1-score were calculated to evaluate the quality of the annotation done by Mazajek.

3.7.2. Confusion Matrix

The confusion matrix is used to validate the classification performance of machine learning models by comparing actual and predicted results obtained by the classifiers used [40].

3.7.3. Threats to Validity

The validity of research includes two domains: internal and external validity. The validity of research implies to what extent the study participants' results are accurate with respect to those of similar individuals outside the study [64]. The applied method is not limited to detecting COVID-19 vaccination sentiment analysis, as it can also be used to detect sentiments concerning any similar cases in the future that require polarity detection in any area. However, several factors can threaten the internal validity of a study. People's opinions, for instance, can change over time. Also, social interaction, the core idea of social network platforms, can influence people's views concerning, for example, vaccines. Certain people, such as social media influencers, might affect people's acceptance and hesitancy about vaccines. The main threat to external validity in the current study was that the opinions expressed through the Twitter platform may not fully represent the opinions of the Saudi population. However, in another study [65], which measured the acceptance of COVID-19 vaccination among the Saudi population via an online cross-sectional survey on individuals aged 18 years and above, the authors found that 52% were uncertain or had not reported any intention to be vaccinated, whereas 48% were willing to be vaccinated. Their results nearly matched our results. In our study, 54% showed negative sentiment, 36% were neutral, and 10% were positive. Both the cross-sectional questionnaire and the Twitter sample yielded close results.

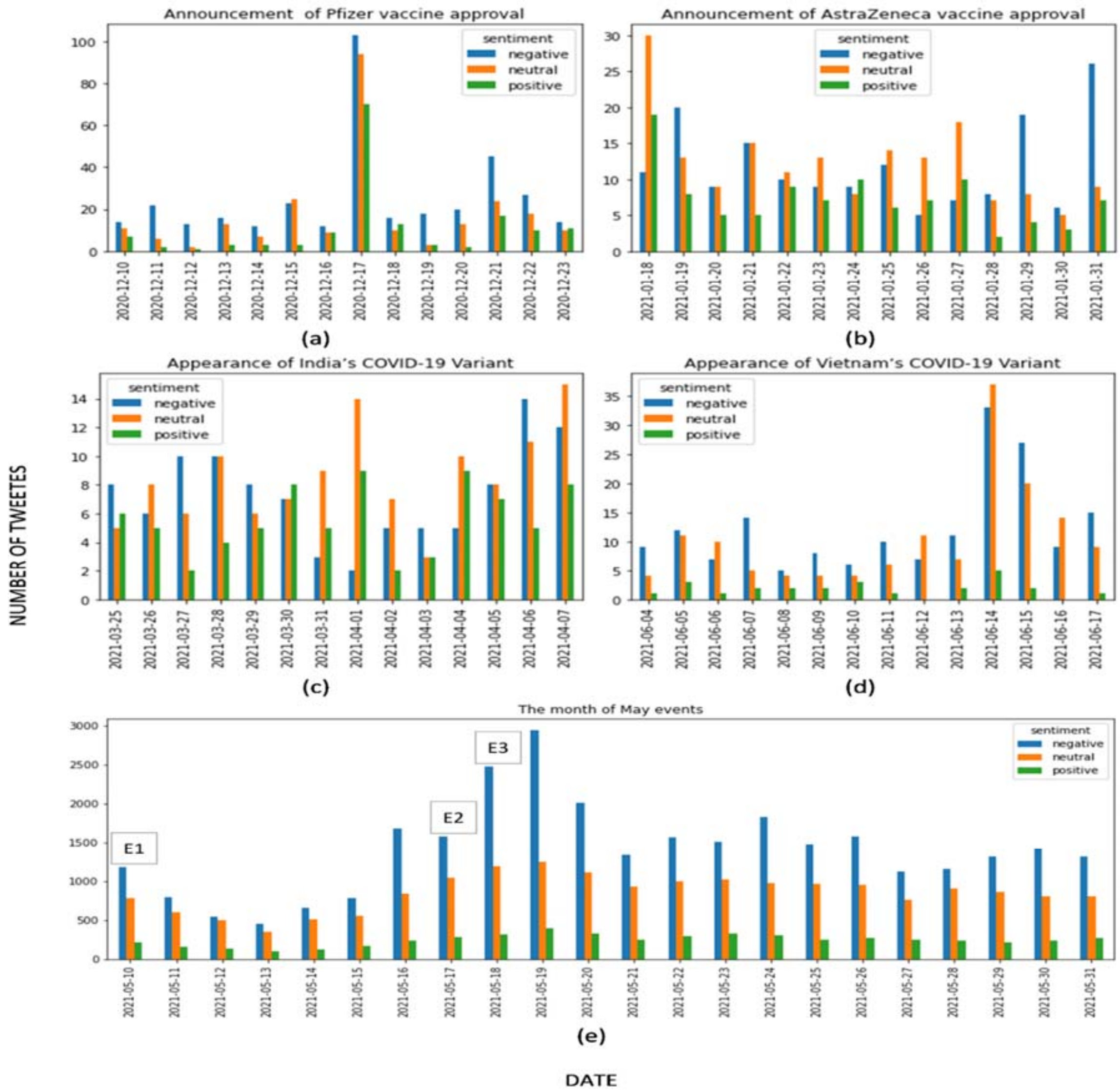

Figure 6. Number and sentiment of tweets posted during the selected events.

4. Results

Our goal was to apply machine learning and deep learning models to identify and classify sentiments about the COVID-19 vaccine in Saudi Arabia from a collected tweet dataset. We also sought to identify which model demonstrated the best performance. The dataset polarity showed that 54% of the tweets were assigned a negative sentiment, 36% were given a neutral sentiment, and 10% were given a positive sentiment. This result supports what was reported in [65]. The per-event sentiment analysis, however, was inconsistent. Next, we analyzed the sentiments contained in tweets posted during each event (Figure 6).

4.1. Tweet Sentiment Analysis

4.1.1. Announcement of Pfizer Vaccine Approval

On December 10, 2020, the Saudi Ministry of Health [66] approved the Pfizer vaccine (the first approved in Saudi Arabia). The dates on which the vaccine would become available were announced on the same day. Figure 6 shows the distribution of tweets and their sentiments during two weeks in December. Figure 6(a) shows a gradual increase in the number of tweets regarding the Pfizer vaccine from December 10 until December 17, when the number of tweets reached its peak. The number of tweets between the 10th and the 16th of December did not exceed 30 tweets per class; while on the 17th, the number of tweets was between 65–100 per class (positive, negative, neutral). This rapid increase can be explained by the fact that the Saudi Minister of Health (Dr. Tawfiq Al-Rabiah) received the first dose of the vaccine in one of the vaccination centers in Riyadh and started the vaccination campaign around the kingdom [67].

4.1.2. Announcement of AstraZeneca Vaccine Approval

On January 18, 2021, the Ministry of Health in Saudi Arabia announced its approval of AstraZeneca's vaccine (the second approved in Saudi Arabia) [68]. Figure 6(b) shows the distribution of tweets and their sentiments posted during two weeks in January. Figure 6(b) shows a fluctuation in the number of tweets regarding the AstraZeneca vaccine. The overall sentiment was negative for several days. However, neutral sentiment can be noticed on the first day of the announcement of the AstraZeneca vaccine, with a total of 30 neutral tweets, 12 negatives, and 18 positives. The number of neutral tweets decreased over the following days, but a clear pattern could not be detected. Compared to the tweets posted regarding the Pfizer vaccine, a less negative reaction can be observed.

4.1.3. Appearance of India's Covid-19 Variant

On March 25, 2021, a new variant (known as the Indian COVID-19 variant) affected multiple countries worldwide [69]. This event became one of the trending topics discussed on Twitter, along with vaccine effectiveness against this variant. Figure 6(c) shows the number of tweets and their polarity over the two weeks following this announcement. The number of positive tweets fluctuated during this time frame. However, the number of negative tweets continued to increase. The set of tweets with a neutral sentiment was dominant during this event.

4.1.4. Appearance of Vietnam's Covid-19 Variant

On June 4, 2021, Vietnam reported a new coronavirus variant, a hybrid of the British and Indian variants. According to Figure 6(d), the number of positive tweets was consistently low during this period, when positives did not exceed five tweets per day. Also, there was an increase in negative tweets during the last few days of the timeframe, between the 14th and 17th. A slight difference can be observed between the number of negative and neutral tweets. A potential reason for less interaction with this event is that this variant did not affect large areas of the world [70] .

4.1.5. Resumption of In-class School Attendance

On May 10, 2021, the Saudi Ministry of Education [71] announced the re-opening of schools and the resumption of in-class attendance for students by the next academic year (fall semester of 2021) [72]. During the pandemic, the Kingdom of Saudi Arabia adopted online education at all levels to reduce the risk of spreading the virus. Figure 6(e) shows the distribution of opinions regarding this event. More than 1,000 tweets were posted about this event on announcement day, which is referred to in Figure 6(e) as (E1). This number decreased gradually over the next few days. The number of negative tweets was always higher than the number of neutral and positive tweets.

4.1.6. Lifting the Temporary Travel Ban and Resuming All International Flights

On May 17, 2021, the Saudi Ministry of the Interior announced the lifting of the travel ban on Saudi citizens and residents, as shown in (E2) in Figure 6(e) [73]. Figure 6(e) shows an increase in the number of tweets that started on May 14 (three days before the official announcement), with more than 500 negative tweets. Another sharp increase occurred on May 16 (one day before the official announcement), with more than 1,500 negative tweets. This indicates that some individuals already knew about this announcement through rumors. Generally, the number of tweets with a negative sentiment was higher than the number of neutral and positive tweets.

4.1.7. Mandating the Vaccine to Enter Any Public or Private Establishment Starting August 2021

On May 18, 2021, the Saudi Ministry of the Interior announced that vaccination would be mandatory in order to enter any government, private, or educational institution. In addition, it also announced that the vaccine would be mandatory for using public transportation starting August 1, 2021 [74,75]. According to Figure 6(e), this event (E3) caused the highest number of posted tweets within the first two days compared to the other events. On the 18th, the number of negative tweets reached 2,500, and the following day, it increased to 3,000, indicating a strong reaction to the decision. However, after the second day of the announcement, the reaction started to decrease, with negative tweets dropping to 2,000 but constantly fluctuating over the 14 days. This fluctuation was more apparent for tweets with a negative sentiment, while neutral and positive tweets showed slight changes in their numbers.

In summary, the event that had the most significant impact based on the number of tweets among the seven events was the announcement of mandatory vaccination. The overall sentiment toward this event was mainly negative. However, most of the remaining tweets showed a neutral sentiment, and a small set showed a positive sentiment. Furthermore, general opinions toward some similar events changed over time. For instance, even though the announcements of Pfizer's vaccine approval and AstraZeneca's vaccine approval were one month apart, a clear pattern in sentiments can be identified during the former but not the latter.

4.2. Performance Analysis of Machine Learning and Deep Learning Algorithms

In Table 4, we have compared the results obtained from training classical machine learning algorithms (RF, LOGR, SVM, and NB) with the BiLSTM. Among the classical machine learning algorithms, both LOGR and SVM had an accuracy of 82% regarding tweet data. Furthermore, their performance among the three classes (negative, neutral, and positive) demonstrated similar performance. On the other hand, NB and RF had lower accuracy than LOGR and SVM: 73% and 78%, respectively. Moreover, BiLSTM exhibited improvement in accuracy compared to the classical machine learning algorithms, at 83%, in addition to robust results in term of precision, recall, and F1-score in all classes for which it was noted that the performance of the machine learning models with the minority class (positive sentiment) had a recall between 3%–46% and an F1-score between 6%–84%. Meanwhile, BiLSTM demonstrated a recall of 64% and an F1-score of 68% for the minority class, which shows that BiLSTM is capable of learning the features of all different classes. Figures 7–11 show the confusion matrices of the machine learning models and BiLSTM that best illustrate the percentage of correctly classified and misclassified samples in each class. BiLSTM correctly classified 56% of positive, 85% of neutral, and 87% of negative sentiments.

Table 4. Classifier Performance

Classifier	Class	Precision	Recall	F1-score	Accuracy
RF	negative	0.77	0.92	0.84	0.78
	neutral	0.79	0.69	0.74	
	positive	0.87	0.33	0.84	
	weighted avg	0.79	0.78	0.77	
LOGR	negative	0.80	0.96	0.87	0.82
	neutral	0.87	0.70	0.78	
	positive	0.85	0.46	0.60	
	weighted avg	0.83	0.82	0.81	
SVM	negative	0.80	0.96	0.87	0.82
	neutral	0.86	0.71	0.78	
	positive	0.87	0.45	0.59	
	weighted avg	0.83	0.82	0.81	

	negative	0.70	0.97	0.81	
NB	neutral	0.83	0.55	0.66	0.73
	positive	0.97	0.03	0.06	
	weighted avg	0.77	0.73	0.68	
	negative	0.87	0.87	0.87	
BiLSTM	neutral	0.80	0.82	0.81	0.83
	positive	0.73	0.64	0.68	
	weighted avg	0.83	0.83	0.83	

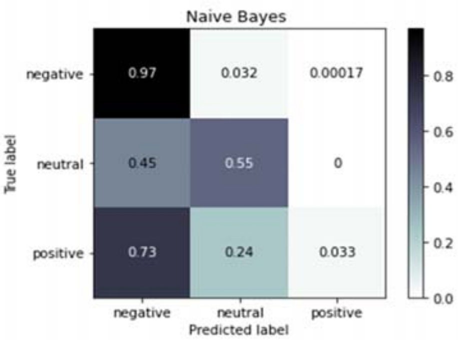

Figure 7. Naïve Bayes (NB) confusion matrix.

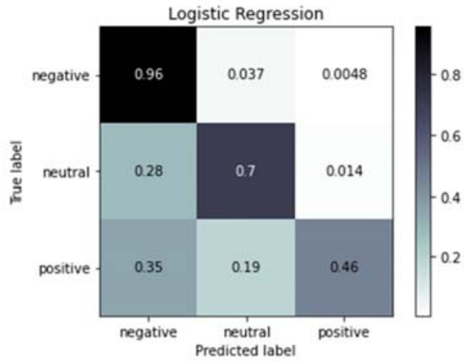

Figure 8. Logistic regression (LOGR) confusion matrix.

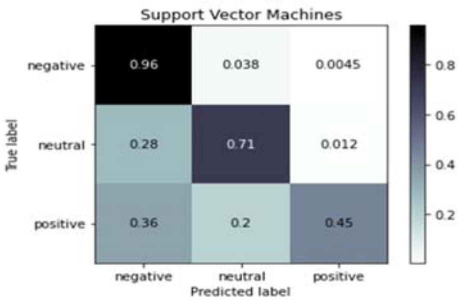

Figure 9. Support vector machine (SVM) confusion matrix.

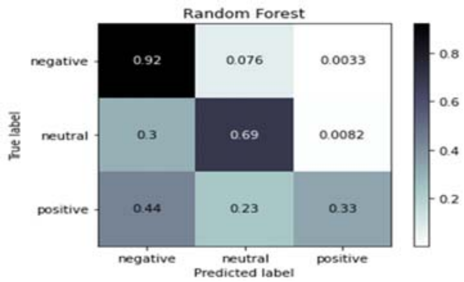

Figure 10. Random forest (RF) confusion matrix.

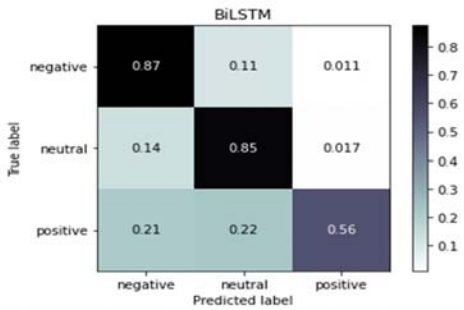

Figure 11. BiLSTM confusion matrix.

The learning curves in Figures 12 and 13 show the training and validation scores of BiLSTM for every epoch. As the number of epochs increased, with every iteration, the training and validation scores started to converge, with the training and validation accuracy increasing to 83% (Figure 12), and the loss\error decreasing to the minimum value of 43% (Figure 13).

Additionally, we experimented with the effect of applying the SMOTE technique, oversampling the minority class, on the dataset since it was imbalanced and discovered its effect on the performance of the machine learning algorithms. Also, we validated the performance of the machine learning algorithms by applying a 10-fold stratified cross-validation method on the data before and after oversampling (Table 5). According to Table 5, in comparison with the weighted average results of all measures presented earlier in Table 4, we observed a slight improvement—a 1% increase—in recall, F1-score, and accuracy, except for SVM, for which the accuracy remained the same. On the other hand, after applying oversampling to handle the imbalanced data issue, the results significantly improved in terms of all measures. Providing a balanced dataset to the machine training algorithms boosted performance. Among the machine learning algorithms, SVM achieved the best results, scoring 91% on accuracy, recall, and F1-score, and 92% on precision.

The results of comparing the Mazajek annotation against manual annotation resulted in a precision of 79.7%, a recall of 74%, and an F1-score of 73.8%. Mazajek showed an acceptable performance above 70% in precision, recall, and F1-score, making it suitable for annotation tasks in sentiment analysis.

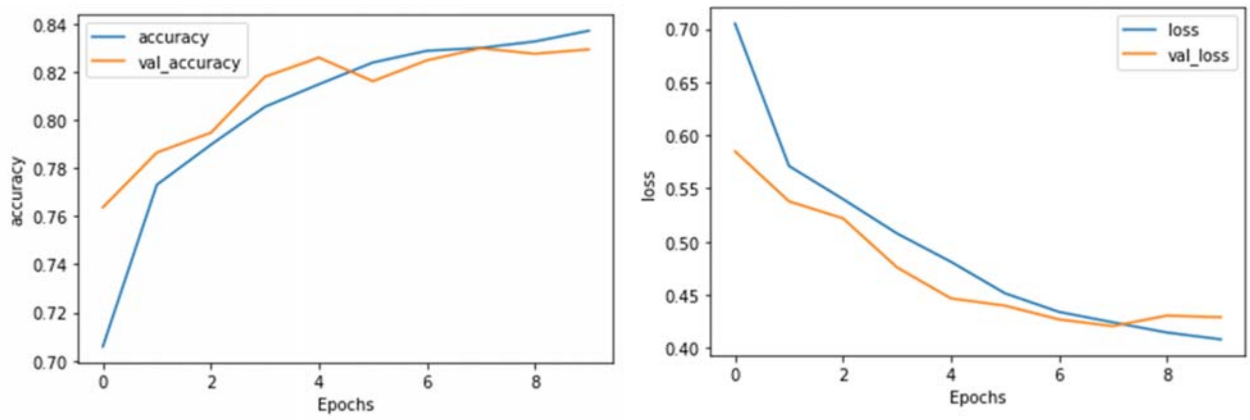

Figure 12. BiLSTM learning curve.

5. Discussion

Handling unbalanced data was addressed differently in [26]. In this work, the authors chose to select an equal number of samples per month. However, opinions toward events can be expected to vary over time, and the number of samples reflects people's reactions toward that event. So, following this method—i.e., taking equal samples—was not suitable for our case, as our goal was to analyze people's reaction to COVID-19 vaccination during certain events and report it as is, which eventually yielded insights into which events influenced people the most. In our study, we chose to use SMOTE, which allowed us to use all of our collected data while handling the issue of imbalanced classes to avoid bias in the training dataset, which might have led us to ignore the minority class and could thus have affected the predictions made via the machine learning algorithms. Significant improvements were achieved after resolving these imbalances, with the accuracy of the SVM increasing from 82% to 91%; similar improvements in precision, recall, and F1-score were also observed.

The comparison we provided in Table 4 between classical machine learning algorithms and BiLSTM demonstrated that BiLSTM performed well in terms of learning and differentiating between classes compared to the classical machine learning algorithms without applying oversampling. BiLSTM provided promising results regarding sentiment analysis. Its remarkable performance is attributable to its architecture—similar to [76], in which remarkable performance was observed when LSTM was compared to BiLSTM. BiLSTM yielded a better recall score for the minority class (positive) than all of the other machine learning algorithms. The recall score in the minority class generated by BiLSTM was 64%; while in the machine learning algorithms, the recall score was between 3%–46% maximum.

Both of the most recent studies [29,30] to discuss COVID-19 vaccination with Arabic tweets suffered from the dataset size limitation, as both had a size between 600 to 3,000. In our study, we addressed this limitation by collecting a larger Arabic dataset of 56,666 tweets.

Although [28] tested multiple deep learning techniques, we noticed that their BiLSTM results were slightly higher—but close—to our results, achieving 85% accuracy, whereas our BiLSTM model achieved 83% accuracy. The difference between their method and ours is that they used word2vec for feature extraction, while we created our list from word embedding.

A variety of techniques have been proposed by authors of related works, with either machine or deep learning techniques, or sometimes a combination of both. The best-obtained results, however, cannot be generalized based on one technique. For instance, the two most recent studies focused on individuals' perceptions about COVID-19 vaccination with Arabic tweets achieved competitive results, with Alabrah et al. [29] scoring 94.61% with Fine-KNN and Ensemble boost, and Khan et al. [30] achieving an accuracy of 95% with LSTM. In comparison, our study demonstrated a great capability to improve the results of machine learning algorithms by handling imbalances, with the SVM achieving an accuracy of 91%, well within high range performance.

Table 5. Comparison of the Performance of the Machine Learning Algorithms with and without Oversampling Technique with 10-fold Stratified Cross-validation

Classifier	**Metrics**	**10-Fold Without Oversampling**	**10-Fold Oversampling**
RF	Precision	0.79	0.89
	Recall	0.79	0.89
	F1-score	0.78	0.89
	Accuracy	0.79	0.89
LOGR	Precision	0.83	0.90
	Recall	0.83	0.90
	F1-score	0.82	0.89
	Accuracy	0.83	0.90
SVM	Precision	0.83	0.92
	Recall	0.82	0.91
	F1-score	0.82	0.91
	Accuracy	0.82	0.91
NB	Precision	0.77	0.83
	Recall	0.74	0.83
	F1-score	0.69	0.83
	Accuracy	0.74	0.83

6. Conclusions

The present study used sentiment analysis of Twitter data related to the COVID-19 vaccine to explore reactions during events in Saudi Arabia. The data classification revealed that a majority (54%) of public perceptions about the COVID-19 vaccine were negative, while 36% were neutral and 10% were positive. Four classical machine learning classifiers were used in addition to the deep learning method using BiLSTM, and the best accuracy was obtained with the SVM classifier, whose accuracy score was 91%. Regardless of the results, the Twitter community does not represent the general population; thus, the opinions expressed on the platform do not fully represent those of the general public in Saudi Arabia. However, [24] found a correlation between the sentiments expressed on social media and the vaccination rates provided by the CDC, suggesting that these data may help generate better understanding of public opinions in Saudi Arabia. In addition, a recent study by Alfageeh et al. [65] measured the acceptance among the Saudi population of the COVID-19 vaccine by distributing an online survey to adult participants. They found that 48% of respondents were willing to be vaccinated, while 52% were not.

In future work, negative tweets can be used to identify emotions associated with user perceptions of the vaccine. These data could be collected over consecutive days to study how opinions change in terms of the three categories over time.

References

[1] Cascella, M.; Rajnik, M.; Aleem, A.; Dulebohn, S.C.; Di Napoli, R. Features, evaluation, and treatment of coronavirus (COVID-19). In *StatPearls*; StatPearls Publishing: Treasure Island (FL), US, 2022.

[2] Zu, Z.Y.; Jiang, M.D.; Xu, P.P.; Chen, W.; Ni, Q.Q.; Lu, G.M.; Zhang, L.J. Coronavirus disease 2019 (COVID-19): A perspective from China. *Radiology* **2020**, *296*, E15–E25, doi:10.1148/radiol.2020200490.

[3] Güner, R.; Hasanoğlu, İ.; Aktaş, F. COVID-19: Prevention and control measures in community. *Turk J Med Sci* **2020**, *50*, 571–577, doi:10.3906/sag-2004-146.

[4] Ebrahim, S.H.; Ahmed, Q.A.; Gozzer, E.; Schlagenhauf, P.; Memish, Z.A. Covid-19 and community mitigation strategies in a pandemic. *BMJ* **2020**, *368*, m1066, doi:10.1136/bmj.m1066.

[5] Lester, G.V.; Brock Baskin, M.E.; Clinton, M.S. Employer-sponsored benefits in the United States: The past, present, and future. *Compensation & Benefits Review* **2021**, *53*, 24–42, doi:10.1177/0886368720947609.

[6] Cajner, T.; Figura, A.; Price, B.M.; Ratner, D.; Weingarden, A. Reconciling unemployment claims with job losses in the first months of the Covid-19 Crisis. *Board of Governors of the Federal Reserve System* 2020, doi:10.17016/FEDS.2020.055.

[7] Alliheibi, F.M.; Omar, A.; Al-Horais, N. Opinion mining of Saudi responses to COVID-19 vaccines on Twitter. *IJACSA* **2021**, *12*, doi:10.14569/IJACSA.2021.0120610.

[8] Alliheibi, F.M.; Omar, A.; Al-Horais, N. Opinion mining of Saudi responses to COVID-19 vaccines on Twitter. *IJACSA* **2021**, *12*, doi:10.14569/IJACSA.2021.0120610.
Schaffer DeRoo, S.; Pudalov, N.J.; Fu, L.Y. Planning for a COVID-19 vaccination program. *JAMA* **2020**, *323*, 2458–2459, doi:10.1001/jama.2020.8711.

[9] Jung, J. Preparing for the coronavirus disease (COVID-19) vaccination: Evidence, plans, and implications. *J Korean Med Sci* **2021**, *36*, doi:10.3346/jkms.2021.36.e59.

[10] Strodel, R.; Dayton, L.; Garrison-Desany, H.M.; Eber, G.; Beyrer, C.; Arscott, J.; Rubenstein, L.; Sufrin, C. COVID-19 vaccine prioritization of incarcerated people relative to other vulnerable groups: An analysis of state plans. *PLoS One* **2021**, *16*, e0253208, doi:10.1371/journal.pone.0253208.

[11] Hardeman, A.; Wong, T.; Denson, J.L.; Postelnicu, R.; Rojas, J.C. Evaluation of health equity in COVID-19 vaccine distribution plans in the United States. *JAMA Network Open* **2021**, *4*, e2115653, doi:10.1001/jamanetworkopen.2021.15653.

[12] Hardeman, A.; Wong, T.; Denson, J.L.; Postelnicu, R.; Rojas, J.C. Evaluation of health equity in COVID-19 vaccine distribution plans in the United States. *JAMA Network Open* **2021**, *4*, e2115653, doi:10.1001/jamanetworkopen.2021.15653.

[13] Barry, M.; BaHammam, A.S. COVID-19 vaccine in the Kingdom of Saudi Arabia: A true operation warp speed. *J Nat Sci Biol Med* **2021**, *4*, 92, doi:10.4103/jnsm.jnsm_8_21.

[14] Alfatease, A.; Alqahtani, A.M.; Orayj, K.; Alshahrani, S.M. The impact of social media on the acceptance of the COVID-19 vaccine: A cross-sectional study from Saudi Arabia. *Patient Prefer Adherence* **2021**, *15*, 2673–2681, doi:10.2147/PPA.S342535.

[15] Almaghaslah, D.; Alsayari, A.; Kandasamy, G.; Vasudevan, R. COVID-19 vaccine hesitancy among young adults in Saudi Arabia: A cross-sectional web-based study. *Vaccines* **2021**, *9*, 330, doi:10.3390/vaccines9040330.

[16] Islam, M.S.; Kamal, A.-H.M.; Kabir, A.; Southern, D.L.; Khan, S.H.; Hasan, S.M.M.; Sarkar, T.; Sharmin, S.; Das, S.; Roy, T.; et al. COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence. *PLoS One* **2021**, *16*, e0251605, doi:10.1371/journal.pone.0251605.

[17] Manguri, K.; Ramadhan, R.; Mohammed Amin, P. Twitter sentiment analysis on worldwide COVID-19 outbreaks. Kurdistan J *J Appl Res* **2020**, 54–65, doi:10.24017/covid.8.

[18] Alamoodi, A.H.; Zaidan, B.B.; Zaidan, A.A.; Albahri, O.S.; Mohammed, K.I.; Malik, R.Q.; Almahdi, E.M.; Chyad, M.A.; Tareq, Z.; Albahri, A.S.; et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Syst Appl* **2020**, 114155, doi:10.1016/j.eswa.2020.114155.

[19] Chandra, R.; Krishna, A. COVID-19 sentiment analysis via deep learning during the rise of novel cases. *PLoS One* **2021**, *16*, e0255615, doi:10.1371/journal.pone.0255615.

[20] Hung, M.; Lauren, E.; Hon, E.; Birmingham, W.; Xu, J.; Su, S.; Hon, S.; Park, J.; Dang, P.; Lipsky, M. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. *J Med Inter Res* **2020**, *22*, e22590, doi:10.2196/22590.

[21] Samuel, J.; Ali, G.G.Md.N.; Rahman, Md.M.; Esawi, E.; Samuel, Y. COVID-19 public sentiment insights and machine learning for tweets classification. *Information* **2020**, *11*, 314, doi:10.3390/info11060314.

[22] Alanezi, M.A.; Hewahi, N.M. Tweets Sentiment Analysis During COVID-19 Pandemic. In Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), October 2020, pp. 1–6.

[23] Du, J.; Xu, J.; Song, H.-Y.; Tao, C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Medical Inform. Decis. Mak* **2017**, *17*, 69, doi:10.1186/s12911-017-0469-6.

[24] Raghupathi, V.; Ren, J.; Raghupathi, W. Studying public perception about vaccination: A sentiment analysis of tweets. *Int J Environ Res Public Health* **2020**, *17*, 3464.

[25] Salathé, M.; Khandelwal, S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Comput Biol* **2011**, *7*, e1002199, doi:10.1371/journal.pcbi.1002199.

[26] Cotfas, L.-A.; Delcea, C.; Roxin, I.; Ioanăş, C.; Gherai, D.S.; Tajariol, F. The longest month: Analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access* **2021**, *9*, 33203–33223, doi:10.1109/ACCESS.2021.3059821.

[27] Praveen, S.; Ittamalla, R.; Deepak, G. Analyzing the attitude of Indian citizens towards COVID-19 vaccine – A text analytics study. *Diabetes Metab Syndr: Clin Res Rev* **2021**, *15*, 595–599, doi:10.1016/j.dsx.2021.02.031.

[28] Nurdeni, D.A.; Budi, I.; Santoso, A.B. Sentiment Analysis on Covid-19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer. In Proceedings of the 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Indonesia, April 2021, pp. 122–127.

[29] Shahriar, K.T.; Islam, M.N.; Anwar, Md.M.; Sarker, I.H. COVID-19 analytics: Towards the effect of vaccine brands through analyzing public sentiment of tweets. *Inform Med Unlocked* **2022**, *31*, 100969, doi:10.1016/j.imu.2022.100969.

[30] Alabrah, A.; Alawadh, H.M.; Okon, O.D.; Meraj, T.; Rauf, H.T. Gulf countries' citizens' acceptance of COVID-19 vaccines—A machine learning approach. *Mathematics* **2022**, *10*, 467, doi:10.3390/math10030467.

[31] Khan, I.U.; Aslam, N.; Chrouf, S.; Atef, I.; Merah, I.; AlMulhim, L.; AlShuaifan, R. Computational intelligence-based model for exploring individual perception on SARS-CoV-2 vaccine in Saudi Arabia. *Comput Intell Neurosci* **2022**, e6722427, doi:10.1155/2022/6722427.

[32] Zacharias, C.; Poldi, F. TWINT—Twitter Intelligence Tool 2021.

[33] Bird, S.; Klein, E.; Loper, E. *Natural Language Processing With Python*, 1st ed.; O'Reilly, 2009; ISBN 978-0-596-51649-9.

[34] Su, Y.; Venkat, A.; Yadav, Y.; Puglisi, L.B.; Fodeh, S.J. Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities. *Comput Biol Med.* **2021**, *132*, 104336, doi:10.1016/j.compbiomed.2021.104336.

[35] Abu Farha, I.; Magdy, W. Mazajak: An Online Arabic Sentiment Analyser. In Proceedings of the Fourth Arabic Natural Language Processing Workshop, Association for Computational Linguistics: Florence, Italy, August 2019, pp. 192–198.

[36] Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y.; Qawasmeh, O. Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Inf Process Manag* **2019**, *56*, 308–319, doi:10.1016/j.ipm.2018.01.006.

[37] Eligüzel, N.; Çetinkaya, C.; Dereli, T. Comparison of different machine learning techniques on location extraction by utilizing geo-tagged tweets: A case study. *Adv Eng Inform* **2020**, *46*, 101151, doi:10.1016/j.aei.2020.101151.

[38] Alhajji, M.; Khalifah, A.A.; Aljubran, M.; Alkhalifah, M. Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19. *Preprints* 2020, 2020040031, doi:10.20944/preprints202004.0031.v1.

[39] Alyami, S.N.; Olatunji, S.O. Application of support vector machine for Arabic sentiment classification using Twitter-based dataset. *J Info Know Mgmt* **2020**, *19*, 2040018, doi:10.1142/S0219649220400183.

[40] Dey, S.; Wasif, S.; Tonmoy, D.S.; Sultana, S.; Sarkar, J.; Dey, M. A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. In Proceedings of the 2020 International Conference on Contemporary Computing and Applications (IC3A), February 2020, pp. 217–220.

[41] Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Inf Process Manag* **2020**, *57*, 102221.

[42] Daeli, N.O.F.; Adiwijaya, A. Sentiment Analysis on movie reviews using information gain and K-nearest neighbor. *J Data Inf Sci* **2020**, *3*, 1–7.

[43] Troussas, C.; Virvou, M.; Espinosa, K.J.; Llaguno, K.; Caro, J. Sentiment Analysis of Facebook Statuses Using Naive Bayes Classifier for Language Learning. In Proceedings of the IISA 2013, IEEE, 2013, pp. 1–6.

[44] Parveen, H.; Pandey, S. Sentiment Analysis on Twitter Data-Set Using Naive Bayes Algorithm. In Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICATCCT), IEEE, 2016, pp. 416–419.

[45] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine learning in Python. *J Mach Learn Res* **2011**, *12*, 2825–2830.

[46] Vangara, R.V.B.; Thirupathur, K.; Vangara, S.P. Opinion mining classification using naive Bayes algorithm. *Int J Eng Innov Technol (IJITEE)* **2020**, *9*, 495–498.

[47] Lorena, A.C.; Jacintho, L.F.O.; Siqueira, M.F.; Giovanni, R.D.; Lohmann, L.G.; de Carvalho, A.C.P.L.F.; Yamamoto, M. Comparing machine learning classifiers in potential distribution modelling. *Expert Syst Appl* **2011**, *38*, 5268–5275, doi:10.1016/j.eswa.2010.10.031.

[48] Zulqarnain, M.; Ghazali, R.; Mazwin, Y.; Rehan, M. A Comparative review on deep learning models for text classification. *Indones J Electr Eng* **2020**, *19*, doi:10.11591/ijeecs.v19.i1.pp325-335.

[49] Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* **2019**, *31*, 1235–1270, doi:10.1162/neco_a_01199.

[50] Rhanoui, M.; Mikram, M.; Yousfi, S.; Barzali, S. A CNN-BiLSTM model for document-level sentiment analysis. *Mach Learn Knowl Extr* **2019**, *1*, 832–847, doi:10.3390/make1030048.

[51] Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst* **2017**, *28*, 2222–2232, doi:10.1109/TNNLS.2016.2582924.

[52] Hameed, Z.; Garcia-Zapirain, B. Sentiment classification using a single-layered BiLSTM model. *IEEE Access* **2020**, *8,* 73992–74001, doi:10.1109/ACCESS.2020.2988550.

[53] Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* **2002**, *16*, 321–357, doi:10.1613/jair.953.

[54] Al-Azani, S.; El-Alfy, E.-S.M. Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text. *Procedia Comput Sci* **2017**, *109*, 359–366, doi:10.1016/j.procs.2017.05.365.

[55] Obiedat, R.; Qaddoura, R.; Al-Zoubi, A.M.; Al-Qaisi, L.; Harfoushi, O.; Alrefai, M.; Faris, H. Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution. *IEEE Access* **2022**, *10*, 22260–22273, doi:10.1109/ACCESS.2022.3149482.

[56] SMOTE — Version 0.10.0. Dev0 Available online: https://imbalanced-learn.org/dev/references/generated/imblearn.over_sampling.SMOTE.html (accessed on 3 July 2022).

[57] Carneiro, T.; Medeiros Da NóBrega, R.V.; Nepomuceno, T.; Bian, G.-B.; De Albuquerque, V.H.C.; Filho, P.P.R. Performance analysis of Google colaboratory as a tool for accelerating deep learning applications. *IEEE Access* **2018**, *6*, 61677–61685, doi:10.1109/ACCESS.2018.2874767.

[58] Wadhe, A.A.; Suratkar, S.S. Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques. In Proceedings of the 2020 International Conference on Industry 4.0 Technology (I4Tech), February 2020, pp. 1–6.

[59] Zeng, Y.; Jiang, K.; Chen, J. Automatic Seismic Salt Interpretation with Deep Convolutional Neural Networks. In Proceedings of the Proceedings of the 2019 3rd International Conference on Information System and Data Mining – ICISDM 2019, ACM Press: Houston, TX, USA, 2019, pp. 16–20.

[60] Alkaff, M.; Rizky Baskara, A.; Hendro Wicaksono, Y. Sentiment Analysis of Indonesian Movie Trailer on YouTube Using Delta TF-IDF and SVM. In Proceedings of the 2020 Fifth International Conference on Informatics and Computing (ICIC), November 2020, pp. 1–5.

[61] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res* **2014**, *15*, 1929–1958.

[62] Ding, B.; Qian, H.; Zhou, J. Activation Functions and Their Characteristics in Deep Neural Networks. In Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), June 2018, pp. 1836–1841.

[63] Jiao, Y.; Du, P. Performance Measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant Biol* **2016**, *4*, 320–330, doi:10.1007/s40484-016-0081-2.

[64] Novaković, J.D.; Veljović, A.; Ilić, S.S.; Papić, Ž.; Milica, T. Evaluation of classification models in machine learning. *Theo Comput Math with Appl* **2017**, *7*, 39–46.

[65] Patino, C.M.; Ferreira, J.C. Internal and external validity: Can you apply research study results to your patients? *J Bras Pneumol* **2018,** *44*, 183, doi:10.1590/S1806-37562018000000164.

[66] Alfageeh, E.I.; Alshareef, N.; Angawi, K.; Alhazmi, F.; Chirwa, G.C. Acceptability of a COVID-19 vaccine among the Saudi Population. *Vaccines (Basel)* **2021**, *9*, 226, doi:10.3390/vaccines9030226.

[67] فريق بوابة وزارة الصحةMinistry of Health – Kingdom of Saudi Arabia. Available online: https://www.moh.gov.sa/en/Pages/Default.aspx (accessed on 18 February 2022).

[68] Alahmed, M. عاجل: وزير الصحة السعودي يتلقى الجرعة الاولى من لقاح كرونا," صحيفة وقع الإلكترونيةUrgent: The Saudi Minister of Health receives the first dose of the Corona vaccine. 2020 صحيفة وقع الإلكترونية.

[69] Al Hadath Saudi Health approves AstraZeneca and Moderna vaccines after Pfizer. Available online: shorturl.at/gpsz3 (accessed on 29 July 2022).

[70] Galluccio, B. New "Double Mutant" Coronavirus Variant Discovered in India. Available online: https://www.iheart.com/content/2021-03-25-new-double-mutant-coronavirus-variant-discovered-in-india/ (accessed on 29 July 2022).

[71] Singh, K. WHO Rules out New "hybrid" COVID Variant in Vietnam - Nikkei | Reuters.Com. Reuters 2021.

[72] Ministry of Education Ministry of Education – Kingdom of Saudi Arabia Available online: https://www.moe.gov.sa/en/pages/default.aspx (accessed on 18 February 2022).

[73] Zarkachat. The Ministry of Education in Saudi Arabia forms a committee to discuss the return of the study in person. Zarkachat 2021.

[74] Ministry of Interior Date for Lifting Travel Suspension for Citizens, Opening All Land, Sea, Air Outlets in Full, Postponed to May 17. Available online: https://bit.ly/3gZa5eI (accessed on 19 February 2022).

[75] Asharq Al-Awsat Saudi Arabia: Vaccination Needed to Enter Govt, Private Offices. Available online: https://english.aawsat.com/home/article/2980751/saudi-arabia-vaccination-needed-enter-govt-private-offices (accessed on 29 July 2022).

[76] Arabian Business Saudi Arabia to Require Covid-19 Vaccine to Enter "all Places in All Regions" - Arabian Business Available online: https://www.arabianbusiness.com/industries/healthcare/466621-saudi-arabia-to-require-covid-19-vaccine-to-enter-all-places-in-all-regions (accessed on 29 July 2022).

[77] Alam, K.N.; Khan, M.S.; Dhruba, A.R.; Khan, M.M.; Al-Amri, J.F.; Masud, M.; Rawashdeh, M. Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. *Comput Math Methods Med* **2021**, e4321131, doi:10.1155/2021/4321131.