Users opinion and emotion understanding in social media regarding COVID-19 vaccine

Abdulqader M. Almars[1] · El-Sayed Atlam[1,2] · Talal H. Noor[1] · Ghada ELmarhomy[1,2] · Rasha Alagamy[1,2] · Ibrahim Gad[2]

Abstract

Online social platforms or social platforms such as Twitter, Facebook and Instagram have become popular platforms for a public discussion about social topics. Recent studies show that there is a growing tendency for people to talk about COVID-19 pandemic in these online channels. The rapid growth of the infected cases by COVID-19 pandemic makes a lots of anxiety and fear among people. With the recent released of Pfizer vaccine, people start posting a lot of rumors regarding the safety concerns of the vaccine, especially among the elderly people. The aim of this study is to bring out the fact that tweets containing all pertinent details about the COVID-19 vaccine and provides an analysis and understanding of users emotions regarding the recent release of COVID-19 vaccine. This study starts with the collection of tweets related to COVID-19 vaccine and then cleaning the dataset from redundant tweets. In this study, we use Twitter API and Web Scraping techniques to obtain a sample of 50,000 tweets talking about COVID-19 vaccine.Further, The analysis of users emotions is achieved by manually labeling and classifying the tweets to either positive or negative. Then, a deep learning based model is used to train the data and classify the people opinion about COVID-19 vaccine. The experimental results illustrate that high percentage of people have shown a positive attitude towards COVID1-19 vaccine. The proposed method is validated over Twitter datasets and the results also demonstrate that use of deep learning classifier can successfully improve the accuracy of people emotions analysis with an accuracy up to 98% for training set and the accuracy for testing set is 73%.

✉ Talal H. Noor
tnoor@taibahu.edu.sa

1 Faculty of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia

2 Faculty of Science, Tanta University, Tanta, Egypt

1 Introduction

In 2020, COVID-19 has been reported as a worldwide pandemic by the World Health Organization (WHO). The pandemic has affected more than 200 countries, with approximately two million death [23, 33]. Also, COVID-19 has resulted in 97,966 of businesses to permanently shut down during the pandemic and destroying millions of job opportunities around the world. Recently, two COVID-19 vaccines have been received an approval from WHO: BioNtec/Pfizer and Moderna [21]. According to the Food and Drug Administration, those COVID-19 vaccines provide strong protection against COVID-19 within about 10 days of the first dose [6, 11, 28].

Social networks such as Twitter, Facebook and Instagram have become an important part of our daily life. it provides interfaces for individuals to use their own terms to discuss emerging topics and express their emotions and feeling towards issues happen all around the world. Twitter, for example, is a free broadcast channel that offers registered users to talk and communicate with others using 140 character texts. According to Twitter latest report [10], A total of 1.3 billion accounts have been created and there are roughly 330 million monthly active users and 145 million daily users who post more than 500 million tweets every day [27].

Recent studies show that there is a growing tendency for people to talk about COVID-19 pandemic in these online channels. With the recent released of Pfizer vaccine, people start posting a lot of rumors regarding the safety concerns of the vaccine especially among the elderly people [9, 32, 37]. In fact, people feel uncomfortable and fair due to vaccine uncertainties and the death of 23 people who received the vaccine [35]. Figure 1 presents an example taken from Twitter that shows a number of users who are depressed and fair toward the Pfizer vaccine.

Fig. 1 A Samples showing sentiment translation for each tweet from the collected dataset

Emotion analysis of users tweets aims to the use of machine learning approaches to analysis the text and the people attitude and feelings toward different topics. Many traditional machine learning approaches has been proposed to detect people opinions on social contents [2]. However, these approaches heavily relies on incomplete handcrafted features, which in result produce an accurate outcomes.

Deep Learning have recently been applied wide range of domains such as computer vision and pattern recognition [1, 7, 18]. Comparing with traditional machine learning approaches, deep learning have already made impressive outcomes in in many fields such as Natural Language Processing (NLP), machine learning and emoticon analysis. Many studies show that deep neural networks outperformed more traditional machine learning in term of accuracy and the detection of people feelings [14, 22].

The primary objective of our study is to provide analysis of the users emotions through the tweets regarding to COVID-19 vaccine. The research begins with the collection of tweets regarding COVID-19 vaccine followed by extensive cleaning of the dataset. The dataset contains around 50,000 tweets that have been published from the release of the vaccine till 1st January 2021. During training the model, the tweets are tokenized into words and then transformed into a feature vector. The main model developed and tested for the framework are Long Short-Term Memory (LSTM), Naive Bayes, decision tree and SVM. Natural language Toolkit (NLTK) library has been utilized which goes about as a suitable content processor for language dealings. Finally, WordCloud is presented to show frequently used words that describe users emotions. The approach is termed as "unigram", for two words it is "bigram", for three words it is "trigram", respectively. To evaluate our model, we compare it with other models as Support Vector Machine (SVM) and Naive Base(NB), the results show that our model is more accurate and gives an efficient results.

The main contributions of our research work include:

- To the best of our knowledge, this is the first paper that analysis and identify the people opinion toward the COVID-19 vaccine.
- We have introduced a deep-leaning model that analysis and understand people feelings, and capture the frequent used words on social content.
- We have executed different state-of-the-art deep learning classifiers to extricate individuals feelings in social media during the new period.
- The main significance of this research work is that since we are using more than one language in predicting users opinion and emotion understanding in social media regarding COVID-19vaccine, it gives more flexibility and accurate results when compared to its single language counter part.

This paper is organized as follows: Section 2 presents the related works. Section 4.1 presents dataset description. Section 3 introduces the proposed methodology. Section 4 presents the experimental observations and detailed discussion. Finally, the conclusions and possible future works are introduced in Section 5.

2 Related work

As Novel Corona Pandemic COVID-19 reports a major warning to the international population health, collaboration is needed by all the nations to combat it [24]. The media and other web-based media channels should morally introduce the pertinent and right reports to expand inspiration among the overall population as opposed to introducing one-sided data; such coverage may only serve to further divide people and cause fear [16, 31].

Emotion analysis is the social occasion of individuals' perspectives with respect to any any event happening in real life [12, 26]. In such circumstances in which the world is at present experiencing, understanding the feelings of the individuals stands critical. The grave situation wherein individuals can't leave their homes requests investigating what the individuals is really being thinking about the entire situation. Thus, the authors have arranged this work on understanding the requesting circumstance particularly on social media [5, 29].

Merchant [25] explained the significance of spreading data through different social media stages during the outbreak of COVID-19. It discussed about that moving right and dependable data is fairly significant for the individuals if there should arise an occurrence of emergencies as online media is the indicative device and the most extreme reference framework. Individuals accept more on presents on when on get tried, how to manage the outcomes, and where to get care. Along these lines, exploring incorrectly and misdirecting data can in some cases be risky to somebody's life. It talked about the significance of correspondence as it is straightforwardly relative to individuals' conviction and response as needs be.

Li [20] explored the effects of COVID-19 on the psychological wellness of individuals by directing different tests on opinion examination utilizing long range informal communication locales. It was shown by tests that the information holes of diminutive term people change in mental conditions after the outbreak.

Sharma [30] conducted an inside and out investigation on shifted subjects comprising of assumptions and patterns with distinguishing bogus data and designated spots in spreading bogus data on pandemic COVID-19 through Twitter. The paper leaves open stamping remark for misrepresentation to improve classification techniques [19].

Previous researchers were focused on developing methods to achieve accurate for prediction of the spread of COVID-19 through twits. The main drawbacks of the previous research works were, they did not mention anything study to provide analysis of the emotions through the tweets regarding to COVID-19 vaccine. Therefore, this paper analysis and identifies the people opinion toward the COVID-19 vaccine and helps decision maker to improve and indicate an ascending trend for the cases in the future.

3 Methodology

Understand the user emotions in social media has gain a lot of attention from researcher recently. The goal of this study is to the impact of social media on the emotions and feelings of people regarding to COVID-19 vaccine. Emotion analysis can be used
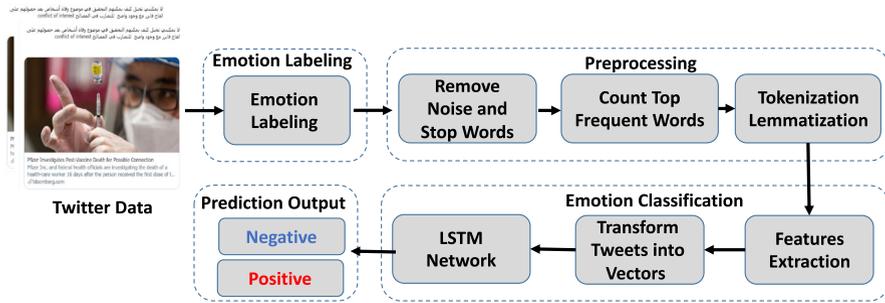
Fig. 2 The system architecture of the proposed model

here to identify expressed emotions from unstructured texts To efficiency understand the social media impacts, we propose a deep learning based model to understand people emotions. Figure 2 illustrates the framework architecture of our model. The proposed model consists of three main components namely: (1) Emotion Labelling; (2) Pre-processing; and (3) Emotion Classification. In the following sections, we will explain those components in details. The detailed steps for the proposed model are shown in Fig. 2.

3.1 Emotion labeling

The problem we address in this paper is to understand the people feeling toward the covid-19 vaccine. As shown in figure, the input to this component is a set of tweets regarding to COVID-19 vaccine crawled from Twitter API using the following hashtags: COVID-19 vaccine, vaccine, and Pfize. Then we asked PHD students to label the data in with 1 or 0 as positive or negative.

3.2 Pre-processing

Pre-processing step play an important role in our mode. To improve the quality of our model, the pre-processing step is done to clean the tweets. In this step, we first combine all tweets in Comma Separated Values (CSV) file has the tweets and it is cleaned by removing duplicate rows or similar tweets. Also, the tweets are cleaned by deleting the associated and redundant symbols such as @, RT, #, URLs, numeric values and punctuation marks. Initially,tokenization for words is done and counting the top frequent words, then the relevant features are extracted from the text available in the tweets to build and train the classification model.

3.3 Emotion classification

The problem we address in the paper is to understand the people emotions expressed regarding COVID-19 vaccine. This step receives the labelling data as input and then the feature extraction process is performed to capture the most important words.

In text mining, TF-IDF, stands for term frequency–inverse document frequency, is a statistic measure that evaluates how relevant a word is to a single document in a corpus of documents. The core idea behind TF-IDF is that each word that appear in a document has two scores: term frequency (TF) score and inverse document frequency (IDF) score. TF computes the number of times a term appears in a document, while IDF is the inverse document frequency of the word across a collection of documents. Multiplying those two score gives the TF-IDF weight of that term [34], where TF-IDF here reflects the weight of a term in a particular document. The higher the score of TF-IDF weight, the more important that term is in that document. In the proposed model, we apply the TF-IDF measure to extract and model different words in our dataset. The formula of TF-IDF is shown by the following equations [3, 34]:

$$TF(t) = \frac{\text{Number of times term t appears in a document}}{\text{Total number of terms in the document}}. \tag{1}$$

$$IDF(t) = \log_e \frac{\text{Total number of documents}}{\text{Number of documents with term t in it}}. \tag{2}$$

Subsequently, the TF-IDF value can be calculated by taking a product of the two statistics: TF and IDF. Intuitively, this means:

$$TF - IDF(t) = TF(t) * IDF(t) \tag{3}$$

As the machine learning or deep learning models do not understand the text data. These algorithms usually deal with numbers, and natural language is text data. So text data need to transform into numbers, this process known as text vectorization. It is a fundamental step in the process of machine learning for analyzing data, and different vectorization algorithms will drastically affect final results. A term will be considered highly relevant in a document and assigned a higher weight when it has a high TF value (in that document) and a low document frequency in the whole corpus. The IDF value in that situation will be greater than 1 (since the log value is greater than 1), and thus the overall TF-IDF value will be large. After words transformed into numbers, the TF-IDF score can be fed to machine learning algorithms such as Naive Bayes and Support Vector Machines.

TF-IDF is useful in our task as it provides a vector representation of tweets. Once we computed the TF-idf for our dataset, the next step aims at analysing the users emotions toward the vaccine. In this work, we have used a deep learning model namely LSTM which is a variant of a Recurrent Neural Network (RNN) and it is a powerful tool in predicting sequence and time-series data-related problems. LSTM model was selected in this work as COVID-19 dataset is also time-bound. The LSTM has a special feature that enables to store previous information and predict the future trend of the virus under consideration. The LSTM model takes in the input features and predicts the class label. The LSTM model that we have used in this study has the following configurations. It uses 300 epochs and a loss function namely sparse categorical crossentropy with *adam* optimizer. The softmax activation function is used in output layer. The total number of tunable parameters used in the proposed model are summarized in Table 1.

Table 1 The description of the proposed LSTM model

Layer (type)	Output Shape	Param #
Embedding (Embedding)	(None, 20, 100)	1000000
lstm (LSTM)	(None, 64)	42240
Dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 2)	130
Total params		1,042,370
Trainable params		1,042,370

Finally, the softmax activation function is used to classify the people emotions to either positive or negative. The mathematical form of this activation function is presented in Eq. 4.

$$S(x_i) = \frac{\exp^{x_i}}{\sum_{j=1}^{n} \exp^{x_j}} \tag{4}$$

where x_i is input vector, n is number of classes in the multi-class classifier and $\exp^{x_i}$ is standard exponential function for output vector. The output of our model is discussed in details in the following section.

4 Experimental results and evaluation

The model aims to classify emotion into positive or negative scores. During training the model, the tweets are tokenized into words and then transformed into a feature vector. The main models developed and tested for the framework are Long Short-Term Memory (LSTM), Naive Bayes, decision tree and SVM. Natural language Tooklit (NLTK) has been widely used recently as it provides an implantation of various traditional and advanced machine leaning algorithm. Hence, in our experiment, we use the available machine leaning mentioned above in NLTK to compare the result of our model.

In the following subsections, we first discuss the dataset used in this paper and then demonstrate the experimental results in term of simulated results and tabular form. We mainly compare the output of model with well-known approaches and the results show that our model can successfully capture the people emotions expressed in social contents

4.1 Dataset description

This section describes the collected dataset from Twitter. The dataset contains around 50,000 tweets that have been published from the release of the vaccine till 1st January 2021. Further, The pre-processing step is done to make text lowercase,remove punctuation emoji's and stop-words. After that, the labeling step is applied to annotate the tweet as positive or negative.
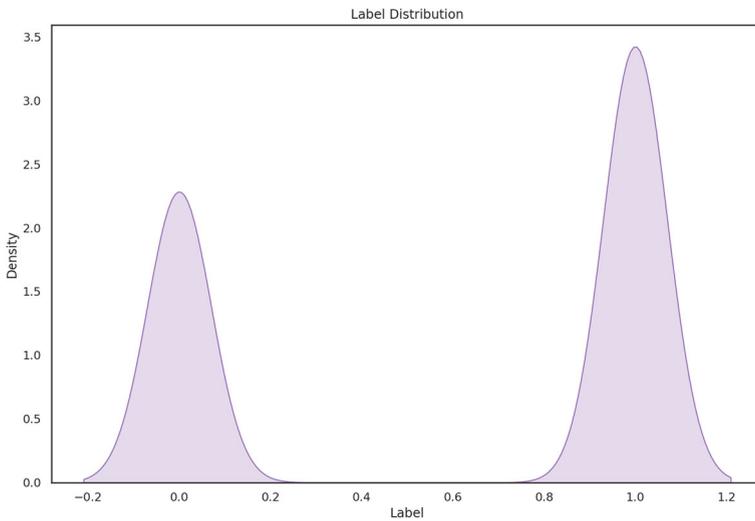
Fig. 3 The wordcloud for the top words of Twitter data

Fig. 4 The wordcloud for top negative Emotion

The most frequent words used in the dataset to show people feelings is visualized using the WordCloud figures [4]. These figures display words that do not prove any efficiency in representing a viable solution during pandemic. Figure 3 shows the wordcould for the top words of Twitter data regards hope and fair toward the Pfizer vaccine. These top frequent words can indicate the common words used in social media to describe people feelings towards COVID-19 vaccine.

Figures 4 and 5 show the wordcloud top words of Twitter datat regard hope and fair toward Pifzer vaccine. The top frequent positive words for example are: آمن (safe), ثقة (trust), and انتصار (victory). While, the top negative words are: خطر (danger), اصابة (infected) and اغماء (fainting).

Table 2 shows the top frequent words in the whole dataset. The top frequent words are لقاح (Vaccine) and كورونا (Corona). While the word فايزر (Pfizer) is repeated 6380 times in the dataset. Figure 6 presents the distribution of the positive and negative labels. From this figure, it is clear that the majority of labels is positive regards the COVID-19's vaccine.

Fig. 5 The wordcloud for top positive words

Table 2 The top 8 words in the dataset

Word	Translation	Count
لقاح	Vaccine	19447
كورونا	Corona	15201
فايزر	Pfizer	6380
الصحة	Health	2832
ستاخذون	They will take	2091
الله	Allah	1633
وزير	Minister	1244
السعودية	Saudi	1226

4.2 System evaluation

The classification accuracy is not the best metric in the case of class imbalanced datasets, and thus we chose to use additional performance metrics to mitigate the issue. Therefore, four additional performance metrics are used as additional performance measurements such as precision, sensitivity (recall), F1-score, and ROC curve. The Eqs. 5, 6, 7, and 8 are the mathematical expressions of Accuracy, F1-score, recall, and accuracy determined on the basis of True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) test data samples. [13].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

$$F1 - score = 2\frac{precision \star recall}{precision + recall} \tag{6}$$

Fig. 6 The distribution of the positive(1) and negative(0) labels

$$Recall = Specificity = \frac{TP}{TP + FP} \qquad (7)$$

$$Precision = \frac{TP}{TP + FN} \qquad (8)$$

4.3 Machine learning results

Experiments are carried out using a dataset obtained through the following hashtags vaccine, corona, coronavirus, and COIVD-19. The collected dataset includes around 50,000 tweets. To fit the model, the final version of the transformed data split into train and test sets such that 70% data for the training set, 30% for the testing set. To evaluate the performance of the selected classifiers, unigram, bigram, trigram and word embedding using word2vec have been performed.

After the texts have been changed into vectors, classification are performed. This study includes hyper-parametric classifiers from k-nearest neighbour (KNN) [17], Naïve Bayes (NB), AdaBoost (ADB), Decision tree (CART) [8], Random Forest (RF) [13], Support Vector Machine (SVM) [36], and Linear Regression Classifier (LR) [15].

In general, Linear classifiers such as SVC, logistic regression, and LR are more efficient for text data classification tasks. Also, the linear method includes linear functions to forecast the input features. Therefore, to achieve high accuracy, a strong classifier must be developed. The K-fold Cross-Validation technique is used to calculate the average accuracy metric of the model by creating many different samples of the data. Moreover, The k-fold is a common form of cross-validation technique in which, for example, if $k = 10$, 9 folds are used for training and 1 fold is used for testing the model, and this process is repeated until all folds get a chance to be the test set one by one as shown in Fig. 7. The performance measure reported by k-fold cross-validation
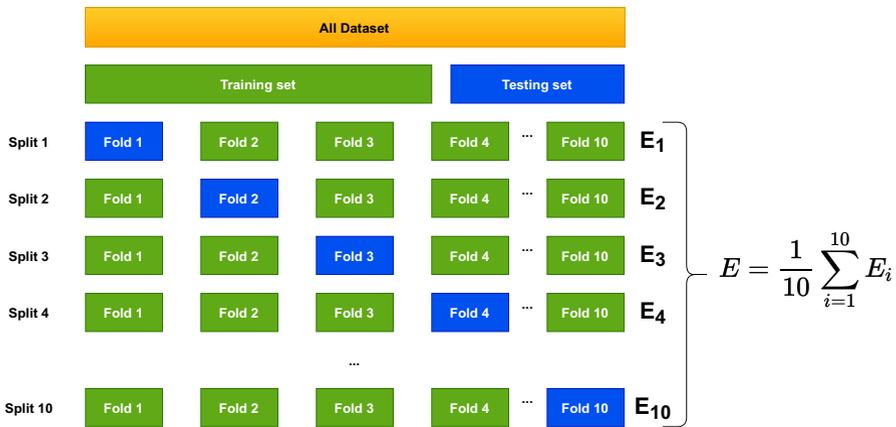
Fig. 7 The K-fold Cross-Validation technique

$$E = \frac{1}{10} \sum_{i=1}^{10} E_i$$

Table 3 The performance of SVM model

	Precision	Recall	F1-score	Support
0	0.75	0.72	0.74	2062
1	0.82	0.84	0.83	3036
Macro avg	0.78	0.78	0.78	5098
Weighted avg	0.79	0.79	0.79	5098
Accuracy			0.79	5098

is then the average of the values computed in the loop. Thus, it gives us a good idea and clear understanding of the model's generalization ability, which is particularly useful when we have small data and can not afford to break it into training and testing sets.

Table 3 shows the values of recall, precision, and F1-score for the SVM model. The proposed SVM model used for the determination of the Emotion for class 1 outperforms class 0. Moreover, the proposed method is biased due to it exhibits high sensitivity towards major class 1. Also, to calculate the comprehensive measurement of the performance, F1-score is used to compare the accuracy of the technique. From the performance measures values, it is clear that the SVM approach provided remarkable results in predicting Emotion from tweets.

Similarly, Table 4 presents the values of Precision, Recall, and F1-score for the experimental results of the Naive Bayes model. The Naive Bayes model used for classifying the Emotion into two classes namely class 1 and class 0. The Naive Bayes method is also biased towards major class 1. Finally, in Table 4 the performance of the Naive Bayes model gives maximum accuracy of 79%, the precision of 78% and F1-score of 79%.

Figure 8 presents the comparison of machine learning methods. The comparative study is carried in the following Machine learning models namely Logistic Regression (LR), K-Nearest Neighbors (KNN), AdaBoost (ADB), Decision Tree (CART), GaussianNB (NB), Random Forest (RF), and SVC (SVM). The performance of each model is assessed utilizing arrangement measurements such as precision, recall, f1-

Table 4 The performance of NB model

	Precision	Recall	F1-score	Support
0	0.74	0.73	0.74	2062
1	0.82	0.82	0.82	3036
Macro avg	0.78	0.78	0.78	5098
Weighted avg	0.79	0.79	0.79	5098
Accuracy			0.79	5098
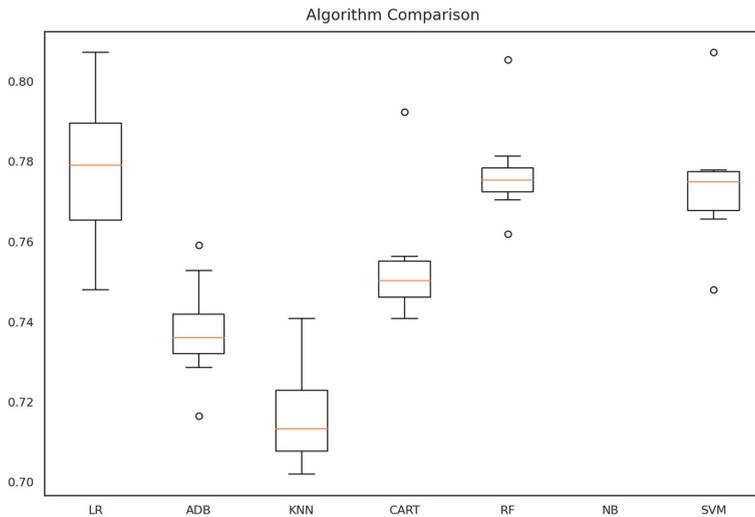
Fig. 8 Comparison of machine learning methods

score and accuracy on the test dataset. Based on the experimental results of the LR model, it is outperforms compared with the other models.

The outcomes in Fig. 8 show an improvement in the accuracy when a comparison is made between the different existing models. Improvement in the results of performance comparison can be enhanced through utilizing grid search strategies to select the most effective of the proposed techniques.

4.4 Long short term memory (LSTM)

Table 5 presents the accuracy achieved by the state-of-the-art LSTM model being actualized on the gathered information. For example, all tweets are represents using tfidf method and implemented with the unigram. Because of the dataset has different numbers of positive, and negative tweets, LSTM model gives the highest accuracy of 98% for training set and 73% testing set with unigram under the Tfidf technique. Figures 9, 10 show the plot for the loss and accuracy of the LSTM model.

Table 5 Long Short Term Memory (LSTM)		Train	Test
	Loss	0.0459	3.6285
	Accuracy	0.9761	0.732

Training and validation loss
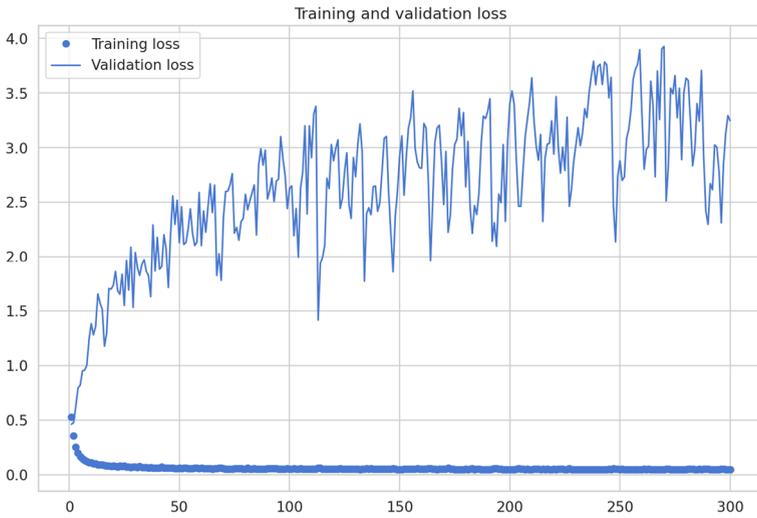
Fig. 9 The loss values of LSTM model for the training and testing sets
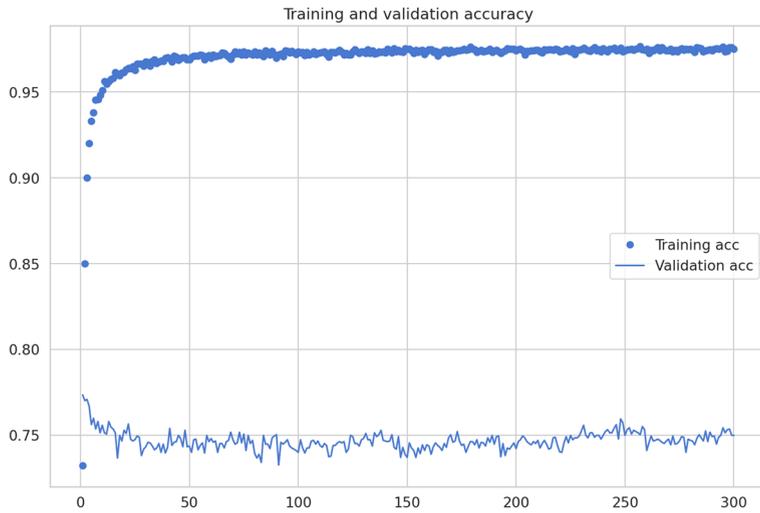
Training and validation accuracy

Fig. 10 The accuracy of LSTM model for the training and testing sets

5 Conclusion

This paper has provided an analysis of the emotions through the tweets regarding to COVID-19 vaccine. The release of COVID-19 vaccine brings out the fact that a lot of people in social media starts talking about safety of vaccine. People often use specific words to express their feelings and emotions towards any topics they discussed in social media. Hence, in this paper, we present a LSTM model that provides an analysis of such emotions expressed by people regarding the COVID-19 vaccine. The proposed model first received the Twitter data as inputs. Then, the classifying process is performed on the tweets to identify the people emotions. A LSTM model then is used to train the data and classify the people opinion about COVID-19 vaccine to either positive or negative. We evaluate our model in Twitter dataset, and the result shows that our model outperforms the current approaches for emotion analysis in term of accuracy. Our proposed deep learning classifiers has achieved accuracy up to 98% for training set and the accuracy for testing set is 73%. In the future work, we will focus on applying the proposed model in other type of vaccine.

References

1. AlphaGo: AlphaGo AI wins contest against go world champion lee sedol. Phys Today (2016). https://doi.org/10.1063/pt.5.029653
2. Apolinario-Arzube Ó, García-Díaz JA, Medina-Moreira J, Luna-Aveiga H, Valencia-García R (2020) Comparing deep-learning architectures and traditional machine-learning approaches for satire identification in spanish tweets. Mathematics 8(11):2075. https://doi.org/10.3390/math8112075
3. Atlam ES, Fuketa M, Morita K, ichi Aoe J (2003) Documents similarity measurement using field association terms. Inf Proces Manag 39(6):809–824. https://doi.org/10.1016/s0306-4573(03)00019-0
4. Bashri MFA, Kusumaningrum R (2017) Sentiment analysis using latent dirichlet allocation and topic polarity wordcloud visualization. In: 2017 5th International Conference on Information and Communication Technology (ICoIC7). IEEE. https://doi.org/10.1109/icoict.2017.8074651
5. Chakraborty K, Bhattacharyya S, Bag R (2020) A survey of sentiment analysis from social media data. IEEE Trans Comput Soc Syst 7(2):450–464. https://doi.org/10.1109/tcss.2019.2956957
6. Chen W (2020) Promise and challenges in the development of COVID-19 vaccines. Human Vaccin Immunotherapeut 16(11):2604–2608. https://doi.org/10.1080/21645515.2020.1787067
7. Ciresan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE. https://doi.org/10.1109/cvpr.2012.6248110
8. Clarin JA (2020) Academic analytics: Predicting success in the licensure examination of graduates using CART decision tree algorithm. J. Adv. Res. Dyn. Control Syst. 12(01–Special Issue):143–151. https://doi.org/10.5373/jardcs/v12sp1/20201057
9. van Dijck J, Alinead D (2020) Social media and trust in scientific expertise: Debating the covid-19 pandemic in the netherlands. Social Media + Society 6(4):205630512098105. https://doi.org/10.1177/2056305120981057
10. Ellison NB, Gray R, Lampe C, Fiore AT (2014) Social capital and resource requests on facebook. New Media Soc 16(7):1104–1121. https://doi.org/10.1177/1461444814543998
11. FDA: Vaccines and Related Biological Products Advisory Committee December 10, 2020 Meeting Announcement. (December 10, 2020). https://www.fda.gov/advisory-committees/advisory-committee-calendar/vaccines-and-related-biological-products-advisory-committee-december-10-2020-meeting-announcement#event-materials
12. Fersini E (2017) Sentiment analysis in social networks. In: Sentiment Analysis in Social Networks. Elsevier, pp 91–111. https://doi.org/10.1016/b978-0-12-804412-4.00006-1

13. Gad I, Hosahalli D (2020) A comparative study of prediction and classification models on NCDC weather data. Int J Comput Appl. https://doi.org/10.1080/1206212x.2020.1766769
14. Gad I, Hosahalli D, Manjunatha BR, Ghoneim OA (2020) A robust deep learning model for missing value imputation in big NCDC dataset. Iran J Comput Sci. https://doi.org/10.1007/s42044-020-00065-z
15. Hosahalli D, Gad I (2018) A generic approach of filling missing values in NCDC weather stations data. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, pp 143–149. https://doi.org/10.1109/icacci.2018.8554394
16. Hui DS, Azhar EI, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A, Petersen E (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in wuhan, china. Int J Infect Diseases 91:264–266. https://doi.org/10.1016/j.ijid.2020.01.009
17. Jivani AG, Shah K, Koul S, Naik V (2016) The adept k-nearest neighbour algorithm—an optimization to the conventional k-nearest neighbour algorithm. Trans Mach Learn Artif Intell. https://doi.org/10.14738/tmlai.41.1876
18. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386
19. Latif S, Usman M, Manzoor S, Iqbal W, Qadir J, Tyson G, Castro I, Razi A, Boulos MNK, Weller A, et al (2020) Leveraging data science to combat covid-19: A comprehensive review. TechRxiv
20. Li S, Wang Y, Xue J, Zhao N, Zhu T (2020) The impact of COVID-19 epidemic declaration on psychological consequences: A study on active weibo users. Int J Environ Res Public Health 17(6):2032. https://doi.org/10.3390/ijerph17062032
21. Lin C, Tu P, Beitsch LM (2020) Confidence and receptivity for COVID-19 vaccines: A rapid systematic review. Vaccines 9(1):16. https://doi.org/10.3390/vaccines9010016
22. Malki Z, Atlam E, Dagnew G, Alzighaibi AR, Ghada E, Gad I (2020) Bidirectional residual LSTM-based human activity recognition. Comput Inf Sci 13(3):40. https://doi.org/10.5539/cis.v13n3p40
23. Malki Z, Atlam ES, Hassanien AE, Dagnew G, Elhosseini MA, Gad I (2020) Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. Chaos Solitons Fractals 138:110137. https://doi.org/10.1016/j.chaos.2020.110137
24. Malki Z, Elsayed, Ewis A, Dagnew G, Reda A, Elmarhomy G, Elhosseini MA, Hassanien AE, Gad I (2020) ARIMA models for predicting the end of COVID-19 pandemic and the risk of a second rebound. Neural Comput Appl. https://doi.org/10.21203/rs.3.rs-34702/v1
25. Merchant RM, Lurie N (2020) Social media and emergency preparedness in response to novel coronavirus. JAMA 323(20):2011. https://doi.org/10.1001/jama.2020.4469
26. MK S (2020) Social media sentiment analysis for opinion mining. Int J Psychosoc Rehabilit 24(5):3672–3679. https://doi.org/10.37200/ijpr/v24i5/pr202075
27. Parvathy SS, Bhuvaneswari M (2019) Handling of voluminous tweets and analyzing the sentiment of tweets. In: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE. https://doi.org/10.1109/icaccs.2019.8728439
28. Pruß BM (2021) Current state of the first COVID-19 vaccines. Vaccines 9(1):30. https://doi.org/10.3390/vaccines9010030
29. Risnantoyo R, Nugroho A, Mandara K (2020) Sentiment analysis on corona virus pandemic using machine learning algorithm. J Inf Telecommun Eng 4(1):86–96. https://doi.org/10.31289/jite.v4i1.3798
30. Sharma K, Seo S, Meng C, Rambhatla S, Dua A, Liu Y (2020) Coronavirus on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309
31. Stechemesser A, Wenz L, Levermann A (2020) Corona crisis fuels racially profiled hate in social media networks. EClinicalMedicine 23:100372. https://doi.org/10.1016/j.eclinm.2020.100372
32. Tanne JH (2020) Covid-19: FDA panel votes to authorise pfizer BioNTech vaccine. BMJ m4799. https://doi.org/10.1136/bmj.m4799
33. (WHO), W.H.O.: Coronavirus Disease 2019 (COVID-19) Situation Report - 97. 2020. (2020 (accessed April 27, 2020)). https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200426-sitrep-97-covid-19.pdf?sfvrsn=d1c3e800_6
34. Wu HC, Luk RWP, Wong KF, Kwok KL (2008) Interpreting TF-IDF term weights as making relevance decisions. ACM Trans Inf Syst 26(3):1–37. https://doi.org/10.1145/1361684.1361686
35. Yoda T, Katsuyama H (2021) Willingness to receive COVID-19 vaccination in japan. Vaccines 9(1):48. https://doi.org/10.3390/vaccines9010048

36. Yuan H (2008) Paper sheet defects classification based on support vector machine method. J Comput Appl 28(2):330–332. https://doi.org/10.3724/sp.j.1087.2008.00330
37. Zhong B, Huang Y, Liu Q (2020) Social media support during COVID-19 measure. https://doi.org/10.1037/t78465-000

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