Sentiment analysis system for COVID-19 vaccinations using data of Twitter

Eman Thabet Khalid[1], Entesar Barges Talal[1], Methaq Khamees Faraj[2], Ali A. Yassin[1]
[1]Department of Computer Sciences, Education College for Pure Sciences, University of Basrah, Basrah, Iraq
[2]Department of Computer Engineering, University of Technology, Baghdad, Iraq

Article Info

ABSTRACT

COVID-19 vaccination topic has been a hot topic of discussions on social media platforms wondering its effectiveness against the SARS-COV-2 virus. Twitter is one of the social media platforms that people widely lunched to express and share their thoughts about different issues touching their daily life. Though many studies have been undertaken for COVID-19 vaccine sentiment analysis, they are still limited and need to be updated constantly. This paper conducts a system for COVID-19 vaccine sentiment analysis based on data extracted from Twitter platform for the time interval from 1st of January till the 3rd of Sep. 2021, and by using deep learning techniques. The introduced system proposes to develop a model architecture based on a deep bidirectional long short-term memory (LSTM) neural network, to analyze tweets data in the form of positive, neutral, and negative. As a result, the overall accuracy of the developed model based on validation data is 74.92%. The obtained outcomes from the sentiment analysis system on collected tweets-data of COVID-19 vaccine revealed that neutral is the prominent sentiment with a rate of 69.5%, and negative sentiment has less rate of tweets reached 20.75% while the positive sentiment has a lesser rate of tweets reached of 9.67%.

Corresponding Author:

Eman Thabet Khalid
Department of Computer Sciences, Education College for Pure Sciences, University of Basrah
Basrah, Iraq
Email: eman.alasadi@uobasrah.edu.iq.

1. INTRODUCTION

Sentiment analysis is the process of examining people's feelings on a particular aspect by using mathematical calculations [1]. Individuals can express themselves genuinely in time via using social media platforms like Twitter. Twitter is a great platform to utilize to analyze the true feelings of the general public. Because users can freely share their location, comments, opinions, and feelings, albeit limited within 280 characters, it is suitable in studies that require opinion analysis [2]. However, since October 2019 COVID-19 outbreak keeps disturbing the ordinary lifestyle of humans on the earth in most domains such as health, education, production, industry, tourism, and so on. To save humanity and cut down the danger of this virus, health organizations and companies that work in the medical section aimed at finding a proper vaccine against COVID-19 virus to preserve people's life as well as decrease of virus's effects on health. Many vaccines have been released recently such as Pfizer/BioNTech, Moderna, Oxford/AstraZeneca, Covaxin, Sputnik V, Sinopharm, and Sinovac. People commenced getting the vaccine all over the world to prevent COVID-19 threat. On the other hand, people throughout the world show different thoughts and sentiments toward the vaccine topic expressing that via their discussions on social media that could be varying between endorsements and oppositions for getting COVID-19 vaccine [3]. Research revealed that discussions

regarded the COVID-19 vaccine on Twitter have a high impact on individuals' thinking positively or negatively that could encourage, be hesitant, or prevent people from taking the vaccine [4].

Therefore, it is essential to analyze general public sentiments toward COVID-19 vaccination to support governments and health organizations in all countries around the world to reach immunity goals in their communities. Public emotion inquiry can help in identifying the public overall sentiment rates toward a vaccine the matter that could assist in investigating issues that make people hesitate to be vaccinated. Besides, this will benefit policymakers and health researchers to perform appropriate initiatives to increase vaccines trust, in addition, to retain people aware enough and safer. Hence, this study conducts a sentiment analysis system for COVID-19 vaccine-based discussions topic on Twitter. The contribution of this work is summarized as follow: (i) introducing a sentiment analysis system for COVID-19 vaccine-related discussions on Twitter investigating the emotion of the public toward COVID-19 vaccines; (ii) deploying a model based on a deep bidirectional long short-term memory (LSTM) network can analyze big text data of Twitter posts regarded SARS-COV-2 vaccine sentimentally in the form of positive, negative, and neutral; (iii) performing data collection via extracting tweets posts-data concerning COVID-19 vaccination from Twitter platform using tweets- scarping API; (iv) gathering 131,268 tweets data from Twitter regarding COVID-19 vaccine for this analysis within a time interval mostly from January 1st, 2021 to September 3rd, 2021.

Recently, several studies have been introduced for COVID-19 vaccine sentiment analysis. Research by Rahman and Islam [5] conducted three machine learning sentiment analysis models for COVID-19 vaccine, which are stacking classifier (SC), the voting classifier (VC), and bagging classifier (BC). The data was collected from Twitter and it is about 12,000 tweets and annotated by three self-governing reviewers. The SC classifier has a significant F1-score that reached 83.5% in comparison with other classifiers. The study by Bonnevie et al. 2021 [6] introduced work to quantify the rise in vaccine opposition on Twitter. The study found that there is a clear fomenting against taking the vaccine, which may encourage those who are questioning vaccine, being hesitant, or mistrust in authorities of health to move toward COVID-19 vaccine opposition. Concluding that, such matter could have impacts on people's health or life. The undertaken work compared the conversations on Twitter for four months before COVID-19 outbreak in United States for the period (10/15/2019 to 2/14/2020) to four subsequent months period (2/15/2020 to 6/14/2020) of the pandemic. Similarly, Rahul et al. [7] performed a method that considers topic modeling and sentiment analysis of COVID-19 vaccine based on Tweets data for the time from November 1 to December 16 in 2020. The total tweets that were extracted and used within the sentiment analysis method were 572,958 tweets, utilizing VADER and TextBlob. The results of VADER presented that 47.3% of tweets were positive, 8.6% of tweets were neutral, and 24.1% of tweets were negative, while according to TextBlob, 48.3% of tweets were positive, 36.1% of tweets were neutral, 15.6% of tweets were negative. It concluded that positive sentiment was the overweight tweets among others for both methods. Melton et al. [8] Achieved public sentiment analysis and latent dirichlet allocation topic modeling on textual data gathered from 13 Reddit communities about COVID-19 vaccine from Dec 1, 2020, to May 15, 2021, using a method of polarity analysis. The outcomes of the analysis revealed that for the mentioned period the positive sentiment was the more sentiment. In the paper, Liu and Liu [9] implemented sentiment analysis on English tweets of COVID-19 vaccine collected between November 1, 2020, and January 31, 2021. The compound score is computed using valence aware dictionary and sentiment reasoner tool to determine the sentiment of each tweet whether were positive, neutral, or negative respectively according to compound score (compound $\geq$ 0.05), (-0.05 < compound < 0.05), or (compound $\leq$ -0.05). The study did latent Dirichlet allocation analysis identified the key topics for positive and negative sentiment tweets in addition to performing temporal analysis to define trends of time a geographic analysis to investigate sentiment changes in tweets posted in diverse locations. The results of sentiment analysis for 2,678,372 COVID-19 vaccine related tweets were 42.8%, 26.9%, and 30.3% for positive, neutral, and negative, respectively. Alam et al. [10] introduced a sentiment analysis study for the topic of COVID-19 vaccine on Twitter using comparison results between Lstm and Bi-Lstm neural networks. However, their data collected from Twitter was for the period from 21 December 2020 till 21 July 2021. Their collected accuracy for deployed model was 90.59% for LSTM, and 90.83% for Bi-LSTM.

However, yet there is a growing need to keep analyzing big data of COVID-19 vaccine discussions on Twitter as long the COVID-19 pandemic is continual all over the world. Especially with the presence of other variants of COVID-19 virus lately such as Delta which put vaccination topic in a hot spot of discussions on social media wondering the efficacy of discovered vaccines of COVID-19 against SARS-COV-2 virus and its variants. Besides, the last conducted studies regarded COVID-19 vaccine has been performed for a period of collected data as well as sometimes for specific geo-location or country. There is a requirement to perform constant analysis for topics regarding SARS-COV-2 vaccination to understand issues that put people away from being vaccinated [11]–[19]. The rest of this paper is ordered as follows, section 2 discusses the proposed COVID-19 sentiment analysis system methodology, section 3 discusses experimental results, and section 4 discusses the conclusion.

2.   METHOD

This study proposed a system based on model deployment in order to perform sentiment analysis for discussions regarded COVID-19 vaccine on Twitter platform. The proposed system uses the techniques of natural language processing (NLP) and deep learning. The methodology of our system consists of two stages. The first stage is responsible for model building, training and validation while the second stage is responsible for model deployment and implementation. The framework of our proposed system is depicted in Figure 1. Each stage of the proposed method commenced with a data pre-processing approach which includes data cleaning, data Tokenization, and pad-sequencing. That's to make the data ready for analysis. The collected results from the Tokenization and pad-sequencing process is cleaned text in the form of numbers since numbers are the language of computer and neural networks can deal with numbers only. The sentiment words are also coded as numbers "neutral" into 0, "negative" into 1, and "positive" into 2. By the end of pre-processing only valid texts with relevant words remain so that the model can work on it.

In the building and training model stage, after the data pre-processing step the pre-processed train data is divided into train data and test data, where test data has 0.33%. Then, we construct our model using a deep bidirectional long short-term memory (LSTM) neural network. The architecture of our model consists of an embedding layer, Bi-LSTM layer, and Softmax dense layer in order to obtain features and train the model to classify text based on pre-defined sentiments (neutral, positive, and negative). In the model implementation and deployment stage, the trained model is saved and used to analyze the collected data of Twitter to identify the public sentiments toward COVID-19 vaccine topic, to help organizations of health and governors to find a policy to deal with matters that mistrust taking the vaccine. The data is pre-processed to be prepared for analyzing and then performing sentiment prediction.
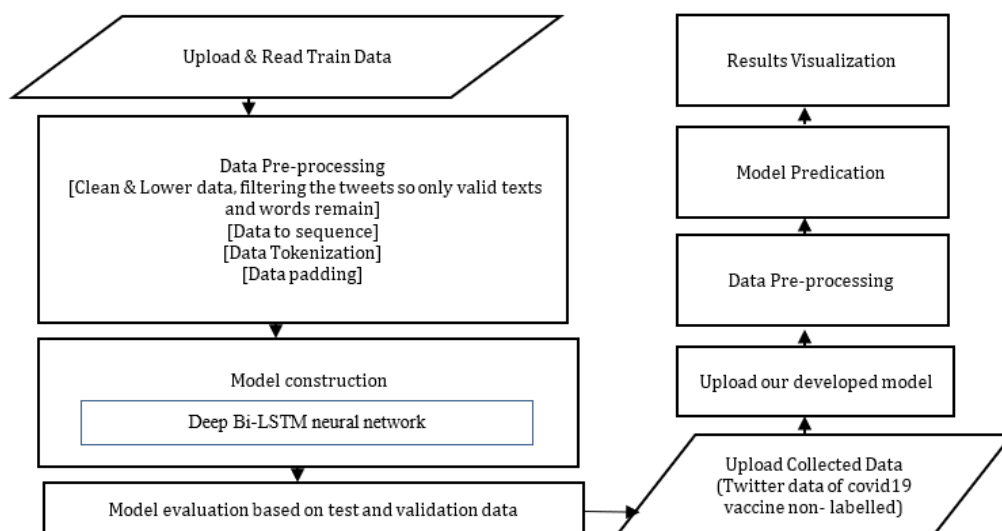

Figure 1. The framework of COVID-19 vaccine sentiment analysis system

2.1. Model construction

The architecture of our model is composed of the pre-processing method, model construction, and model validation and evaluation. In the pre-processing method, based on previous studies [7]–[10], we cleaned the training dataset to remove the hyperlinks, numbers, punctuations, and particular characters that are not desired for building our sentiment analysis model. We performed data lowering and removing stop words that are represented in articles and prepositions that may not have particular meaning on their own or may not apply any kind of value to the analysis. We compute the max-length of the input sentences to perform Tokenization and padding. Tokenization is an operation for splitting texts into words via considering delimiters such as white-space, comma, semicolon, quotes, and periods. Tokens could be single words of type (verb, noun, pronoun, article, conjunction, preposition, punctuation, numbers, and alphanumeric), where the tokens list can be input for extra processing. Tokenizer converts each sentence into a sequence of integers via using python Keras's pre-processing library for text, where each word in the sentence is replaced with an integer value from the vocabulary index. A whole sentence can be presented as a vector of size "max_len", where "max_len" represents the number of words in the sentence via computing the lengthiest sentence in the

text. We follow that with a post-zero padding strategy in which all sentences in the input text (tweets) will have the same vector dimension (max_len). The proposed model architecture is composed of an input layer, embedding layer, Bi-LSTM layer with dropout strategy, and softmax dense layer, the model architecture illustrated Figure 2, in which 117 represent model input layer shape, 64 represents embedding layer dimension, Bi-LSTM network number of units is 64, the number of neurons for the dense layer of a deep neural network is 128, and 3 represents the dense layer for output layer based on Softmax activation function to classify text data features into 3 categories (positive, negative, and neutral).
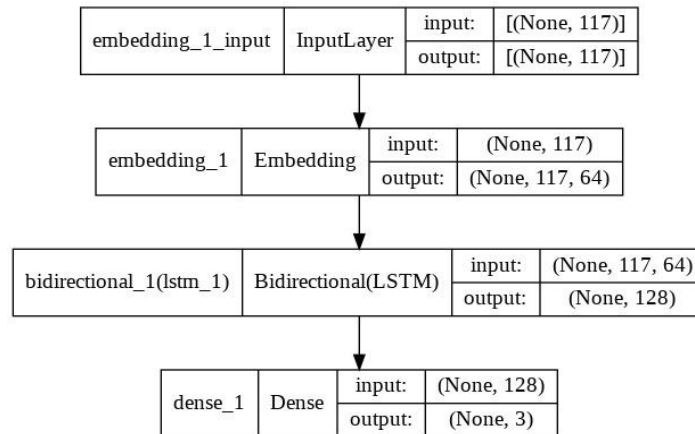
Figure 2. The proposed model of bidirectional LSTM for COVID-19 vaccine sentiment analysis

The input layer is used to define the input shape based on the lengthiest sentence in text or document "max_len" which is equal to 117 in our study. The embedding layer is a method to represent a document word by an integer vector that can capture the context of a word in a text [20]. In this work, we use the embedding layer of the Keras library to map words into vectors. Embedding layer receives integer as input to be looked up in the internal dictionary to then get an associated dense vector. The initial weights for the embedding layer are generated randomly. The word vector is adjusted gradually using the backpropagation algorithm similar to the training algorithm of an artificial neural network. After the training is complete, an ideal word vectors matrix is produced. We set the input length of the word embedding layer to" max_len", and the embedding dimension to 64. Dropout is a strategy to avoid a problem of overfitting via using the dropout parameter to randomly pick some neurons in the network layer and ignore their input and output features (turned to 0) [20], [21]. Because of the dropout ability to capture more randomness, the overfitting could be reduced. In our study, dropout with Bi-LSTM layer is added. In this work, the rate of dropout is set to 0.9%, where the number of epochs is set to 16.

2.2. Bidirectional LSTM neural network

Long short-term memory (LSTM) is a special form of recurrent neural networks (RNN) architecture that is specializing in natural language processing at its initial stage. It was proposed by Sepp Hochreiter and Jürgen Schmidhuber in 1997 to avoid the long-term dependency issue that plagued ordinary [22]. LSTM units come in a variety of topologies. A common architecture consists of a cell (the LSTM unit's memory) and three "regulators," or gates, that control the flow of knowledge within the unit: an input gate, an output gate, and a forget gate [23]. However, bidirectional LSTM models process data in both directions, rather than using only the prior context state to predict the next states as bidirectional RNN models do. At each time step, two separate LSTM models provide both backward and forward information about the sequence. This improves access to long-range state information, which has been effective for word embedding and a variety of other sequence processing issues [24]. In our work, we proposed to use one Bi-LSTM layer with 64 units same as the embedding dimension.

2.3. Softmax dense layer

In this work, in order to classify text or tweets and predicate sentiment, we use a dense layer with Softmax activation function, the number of neurons is 3. The dense layer is the neural network layer deeply connected in which every neuron in the dense layer receives input from all neurons in the earlier layer [20].

The dense layer has a weight-matrix (w), a vector of bias (b), and the activation function (a), the dense layer can be defined as:

$$y = a\big((X.W) + b\big) \tag{1}$$

In this paper, the classification layer with Softmax activation function is utilized to predicate sentiment of 3 classes in the form of positive, neutral, and negative. The softmax layer is defined as follows:

$$P\left(\frac{z}{z}\right) = softmax(w^s h_k + b^s) \tag{2}$$

$$z = argmax\left(p\left(\frac{z}{x}\right)\right) \tag{3}$$

where z is the predicated sentiment, x is the given input, h represents hidden stats. Finally, the model is built by mapping between inputs and outputs.

2.4. Model fitting and training

To train the model, we compile the model using a loss function called categorical cross-entropy. We used the "Adam" optimization algorithm to optimize the optimizer and minimize the loss function. The evaluation metric used is accuracy.

2.5. Model prediction

After the model training and validation stage, the developed model is used to make predication on our collected tweet-data of vaccination. Initially, our data of the COVID-19 vaccine that saved in the form of a ".csv" file is uploaded to be analyzed and pre-processed. The data is cleaned by removing NAN tweets, Non-English tweets, and empty tweets. The Duplicate tweets are removed in order to get rid of any confusion. The second stage of preparing tweets data for model prediction is using the pre-processing method that we did with the training dataset via removing URLs, emails, digits, punctuations, and other non-relevant characters in order to clean data and convert it into sequences through data Tokenization and pad-sequence operations. To then make predication on the gathered data of COVID-19 vaccine using our deployed model. After sentiment prediction, we conducted a result visualization approach to understand the outcomes.

3. RESULTS AND DISCUSSION

This study introduced a sentiment analysis system that can analyze big data of Twitter posts (tweets) related to COVID-19 vaccine discussions in the form of positive, neutral, and negative. This is because Twitter is one of the social media platforms that have a vital impact on people's behaviors in such a way that could convince people to take the vaccine, be hesitant, or prevent them. Besides, it can be beneficial for health organizations and governments to understand the issues that made the public around the world question COVID-19 vaccines, to work on such issues, and try to find solutions.

The system is constructed based on developing a deep learning model using the bidirectional LSTM neural network. To analyze vaccine data, we made data scraping from Twitter for the topics that regarded COVID-19 vaccine. The data was collected for the period from January 1st, 2021 to September 3rd, 2021 by using Python Twint tool library, and for general geo-tag. The keywords or search terms that were used for data extraction are: COVID-19 vaccine, COVID-19 vaccination, COVID-19 vaccine side effects, COVID-19 vaccine risks, Pfizer, Oxford, AstraZeneca, Moderna, and Sinovac. These terms are utilized to search for tweets that discuss COVID-19 vaccine from different aspects. We extracted 131, 268 tweets, only 119, 929 unique tweets were utilized after refining and removing duplicates, also only English tweets were chosen. Besides, there is no annotated tweet-dataset related to COVID-19 vaccine yet. Thus, based on previous studies from the literature [25], we used two annotated public tweet datasets which are the Apple tweeter sentiment dataset and Tweet extraction dataset that are available on Kaggle repository for sentiment analysis issues. The two datasets together composed in total 32,638 Tweets after cleaning, labeled as positive, negative, and neutral. Both data are utilized to train and validate the proposed model. We balanced and refined the training dataset via removing bias in tweets for sentiment over another by deleting excess tweets. However, because the size of our extracted data was big and required a considerable time to get annotated. Then, we used the deployed model to make predication on our tweets-dataset of COVID-19 vaccine.

The proposed system was implemented using the programming language of Python 3.7 and its deep learning libraries for NLP including Keras and TensorFlow library, utilizing GPU of Google Colab. The train

data has been randomly split into 33% test data and 67% train data, the test data has been divided into 20% validation data and 13% test data. The evaluation metric that has been utilized to evaluate our work is the accuracy metric that based on "ModelCheckpoint" of "keras. callbacks", in which we set monitor property equal to the accuracy of validation data, and based on that we set to save the best-acquired model during training process based on validation accuracy. The best model is set to be saved in the form ".h5" file to be utilized in the prediction process. A confusion matrix is also used to assess the results for every sentiment category. For comparison and performance evaluation, the performance of our developed model is compared with the performance of the LSTM neural network model. Based on the accuracy metric, our model has obtained higher accuracy see Table 1.

Table 1. The comparison results-based accuracy

Model name	Training accuracy	Training_loss_erorr	Validation accuracy	Validation _loss_erorr
LSTM	0.7544	0.5982	0.7171	0.6988
Bi-LSTM	**0.7624**	**0.5956**	**0.7492**	**0.6360**

The confusion matrix is calculated for both train and test dataset and the obtained accuracy for each category (negative, neutral, and positive) using train data is 0.88, 0.72, and 0.78, respectively, while the obtained accuracy for each category using test data is 0.97, 0.65, and 0.75 for negative, neutral, and positive sentiments, as illustrated in Figure 3.
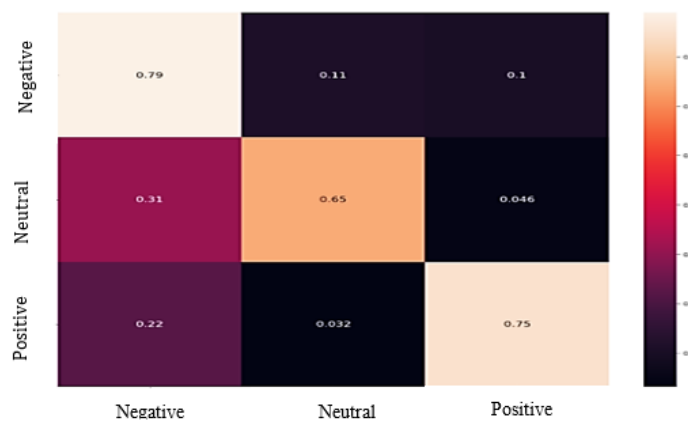
Figure 3. The accuracy of each category based on confusion-matrix on test data

After the model training and validation stage, the developed model is used to make sentiment predictions on the collected tweets data of vaccination. Figure 4 shows the timeline explains the number of tweets for each sentiment category based on the date of each tweet. Consequently, we can notice that the sentiment "Neutral" is the overall sentiment for the detected interval from 1-1-2021 to 3-9-2021. More results about COVID-19 vaccine sentiment analysis is shown in Figure 5, from which we can figure out that the sentiment "Neutral" has the most sentiment rate (69.5%) for all collected data within the time-frame, whereas the "Negative" sentiment has second high-rate (20.75%) sentiment for the mentioned interval, and the "Positive" sentiment has less rate of sentiment (9.67%) among all the collected tweets.

To investigate and visualize the obtained results, we performed more investigations on the collected results from the prediction model to explore the sentiment toward some COVID-19 vaccinations such as Pfizer, Moderna, Sinovac, and AstraZeneca vaccines, the obtained results have shown in Figures from 6 to 9. As we can observe from the visualized results, the neutral sentiment is the relevant one through the addressed time interval, followed by negative sentiment for the same period while positive sentiment has fewer rates of tweets, for all pre-named vaccines. Thus, investigating and analyzing the sentiments of people tweets around the world for some time is beneficial and could contribute to realizing the people emotions toward the COVID-19 vaccinations to help governments and health organizations to highlight and limit the obstacles that may make people afraid form getting vaccinated as well as to make public more aware and trust COVID-19 vaccine.
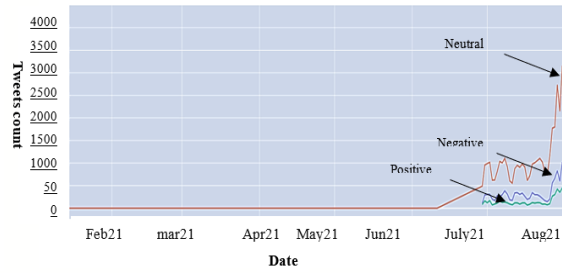
Figure 4. The timeline shows the public sentiment toward COVID-19 vaccine according to date
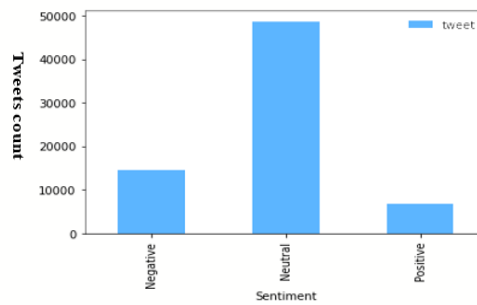

Figure 5. The outcomes of predication regarding public sentiment toward COVID-19 vaccine
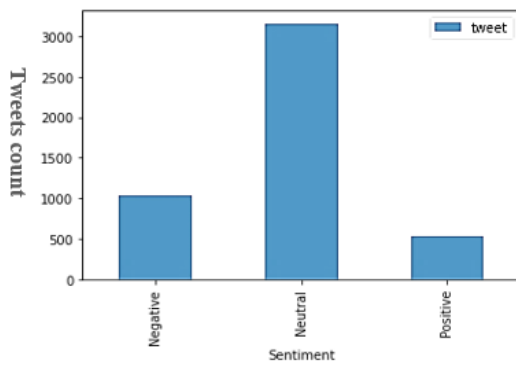

Figure 6. The predicated public sentiment toward Pfizer vaccine using Twitter data
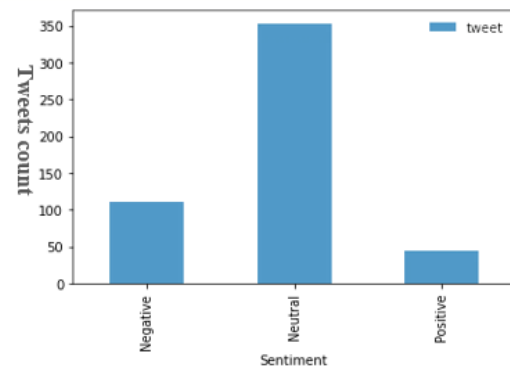

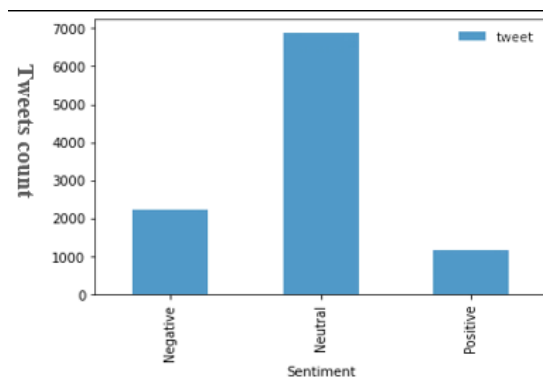Figure 7. The predicated public sentiment toward Sinovac vaccine using Twitter data

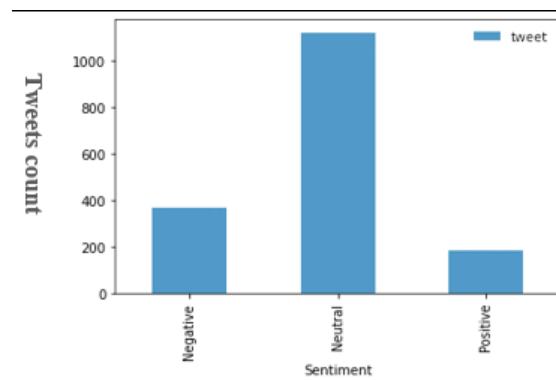Figure 8. The predicated public sentiment toward AstraZeneca vaccine using Twitter data

Figure 9. The predicated public sentiment toward Moderna vaccine using Twitter data

4.   CONCLUSION

This paper developed a system for COVID-19 vaccine sentiment analysis based on data extracted from Twitter platform for the time interval from 1st of January till the 3rd of Sep. 2021, 131, 268 tweets were scraped to implement this study. Since there is a requirement to perform constant analysis for topics regarding SARS-COV-2 vaccination on social media to understand issues that put people away from being vaccinated. This paper proposed a system that used deep learning techniques to deploy a model based on a bidirectional LSTM neural network in order to achieve promising results. The results of the proposed system have been obtained in the form of positive, negative, and neutral. The evaluation metric that we used in our study is accuracy for testing and validating our proposed model. The obtained accuracy for validation is 74.92% within the model training and fitting process. The accuracy of each sentiment category depending on train and test data is computed by confusion-matrix. The outcomes of the prediction on COVID-19 vaccine tweets showed neutral sentiment with a high rate of tweets followed by negative and positive sentiments with less and lesser rates respectively.

REFERENCES

[1]   C. Shofiya and S. Abidi, "Sentiment analysis on covid-19-related social distancing in Canada using twitter data," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p. 5993, Jun. 2021, doi: 10.3390/ijerph18115993.
[2]   C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Information (Switzerland)*, vol. 12, no. 5, p. 204, May 2021, doi: 10.3390/info12050204.
[3]   J. C. Lyu, E. Le Han, and G. K. Luli, "Covid-19 vaccine-related discussion on twitter: Topic modeling and sentiment analysis," *Journal of Medical Internet Research*, vol. 23, no. 6, p. e24435, Jun. 2021, doi: 10.2196/24435.
[4]   A. Hussain *et al.*, "Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study," *Journal of Medical Internet Research*, vol. 23, no. 4, p. e26627, Apr. 2021, doi: 10.2196/26627.
[5]   M. Rahman and M. N. Islam, "Exploring the performance of ensemble machine learning classifiers for sentiment analysis of covid-19 tweets," in *Sentimental Analysis and Deep Learning*, Springer Singapore, 2022, pp. 383–396.
[6]   E. Bonnevie, A. Gallegos-Jeffrey, J. Goldbarg, B. Byrd, and J. Smyser, "Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic," *Journal of Communication in Healthcare*, vol. 14, no. 1, pp. 12–19, Dec. 2021, doi: 10.1080/17538068.2020.1858222.
[7]   K. Rahul, B. R. Jindal, K. Singh, and P. Meel, "Analysing public sentiments regarding COVID-19 vaccine on Twitter," in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, Mar. 2021, pp. 488–493, doi: 10.1109/ICACCS51430.2021.9441693.
[8]   C. A. Melton, O. A. Olusanya, N. Ammar, and A. Shaban-Nejad, "Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence," *Journal of Infection and Public Health*, vol. 14, no. 10, pp. 1505–1512, Oct. 2021, doi: 10.1016/j.jiph.2021.08.010.
[9]   S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," *Vaccine*, vol. 39, no. 39, pp. 5499–5505, Sep. 2021, doi: 10.1016/j.vaccine.2021.08.058.
[10]   K. N. Alam *et al.*, "Deep learning-based sentiment analysis of COVID-19 vaccination responses from twitter data," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–15, Dec. 2021, doi: 10.1155/2021/4321131.
[11]   R. L. Abduljabbar, H. Dia, and P. W. Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction," *Journal of Advanced Transportation*, vol. 2021, pp. 1–16, Mar. 2021, doi: 10.1155/2021/5589075.
[12]   F. M. J. M. Shamrat *et al.*, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 463–470, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp463-470.
[13]   A. D. Dubey, "Public sentiment analysis of COVID-19 vaccination drive in India," *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3772401.
[14]   N. S. Sattar and S. Arifuzzaman, "Covid-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the usa," *Applied Sciences (Switzerland)*, vol. 11, no. 13, p. 6128, Jun. 2021, doi: 10.3390/app11136128.
[15]   S. Gulati, "Decoding the global trend of 'vaccine tourism' through public sentiments and emotions: does it get a nod on Twitter?," *Global Knowledge, Memory and Communication*, Sep. 2021, doi: 10.1108/GKMC-06-2021-0106.
[16]   L. A. Cotfaş, C. Delcea, I. Roxin, C. Ioanăş, D. S. Gherai, and F. Tajariol, "The longest month: analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement," *IEEE Access*, vol. 9, pp. 33203–33223, 2021, doi: 10.1109/ACCESS.2021.3059821.
[17]   S. Das and A. Dutta, "Characterizing public emotions and sentiments in COVID-19 environment: A case study of India," *Journal of Human Behavior in the Social Environment*, vol. 31, no. 1–4, pp. 1–14, Jul. 2020, doi: 10.1080/10911359.2020.1781015.
[18]   T. Na, W. Cheng, D. Li, W. Lu, and H. Li, "Insight from NLP analysis: COVID-19 vaccines sentiments on social media," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.04081
[19]   T. Hu *et al.*, "Revealing public opinion towards covid-19 vaccines with twitter data in the united states: Spatiotemporal perspective," *Journal of Medical Internet Research*, vol. 23, no. 9, Jun. 2021, doi: 10.2196/30854.
[20]   K. Sangeetha and D. Prabha, "Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4117–4126, Mar. 2021, doi: 10.1007/s12652-020-01791-9.
[21]   F. Long, K. Zhou, and W. Ou, "Sentiment analysis of text based on bidirectional LSTM with multi-head attention," *IEEE Access*, vol. 7, pp. 141960–141969, 2019, doi: 10.1109/ACCESS.2019.2942614.
[22]   M. Heikal, M. Torki, and N. El-Makky, "Sentiment analysis of Arabic tweets using deep learning," *Procedia Computer Science*, vol. 142, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.10.466.
[23]   B. Saberi and S. Saad, "Sentiment analysis or opinion mining: A review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 5, pp. 1660–1666, Oct. 2017, doi: 10.18517/ijaseit.7.5.2137.

[24] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS ONE*, vol. 16, no. 8 August, p. e0255615, Aug. 2021, doi: 10.1371/journal.pone.0255615.

[25] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," *Knowledge-Based Systems*, vol. 228, p. 107242, Sep. 2021, doi: 10.1016/j.knosys.2021.107242.

BIOGRAPHIES OF AUTHORS

Eman Thabet Khalid ⓘ 🔾 sc Ⓟ is an assistant lecturer at University of Basrah. She received her BSc in computer science form from Basra University, College of Education for Pure Sciences, Department of Computer Science, Iraq in 2007, and the MSc in Computer Vision from University of Putra Malaysia, Malaysia in 2017. Her research interests include computer vision, machine learning, and deep learning. She can be contacted at email: eman.alasadi@uobasrah,edu.iq.

Entesar Barges Talal ⓘ 🔾 sc Ⓟ is an assistant lecturer at University of Basrah. She received her BSc in Computer Science from Basra University, College of Education for Pure Sciences, Department of Computer Science in 2006, the MSc in Computer science from Basra University in 2015. Her research interests include computer vision and Biometrics. She can be contacted at email: entesar.barges@gmail.com.

Methaq Khamees Faraj ⓘ 🔾 sc Ⓟ received the B.SC degree in computer engineering from Al-Mustansiriay University. The M. Sc degree in computer engineering from university of technology, Baghdad, Iraq, 2020. Her research interest includes machine learning, SDN, and network security. She can be contacted at email: En.metho27@gmail.com.

Prof. Dr. Ali A. Yassin ⓘ 🔾 sc Ⓟ is a Professor with the Department of Computer Science, College of Education for Pure Science, University of Basrah. He received the bachelor's and master's degrees from the University of Basrah, Basrah, Iraq, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China. His research interests include the security of cloud computing, image processing, pattern recognition, biometrics, data integrity, DNA cryptography, steganography, sharing data, graphical password, QR code, and soft computinHe can be contacted at email: alihas@upm.edu.my or Ali.Yassin@uobasrah.edu.iq.