INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 31-05-2022

Accepted: 05-02-2023

Published: 02-03-2023

Citation: Sujatha E, Radha R (2023) New Education Policy 2020: A Sentiment Classification. Indian Journal of Science and Technology 16(9): 614-621. https://doi.org/10.17485/IJST/v16i9.1164

* **Corresponding author**.

sujatha.ravi19@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

New Education Policy 2020: A Sentiment Classification

E Sujatha[1]*, R Radha[2]

1 Research Scholar, Research Dept. of Computer Science, SDNBV College for Women, University of Madras, Chrompet, Chennai, 600044, India
2 Associate Professor, Research Dept. of Computer Science, SDNBV College for Women, Chrompet, Chennai, 600044, India

Abstract

Objectives: To develop a model of multi-class classification which provides better performance for the large dataset. To reduce complexity of the model and to analyse the sentiments of twitter data in an efficient way. **Methods:** The sentiment analysis has been performed on the New Education Policy 2020. Totally, 105045 tweets were collected from the Twitter database using Tweepy library in python. The sentiment analysis was done on English tweets. The pre-processing and feature extraction was done by using pyspark packages. The hybrid of unigram and bigrams feature sets was used. To extract the labelled dataset, AFINN dictionary was used. The classifiers such as Random Forest in Machine Learning and Convolutional Neural Network, Bidirectional Long Short-Term Memory in Deep Learning were used to determine positive, negative and neutral sentiments of tweets. **Findings:** The Accuracy (97%), Precision (97%), Recall (97%), F-Measure (97%) and 99% of ROC-AUC with the minimum Log Loss 0.10 was obtained by the hybrid of Convolutional Neural Network and Bidirectional Long Short-Term Memory. **Novelty :** The complexity of the model was reduced by using Convolutional Neural Network which selects the relevant features. The performance of the model was evaluated by using the various metrics such as accuracy, precision, recall, f-score, log loss and roc-auc whereas in the existing works only limited metrics were used. The efficiency of the proposed model can be proved in any case.

Keywords: Random Forest Classifier (RF); Convolutional Neural Network (CNN); Bidirectional Long Short-Term Memory (BLSTM); Support Vector Machine (SVM); Term Frequency – Inverse Document Frequency (TF-IDF)

1 Introduction

As the Machine Learning, Artificial Intelligence and Natural Language Processing technologies are flourishing in recent day's sentiment analysis is becoming very popular. Dealing with the huge amounts of data, analysing the sentiments will be a tedious one. But with the Artificial Intelligence techniques insights can be gained quickly from a large volume of texts. There exist numerous applications of sentiment analysis such as marketing, e-commerce, research and politics. But this field is still in its infancy.

The sentiment of 7,345 reviews has been analysed by using SVM classifier. From the reviews the important aspects have been extracted and analyse the corresponding sentiments. The proposed model of aspects-based classification obtains the better results of 93% accuracy, 93% precision, 84% recall and 88% f-score[1]. The sentiment classification has been done on Hindi tweets by using RF. The tweets were classified as positive and negative. 90.24% of accuracy and 66.35% of f1-score were achieved[2]. The various hybrid models that combine different deep learning algorithms such as LSTM, GRU, BLSTM and CNN along with different word embeddings were proposed. The proposed method M-Hybrid that combines the CNN+BLSTM performs better than the other models and achieves a higher accuracy of 82.14%[3].

An ensemble of unsupervised dictionary-based classifiers and deep CNNto classify the sentiments of tweets such as positive, negative or neutral was proposed. In this paper, the sentiment scores were extracted by using a dictionary-based classifier and a three-layer sequential CNN model was used to classify the tweets. The results show that the accuracy was enhanced to 93.25% by using the proposed model[4]. The author analysed the people's sentiments about the Covid virus by using ten ML classifiers. TF-IDF was used to select features from 65,854 tweets. Among the ten classifiers Logistic Regression, Ridge Regression and Linear SVC performs well by gaining the accuracy of 67% whereas RF obtains 56% accuracy[5].

The analysis was performed by using LSTM on 25,000 movie reviews from the IMDB dataset. It gains 86.85% accuracy, 88% precision, 87% recall and 87% f1-score. The accuracy was greater than the other ML algorithms when using this dataset[6]. To analyse the sentiments of people expressed in social media regarding COVID-19 implemented an attention mechanism, BLSTM. The results revealed that the proposed mechanism performed well and achieved the highest f1-score of 72.09%[7]. The model suggests a better way to pad input sequences. LSTM and CNN were used to analyse 157,860 tweets with different padding to show the variances. LSTM with pre-padding sequences gains higher accuracy of 88.32% than CNN[8].

The proposed method includes the generation of word vectors by using TF-IDF and applying BLSTM on these vectors for the effective sentiment analysis. The result was compared with other classifiers such as RNN, CNN, LSTM and NB. The result shows that the proposed method gains the highest precision of 91.54%, 92.82% recall and 92.18% f1-score[9]. The Dialectal Arabic Sentiment Analysis was performed by using DL algorithms. The proposed method uses the hybrid of CNN + LSTM to yield better accuracy than the individual algorithms. For binary classification the proposed model obtains the accuracy between 81% and 93%. For three way classification it obtains the accuracy between 66% and 76%[10]. An ensemble of CNN and LSTM was used to classify the sentiments on two datasets. The IMDB dataset has 50,000 reviews and the SST2 dataset has 16,000 reviews. Each word of reviews was represented by Glove embedding and then the embedding was fed into the model. The proposed CNN + LSTM perform better in both the datasets. A higher accuracy of 90% was obtained[11].

The DL models such as CNN, LSTM, BLSTM, hybrid of CNN + LSTM and CNN + BLSTM were used to analyse the text into positive, negative and neutral. The dataset has 2003 French articles from international and national newspapers. An average of 4000 words has been presented in each article. The higher accuracy of 90.66% was obtained by CNN + BLSTM whereas the accuracies of 88% by CNN, 85.87% by LSTM, 86.40% by BLSTM and 90.13% by CNN + LSTM were obtained[12]. A hybrid deep learning model of CNN and stacked BLSTM was proposed to perform long term sentiment analysis. The proposed model was implemented in two datasets such as IMDB and SST2. The proposed performance performed better than other models such as CNN, LSTM and ensemble of CNN-LSTM that achieves the higher accuracy of 94.1%[13].

The author also compared the Textblob and AFINN dictionary for labelling the sentiments of #Swachh Bharat tweets. There exists a more or less similar result; it shows that the majority of the tweets are positive. Thus, gives them more confidence about sentiment labelling. The highest accuracy of 83.57% has been obtained by Passive aggressive Classifier with TFIDF – bigrams and LinearSVC with TFIDF – trigrams. RF also performs well with an accuracy of 78.21%[14]. Highlight the content that promotes violence or hatred against individuals or groups based on religion, gender or ethnicity by analysing the sentiments of tweets. Logistic regression algorithm was used to detect the appropriate sentiments with 83.98% accuracy[15].

The sentiments of people regarding vaccines of all sorts were assessed using LSTM and Bi-LSTM. This study improves understanding of the public's opinion on COVID-19 vaccines. LSTM achieves an accuracy of 90.59% and BLSTM achieves 90.83%[16].

The sentiments of consumer reviews were classified as either positive or negative by using Naive Bayes, Maximum Entropy, and SVM, as well as the Semantic Orientation based WordNet, which extracts synonyms and similarity. Finally, the proposed models were evaluated in terms of recall, precision, and accuracy[17]. The feature work is combined with tweet words, word2vec, stop words and integrated into the deep learning techniques of CNN and LSTM. Those two models are well trained and applied for IMDB dataset which contains 50,000 movie reviews. With huge amount of twitter data is processed for predicting the sentimental tweets for classification. The result of Deep Learning algorithms aims to rate the review tweets and also able to identify movie review with testing accuracy as 87.74% and 88.02%[18].

The main problems that exist in the current techniques are:

- inadequate accuracy,

- performance in sentiment analysis based on insufficient labelled dataset,
- limited performance metrics.

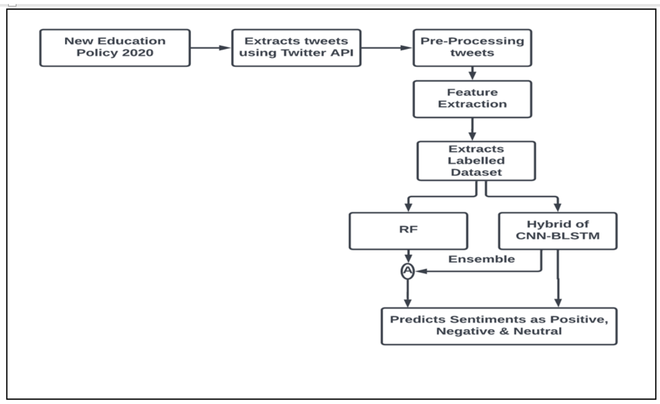

Fig 1. Architecture Diagram

The major contributions of this study are: a hybrid of CNN-BLSTM has been proposed to analyse the sentiments for large dataset, the performance of the proposed model is evaluated using various metrics and the labelled dataset was extracted using AFINN dictionary. Figure 1 represents the architecture diagram of the proposed model. Figure 2 represent how the features were selected in each layer of the proposed model.

2 Methodology

The proposed model has been described in two phases:

2.1. Phase 1

In this phase, the tweets were classified into positive, negative and neutral by using Random Forest classifier and calculated their corresponding percentages. Totally 105045 tweets of New Education Policy 2020 were extracted by using Twitter API. Here the re-tweeted tweets are also considered in order to count the corresponding percentages. In case of re-tweeted tweets the duplicated tweets were removed by using its user-id attribute. After removing the null and duplicated tweets the total of 126619 tweets were obtained. The pre-processing has been done on the extracted tweets which include tokenization; removal of links, symbols and any other special characters Table 1 given below describes the total number of tweets. The algorithm1 in [19] describes how the tweets were pre-processed and classified by using Random Forest classifier.

After pre-processing, each token was rated with an integer between -5 to +5 according to their polarity by using AFINN dictionary. Then each tweet was classified as positive, negative and neutral according to their scores. The feature vectors have been generated from the hybrid of unigram and bigrams feature sets by using the TF-IDF method. The 70% of data was taken as a training set (i.e. 88633 samples) and 30% of data as a test set (i.e. 37986 samples).

Table 1. Total number of tweets of Dataset 2

Scheme	Total no. of tweets	No. of original tweets without duplication	No. of re-tweeted tweets without duplication	No. of tweets	Total no. of tweets without null values
Education Policy	105045	93642	32979	126621	126619

After preprocessing the data was reshaped into a Dataframe that has User_id, User_name, Screen_name, Text, Full_text and Txt_msg as columns. Table 2 describes the sample data with the corresponding columns. The Full_text column was used in case of longer messages.

Table 3 describes the labelled dataset which has been obtained based on their corresponding scores. The tweet was labelled as 1 for positive if the score value exceeds zero, 2 for negative if it is lesser than 0 and 0 for neutral if it is equal to zero. From this labelled dataset the number of tweets in each polarity and their corresponding percentages were calculated. Here the positive percentage is higher than the negative percentage which is described in Table 4.

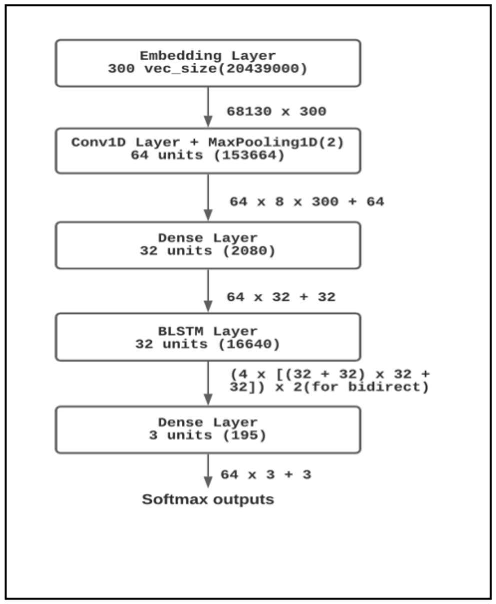

Fig 2. Architecture Diagram of DL Classification

Table 2. Sample Data of Dataset 2

User_id	User_name	Screen_name	Text	Full_text	Txt_msg
6166042	Nicky Penttila	NickyPenttila	We know kids aren...	We know kids aren...	We know kids aren...
13151882	Thacknology	DaveThackeray	If you have a URL...	null	If you have a URL...
15843023	EvielKhon	EvielKhon	RT @dante-mendes: ...	F29 RETALIATOR on...	F29 RETALIATOR on...
16615241	Edouard Stenger	EdouardStenger	Reading the react...	null	Reading the react...
19332805	WhatCulture	WhatCulture	Grab yourself a t...	null	Grab yourself a t...

Table 3. Labelled Dataset of Dataset 2

Id	Token	Token_clean	Score	Sentiment	Label
6166042	[know, kids, aren...	[know, kids, aren...	2	Positive	1
13151882	[ends, theres, fi...	[url, ends, say, ...	0	Neutral	2
15843023	[retaliator, amig...	[retaliator, amig...	2	Positive	1
16615241	[reading, reactio...	[reading, reactio...	-4	Negative	0
19332805	[grab, tall, glas...	[grab, tall, glas...	5	Positive	1

Table 4. Percentages of New Education Policy

Scheme	Total no. of tweets	No. of Positive tweets	No. of Negative tweets	No. of Neutral tweets	Positive %	Negative %	Neutral %
Education Policy	126619	44713	27137	54769	35.32%	21.43%	43.25%

Table 5 shows that the performance of the model decreases as the number of features increases. The model performance was evaluated by various metrics such as accuracy, log loss, precision, recall, f1-score and roc-auc. It shows the more or less same result for 300 and 160 features. So in this case the minimum features 160 were taken for the further classification purpose. The least minimum 160 features obtain the better results with 80% of accuracy, 81% of precision, 80% of recall, 80% of f1-score and 94% of roc-auc.

Table 5. Number of Features (Using RF)

No of features	Accuracy	Log Loss	Precision	Recall	F1-Score	ROC-AUC
1000	0.372	1.582	0.362	0.372	0.365	0.51
500	0.371	1.550	0.358	0.371	0.361	0.50
300	0.804	0.545	0.809	0.805	0.803	0.94
160	0.802	0.508	0.814	0.802	0.801	0.94

2.2. Phase 2

In this phase, the various combinations of Deep learning techniques were carried out. Before classification the text messages need to be converted into numerical form. After tokenization each token of text was replaced by its index value. Every sequence of text must be in a fixed length; in this case maxlen parameter was set to 160. Thus the text data was converted to a numerical list of equal length 160. In Table 6, comparison of results shows that the hybrid of CNN and BLSTM performs better when compared with other approaches. The higher accuracy of 97% with the minimum log loss 0.10 was obtained. Figure 3 shows the roc-auc for each class which was obtained by the proposed method. Algorithm1 (Table 7) given below was used to build this proposed model.

Table 6. Deep Learning approach

Hybrid Classifiers	Val_Accuracy	Val_ Loss	Precision	Recall	F1-Score	ROC-AUC
CNN	0.956	0.171	0.957	0.956	0.956	0.99
BLSTM	0.961	0.112	0.961	0.961	0.961	0.99
CNN + BLSTM	0.968	0.103	0.968	0.968	0.968	0.99
RF + CNN + BLSTM	0.934	0.298	0.935	0.934	0.934	0.98

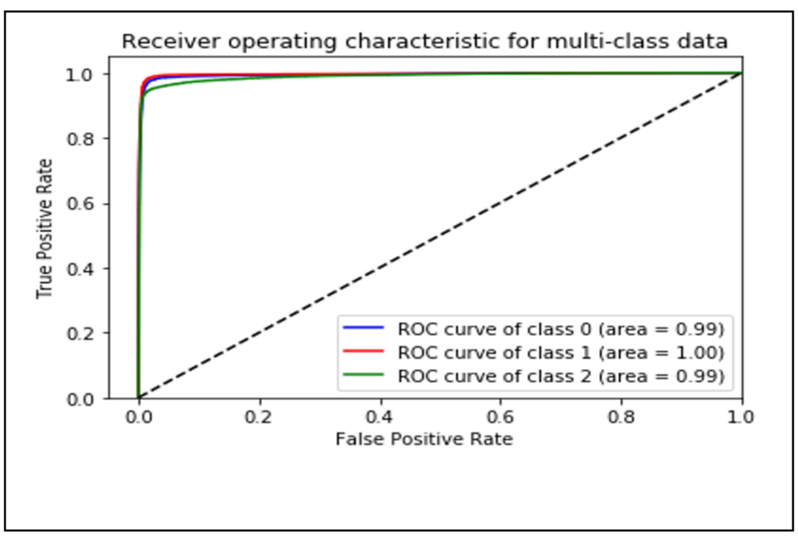

Fig 3. ROC-AUC for CNN + BLSTM

Table 7. Algorithm 1: Classifying Tweets using Hybrid of CNN + BLSTM

Input: Given the feature vectors. Output: Classify the tweets to positive, negative and neutral. 1. A Sequence of eight layers including input and output layers was created. 2. An Embedding layer was used to train on vocabulary size of 68130 with a vector of 300 features each witha maximum length of 160. 3. Then the input data was sent to a Convolutional layer of 64 feature maps and kernel size 8. 4. To reduce the dimensionality select the most salient features from the feature map by using MaxPooling layer. 5. By using the Dropout layer 20% of selected features were removed randomly. 6. Pass the learned features to the fully connected layer that uses RELU activation function. 7. Again by using the Dropout layer 50% of selected features were removed randomly. 8. Then the selected features were passed to the BLSTM layer in which it passes the input sequence in forward direction as well as in the backward direction. The outputs from both the directions were concatenated. 9. To perform the required classification into polarities the concatenated output from BLSTM layer was passed to fully connected SoftMax layer. 10. Because of multi-class classification, sparse_categorical_crossentropy loss function was used. 11. By using the training dataset fit the model. 12. Measure the various evaluation metrics such as accuracy, log loss, precision, recall, f1-score and roc-auc to find the performance of the model.

3 Results and Discussion

In this study, 105045 numbers of tweets were collected regarding the New Education Policy 2020. After removing the duplicate and null tweets a total of 126619 tweets were used for classification. The hybrid of CNN and BLSTM obtains the higher accuracy of 97%. The performance of the model was evaluated by using various metrics. Table 8 given below compares the performance of proposed work with the existing methods. Thus proves the efficiency of the proposed model.

Table 8. Comparison of various Existing ML and DL techniques

Papers	Year Published	Dataset size	Classifiers	Accuracy %	Precision %	Recall %	F-Score %
(6)	2019	25,000 movie reviews	LSTM	86.85	88.0	87.0	87.0
(8)	2019	157,860 tweets	LSTM	88.32	-	-	-
(10)	2019	-	CNN + LSTM	76.0	-	-	-
(9)	2019	-	TF-IDF + BLSTM	-	91.54	92.82	92.18
(12)	2019	2003 French articles	CNN + BLSTM	90.66	-	-	-
			CNN + LSTM	90.13	-	-	-
			CNN	88.0	-	-	-
			LSTM	85.87	-	-	-
			BLSTM	86.40	-	-	-
(11)	2019	50,000 + 16,000 reviews	CNN + LSTM	90.0	-	-	-
(14)	2019	-	PAC	83.57	-	-	-
			RF	78.21	-	-	-
(1)	2019	7,345 reviews	SVM	93.0	93.0	84.0	88.0
(2)	2020	-	RF	90.24	-	-	66.35
(3)	2020	-	CNN + BLSTM	82.14	-	-	-
(4)	2020	-	CNN	93.25	-	-	-
(5)	2020	65,854 tweets	LR	67.0	-	-	-
			RR	67.0	-	-	-
			Linear SVC	67.0	-	-	-
			RF	56.0	-	-	-
(13)	2020	-	CNN + LSTM	94.1	-	-	-
(7)	2021	-	BLSTM	-	-	-	72.09
(15)	2021	-	Logistic Regression	83.98	-	-	-
(16)	2021	-	LSTM	90.59	-	-	-
			BLSTM	90.83			

Continued on next page

Table 8 continued

(17)	2021	-	NB	88.3	-	-	-
			Max Ent	83.9			
			SVM	85.5			
			Semantic Analysis (WordNet)	89.8			
(18)	2022	50,000 movie reviews	CNN	87.74	-	-	-
			LSTM	88.02			
Proposed Model		126619	CNN + BLSTM	97.0	97.0	97.0	97.0

Table 9 represents the confusion matrix. By using this confusion matrix, various evaluation metrics such as accuracy, precision, recall and f-measure were calculated. From Table 9, we see that 36,785 reviews were correctly classified among 37986, and 1,201 reviews were misclassified.

Here the Actual values were represented as columns and the Predicted values were represented as rows. The main diagonal (7843, 13097, 15845) gives the correct predictions, this is because the actual and predicted values are the same.

The proposed model was analysed by using various parameters since the completeness and exactness are more important than the high accuracy. The F1-score has been widely used in the Natural Language Processing literature. ROC-AUC is used when the dataset is imbalanced since accuracy is not a reliable performance metric for imbalanced data. These results were represented in Table 10.

Table 9. Confusion Matrix (Hybrid of CNN and BLSTM)

Class	Negative	Positive	Neutral
Negative	7843	36	242
Positive	65	13097	307
Neutral	324	227	15845

Table 10. Classification Report (Hybrid of CNN and BLSTM)

Class	Precision	Recall	F1-Score	Support
Negative	0.95	0.97	0.96	8121
Positive	0.98	0.97	0.98	13469
Neutral	0.97	0.97	0.97	16396
Total/avg	0.97	0.97	0.97	37986

4 Conclusion

In this study, a hybrid of CNN and BLSTM has been implemented for the multi-class classification that classifies the tweets into positive, negative and neutral on New Education Policy 2020. TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. In [14], the author used TextBlob and AFINN for sentiment assigning to tweets. The proportion of the number of tweets from each sentiment in each classifier is consistent. Thus, gives more confidence about sentiment labelling by using AFINN dictionary. Finally, the performance of the model was analysed by using a confusion matrix yielding the Accuracy (97%), Precision (97%), Recall (97%), F-Measure (97%) and 99% of roc-auc with the minimum log loss 0.10. Thus, the above results show that our approach works better compared to existing systems (proved in Table 8) in terms of various measures. By using the same approach, we want to develop a model for the big data and thereby analyse the sentiments of Tamil tweets.

References

1) Alqaryouti O, Siyam N, Monem AA, Shaalan K. Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics*. 2019. Available from: https://doi.org/10.1016/j.aci.2019.11.003.
2) Munshi A, Sanchitsapra M, Arvindhan. A Novel Random Forest Implementation of Sentiment Analysis". *International Research Journal of Engineering and Technology (IRJET)*. 2020;(7):2821–2824. Available from: https://www.irjet.net/archives/V7/i6/IRJET-V7I6532.pdf.

3) Mehmet U, Salur I, Aydin. A Novel Hybrid Deep Learning Model for Sentiment Classification. *IEEE Access*. 2020;8:58080–58093. Available from: https://ieeexplore.ieee.org/document/9044300.
4) Rani S, Gill NS. Tweet Sentiment Classification Using an Ensemble of Unsupervised Dictionary Based Classifier and Deep Convolutional Neural Network. *International Journal of Advanced Science and Technology*. 2020;29(3):4904–4912. Available from: http://sersc.org/journals/index.php/IJAST/article/view/5708.
5) Effrosynidis D. How I created a Real-Time Twitter Sentiment Analysis tool for COVID. 2020. Available from: https://towardsdatascience.com/how-i-created-a-real-time-twitter-sentiment-analysis-tool-for-covid-292ff6a6323b.
6) Tholusuri A, Anumala M, Malapolu B, Lakshmi GJ. Sentiment Analysis using LSTM". *International Journal of Engineering and Advanced Technology (IJEAT)*. 2019;(8):1338–1340. Available from: https://www.ijeat.org/wp-content/uploads/papers/v8i6S3/F12350986S319.pdf.
7) Kastrati Z, Ahmedi L, Kurti A, Kadriu F, Murtezaj D, Gashi F. A Deep Learning Sentiment Analyser for Social Media Comments in Low-Resource Languages. *Electronics*. 2021;10(10):1133. Available from: https://www.mdpi.com/2079-9292/10/10/1133.
8) Reddy NVDM, Reddy S. Effects Of Padding On LSTMS And CNNS. 2019. Available from: https://arxiv.org/abs/1903.07288.
9) Xu G, Meng Y, Qiu X, Yu Z, Wu X. Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access*. 2019;7:51522–51532. Available from: https://doi.org/10.1109/ACCESS.2019.2909919.
10) Kwaik KA, Saad M, Chatzikyriakidis S, Dobnik S. LSTM- CNN Deep Learning Model for Sentiment Analysis of Dialectal Arabic. *International Conference on Arabic Language Processing: Arabic Language Processing: From Theory to Practice*. 2019;p. 108–121. Available from: https://www.researchgate.net/publication/336267517_LSTM_CNN_Deep_Learning_Model_for_Sentiment_Analysis_of_Dialectal_Arabic.
11) Minaee S, Azimi E, Amiraliabdolrashidi. Deep-Sentiment. *Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models*. 2019. Available from: https://arxiv.org/abs/1904.04206.
12) Rhanoui M, Mikram M, Yousfi S, Barzali S. A CNN-BiLSTM Model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction*. 2019;1(3):832–847. Available from: https://doi.org/10.3390/make1030048.
13) Guha T, Mohan KG. A Hybrid Deep Learning Model for Long-Term Sentiment Classification. *Webology*. 2020;17(2):663–676. Available from: https://www.webology.org/abstract.php?id=338.
14) Sai S. Sentiment Analysis on Swachh Bharat using Twitter. 2019. Available from: https://towardsdatascience.com/sentiment-analysis-on-swachh-bharat-using-twitter-216369cfa534.
15) Raut P, Rathod R, Tidke R, Pande R, Rathod N, Kulkarni N. Sentiment Analysis of Twitter. *International Journal for Research in Applied Science and Engineering Technology*. 2022;10(12):621–627. Available from: https://doi.org/10.22214/ijraset.2022.47954.
16) Alam KN, Khan MS, Dhruba AR, Khan MM, Al-Amri JF, Masud M, et al. Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. *Computational and Mathematical Methods in Medicine*. 2021;2021:1–15. Available from: https://doi.org/10.1155/2021/4321131.
17) Ashique M, Kumar S, Aanchalvij, Panwar S. Sentiment Analysis Using machines Learning Approaches of Twitter Data and Semantic Analysis". *Turkish Journal of Computer and Mathematics Education*. 2021;12(6):5181–5192. Available from: https://turcomat.org/index.php/turkbilmat/article/view/8771.
18) Gandhi UD, Kumar PM, Babu GC, Karthick G. Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). *Wireless Personal Communications*. 2021. Available from: https://doi.org/10.1007/s11277-021-08580-3.
19) Sujatha RER. A Sentiment Classification on Indian Government Schemes Using PySpark". *International Journal on Emerging Technologies*. 2020;11(2):25–30. Available from: https://researchtrend.net/ijet/pdf/A%20Sentiment%20Classification%20on%20Indian%20Government%20Schemes%20Using%20PySpark.pdf.