Sentiment Analysis of Financial Textual data Using Machine Learning and Deep Learning Models

Hero O. Ahmad[1], Shahla U. Umar[2]
[1]Presidency of the University of Kirkuk, University of Kirkuk, Iraq
[2]College of Computer Science and Information Technology, University of Kirkuk, Iraq
E-mail: heroomarahmad1984@gmail.com

Keywords: social media comments, MNB, LR, RNN, LSTM, GRU

Received: February 9, 2023

Recently, extensive research in the field of financial sentiment analysis has been conducted. Sentiment analysis (SA) of any text data denotes the feelings and attitudes of the individual on particular topics or products. It applies statistical approaches with artificial intelligence (AI) algorithms to extract substantial knowledge from a huge amount of data. This study extracts the Sentiment polarity (negative, positive, and neutral) from financial textual data using machine learning and deep learning algorithms. The constructed machine learning model used Multinomial Naïve Bayes (MNB) and Logistic regression (LR) classifiers. On the other hand, three deep learning algorithms have been utilized which are Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU). The results of the MNB and LR obtained a good and very good rate of accuracy respectively. Likewise, the results of RNN, LSTM, and GRU obtained an excellent rate of accuracy. It can be concluded from the outcomes that the used preprocessing stages made a positive impact on the accuracy rate.

Povzetek: Raziskava analizira finančna besedila z uporabo klasičnega strojnega učenja in globokega učenja z namenom iskanja sentimenta (negativno, pozitivno, nevtralno).

1 Introduction

Recently, the advancement of technology produces an enormous number of social media users that generate huge data [1]. The data of the users consist of a huge number of writings in the form of text comments. This constructive information from social media is an emerging area for research [2]. The term "Deep Learning" (DL) pertains to the latest advancement in technology and current research emphasis in the field of Machine Learning. DL has become ubiquitous in our daily lives, offering solutions that were once considered the subject matter of science [3]. The widespread use of machine learning and deep learning has made it possible to apply them in various fields, such as computer vision, machine translation, recommendation systems, cybersecurity, and sentiment analysis [4]. Sentiment analysis (SA) is also called Opinion analysis or Opinion mining, it is a subfield of natural language processing (NLP) that evaluates the degree of polarity in the sentence to analyze and extracts feelings from text data. Several types of research have been carried out by establishments for finding people's sense of a given matter [5]- [6]. The polarity involves determining the attitude or emotion expressed in a piece of text, which can be positive, negative, or neutral. It is a subfield of text classification, which involves analyzing people's opinions, emotions, and attitudes towards entities and their characteristics as expressed in a written text [7]. SA can be challenging because it can be difficult for humans to determine the emotion behind the text. Additionally, a single text can contain multiple emotions, which can make it difficult to achieve high accuracy in the analysis process. Hence, identifying the appropriate features or markers to effectively distinguish between different classes is a significant obstacle in any classification task that involves text [8] [9]. Assessing the emotions expressed through digital channels like financial news comments can be advantageous for creating trading plans. Furthermore, the emotional content conveyed through financial news has the potential to predict future trends, which can be beneficial for those managing portfolios and risks [10]. Furthermore, SA can be conducted at three levels which are the aspect level, the sentence level, and the document level. With the rise of social media and other platforms that allow users to share their thoughts and feelings about various things, such as entities, products, people, and organizations, it has become possible to analyze the sentiments expressed in these reviews and other online content [11]. Throughout the last decade, Sentiment analysis methods have grown greatly and evolved from basic statistics rules to advanced machine learning methods such as deep learning, which has become an outstanding technology in various NLP projects. Likewise, these machine and deep learning systems have achieved impactful outcomes on the SA data [12]. Deep learning is a method of representation learning that uses nonlinear neural networks to learn multiple levels of representation. These representations are created by transforming the representation at one level into a more abstract representation at a higher level. These learned representations can be used as features in detection or classification jobs [13]. The main goal when using classification algorithms is to properly prepare and preprocess the dataset and, ideally, use a large number of data points to train the model. Different supervised and unsupervised algorithms with their parameters can be tested for finding the sentiment to achieve the best results as shown in figure 1 [14]. Hence, SA can be used in a variety of contexts beyond just product reviews, such as in the analysis of stock markets, news articles, political debates, advertising, elections, etc. [15].

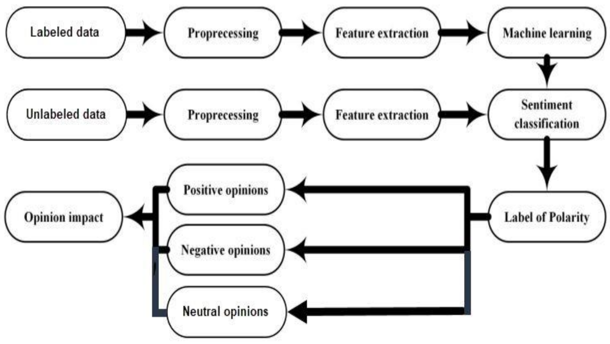

Figure 1: Organization of sentiment analysis [16].

In this study, the aim is to extract the sentiment polarity from financial textual data using both machine learning and deep learning algorithms. Specifically, Multinomial Naïve Bayes (MNB) and Logistic regression (LR) classifiers are used as machine learning models and recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent units (GRU) as deep learning models. These models can capture the contextual dependencies and temporal dependencies present in text data, which are crucial for accurately identifying the sentiment expressed in a given text. We also examine the impact of the utilized preprocessing stages on the model's accuracy and also on the heterogeneous data collected from different sources.

Overall, this research aims to contribute to the growing body of knowledge on sentiment analysis in the field of finance from social media and provide insights into the effectiveness of different machine learning and deep learning approaches for this task.

This article is organized as follows: Section two explains related work for sentiment analysis, Section three illustrates the proposed system architecture, Section four portrays the experimental and results, and Section five demonstrates the discussion. Finally, the conclusion and future work are in section 6.

2 Related work

Nowadays, many articles concentrated on addressing the problem of sentiment analysis classification by using different techniques for classifying the sentiment of individuals at several stages. In this section, the focus will be on the latest contributions of this field.

G. Mostafa et al, 2021 [17] Used a few algorithms of machine learning on the Twitter data by utilizing many steps of preprocessing and encoding methods for increasing the accuracy rate. Then a comparison among the attained accuracies was presented. Their experiments demonstrate that the Neural Network algorithm offers outstanding accuracy compared with other algorithms. A. Al Shamsi et al, 2022 [18] Constructed the Emirati dialect dataset for the Instagram platform according to three comment polarities. They evaluate the quality of their corpus utilizing Cohen's kappa coefficient and also assessed the quality of the corpus by employing eight machine learning algorithms. Finally, they compared the performance of the utilized algorithms to find out the accurate classifiers. A. FiIzollah et al, 2019 [19] concentrated on halal products from the Twitter dataset to extract sentiment polarity from the tweets by utilizing deep learning in English and Malay languages. The highest accuracy results were scored by using CNN and LSTM algorithms. B. Fazlija et al, 2022 [20] discovered the financial sentiment knowledge from the news article. Then apply the predicted sentiment scores to estimate the stock market price direction by using the BERT algorithm. H. Fouadi et al, 2021 [21] Their work attempts to compare the performance of Arabic sentiment analysis by employing several machines and deep learning algorithms to extract the polarity of the sentiment. The machine learning algorithms include Support Vector Machines (SVM), Logistic Regression (LR), and K-Nearest Neighbors (KNN). The deep learning algorithm is used the Long Short-Term Memory (LSTM) model. These algorithms are applied to a dataset called the Arabic Review, which consists of manually annotated text data collected from various Arabic sources. H. Shehu et al, 2021 [22] Their study uses three techniques to expand the size of the training data: Shift, Shuffle, and Hybrid. Then, we employ three deep learning models which are recurrent neural network (RNN), convolution neural network (CNN), and hierarchical attention network (HAN) to classify stemmed Turkish data in Twitter for sentiment analysis and compare the performance of these models to traditional machine learning models. S. Liu et al, 2020 [23] evaluated the ability of various machine and deep learning models to predict user sentiment polarities and found that certain techniques, such as using binary bag-of-word, incorporating bi-grams, and normalizing text, improved the performance of machine learning models. For deep learning models, discovered that using pre-trained word embeddings and limiting maximum length often enhanced model performance. Also, he found that simpler models such as LR and SVM were more effective at predicting sentiments than more complex models like Gradient Boosting, LSTM, and BERT. G. Kaur1 et al, 2023 [24] introduce a method for sentiment analysis that combines different approaches. The process involves three main steps: pre-processing, feature extraction, and sentiment classification. To eliminate unwanted data from text reviews, the pre-processing stage uses NLP techniques. To extract features effectively, the authors introduce a hybrid approach that combines review-related and aspect-related features to create a unique hybrid feature vector for each review. Finally, sentiment classification is carried out using LSTM deep learning classifier.

The main goal of this research is to expand the existing knowledge about sentiment analysis in the field of finance, particularly from social media sources. Additionally, the study aims to provide valuable insights into the effectiveness of various machine learning and deep learning techniques for this specific task and then find out the most suitable algorithm for this area. However, testing the impact of various preprocessing steps in the process of increasing the accuracy in each model.

3 Proposed system architecture

In this section, the proposed system framework for financial sentiment analysis using machine learning and deep learning algorithms is outlined. The system architecture includes five types of algorithms, the overall framework follows a set of steps to achieve its goal and it is depicted in Figure 2. As can be seen, the proposed system steps consist of six stages:

1- Data Collection: Financial data is collected from multiple social media sources, including Facebook, Twitter, and financial blogs.
2- Data Aggregation: The data is aggregated and filtered by using a Python algorithm for detecting and collecting social comments in the English language only.
3- Preprocessing: The collected data undergoes several preprocessing steps to improve the accuracy of the models. These steps include feature selection, data cleaning, data balancing, tokenization, stemming, etc.
4- Model Selection: Five different algorithms are evaluated for sentiment analysis: a Multinomial Naïve Bayes (MNB) and Logistic regression (LR) classifiers as machine learning models, and also Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and Long Short Term Memory (LSTM) as deep learning models. All the utilized deep learning models are implemented by using the Keras library on top of the TensorFlow.
5- Model Training and Evaluation: The data is split into a training set and a testing set. The training set is used to train the models, and the testing set is used to evaluate their performance. Evaluation metrics, including precision, recall, and F1 measure score, are used to measure the accuracy of the models.
6- Model Comparison: Comparison between the results of the models based on both accuracy and time performance.

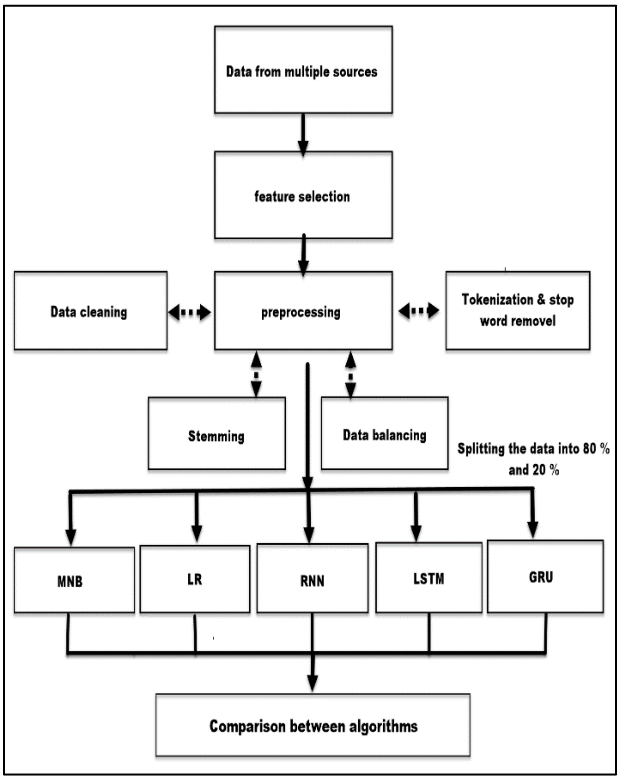

Figure 2: The proposed system steps

4 Experimental and results

In the following subsections, the experiments and results have been shown. An experimental setup is introduced in subsection 4.1, and the dataset description in detail is demonstrated in subsection 4.2. Lastly, the results and analysis of all the classifiers are presented in subsection 4.3.

4.1 Experimental setup

The experiments were applied in the Windows 64-bits operating system. Additionally, anaconda with python 3.8.8 has been used. Also, to implement deep learning algorithms TensorFlow and Keras frameworks have been installed. The rest environmental information's shown in table 1.

Table 1. The environment of tests.

No.	Used resource	Resource information
1	Operating System	Windows 10, 64-bits
2	Computer CPU	Intel(R) Core (TM) i7-4600U CPU @ 2.10 GHz 2.70 GHz
3	Type of hard disk drive	SSD
4	Tensor Flow Version:	2.3.0
5	Keras Version:	2.4.0
6	Pandas:	1.2.4
7	python Version	The latest release of anaconda with python 3.8.8

4.2 Dataset

The dataset in the experiments is based on the different social media platforms. Many texts are not appropriate for the experiments, such as being either too long, having complex emotional tendencies, or having too many special characters. Hence, the original textual data are preprocessed. The dataset was taken from various social media site API which presents the financial comments. It consists of two attributes the first one is the comments and the second one is the label. Lastly, there are about 70000 records that meet the criteria of the algorithms which consist of positive, negative, and neutral data. The dichotomy of sentiment research in this paper is, negative which is represented by 0, neutral by 1 and, positive by 2. However, the selected 80% of the dataset is the training set, and 20% is the test set. Figure 3 shows the screenshot of the data before and after preprocessing.

	Sentence	Sentiment
0	The GeoSolutions technology will leverage Bene...	positive
1	*ESIonlows*, *down*1.50 to $2.50 BK a real po...	negative
2	For the last quarter of 2010 , Componenta 's n...	positive
3	According to the Finnish-Russian Chamber of Co...	neutral
4	The Swedish buyout firm has sold its remaining...	neutral

	Sentence	Sentiment
0	geosolutions technology leverage benefon gps s...	2
1	esi low 150 250 bk real possibility	0
2	last quarter 2010 componenta net sale doubled ...	2
3	according finnishrussian chamber commerce majo...	1
4	swedish buyout firm sold remaining 224 percent...	1

Figure 3: Snapshot of data before & after preprocessing.

4.3 Results and analysis

As mentioned before, the proposed work is consisting of two machine learning algorithms and three types of deep learning algorithms. Since the data is not balanced we did an up-sampling process on the minority classes and then concatenate the up-sampled data frames with the higher-class data frame. For the MNB we create a grid search object using the classifier and the hyperparameter dictionary and sets the number of folds for cross-validation to 5. Also, for LR the maximum iteration was set at 500. On the other hand, the deep learning algorithms are applied in 10 epochs with 128 batch sizes. Finally, for each algorithm, the measures of accuracy, precision, recall, F1 score, and execution time performance have been computed as shown in the following table 2 and figures 4, and 5.

Table 2. Results of the algorithms

Algo.	Preci.	Recall	F1-score	Acc.	Exe. Time in Min.
MNB	0.75	0.77	0.76	0.74	1.2
LR	0.88	0.84	0.86	0.85	1.3
RNN	0.96	0.92	0.94	0.94	12.6
LSTM	0.96	0.96	0.96	0.96	121.1
GRU	0.96	0.97	0.96	0.95	93.9

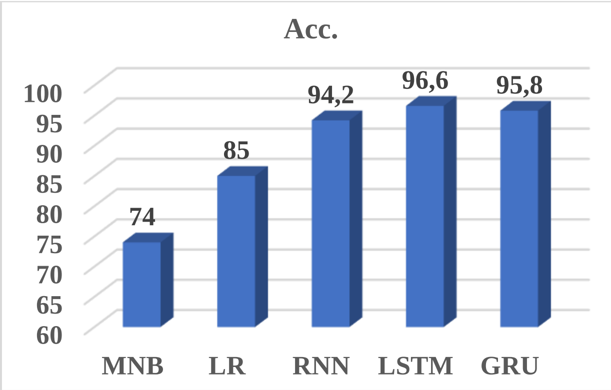

Figure 4: Accuracy results of the algorithms.

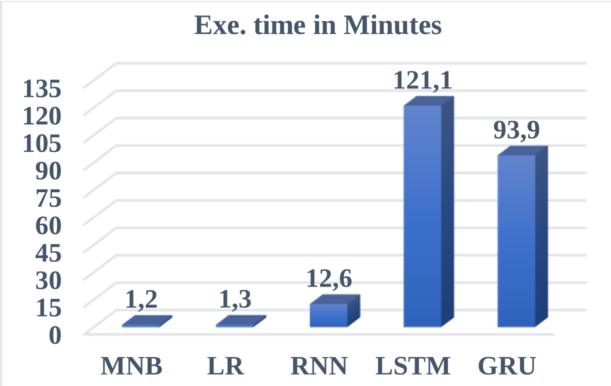

Figure 5: Execution time results of the algorithms.

The MNB classifier achieved an accuracy of 74% and LR achieved 85% while the RNN, GRU, and LSTM models achieved accuracies of 94.2%, 95.8%, and 96.6%, respectively. These results suggest that the deep learning models outperformed the machine learning classifiers but their time performance relatively are high. In addition, the LSTM model achieves the highest accuracy with the highest execution time, on the other hand, the MNB classifier scored the lowest accuracy rate with the lowest execution time compared with the other algorithms.

5 Discussion

The results of this study demonstrate the potential of machine learning and deep learning algorithms for sentiment analysis of financial data. The deep learning models, in particular, showed significantly higher accuracy compared to the MNB and LR classifiers. This suggests that the use of more advanced algorithms can improve the performance of sentiment analysis for financial data. However, it is important to note that the accuracy of the models may be affected by various factors, such as the quality and quantity of the data, the preprocessing steps taken, and the specific parameters of the algorithms.

Further optimization of the models, using techniques such as swarm optimization or bat optimization, may be necessary to achieve even higher accuracy in the future [25].

6 Conclusion

This study collects and aggregates a financial dataset from multiple social media sources for sentiment analysis to inspect how the utilized preprocessing and the models are performed with the heterogeneous data. Python algorithm has been utilized to detect and collect the social comments in English language only. The experiments are classified the financial data into three polarities (positive, negative, and neutral) by using machine and deep learning algorithms. The paper made a comparison between five types of algorithms to find out which one provides more accuracy and less time execution. The models of deep learning are constructed by using the Keras package on top of TensorFlow. The dataset went through many preprocessing steps to increase the accuracy of the models such as feature selection, data cleaning, data balancing and etc. The process of data resampling and data balancing has made a significant impact on the accuracy of all the utilized algorithms. The scored accuracy results for all the constructed models in order are MNB 74%, LR 85% RNN 94%, GRU 95%, and, LSTM 96%.

The future work will focus on gaining better accuracy results by applying some optimization algorithms such as Swarm, Bat, etc.

References

[1] David Zimbra Ahmed Abbasi, Daniel Zeng, Hsinchun Chen, "The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation," *ACM Transactions on Management Information Systems,* vol. 9, no. 2, p. 1–29, June 2018.

[2] Arif Ullah, Sundas Naqeeb Khan, Nazri Mohd Nawi, "Review on sentiment analysis for text classification techniques from 2010 to 2021," *Multimedia Tools and Applications,* pp. 2-58, October 2022.

[3] Inam Abdullah Abdulmajeed, Idress Mohammed Husien, "MLIDS22- IDS Design by Applying Hybrid CNN-LSTM Model on Mixed-Datasets," *Informatica,* vol. 46, p. 121–134, 2022.

[4] Sensen Guo, Xiaoyu Li, Zhiying Mu, "Adversarial Machine Learning on Social Network: A Survey," *Frontiers in Physics,* vol. 9, pp. 1-18, November 2021.

[5] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review,* vol. 55, p. 5731–5780, February 2021.

[6] Yili Wang, Jiaxuan Guo , Chengsheng Yuan, Baozhu Li, "Sentiment Analysis of Twitter Data," *Applied Sciences MDPI,* pp. 2-14, November 2022.

[7] Zenun Kastrati, Fisnik Dalipi, Ali Shariq Imran, Krenare Pireva Nuci, Mudasir Ahmad Wani, "Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study," *Applied Sciences MDPI,* vol. 11, pp. 1- 23, April 2021.

[8] Lai Po Hung, Suraya Alias, "Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection," *Journal of Advanced Computational Intelligence and Intelligent Informatics,* vol. 27, no. 1, pp. 84 - 95, 2023.

[9] Rokas Štrimaitis 1, Pavel Stefanovi, Simona Ramanauskaite, Asta Slotkiene, "Financial Context News Sentiment Analysis for the Lithuanian Language," *Applied Sciences MDPI,* vol. 11, pp. 1 - 13, May 2021.

[10] Justina Deveikyte, Helyette Geman, Carlo Piccari, Alessandro Provetti, "A sentiment analysis approach to the prediction of market volatility," *Frontiers in Artificial Intelligence,* pp. 1- 10, December 2022.

[11] Ali Shariq Imran, Sher Muhammad Daudpota, Zenun Kastrati, Rakhi Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," *IEEE Access,* vol. 8, pp. 181074 - 181090, September 2020.

[12] Mohammed Nazim Uddin, Md. Ferdous Bin Hafiz, Sohrab Hossain, Shah Mohammad Mominul Islam, "Drug Sentiment Analysis using Machine Learning Classifiers," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 13, no. 1, pp. 92- 100, 2022.

[13] Duyu Tang, Bing Qin, Ting Liu, "Deep learning for sentiment analysis: successful approaches and future challenges," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 5, p. 292–303, October 2015.

[14] Fatemeh Hemmatian, Mohammad Karim Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif Intell Rev,* vol. 52, p. 1495–1545, 2019.

[15] Osama M. Rababah, Ahmad K. Hwaitat, Dana A. Al Qudah, "Sentiment analysis as a way of web optimization," *Scientific Research and Essays,* vol. 11, no. 8, pp. 89 - 96, April 2016.

[16] Anastasia Giachanou, Fabio Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys,* vol. 49, no. 2, p. 1–41, June 2017.

[17] Golam Mostafa, Ikhtiar Ahmed, Masum Shah Junayed, "Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data," *I.J. Information Technology and Computer Science,* vol. 2, pp. 38-48, 2021.

[18] Arwa A. Al Shamsi, Sherief Abdallah, "Sentiment Analysis of Emirati Dialects," *Big Data and Cognitive Computing,* Vols. 6-57, pp. 1-18, May 2022.

[19] Ali Feizollah, Sulaiman Ainin, Nor Badrul Anuar, "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms," *IEEE Access*, Vols. 7, 2019, pp. 83354-83362, June 17, 2019.

[20] Bledar Fazlija * and Pedro Harder, "Using Financial News Sentiment for Stock Price," *MDPI Mathematics*, pp. 1-20, June 2022.

[21] Hassan Fouadi, Hicham El Moubtahij, Hicham Lamtougui, Ali Yahyaouy, "Sentiment Analysis of Arabic Comments Using Machine Learning and Deep Learning Model," *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 13, no. 3, pp. 598 - 606, June 2022.

[22] Harisu Abdullahi Shehu, Md. Haidar Sharif, MD. Haris Uddin Sharif, Ripon Datta, Sezai Tokat, Sahin Uyaver, Huseyin Kusetogullari, Rabie A. Ramadan, "Deep Sentiment Analysis: A Case Study on Stemmed Turkish Twitter Data," *IEEE Access*, vol. 9, pp. 56836 - 56854, April 2021.

[23] S. Liu, "Sentiment Analysis of Yelp Reviews: A Comparison of Techniques and Models abs/2004.13851 (2020): n. pag.," *ArXiv*, April 2020.

[24] Gagandeep Kaur, Amit Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *Journal of Big Data*, vol. 10, no. 5, pp. 1 - 23, 2023.

[25] Shahla U. Umar, Tarik A. Rashid, "Critical Analysis: Bat Algorithm based Investigation and Application on Several Domains," *World Journal of Engineering*, vol. 18, no. 4, pp. 606-620, July 2021.