ISSN No:-2456-2165

# Deep Learning Based Technique for Covid-19 Vaccination Sentiments Prediction

[1]Ahmed Mohammed, [2]Dr. A. Pandian
Department of Computing Technologies
SRM institute of Science and Technology,
Kattankulathur, Tamil Nadu, India

**Abstract:- The COVID-19 pandemic has affected a large number of people, causing great worry, fear, and conflicting feelings or emotions. It has elevated our understanding of the world to unprecedented heights. COVID-19 is rapidly spreading, and the only way to halt it is for the entire population to be vaccinated. However, there is still concern about vaccinations among the general public. From the beginning of vaccinations, many people have refused to have vaccines injected into them. Using survey data acquired via Google form, deep learning techniques were used to build a model for sentiment classification and prediction of COVID-19 vaccination. Public perceptions towards the COVID-19 vaccine were analyzed using natural language processing (NLP) and deep learning techniques. The dataset's responses were 42.60% positive, 35.74% negative, and 21.66% neutral. A convolutional neural network (CNN) and long short-term memory (LSTM) were used. The LSTM algorithm performed better than the CNN algorithm. The average accuracy scores obtained for CNN and LSTM sentiment classification and prediction models were 68% and 93%, respectively. As evaluation metrics, accuracy, precision, recall, and f-measure were used. This research demonstrates the application of deep learning techniques to sentiment analysis tasks involving the COVID-19 vaccine.**

***Keywords:-*** *Deep Learning, Natural Language Processing, Sentiment Prediction, Covid-19 Vaccination, Convolutional Neural Network, and Long Short-Term Memory.*

## I. INTRODUCTION

Sentiment analysis is the computer study of people's opinions, feelings, emotions, evaluations, and attitudes about things such as goods, services, organizations, people, topics, events, themes, and their attributes. Because we now have a vast amount of opinionated data recorded in digital formats for the first time in human history, the field's genesis and rapid development parallel those of social media on the Web, such as reviews, forum debates, blogs, microblogs, Twitter, and social networks. Since its start in the early 2000s, sentiment analysis has been one of the most active research domains in natural language processing [1].

## II. LITERATURE SURVEY

If you want to use a backpropagation method to train an artificial neural network, you can use a multi-layer customer feedback architecture that includes word embedding and compositional vector models to do that [2].

Sentiment analysis of Sinovac and Pfizer vaccines in Indonesia will be carried out through Twitter in this study, according to the authors. Data scraping and analysis were done in October and November 2020 in order to better understand user perceptions. Each dataset was manually tagged as either positive, negative, or neutral. Using Twitter crawling, a 10-fold cross validation, and the right labeled predictions, we used SVM, Naive Bayes, and Random Forest to look at the results. Then, we gave the results with the right labeled predictions [3].

For example, a feed forward neural network model may be used to learn word associations from a dataset, allowing it to locate synonyms and suggest new words given a partial sentence. It uses continuous bag-of-words (CBOW) or continuous skip-gram model architectures to build a distributed representation of words. Using this method, each word in the corpus is represented by a large vector using this method in order to save all of the semantic information and word linkages. Word2vec has been shown to be able to tell the meaning of two phrases that only have a few words in common [4].

Twitter data was utilized to study public opinion. The purpose of this study was to use machine learning techniques to identify the most important concerns and emotions expressed on Twitter about the COVID-19 vaccination. COVID-19 and vaccination sentiments, COVID-19 infection control measures advocacy, and COVID-19 control misconceptions and complaints were the three main areas of focus for the team. [5] Between January and October 2020, they got 31,100 English tweets from Australian Twitter users who used COVID-19-related keywords in them.

At the time they were established, Pfizer, AstraZeneca, and Johnson & Johnson were the three most renowned vaccine research businesses. From the start of the immunization, researchers collected data from Instagram users who used their own hashtags to post information about their experiences. The photos were classified using VGG-16, Inception V3, and ResNet50 after being extracted with Instaload. Algorithms were tested in a controlled experiment in order to validate their accuracy and select the two top performers. Users who post controversial

comments are more likely to have neutral to negative opinions, according to research that used a convolutional neural network (CNN) [6].

## III. METHODOLOGY

Sentiment analysis may identify and predict opinions as positive, negative, or neutral, in addition to text classification. Because traditional text classification techniques fall short of revealing hidden features, as a result, deep learning algorithms can accurately classify the perception. The use of sentiment analysis in a variety of industries, including market research, corporate intelligence, e-government, email filtering, social media content moderation, and others, is driving up demand. Deep learning is increasingly being used to overcome this challenge.

Our proposed deep learning model for sentiment classification and prediction on public perception text data is illustrated in Figure 1.

### *A. Proposed Model Block Diagram*

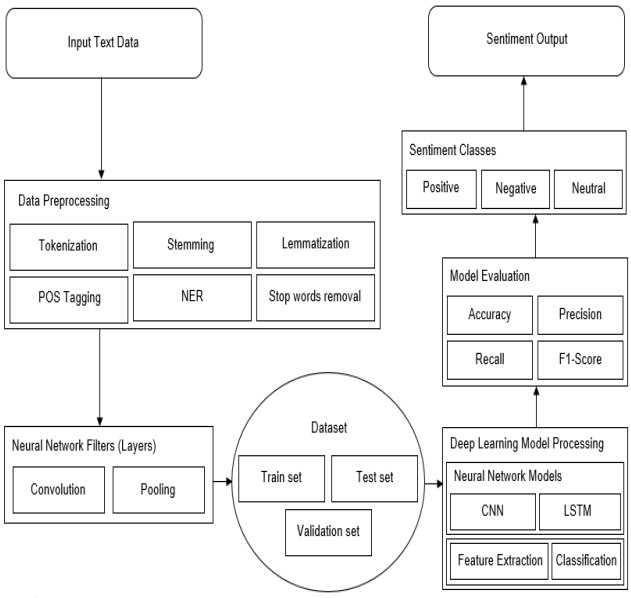


Fig. 1: Deep learning sentiment analysis proposed model block diagram

### *B. Dataset*

In this paper, the dataset used was gathered through a Google form. The dataset contains a collection of perception texts with negative, positive, and neutral classes, respectively.

### *C. Data Pre-processing*

Data preprocessing is crucial in deep learning [7]. Most perception texts are unstructured in nature and made up of incomplete phrases with a lot of noise and phrasing that has bad structure, such as bad syntax, imperfect vocabulary, and a lot of repetition. Unstructured data also has an impact on sentiment classification performance. To begin, we must perform a series of preprocessing operations on the perception text in order to reduce errors and maintain a consistent structure.

### *D. Neural Network*

Neural networks are mathematical models that store data using brain-inspired learning processes. The phrase "artificial neural network" is used to refer to neural networks that are used in machines. Machine learning is a scientific subject that focuses on the research and development of algorithms that enable computers to learn from data, such as sensor data or databases, and is widely used in a wide variety of industries today. Machine-learning research is focused on automatically learning to detect complex patterns and make intelligent judgments based on such data [8].

## IV. RESULT AND DISCUSSIONS

After data analysis, Python and the Keras framework were used to develop models and test them. The model was deployed using the Streamlit framework. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) were the algorithms used.

### *A. Experimental Procedure*

Convolutional Neural Networks and Long Short-Term Memory were utilized in conjunction with Python and the Keras framework to analyze sentiments regarding the COVID-19 vaccine. The dataset, which comprises three classes: positive, negative, and neutral, was used to train and test the proposed sentiment classification model. The dataset was split into training and test sets in a ratio of 80% for training and 20% for validation. The experiments were run on Google Colaboratory, a free Jupyter notebook. A deep learning model was trained, tested, and saved in the Google Colaboratory using Python and other libraries such as Keras. The following are the experimental steps for training and testing the proposed deep learning model:

- The length of the input sequences is specified by the input layer.
- Embedding layer has a vocabulary size of 3000.
- There are two convolution1D layers, each with its own set of filters and kernel size.
- There is single LSTM layer with 64 neurons.
- Filters and a kernel size for a convolution1D layer.
- Use the MaxPooling1D layer to stabilize the convolution layer's output.
- 0.5 was used as dropout.
- Obtain the output by flattening the output layer.
- Number of epochs used were 2500.

This channel is processed by a dense layer with a number of neurons and a Softmax activation function, as well as an output layer with one neuron and a Sigmoid activation function, which all reach into a single vector. This model uses an embedding layer as the hidden layer after the input layer. The vocabulary size, real value vector space size, and maximum length of input documents are all required by the embedding layer. With kernel size 4 and the Relu activation function, we employed two convolution1D layers with 64 and 32 filters for CNN and 64 filters for LSTM, respectively.

ISSN No:-2456-2165

| Sentiment Class | Precision | Recall | F1-measure |
|---|---|---|---|
| Positive | 0.68 | 0.88 | 0.77 |
| Negative | 0.69 | 0.64 | 0.67 |
| Neutral | 0.67 | 0.29 | 0.40 |

Table 1: Evaluation Report for Convolutional Neural Network Model.

| Sentiment Class | Precision | Recall | F1-measure |
|---|---|---|---|
| Positive | 0.94 | 0.88 | 0.77 |
| Negative | 0.94 | 0.94 | 0.94 |
| Neutral | 0.89 | 0.84 | 0.86 |

Table 2: Evaluation Report For Long Short-Term Memory Model

*B. Convolutional Neural Networks*

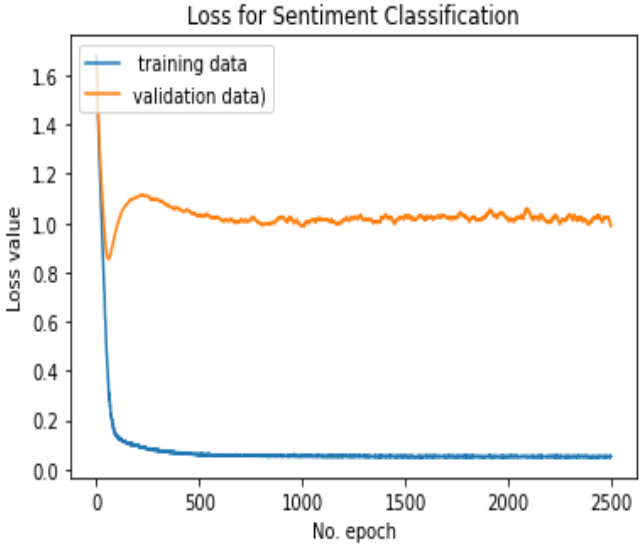


Fig. 2: Loss result for training and validation set for CNN model

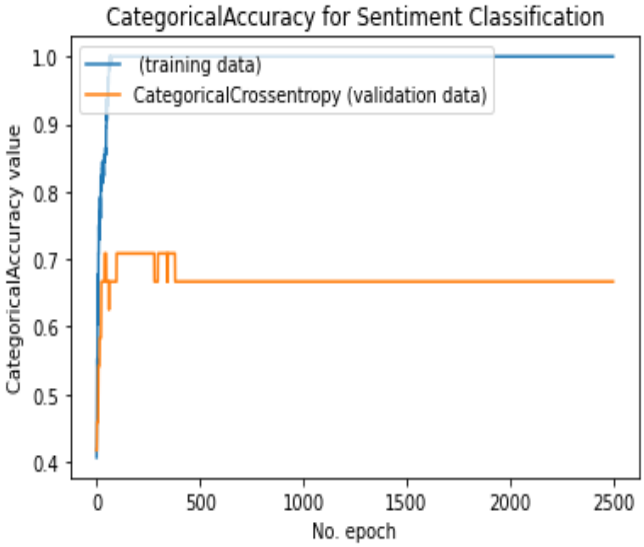


Fig. 3: Accuracy result for training and validation set for CNN model

*C. Long Short-Term Memory*

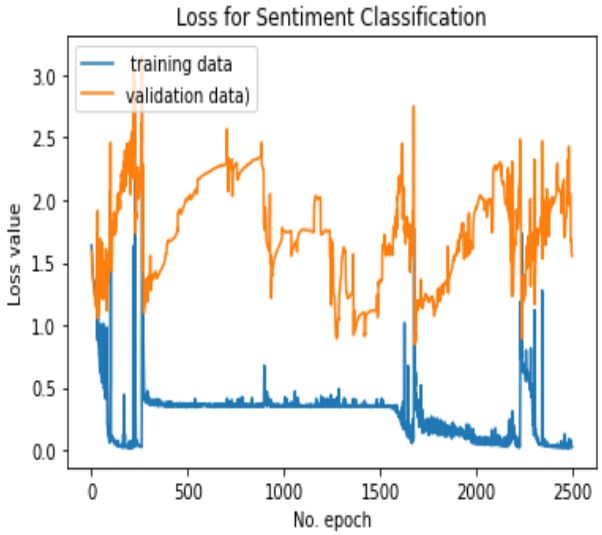


Fig. 4: Loss result for training and validation set for LSTM model.

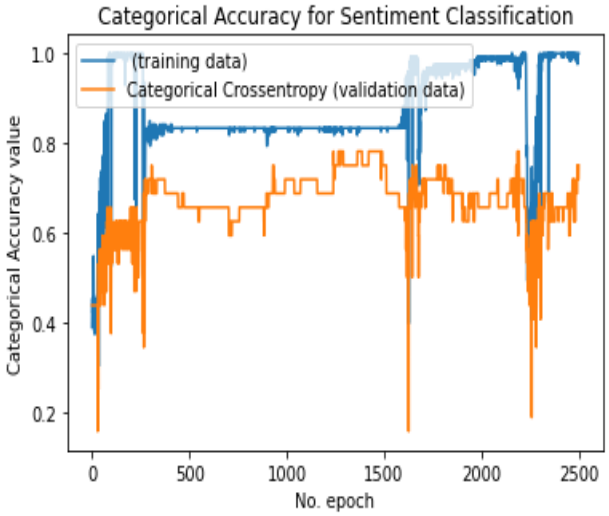


Fig. 5: Accuracy result for training and validation set for LSTM model

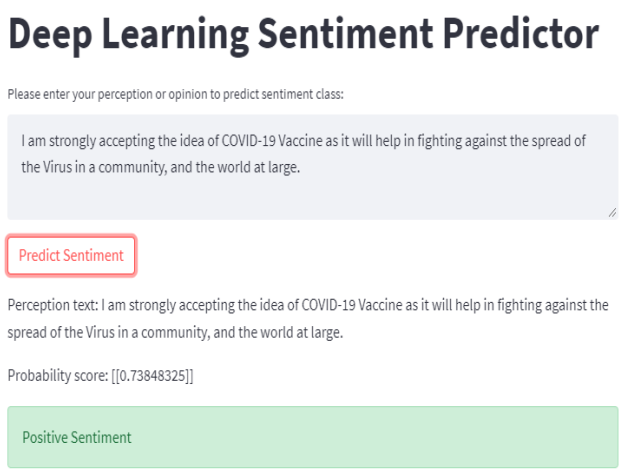


Fig. 6: Positive sentiment predicted with 0.74 probability threshold value

ISSN No:-2456-2165

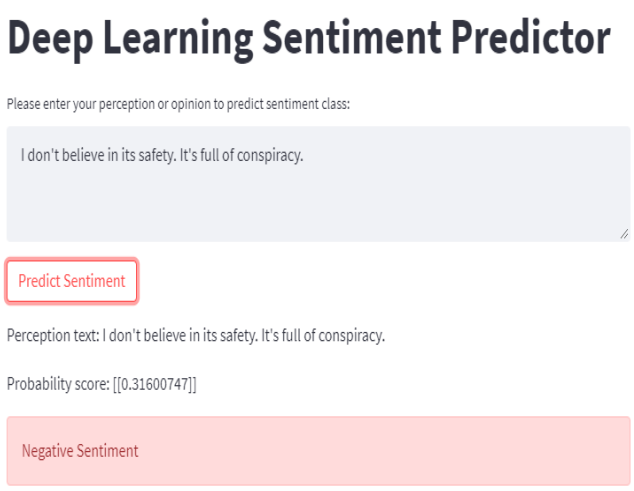


Fig. 7:Negative sentiment predicted with 0.32 probability threshold value

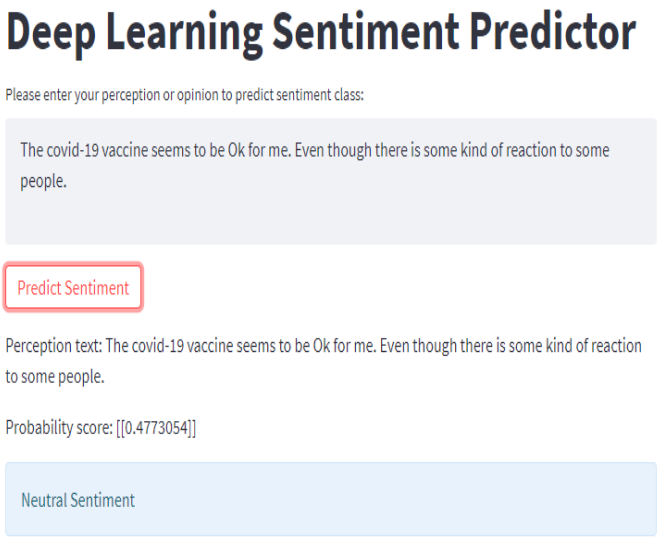


Fig. 8: Neutral sentiment predicted with 0.48 probability threshold value

## V. CONCLUSION

Our study demonstrates the use of deep learning approaches to sentiment analysis tasks regarding COVID-19 vaccination. Public perceptions towards the COVID-19 vaccine were analyzed using natural language processing (NLP) and deep learning techniques. The dataset's responses were 42.60% positive, 35.74% negative, and 21.66% neutral. A convolutional neural network (CNN) and long-term memory (LSTM) were used. The LSTM algorithm performed better than the CNN algorithm. The average accuracy scores obtained for CNN and LSTM sentiment classification and prediction models were 68% and 93%, respectively. As evaluation metrics, accuracy, precision, recall, and f-measure were used.

## ACKNOWLEDGMENT

I am grateful to Dr. A. Pandian, who served as my project guide, for sharing his expertise with me during this work. We would like to express our gratitude to the SRM Institute of Science and Technology in Chennai, India, for providing a suitable environment for conducting the study.

## REFERENCES

[1.] L. Zhang, S. Wang and B. Liu, "Deep learning for sentiment analysis: A survey", WIREs Data Mining and Knowledge Discovery, vol. 8, no. 4, 2018. Available: 10.1002/widm.1253.

[2.] D. Pham and A. Le, "Learning multiple layers of knowledge representation for aspect based sentiment analysis", Data & Knowledge Engineering, vol. 114, pp. 26-39, 2018. Available: 10.1016/j.datak.2017.06.001.

[3.] D. A. Nurdeni, I. Budi, and A. B. Santoso, “Sentiment analysis on Covid19 vaccines in Indonesia: from the perspective of Sinovac and Pfizer,” in 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), pp. 122–127, Surabaya, Indonesia, 2021.

[4.] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. “Distributed representations of words and phrases and their compositionality”. arXiv preprint arXiv:13104546. 2013. https://doi.org/10.48550/arXiv.1310.4546.

[5.] S. W. H. Kwok, S. K. Vadde, and G. Wang, “Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: machine learning analysis,” Journal of Medical Internet Research, vol. 23, no. 5, article e26953, 2021.

[6.] D. Amanatidis, I. Mylona, I. E. Kamenidou, S. Mamalis, and A. Stavrianea, “Mining textual and imagery Instagram data during the COVID-19 pandemic,” Applied Sciences, vol. 11, no. 9, p. 4281, 2021.

[7.] N. Karimimehr, A. Beheshti Shirazi and M. Ba haghighat, "Fingerprint image enhancement using Gabor wavelet transform", 2010 18th Iranian Conference on Electrical Engineering, 2010. Available: 10.1109/iraniancee.2010.5507055.

[8.] N. Keijsers, "Neural Networks", Encyclopedia of Movement Disorders, pp. 257-259, 2010. Available: 10.1016/b978-0-12-374105-9.00493-7.