2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9395802

Sentiment analysis for product rating using a deep learning approach

Krishna Kumar Mohbey
Department of Computer Science
Central University of Rajasthan, Ajmer, India
kmohbey@gmail.com

Abstract- Sentiment analysis is a study about opinions, emotions, and attitudes of the people towards an event or issue. Social networking is an invaluable medium for individuals to express their thoughts and views about any subject or topic, contributing to massive quantities of unstructured knowledge. These emotions can be processed and examined to analyze and obtain insights. Therefore, several machine learning and Natural Language Processing (NLP) based methods were used to examine these opinions. Because of the shifts in the sequential order, stream lengths, and complex logic, the exact sentiments in the consumer feedback are still difficult. Recently, deep learning methods have been used to attain improved results. The most common forms of deep learning method used are the Recurrent Neural Network (RNN) and the Convolutional Neural Network (CNN). Long short-term memory (LSTM), particularly with attentive layers, pays greater focus to the sentiment impact. LSTM has an advantage over alternative RNNs and other deep learning approaches because of relative insensitivity to gap length. Deep learning approaches have better accuracy over existing state-of-the-art approaches because of their ability to handle extensive real-time data and the power of feature extraction, but there is still room for improvement. This paper has used Long Short-Term Model (LSTM) model to predict the customer review's opinion, attaining an accuracy of 93.66%. Furthermore, comparative analysis of the deep LSTM model with existing has been presented.

Keywords- Sentiment Analysis, Deep Learning, LSTM, Machine Learning

I. INTRODUCTION

Sentiment analysis is an original work in Natural Language Processing (NLP). The main goal of sentiment analysis or opining mining is to extract the attributes or views regarding political opinion on social media [1], draft trading strategies from market sentiment [2] , or collect product insights from customer review [3]. To enhance decision-making, business and academic researchers have suggested a variety of ways to address the issue. The typically used techniques of such traditional approaches are the POS (part of speech) tags, phrase sentiment, words position, occurrence and frequency, negative terms, and lexical dependence. The conventional approaches to opinion processing are split into two classes: lexicon-based and machine-learning.

The lexicons' methods use preprocessed lexicons that contain various terms and their polarity to identify a specific word into different classes [4]. Various prevalent sentiment mining approaches include n-gram methods and bag-of-words feature, which could be used with machine learning algorithms stacked together as sentiment classifiers [5]. However, such approaches do not work well with long word sequences and effectively deal with negation. Word embedding pioneering approaches such as Word2Vec [6] and GloVe [7] helps to override such issues. Moreover, using a word embedding with more advanced algorithms can manage long sequential input data such as Recurrent Neural Network (RNN). The conventional approaches are unsuitable for handling a new data pattern with a diverse linguistic existence, increased high dimensional data, and short texts such as tweets with systemic and cultural complexities. These methods are focused on the features extraction that proved to be a difficult challenge with existing data. Researchers have also pointed to the need for new approaches to addressing these problems. Researchers have discovered that Deep Learning (DL) methods yield impressive efficiency and accuracy in opinion mining.

In recent years, we have seen developments in the area of natural language analysis in deep learning models. Deep neural networks (DNNs) consist of the artificial neural network with the input layer to the output layer being multi-hidden. Incorporating DL methods in opinion mining was motivated by automated feature engineering, allowing them to train and explore input features automatically. In addition, their integration was inspired by the growing training data and the progress of word embedding in the multiclass classification [4]. CNN (Convolutional neural networks) are a neural network of feed-forward type. The CNN consists of three layers, input layer, extraction layers, and the layer of classification. Embeddings are generated at the input layer.

Then learn about the related features in extraction layers like convolution and pooling layers. The function layers' features are then moved to the classification layer, which consists of a network entirely attached to a classification scheme. CNN has demonstrated significant progress and creativity in computer vision and image recognition. One of the variants of RNN is Long Short-Term Memory (LSTM), which has shown excellent performance result in handwriting recognition, speech recognition, machine learning, and many sequence problems. RNNs are made up of feed-forward neural networks.

However, RNNs are subtly different by guided loops, which leads to improvement in propagating the activation function into the input sequence from other feed-forward neural networks. Therefore, it has a memory as the status of prior calculations can be conveniently recalled. The RNNs assume that the performance depends on the prior calculation. LSTM Long short-term memory) were suggested to resolve the disappearance gradient problem by incorporating a gating system into classic RNN. They invented the forget gate that helps the memory cell to retain knowledge for a long time or generate previous calculations.

Recently, deep learning has gained immense popularity from researchers and practitioners in various domains, including sentiment analysis. In this paper, LSTM deep learning model is implemented on the product review dataset to gain perception about product quality and performance, thereby aids in increasing sales. Comparative analysis of the deep model with SVM, naïve Bayes, decision tree, and logistic regression is presented where LSTM generates better accuracy results. This paper has used an online product review dataset of ratings and deep learning models to learn and predict the user's review sentiment.

Related work is presented in section 2. The description of data, along with data processing, is shown in section 3. Section 4 is about the approaches used, such as Naïve Bayes, SVM, decision tree, and LSTM. Section 5 lists the performance results of the approaches. The conclusion is drawn in section 6.

II. RELATED WORK

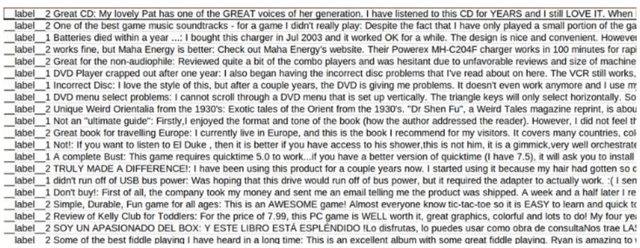
In sentiment classification, deep learning is being widely incorporated by the research community. To illustrate the motive behind the study of sentiments, we briefly examine the esteemed best opinion analysis works. Bhardwaj et al. [8] introduced a method that picks up on stocks' live prices, acting as a key stock market predictor. Haque et al. [9] used Amazon customer feedback to classify the sentiments via Linear SVM, stochastic Gradient Descent, Multinomial Naïve Bayes, Logistic regression, Random Forest, and Decision Tree. Socher et al. [10] have used Recursive autoencoder (RAE) for sentiment distribution prediction implemented at the sentence level. In [10] has designed a sentiment treebank using the RNTN model. In [11] has used CNN trained on word vectors for classification of sentences. The combination of CNN and LSTM is used by Zhou et al. [12] for text classification. Feizollah et al. [13] have used a movie review dataset using deep models to attain 87.7% accuracy in CNN and 86.64 in LSTM. Abdi et al. [14] have used LSTM to evaluate text sentiment considering sentence type, polarity, and word sense. Zhang and Wallace [15] presented an approach that utilizes one-layer CNN for sentence sentiment analysis. They discussed how the model's efficiency could be influenced by a change in its implementation (hyper-parameters, dropouts, filters Size, regularization, etc.). Kumar et al. [16] present a bidirectional long-short architecture based on multi-headed attention to identifying satirical remarks. The model has two key components: word encoder and multi-head attention at the phrase level. Li et al. [17] propose a sentiment analysis method based on a lexicon integrated two-channel CNN-LSTM model that combines branches of CNN and LSTM/BiLSTM in parallel. It uses a new padding mechanism to create a standardized sampling of data inputs and an improvement in each review's amount of emotion. Zhao et al. [18] introduced a model by integrating multiple types of LSTM system to estimate the online user's personality traits. This model transforms the themes interests of consumers and textual opinion into attention information. Zhang et al. [19] recommend a short-term interactive (LSTM) network for an interpretation of conversational opinion in order to model dialog experiences between participants by incorporating (1) a trust layer before every LSTM hidden layer for an observation of prior participants' credibility and (2) combining the output layer with the trained impact to integrate the impacts of former participants. RK Behera et al. [20] suggested Co-LSTM paradigm suggested aims specifically at two sentimental research goals. Firstly, it is highly adaptive to analyze broad social networking data, taking into account the scalability and, secondly, domain dependency-free model, unlike traditional machine training approaches.

III. DATA

Data description and preprocessing steps are discussed in this section.

A. Description of data

Amazon Review for sentiment analysis[1] the dataset is obtained from Kaggle.com that contains user reviews, out of which we have considered 1,32,407 for training and 33,102 for testing. Dataset sample is shown in figure 1.



__label__2 Great CD: My lovely Pat has one of the GREAT voices of her generation. I have listened to this CD for YEARS and I still LOVE IT. When
__label__2 One of the best game music soundtracks - for a game I didn't really play: Despite the fact that I have only played a small portion of the ga
__label__1 Batteries died within a year ...: I bought this charger in Jul 2003 and it worked OK for a while. The design is nice and convenient. Howeve
__label__2 works fine, but Maha Energy is better: Check out Maha Energy's website. Their Powerex MH-C204F charger works in 100 minutes for rapi
__label__2 Great for the non-audiophile: Reviewed quite a bit of the combo players and was hesitant due to unfavorable reviews and size of machine
__label__1 DVD Player crapped out after one year: I also began having the incorrect disc problems that I've read about on here. The VCR still works,
__label__1 Incorrect Disc: I love the style of this, but after a couple years, the DVD is giving me problems. It doesn't even work anymore and I use m
__label__1 DVD menu select problems: I cannot scroll through a DVD menu that is set up vertically. The triangle keys will only select horizontally. So
__label__2 Unique Weird Orientalia from the 1930's: Exotic tales of the Orient from the 1930's. "Dr Shen Fu", a Weird Tales magazine reprint, is abou
__label__1 Not an "ultimate guide": Firstly,I enjoyed the format and tone of the book (how the author addressed the reader). However, I did not feel th
__label__2 Great book for travelling Europe: I currently live in Europe, and this is the book I recommend for my visitors. It covers many countries, co
__label__1 Not!: If you want to listen to El Duke , then it is better if you have access to his shower,this is not him, it is a gimmick,very well orchestrate
__label__1 A complete Bust: This game requires quicktime 5.0 to work...if you have a better version of quicktime (I have 7.5), it will ask you to install
__label__2 TRULY MADE A DIFFERENCE!: I have been using this product for a couple years now. I started using it because my hair had gotten so c
__label__1 didn't run off of USB bus power: Was hoping that this drive would run off of bus power, but it required the adapter to actually work. :(I sen
__label__1 Don't buy!: First of all, the company took my money and sent me an email telling me the product was shipped. A week and a half later I re
__label__2 Simple, Durable, Fun game for all ages: This is an AWESOME game! Almost everyone know tic-tac-toe so it is EASY to learn and quick t
__label__2 Review of Kelly Club for Toddlers: For the price of 7.99, this PC game is WELL worth it, great graphics, colorful and lots to do! My four ye
__label__2 SOY UN APASIONADO DEL BOX: Y ESTE LIBRO ESTÁ ESPLÉNDIDO !Lo disfrutas, lo puedes usar como obra de consultaNos trae LA
__label__2 Some of the best fiddle playing I have heard in a long time: This is an excellent album with some great fiddle playing. Ryan is amazing to

Fig. 1. Sample dataset

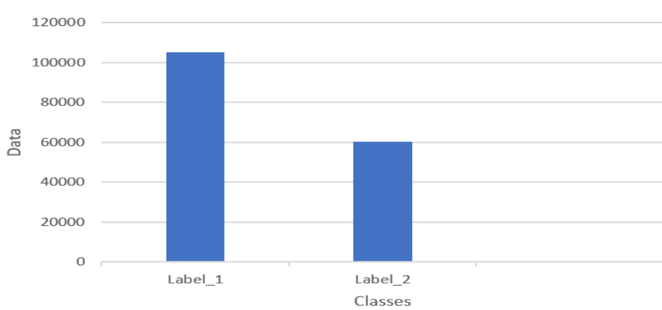


Fig. 2. Rating data

In figure 2, label_1 represented class 1, and it symbolizes to 1 or 2 stars rating given by the service. Another class, represented by label_2, and indicates to 4- or 5-star rating. However, label_3 is of a 3-star rating indicating neutral but is ignored.

Label_1: 1- or 2-star rating
Label_2: 4- or 5-star rating

B. Data pre-processing

In this, we clean the data and smooth the noise. We perform the following data preprocessing steps:

- Normalize words and remove unwanted symbols and punctuation marks.
- Stop words are eliminated, commonly used words such as this, is, are, that, etc.
- Tokenization is performed that is breaking of character sequence into tokens (pieces).
- Vectorization is conducted to convert text into numeric form for easy and fast processing.

IV. APPROACHES

In this section, various sentiment analysis approaches are discussed below.

A. Naïve Bayes

It is one of the most popular probabilistic classifiers under machine learning. This model reads the review and computes the posterior probability for each class, and assigns the class to the review, which has a maximum posterior probability. It utilizes the concept of Bayes conditional probability theorem [21]. In many practical applications such as text classifications, medical problems, and recommendation systems, naive Bayes have demonstrated effectiveness. The computational equation is shown below:

$$\mathrm{P}\left(\frac{Cl}{x}\right) = \frac{\mathrm{P}\left(\frac{x}{Cl}\right).P(Cl)}{P(x)} \tag{1}$$

Where, Clis specified class, x is review used for classification. $P(Cl)$ and $P(x) are$ prior probabilities and $\mathrm{P}\left(\frac{Cl}{x}\right)$ $is\ a$ posterior probability.

For a data point Y={y_1, y_2, y_3, ….., y_i}, the equation can be rewritten as:

$$\mathrm{P}(Cl/x) = P(Cl).\Pi\, P(yi/Cl) \tag{2}$$

B. Support Vector Machine (SVM)

It is a well-known supervised classification model introduced by Vapnik [22]. It is commonly used for classification, regression, and outlier detection. The primary objective of SVM is to establish the best boundary which can divide n-dimensional data into groups. The best-fitted line is called a hyperplane. It identifies the optimal hyperplane that maximizes the margin of two classes. It evaluates the extreme dots to help establish a better hyperplane. These points are referred to as support vectors. This algorithm can be used in medical science, face detection, image classification, pattern classification, and many more. SVM can be of two types: linear and non-linear SVM, and use on the separable property of data. Each training review is marked with one of the available categories, and then the SVM model is trained to allocate the new review. Any hyperplane can define as below:

[1] Amazon Review for Sentiment Analysis
https://www.kaggle.com/bittlingmayer/amazonreviews#train.ft.txt.bz2

$W^T X - b = 0,$ - W is a weight vector
- X is an input vector
- b is bias

C. Decision Tree

It is one of the most widely-applied supervised classification methods in the predictive analytics system. It has an internal flow chart structure, which reflects a "test" state by its internal nodes, each branch shows the test condition outcome, and every leaf node shows a label selected after analyzing all domains. It follows the grouping concept top-to-bottom approach. A decision tree classifier is a non-parametric approach that does not rely on assumptions regarding probability distribution. It can accommodate high-dimensional data with great precision. It produces a sequence of the rule used to determine the class of the new review [23]. It handles overfitting by using the post pruning methods. To choose the splitting criteria used for the best split of data, it uses different attribute selection measures commonly known as splitting rules. Information Gain, Gain ratio, and Gini index are the most common selection measures.

D. Logistic Regression

Logistic regression is a well-known method for statistical modeling used for classification problems. They are used to interpret the results and the relationship between a single binary dependent variable and several ordinals, or independent interval variables. Dependent data have values of win/lose measurements, on/off, true/false [24]. The linear model has log-odds for the value tagged as "1," which implies a merger with another independent variable. A well-known statistical analysis technique used for classification problems is logistic regression. The logistic regression method attempts to limit the cost function to 0-1. It provides us with a collection of probability-based groups when the input data passes through a prediction function, giving a probability score between 0 and 1. It uses a non-linear logistic function or sigmoid function. The sigmoid function can be defined as:

$$y = 1/(1 + e^{-x}) \quad (3)$$

X is the input vector, y is the class vector, and e is the exponential function.

E. Long Short-Term Memory (LSTM)

LSTM [25] is divergent of RNN, which can handle long-term sequences. It keeps the context of previous words. It has input, output, and forget gates that control the flow of information. Every review is passed to LSTM, where firstly forget gate decide what information is to be retained or forget then, the input gate decides what new information is to be stored where the *tanh* layer creates a new vector for the candidate values, and finally, output gate manage the result that we want to show. The sigmoid activation function of the fully connected layer is also used. A dropout layer is put after the LSTM. Figure 3 presents the model of LSTM, where symbol X is an element-wise operator and σ is the activation function.

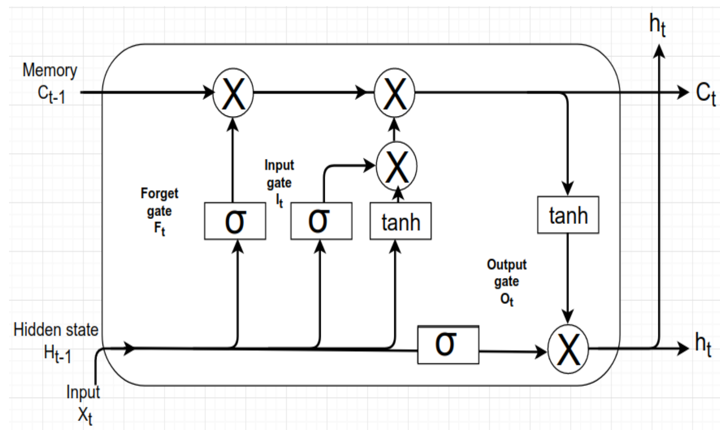


Fig. 3. LSTM model

V. Experimental Result

A. Performance evaluation

Precision, recall, accuracy, AUC, and f-measure are used to assess the performance of the models [23]. Table 1 depicts various performance parameters with their definitions.

Table I: Performance evaluation measures

Matric Name	Definition
Accuracy	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\text{Precision} = \frac{TP}{TP + FP}$
Recall	$\text{Recall} = \frac{TP}{TP + FN}$
F1-Measure	$\text{F1} - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$
AUC	$AUC = \left(Recall - \frac{FP}{FP + TN} + 1\right)/2$

B. Result analysis

The results of various applied models are displayed in this section. For LSTM design, parameter tuning is a must to yield better results. Table 2 shows the hyperparameters tuned for LSTM.

Table II: Hyperparameters values

S No.	Parameter	Value
1	Batch Size	64
2	Dimensions	300
3	Activation	Sigmoid
4	Optimizer	adam
5	Learning rate	0.001
6	No of epochs	10
7	Dropout	0.2

The metrics used is the Area Under ROC Curve (AUC), accuracy, and F1 score for various approaches. Table 3 shows comarative results of various appraoches.

Table III: Results comparision

Methods	Accuracy	Precision	Recall	F1-Score	AUC
LSTM	**0.93**	**0.94**	**0.94**	**0.95**	**0.97**
Naïve Bayes	0.50	0.56	0.49	0.51	0.52
SVM	0.45	0.44	0.42	0.43	0.50
Logistic Regression	0.63	0.56	0.63	0.77	0.50
Decision Tree	0.74	0.73	0.73	0.80	0.73

Figure 4 displays the accuracy of different models where the LSTM model has a maximum accuracy of 93.6%, while SVM and NB are on the lower end with 45.9% and 50%, respectively.

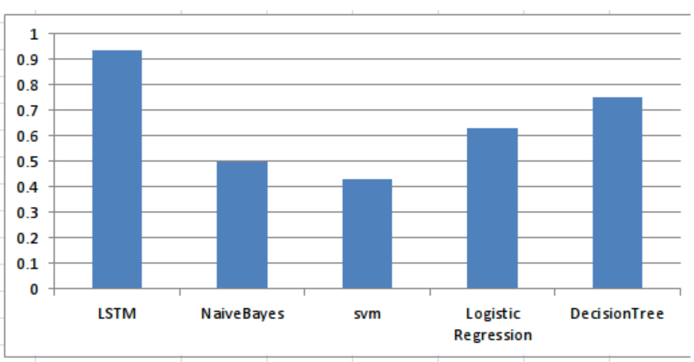


Fig. 4. Accuracy comparison of various approaches

AUC is highest in LSTM with 97.4% while moderate at 73.1% in the decision tree and least in SVM with 50.8%. Figure 5 presents the AUC measure.

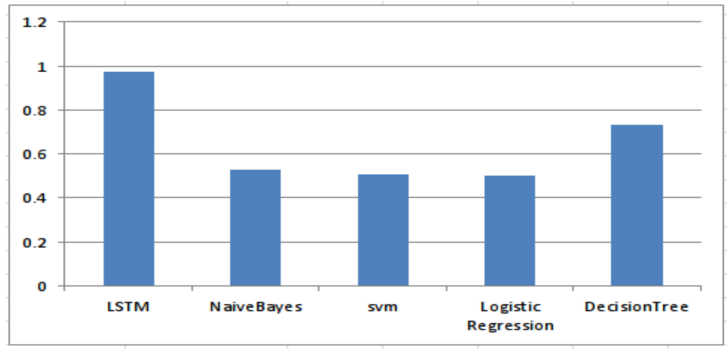


Fig. 5. AUC comparison of various approaches

F1 score is a commonly used parameter for comparison. Figure 6 shows an F1 measure where LSTM has 97.4%, the decision tree has 80.1%, and the least is SVM with 73.1%.

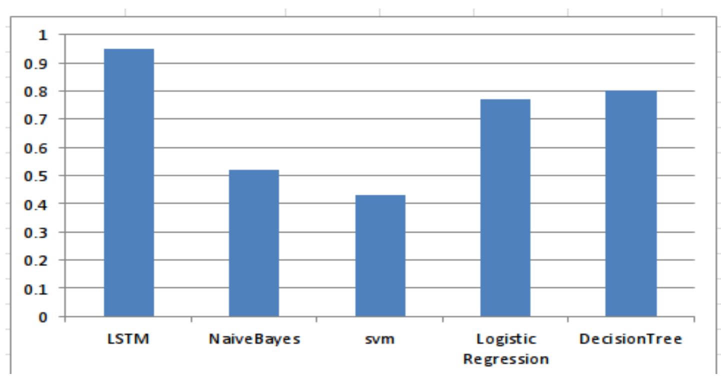


Fig. 6. F1-measure comparison of various approaches

VI. Conclusion

Deep learning results are better on the performance of sentiment analysis for the user's review. The LSTM model does not use any feature extraction method or complex structure such as sentiment treebank. LSTM only depends upon the pre-trained word vector form. The model generates an accuracy of 93.6%, F1-score of 95%, and AUC of 97.4%, significantly surpassing all the existing models. Naïve Bayes and SVM do not show a good result on extensive data and require more training time. The decision tree has obtained the highest accuracy of 74.9% among all machine learning approaches due to getting the best split factor, but it is still less than the deep LSTM.

As for the future work lines, more focus could be done on the hybrid approaches where a combination of CNN and LSTM can be implemented. LSTM could be run on other more massive datasets.

REFERENCES

[1] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, "Analyzing political sentiment using Twitter data," in Information and Communication Technology for Intelligent Systems, Springer, 2019, pp. 427–436.

[2] A. Mudinas, D. Zhang, and M. Levene, "Market trend prediction using sentiment analysis: lessons learned and paths forward," arXiv Prepr. arXiv1903.05440, 2019.

[3] J. Barry, "Sentiment Analysis of Online Reviews Using Bag-of-Words and LSTM Approaches.," in AICS, 2017, pp. 272–274.

[4] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, "Sentiment analysis using deep learning approaches: an overview," Sci. China Inf. Sci., vol. 63, no. 1, pp. 1–36, 2020.

[5] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," arXiv Prepr. cs/0205070, 2002.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv Prepr. arXiv1301.3781, 2013.

[7] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[8] A. Bhardwaj, Y. Narayan, M. Dutta, and others, "Sentiment analysis for Indian stock market prediction using Sensex and nifty," Procedia Comput. Sci., vol. 70, pp. 85–91, 2015.

[9] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in 2018 IEEE international conference on innovative research and development (ICIRD), 2018, pp. 1–6.

[10] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in Proceedings of the 2011 conference on empirical methods in natural language processing, 2011, pp. 151–161.

[11] A. Rakhlin, "Convolutional Neural Networks for Sentence Classification," GitHub, 2016.

[12] C. Zhou, C. Sun, Z. Liu, and F. Lau, "A C-LSTM neural network for text classification," arXiv Prepr. arXiv1511.08630, 2015.

[13] A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, and M. Hazim, "Halal products on Twitter: Data extraction and sentiment analysis using stack of deep learning algorithms," IEEE Access, vol. 7, pp. 83354–83362, 2019.

[14] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion," Inf. Process. \& Manag., vol. 56, no. 4, pp. 1245–1259, 2019.

[15] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," arXiv Prepr. arXiv1510.03820, 2015.

[16] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional LSTM," Ieee Access, vol. 8, pp. 6388–6397, 2020.

[17] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN--LSTM family models," Appl. Soft Comput., vol. 94, p. 106435, 2020.

[18] J. Zhao, D. Zeng, Y. Xiao, L. Che, and M. Wang, "User personality prediction based on topic preference and sentiment analysis using LSTM model," Pattern Recognit. Lett., vol. 138, pp. 397–402, 2020.

[19] Y. Zhang et al., "Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis," Neural Networks, vol. 133, pp. 40–56, 2021.

[20] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," Inf. Process. \& Manag., vol. 58, no. 1, p. 102435, 2021.

[21] A. Rane and A. Kumar, "Sentiment Classification System of Twitter Data for {US} Airline Service Analysis," Jul. 2018, doi: 10.1109/compsac.2018.00114.

[22] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[23] S. Kumar, V. Koolwal, and K. K. Mohbey, "Sentiment analysis of electronic product tweets using big data framework," Jordanian J. Comput. Inf. Technol., 2019, doi: 10.5455/jjcit.71-1546924503.

[24] D. R. Cox, "The Regression Analysis of Binary Sequences," J. R. Stat. Soc. Ser. B, vol. 20, no. 2, pp. 215–232, Jul. 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.