Grenze International Journal of Engineering and Technology, Jan Issue

Sentiment Analysis Classification using Deep Learning Techniques

C. Nalini[1], B. Dharani[2] and Tamilarasu Baskar[3]
[1-2]Kongu Engineering College, Erode, India
Email: nalinikec@gmail.com, dharanib.21mit@kongu.edu
[3]Infoway Technosoft Solutions, Bangalore, India
Email: tamilarasubaskar@gmail.com

***Abstract*— People are increasingly expressing their ideas on websites in this day and daily, enormous amounts of client feedback are produced. These unstructured text reviews and customer comments are frequently used by customers when making decisions. However, because there is so much consumer feedback, it is time-consuming to read them all. The ability to forecast the precise sentimental polarity of user textual feedback evaluations for particular entity is still a difficult task because of phrase length constraints, textual order variations, and logical complexities. Bidirectional Recurrent Neural Network based on Self Attention Mechanism (BRNNA) for subjectivity categorization of reviews was developed to overcome these issues. The described method uses pre-trained word embedding to minimize text representations, avoiding problems with data sparsity. Moreover, the attention mechanism uses multiple word and phrase weights to capture n-gram attributes and concentrate on the most important context-relevant information. The suggested model automatically learns classification features and captures the semantic and spatial information that is essential in detecting the polarities of sentiments utilizing Bidirectional Recurrent Neural Network (BRNN) and Self Attention Mechanism. The proposed BRNNA model is evaluated in comparison to various benchmark techniques. The accuracy of the proposed approach is 91.06%.**

***Index Terms*— Natural Language Processing, Sentiment analysis, Subjectivity classification, Deep learning, Bidirectional Recurrent Neural Network, Neural network, Self-attention.**

I. INTRODUCTION

A. *Sentiment Analysis*

Sentiment analysis is a technique used to identify information by analyzing the subjectivities presented in a sentence. In sentiment analysis, text analytics is used to create classification models and estimate sentiment scores, which combines natural language processing with machine and deep learning approaches. Text can be classified as positive, neutral, or negative, and is often represented with a score to reflect the strength of the sentiment. Social media tracking, customer support control, and comments analysis are examples of popular sentiment analysis applications. Using natural language processing techniques, deep learning algorithms can categorize unstructured textual content through emotion and opinion. Sentiment analysis in business allows companies to identify negative or positive views toward their product or service and take the necessary steps to address such concerns. It determines whether the provided sentences are positive, negative, or neutral. It also recognizes and extracts reviews

Grenze ID: 01.GIJET.10.1.328

and attitudes from the text.
The goal of sentiment analysis is to assess customers or users' opinion on a target object by analyzing a vast volume of text from many sources. Due to the rapid evolution of sites, interpersonal organizations, web journals, online entries, surveys, sentiments, ideas, evaluations, and criticism, the amount of data on the internet is growing by the day. It is challenging to discover the appropriate sentiment polarity and analyze proper sentiment meaning from large amount of information.
Three main levels can be used to analyze attitudes in text reviews: categorization at the document, phrase, and aspect levels. Document level sentiment analysis module at the document level examines a passage of text and determines whether it has a positive or negative sentiment. It supports all texts with emotional content (that is, texts which are not too short, numeric, include only a street address, or an IP address). A text analysis technique called aspect level sentiment analysis divides data into categories and determines the sentiment that goes with each one. Aspect-based sentiment analysis can be utilized to analyze feedback from customers by combining particular sentiments with various characteristics of a goods or service. Sentiment analysis at the sentence level identifies the character of opinions stated in sentences to ascertain whether they are neutral, positive, or negative. Text frequently uses modality. Around 18% of sentences in a typical corpus have modality. It may be challenging to ascertain the sentiment a modality carries due to its unique qualities.

B. *Deep Learning*

Deep learning is the field of study that imitates the working of human brains which will process the data for decision making. It continuously learns by analyzing data to find patterns and classify the image. It is used for image classification, language transaction and speech recognition. Deep neural networks are included in this category of networks, where each layer is capable of carrying out complicated operations like modeling and synthesis that help text, voice, and image content. It can be used for any type of recognition problem. The main goal is to let computers learn according to their own and alter their behavior accordingly.

C. Working of Deep Learning

Together with the digital world, which has resulted in an enormous amount of data in all forms and from every region of the world, deep learning has developed. Big data, as it is commonly known, is gathered from a variety of sources, including social media, online search engines, e-commerce platforms, and online movie theatre. The sharing of this large datasets is possible due to financial tools like cloud computing. However, because the data is usually unstructured and large, it could require decades for people to understand it and determine the necessary information. Companies are increasingly utilizing AI systems for automated help as they recognize the great potential that can be attained from unlocking this wealth of knowledge.
In section 2, numerous neural network-based sentiment analysis techniques are discussed. A detailed explanation of the suggested technique is included in section 3. In section 4, experimentation and comparative outcomes are covered. The outcome of the entire work is presented in section 5, along with the potential of further investigation.

II. Literature Review

Amplayo et.al., [1] a fine-grained sentiment analysis that can be used to summaries multiple short online reviews. An interesting approach for assembling lists of different web-based surveys that is based on a highly small-grained sentiment utilization model for short texts and is applicable to various fields and languages. They collect variety of data to accurately assess our approach's adaptability. They filter the data after it has been collected if the character length exceeds 140. The English sites Rotten Tomatoes and Amazon PC items accumulated information from two separate spaces, movies and items. Similarly, for Korean data, data on non-English movie reviews is collected from Naver movie, and for Chinese data, from Douban. Final classifier accuracy of amazon (English product) is 0.98%.
Elmurngi, E et.al., [2] proposed supervised learning techniques are used to detect fake movie reviews using sentiment analysis. One of the most pressing difficulties facing sentiment analysis is determining how to extract emotions from opinions and detecting good and negative evaluations from opinion reviews. The first dataset, movie reviews dataset V1.0, contains 700 positive and 700 negative surveys out of 1400 aggregate. The second dataset is the movie reviews dataset V2.0, which includes a total of 2000 film audits, 1000 of which were favorable and 1000 of which were unfavorable. The V3.0 movie reviews dataset, which contains 10662 ratings for movies, with 5331 positive ratings and 5331 negative ratings, is the final dataset. Sentiment analysis and text categorization techniques are tested using a dataset of movie reviews.

Nguyen et.al., [3] on the feedback corpus of Vietnamese students, deep learning was compared to traditional classifiers. Student input is a valuable source of information obtaining feedback from students in order to enhance teaching quality activities. Using sentiment analysis in the classroom may determine sentiment polarity using feedback data, this encapsulates all issues that have arisen in the institution as a result of the changes efforts will be made to increase the quality of instruction as well as education. The Vietnamese students' feedback corpus was gathered from a college and exposed to machine learning and regular language handling calculations. Based on a few appraisal standards, the eventual outcomes were contrasted and dissected all together with pick the best model.

Jitendra Kumar Rout et.al., [4] using semi-supervised learning to detect deceptive reviews online. As additional purchasers utilize online sentiment assessments to illuminate their administration direction, assessment reviews influence the fundamental concern of organizations. The 'gold standard' dataset is used. The dataset includes 1,600 review words from 20 residences in the northwest region of the United States, including 800 false and 800 genuine reviews. Using PU Learning-based categorization, they able to acquire an F-score of 0.837. They showed how to enhance the F-score metric in classification using four standard semi-supervised learning algorithms. They improved the findings by including additional dimensions in the feature vector such as POS, Sentimental Content, Linguistic and Word Count features.

Imran, A.S et.al., [5] mainly focused on cross-cultural polarity and sentiment analysis of COVID-19-related tweets using sentiment analysis and deep learning. The public's standards, as well as political will, are prevalent in how various societies respond and respond in the face of an emergency. The datasets are taken from Kaggle. The lack of grouped tweets pulled from Twitter and a state-of-the-art dataset with physical labels. Preparing and approval precision on the baseline models were 96% and 81%, individually. A profound learning to detect both opinion extremity and close to home state from clients' tweets on Twitter, LSTM designing using prearranged integrating models that used the sentiment 140 data source and the passionate tweet data source to achieve state-of-the-art precision.

Gichang Lee et.al., [6] sentiment classification with word localization using a convolutional neural network (CNN) and weakly supervised learning. SA datasets, simply give the feeling mark to every message or sentence. To put it another way, there is a shortage of information based on which conditions are critical in sentiment classification. They utilized two unique arrangements of film audits, one in English and one in Korean. In comparison to other formal writings such as news stories, movie reviews not only feature clear sentiment markers (ratings or stars), but they also have more subjective language. In this model, each phrase is treated as a consistent esteemed vector, and each sentence is treated as a framework which lines correspond to the sentence's assertion vector. Then, with these expression lattices as information sources and sentiment labels as the CNN model is prepared.

Kastrati, Z et.al., [8] deliberate student review aspect-based sentiment analysis of Massive Open Online Courses using a weakly supervised framework. Criticism from understudies is a significant apparatus for acquiring essential bits of knowledge into the educating learning process. The dataset includes student reviews that include comments on various parts of a Massive Open Online Courses. The framework, specifically, depends on viewpoint level sentiment analysis with the objective of naturally identifying opinion or assessment extremity communicated toward a specific component of the Massive Open Online Courses.

Colon-Ruiz, C et.al., [7] comparing deep learning architectures for drug review sentiment analysis. Sentiment analysis has been concentrated in an assortment of fields, including financial matters, legislative issues, and medication, among others. Programmed investigation of online client assessments in the drug business takes into consideration the assessment of tremendous quantities of client sentiments and the extraction of helpful data about drug viability and unfriendly impacts, which may be utilized to further develop pharmacovigilance frameworks. These are reviews of various medications that were taken from the Drugs.com website. A score, ranging from 0 to 9, is provided for each medicine evaluation, indicating how satisfied the patient is with the treatment.

According to Hongjie Deng et al., [15] text is the fundamental way to communicate language, and both the details of each word and the whole text should be given attention. In order to explore classification prediction and semantic sentiment analysis is utilized to analyze both the positive and negative opinion associated with banking news, social media, etc. For the standard encoder-decoder, the input source's semantic encoder C is the same. Each word in the phrase "source based on encoder" is encoded to produce the semantic code C. A text sentiment analysis method that combines multichannel Bi-LSTM with a mechanism for attention and CNN, and compares the prediction outcomes to other sentiment models.

Yang et.al., [11] polar flipping and lexicon embedding are included in a two-level LSTM for sentiment analysis. In order to use the two labelled data in two-stage labelling technique, a two-level network was subsequently proposed. The Stanford Sentiment Treebank and movie review datasets are utilized as two benchmark datasets for analyzing the suggested algorithms. The latter contains a lot of 11885 sentences with five classifications-positive, very positive, very negative, negative, neutral, in contrast to the former's 10662 phrases with binary classes (positive and negative).

Various investigations on the fundamental structure of neural networks have been performed. We gathered from the literature review described above that integrating two or more neural networks can lead to improved performance models. The weaknesses of one model are filled with the strengths of another network by merging the networks. In this direction, more interesting combinations that could produce a system that performs better need to be investigated. We provide a BRNNA neural network model to solve the feature space sparsity issue and obtain the long-distance contextual opinion dependency. And the majority of the latest research on sentiment analysis have simply looked at how the suggested model is accurate.

III. Methodological Approach

Due to a number of factors, it might be challenging to determine the actual opinion in a review. The subjectivity of the texts and the requirement to identify perspective words from non-opinionated phrases are two major contributing factors. Therefore, there ought to be more effective techniques for choosing features automatically. To that end, a movie review dataset used to train and evaluate the classifier, as well as to pre-processing data and embedding. Fig. 1 shows the suggested work's workflow in process.

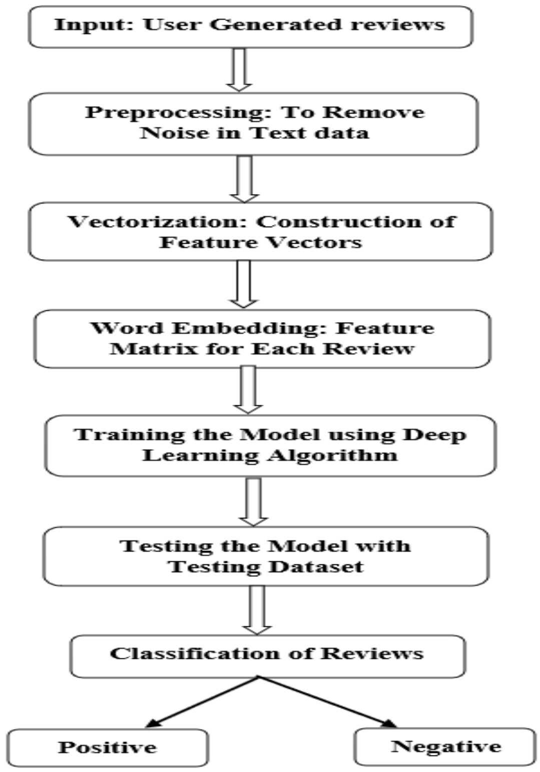

Figure 1. Process Flow Diagram

A. Pre-processing and Word Embedding

Slang phrases and potential errors are frequently found in online comments. An algorithm's performance may decrease the more noise that exists in the data. Pre-processing is the technique used to clean up data and also get text prepared for categorization. In the first phase of the process, significant key terms are chosen from the incoming data and eliminates any that don't help classify sentiment. By contributing in sentiment analysis, it also decreases the more complex dimensionality issue and improves the classifier's performance. Table I provides a summary of some of the key pre-processing techniques used in this proposed study. Each review has undergone pre-processing followed by text vectorization. Each part of the review's vector representation corresponds to a word's index in the vocabulary dictionary. The vector's length is currently 100. Word embeddings or word vectors, are produced by retraining with the text corpus using the word2vec skip gramme framework with a window's size of five in order to preserve the contextual relevance of the words, which is important for sentiment analysis.

B. Modelling and Analysis Aspects

The input layer receives a word vector index of size 100 x 300, and then passes on to the BRNN layer. The output of the 100-units BRNN layer, which averages rather than combining the forward and backward hidden state vectors at a given time step is 100 x 64 matrices. All word is given a context vector by the self-attention mechanism, which reflects the interior spatial information between each sentence and the other sentences. A context vector is generated due to newly added self-attention layer, which is connected to a specified actual average hidden state variable. As a result, context vector matrix with sizes of 10 x 64 is the self-attention layer's output.

TABLE I. PRE-PROCESSING TECHNIQUES

S.No	Pre-Processing Techniques
1	Removal of URL and Non-English Words
2	Change to Lower Case
3	Eliminating Extra Punctuation
4	Substituting Equivalent Numbers
5	Emotional Handling
6	Parts of Speech Tagging
7	Stemming
8	Alternatives to Negative Remarks
9	Spelling and Grammar Check
10	Shortening Long Words

Each feature vector is subjected to a max-pooling function with a duration of 4, and the outcome is a feature vector of length 25. Combining the extracted features from the max-pooling layer provides a 25 x 6 matrix, then reduced to produce a vector size 150. To establish the probability distribution of each segment, an interconnected sigmoid layer receives the feature vectors that were recently collected. The following is how it can be expressed mathematically.

$$P_{sigmoid}\ (Cj)\ = \frac{e^{oj}}{1+e^{oj}} \qquad (1)$$

Where Psigmoid (Cj) is the category j probability distribution. and oj indicate the output related to the category j. The confidence value of the classifier is standardized between zero and one using the Sigmoid activation function. After obtaining the probability distribution from the sigmoid layer, the difference between the actual and predicted sentiments is calculated using binary cross-entropy as the loss function.

$$Loss = -\sum_{i=1}^{k} R(Ci) * \log P_{sigmoid}(Ci) \qquad (2)$$

Where R(Ci) is the actual sentiment associated to the content, and k is the number of categories. It can accept discrete numbers from the range L = 0 to 1, where L denotes the review text's sentiment label (Negative, Positive). For the proposed method to generate the backward and forward context into the entirety of the phrase for purposes of classification, the attention layer and BRNN layers are used in the classification of the sentiments for the proposed approach. The prediction about whether to record the data as positive or negative values is identified using the mean probability scores.

C. Bidirectional Recurrent Neural Networks

Two hidden layers with opposing directional connections are connected to the same output by bidirectional recurrent neural networks (BRNN). The output layer can simultaneously receive data from past (backwards) and future (forward) states with this type of generative deep learning. To increase the network's input information, BRNNs were developed. Recurrent neural networks also have limitations because they can't access future input data the current state. BRNNs, on the other hand, don't need their input data to be fixed. Additionally, they can access their future input data from the present situation. BRNN are especially helpful when the input's context is required. Recurrent neural networks that are bidirectional are simply two separate RNNs combined. For one network, the input sequence is fed in normal time order; for a different network, it is fed in reverse time order.

Even though there are other options, such as summation, the outputs of the two networks are typically concatenated at each time step.
The following are the standard training procedures: Forward pass involves passing output neurons after passing forward and backward states. Output neurons are passed first during a backward pass, followed by forward and backward states. The weights are updated after both forward passes and backward passes have been completed. A multi-layer Bidirectional Recurrent Neural Networks will be used. Which indicates that several RNN layers will be stacked on top of one another.
The advantage of having more context than a single directional network is found in bidirectional RNNs. For instance, if a model needs to indicate the next word in a sentence that is flowing through it, it would do so based on prior knowledge. However, because two networks flow in opposite directions and are stacked on top of one another in a bidirectional network, it will also know what comes next. Both the sequence and the direction of the input are the opposite.

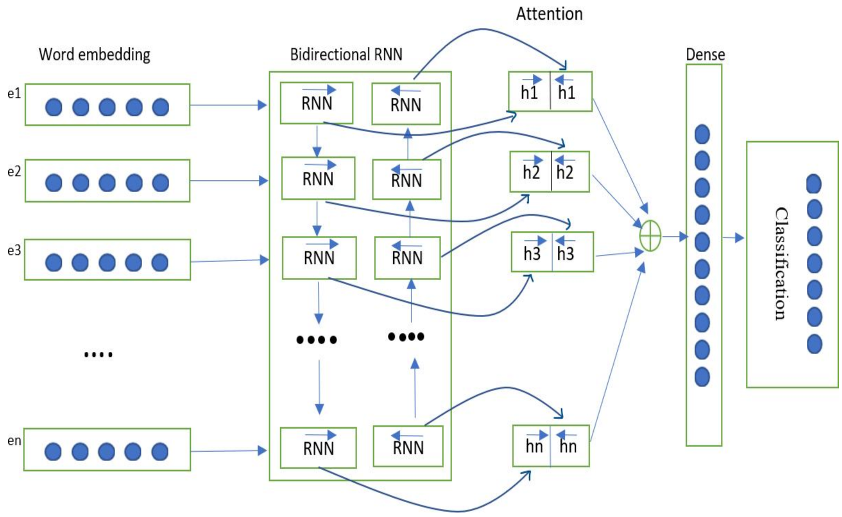

Figure 2. The Proposed BRNNA Architecture Model

IV. Results And Discussion

A. Dataset

The movie review dataset, which includes reviews with both positive and negative sentiment polarity, was used in the current investigation. Identifying whether the movie reviews are positive or negative is the task. The dataset contains 50,000 opinion reviews in total, of which 35,000 (or nearly 80%) were used for training and the remaining 15,000 (or nearly 20%) for testing. There are 25 K reviews for both the training and testing sets.

B. Experimental Setup

Table II provides an illustration of the experimental design used for this study.

TABLE II. Environment For Experiments

Experimental Environment	Configuration
Operating System	Windows 11
DL Framework	Tensorflow 1.12
Computer Memory	8GB RAM
Processor	INTEL CORE @ 2.40 GHz
Programming Language	Python 3.7

C. Method of Evaluation and Experimental Verification

Determining the current sentiment review's subjective polarity. Precision, recall, accuracy, and F1-score are used as evaluation measures to gauge how an algorithm is effective. Accuracy is calculated for all measures using the Confusion Matrix, which includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A model's accuracy is determined by comparing the proportion of reviews that predicted correctly to the total amount of comments. The performance analysis of the classifier whose true values are known for a table describes the test dataset called confusion matrix. The expected results for the categorization task are enumerated in a confusion matrix.

a. Accuracy

The ratio of accurately predicted examples to all examples is how accuracy is measured.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (3)$$

b. Precision

It calculates the percentage of correctly identified events among those that are positive.

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

c. Recall

The term analyses the percentage of correct positive predictions among every possible positive prediction.

$$Recall = \frac{TP}{TP+FN} \qquad (5)$$

d. F1-Score

It offers a technique to combine recall and precision into a single score that factors for both characteristics.

$$FI - Score = \frac{(2*Precision*Recall)}{Precision+Recall} \qquad (6)$$

Lists of all evaluating accuracy and F-1 scores for deep learning and machine learning techniques are shown in Tables III and IV, respectively. According to experimental findings, deep neural networks are more effective than other models at classifying sentiments, and the proposed model has a higher accuracy rate. The suggested BRNNA model outperforms the competing models by a large margin when the experimental F1 score results are compared, demonstrating that our model is better suited to handle the text classification challenge. This result implies that the low learning ability of the single model can be partially compensated for with the hybrid techniques.

TABLE. III ACCURACY OF EACH MODEL IN THE DATASET OF MOVIE REVIEW

Algorithms	Optimizer	Activation Function	Accuracy	
			Without Attention Mechanism	With Attention Mechanism
CNN	RMSprop	tanh	86.31%	87.42%
	Adam	Sigmoid	88.41%	**88.72%**
RNN	RMSprop	tanh	83.23%	85.12%
	Adam	Sigmoid	87.87%	**88.40%**
BRNNA	RMSprop	tanh	84.20%	85.68%
		Sigmoid	86%	**86.34%**
	Adam	Sigmoid	86.25%	**91%**
		tanh	84.03%	84.93%

TABLE. IV F1-SCORE OF EACH MODEL IN THE DATASET OF MOVIE REVIEWS

Algorithms	Optimizer	Activation Function	F1-Score	
			Without Attention Mechanism	With Attention Mechanism
CNN	RMSprop	tanh	80.3%	**82.1%**
	Adam	Sigmoid	85%	82%
RNN	RMSprop	tanh	81%	82.4%
	Adam	Sigmoid	80.2%	**84%**
BRNNA	RMSprop	tanh	80%	82.2%
		Sigmoid	84%	**84.52%**
	Adam	Sigmoid	84.72%	**88.4%**
		tanh	83.02%	84.35%

TABLE. V THE PROPOSED MODEL'S HYPERPARAMETERS

Hyperparameter	Movie Review Dataset
Word embedding dimension	200
Optimizer	Adam
Loss function	binary_cross_entropy
Activation function of hidden layer	Relu
Activation function of output layer	Sigmoid
Batch size	64
Epochs	50
Dropout	0.2

V. CONCLUSIONS

The detection of sentiment polarity in opinion review is proposed using a Bidirectional Recurrent Neural Network based on Self-Attention Mechanism model. It is created using Keras APIs within the Tensor Flow framework. Our model concatenates the features with the aid of pre-trained embedded words and then sends these features into various layers of the Bidirectional Recurrent Neural Network based on Self-Attention Mechanism (BRNNA) Model. While simultaneously developing a summary of the term value and its connection to the text, it preserves the characteristics of word context-specific ideas and connections in text sequences. Performance testing was carried out by contrasting the proposed model with different baseline models to assess its performance. The results of the experiment make it obvious that our model has a high classification performance value. Since stickers and memes are also emotional objects, it is planned to use the images as a source of information for sentiment analysis in the future.

REFERENCES

[1] Amplayo, R.K. and Song, M., 2017. An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. Data & Knowledge Engineering, 110, pp.54- 67.

[2] Elmurngi, E. and Gherbi, A., 2018. Fake reviews detection on movie reviews through sentiment analysis using supervised learning techniques. International Journal on Advances in Systems and Measurements, 11(1 & 2), pp.196-207.

[3] Nguyen, P.X., Hong, T.T., Van Nguyen, K. and Nguyen, N.L.T., 2018, November. Deep learning versus traditional classifiers on Vietnamese students' feedback corpus. In 2018 5th NAFOSTED conference on information and computer science (NICS) (pp. 75-80). IEEE.

[4] Rout, J.K., Dalmia, A., Choo, K.K.R., Bakshi, S. and Jena, S.K., 2017. Revisiting semi-supervised learning for online deceptive review detection. IEEE access, 5, pp.1319-1327.

[5] Imran, A.S., Daudpota, S.M., Kastrati, Z. and Batra, R., 2020. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. Ieee Access, 8, pp.181074-181090.
[6] Lee, G., Jeong, J., Seo, S., Kim, C. and Kang, P., 2018. Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network. Knowledge-Based Systems, 152, pp.70-82.
[7] Colon-Ruiz, C. and Segura-Bedmar, I., 2020. Comparing deep learning architectures for sentiment analysis on drug
[8] Kastrati, Z., Imran, A.S. and Kurti, A., 2020. Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. IEEE Access, 8, pp.106799-106810.
[9] Utami, T. A. (2021). Sentiment Analysis of Hotel User Review using RNN Algorithm. International Journal of Informatics and Computation, 3(1), 30-39.
[10] Alroobaea, R. (2022). Sentiment Analysis on Amazon Product Reviews using the Recurrent Neural Network (RNN). International Journal of Advanced Computer Science and Applications, 13(4).
[11] Wu, O., Yang, T., Li, M., & Li, M. (2020). Two-level LSTM for sentiment analysis with lexicon embedding and polar flipping. IEEE Transactions on Cybernetics.
[12] Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z., & Pu, B. (2021). Aspect-level sentiment analysis approach via BERT and aspect feature location model. Wireless Communications and Mobile Computing, 2021.
[13] Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Generation Computer Systems, 115, 279-294.
[14] Tam, S., Said, R. B., & Tanriöver, Ö. Ö. (2021). A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification. IEEE Access, 9, 41283-41293.
[15] Deng, Hongjie, et al. "Text sentiment analysis of fusion model based on attention mechanism." Procedia Computer Science 199 (2022): 741-748.
[16] Prottasha, N.j., Sami, A.A., Kowsher, M., Murad, S.A., Bairagi, A.K Masud, M., & Baz, M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors, 22(11), 4157.