Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 244 (2024) 1–8

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on AI in Computational Linguistics

Improving sentiment analysis of financial news headlines using hybrid Word2Vec-TFIDF feature extraction technique

Meera George[a,*], Dr. R. Murugesan[b]

[a]Research Scholar, Department of Humanities & Social Sciences, National Institute of Technology Tiruchirappalli, Tamil Nadu 620015, India
[b] Professor, Department of Humanities & Social Sciences, National Institute of Technology Tiruchirappalli, Tamil Nadu 620015, India

Abstract

With the evolution of big data and information technology, sentiment analysis has become a research hotspot in the financial market. Researchers are increasingly focused on improving the efficiency of sentiment analysis using different machine learning and deep learning architectures. Feature extraction is a fundamental process in sentiment analysis that enhances text representation and classification. Though various feature extraction techniques are present in the field, limited attention has been given to hybrid feature extraction techniques. The primary objective of this study is to improve the sentiment analysis of financial news headlines using a hybrid Word2Vec-TFIDF feature extraction technique. The study evaluates the performance of Word2Vec, Doc2Vec, TFIDF, Word2Vec-TFIDF, and Doc2Vec-TFIDF with six machine learning classifiers. The results find that the hybrid feature extraction technique, Word2vec-TFIDF with SVM classifier, outperforms all other sentiment analysis models with an accuracy of 82 percent.

Keywords: Sentiment analysis; Feature extraction; Financial news; Word2Vec-TFIDF; SVM classifier

1. Introduction

With the emergence of digital communication platforms and the accessibility of enormous text data, sentiment analysis has gained significant prominence over the past decade [1]. Sentiment analysis is a Natural Language Processing (NLP) Tool that identifies human emotions, opinions, and sentiments from text data [2]. This analysis

* Corresponding author. Tel.: +91 8590882793.
E-mail address: 409121002@nitt.edu

1877-0509
10.1016/j.procs.2024.10.172

classifies the text into positive, negative, and neutral sentiments. A fundamental task in sentiment analysis is feature extraction, which enhances classification by extracting crucial information from the text data [3]. This method associates words from the unstructured data into vectors of continuous real numbers [4].

Over the years, researchers have employed various feature extraction techniques, such as Word2Vec, Doc2Vec, and TF-IDF, for sentiment analysis of text data. TF-IDF is computationally efficient and easy to calculate. Word2Vec determines semantic relationships between words, while Doc2Vec produces precise word vectors by considering word order. However, relying on a single technique can compromise text representation. For example, TF-IDF cannot unearth word relationships, and Word2Vec cannot handle out-of-vocabulary words. Therefore, this study employs a hybrid feature extraction technique to utilize the strengths of each method for generating high-quality vectors.

With the rise in news consumption, the importance of news headlines has increased more than ever [9]. Financial news headlines are a source of sentiment in the financial market [10]. Market participants and policymakers rely on these sentiments to enhance decision-making. Henceforth, it is necessary to improve the performance of sentiment analysis on financial news headlines.

The primary aim of this study is to improve the sentiment analysis of financial news headlines using a hybrid feature extraction technique. The study evaluates the performance of five feature extraction techniques, Word2Vec, Doc2Vec, TFIDF, Doc2Vec-TFIDF, and Word2Vec-TFIDF, with six machine learning classifiers, Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Gradient Boosting (GB), Extra Trees (ET), and Logistic Regression (LR). The results find that the accuracy of sentiment analysis can be significantly improved by combining Word2Vec-TFIDF hybrid feature extraction with the SVM classifier.

2. Related works

Over the past decade, sentiment analysis has gained prominence due to the availability of enormous textual data on various news and social media platforms. Researchers used multiple feature extraction techniques for sentiment analysis on news datasets. Traditional models of feature extraction and text representation involve applying statistical methods such as the Term Frequency (TF) approach. With the advancements in NLP, learning-based methods such as Word2Vec, GloVe, and FastText were utilized [11]. [12] proposed a sentiment analysis model that combines BERT and Word2Vec feature extraction techniques with POS tagging to extract the feature vectors from the financial news dataset. These vectors were fed into a CNN classifier, resulting in improved performance compared to other baseline models. [13] performed sentiment analysis on a large financial news dataset using Doc2Vec for feature extraction, followed by classification with LSTM, Bi-LSTM, and stacked Bi-LSTM. To obtain the sentiments of comments under electronic news data, [14] used FastText for feature extraction and NB, SVM, and MPL for classification. Results show that the SVM classifier performs best with Fast Text. [15] employed Word2Vec and FastText feature extraction techniques for performing a sentiment analysis on Sinhala news text, concluding that FastText with LSTM yields higher prediction results. [16] analyzed the performance of Word2Vec and CNN on real news articles and tweets and found that using Word2Vec significantly improves the prediction accuracy of the classifier. [17] employed Word2Vec, Glove, and FastText to extract the feature vectors from the financial news headlines, which were then classified using BiGRU-Attention and CNN. The study found that the model using Glove and BiGRU-Attention outperformed all others. [18] performed a sentiment analysis on stock posts and financial news data using Word2Vec and later combined them with an LSTM model to predict the China Shanghai A-share market. [19] created embeddings on financial news using Doc2Vec, which were then fed into CNN and LSTM models for classification. The results showed that the performance improved when Doc2Vec vectors were used with LSTM. [20] classified Hindi news by employing count vectorizer, TFIDF, and Doc2Vec as feature extraction techniques, and the results found Doc2Vec vectorizer with linear regression as the best performing model with an accuracy of 97.04 percent.

From the literature, it is observed that studies employing hybrid feature extraction techniques are scarce. Limited studies have assessed the performance of different feature extraction techniques, including hybrid models on financial news data. This study attempts to fill this gap. The study evaluates the performance of different feature extraction techniques to enhance text representation and, thereby, the performance of sentiment analysis.

3. Methodology

The primary objective of the study is to enhance the performance of sentiment analysis using a hybrid feature extraction technique. For this, the study evaluates the performance of different feature extraction techniques (Word2Vec, TFIDF, Doc2Vec, hybrid Word2Vec-TFIDF, hybrid Doc2Vec-TFIDF) and machine learning classifiers (RF, SVM, DT, GB, ET, and LR). The 200501 financial news headlines are collected from the public platform Kaggle for 2017-2021[1]. The financial news data is primarily pre-processed using seven pre-processing techniques: lowercase conversion, white space removal, tokenization, stop words removal, punctuation removal, stemming, and lemmatization. Lowercase conversion reduces vocabulary by converting all text to lowercase. White space removal eliminates any spaces or tabs from a string. Tokenization breaks down text into smaller tokens, while stop words removal eliminates common words that do not provide any value to text analysis. Punctuation removal eliminates any punctuation in the text. Stemming and lemmatization reduce the vocabulary size by converting words to their base forms. While these pre-processing techniques improve the quality of the data, certain pre-processing techniques when applied alone can reduce the accuracy of classification. Hence, the pre-processing techniques were selected after analyzing the impact of various combinations of these techniques on the classification accuracy. The study then uses three feature extraction techniques: Word2Vec, Doc2Vec, TFIDF, and two hybrid feature extraction techniques- Word2Vec-TFIDF and Doc2Vec-TFIDF. The schematic diagram of the methodology is shown in Fig. 1.

Word2Vec is a two-layer neural network representing words into vectors by capturing their similarity. It helps efficiently implement the continuous bag of words and Skip-Gram for calculating the word vectors from the text data. On the other hand, Doc2Vec is an extension of Word2Vec that extracts the word vectors from a paragraph or document. It is instrumental in preserving the semantic information in large textual data. TFIDF (Term Frequency – Inverse Document Frequency) is another feature extraction technique that determines the significance of words in text data. The extracted vectors from Word2Vec and Doc2Vec are individually concatenated to the TFIDF vectors to create two hybrid feature extraction techniques, Word2Vec-TFIDF and Doc2Vec-TFIDF. These hybrid vectors for Word2Vec-TFIDF and Doc2Vec-TFIDF can be represented in Eq.1 and Eq. 2, respectively.

$$S(h_i) = \emptyset(TFIDF(h_i), w2v(h_i) \quad (1)$$
$$S(h_i) = \emptyset(TFIDF(h_i), d2v(h_i) \quad (2)$$

Given a news document h, $S(h_i)$ is the total vector of the i^{th} news headline in the document. $TFIDF(h_i)$ represents the vectors of the i^{th} news headline extracted using the TF-IDF approach. $w2v(h_i)$ and $d2v(h_i)$ are the vectors extracted using the Word2Vec approach and Doc2Vec approach respectively. ∅ represents the concatenation process.

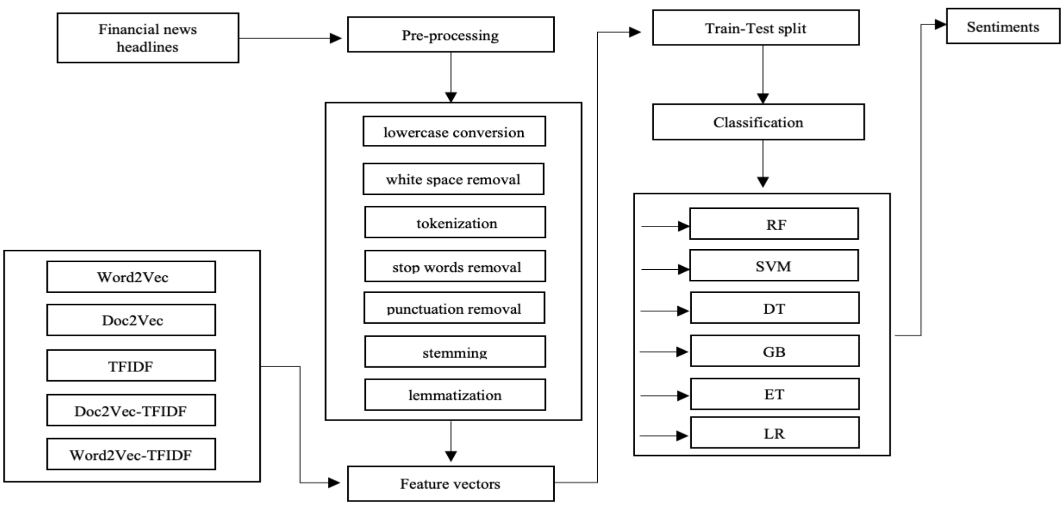

Fig. 1. Schematic diagram of methodology.

The vectors obtained from each feature extraction technique are fed into the classifiers to classify the news sentiments. Machine learning classifiers, such as SVM, yields high accuracy and optimization in sentiment classification while deep learning architectures are highly complex in capturing word compositionality [21]. Hence, the study uses six machine learning classifiers viz. RF, SVM, DT, GB, ET, and LR. RF is an ensemble machine learning model used for regression and classification tasks. RF is a collection of decision trees where each tree provides a classification result based on the input data [22]. It aggregates these classifications and selects the one with the most votes the final prediction. SVM is an effective tool for solving binary classification problems. It performs classification based on a hyperplane that separates the data into distinct classes. The primary aim of this algorithm is to maximize the hyperplane margin while minimizing the number of incorrectly classified data points [23]. DT is a supervised machine learning algorithm that makes predictions based on input features. It has a tree-like structure, where branches represent possible feature values and leaf nodes represent labels [24]. GB, like RF, makes predictions based on multiple DTs. It combines weak learners, typically decision trees, to create a strong predictive model [25]. ET is a tree-based model that generates multiple DTs from the training dataset [26]. For each leaf node, Gini index values are calculated. LR is a supervised machine learning algorithm that classifies binary response variables [27]. It can be extended to categorical variables by introducing dummy variables.

The performance of these models is evaluated using four valuation metrics: precision, recall, F1-Score, and accuracy. Given true positives as T_p, false positives as F_p, true negatives as T_n, and false negatives as F_n, the formula for the valuation metrics is given in Eq. 3, Eq.4, Eq.5, and Eq.6.

$$Precision = \frac{T_p}{T_p+F_p} \tag{3}$$

$$Recall = \frac{T_p}{T_p+F_n} \tag{4}$$

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

$$Accuracy = \frac{T_p+T_n}{T_p+T_n+F_p+F_n} \tag{6}$$

4. Results and discussions

The study performs sentiment analysis on five-year financial news data. The sample of raw data is presented in Table 1. A total of 200501 new headlines are present in the dataset. The characteristics of the data are represented in Table 2. It is observed that the dataset has a significant vocabulary size, and the average length of headlines sums to approximately 12 words.

Table 1. Sample of raw data

Date	News Headline
05/01/17	Eliminating shadow economy to have positive impact on GDP : Arun Jaitley
05/01/17	Two Chinese companies hit roadblock with Indian investments
05/01/17	SoftBank India Vision gets new $100
05/01/17	Nissan halts joint development of luxury cars with Daimler : Sources
05/01/17	Despite challenges Rajasthan continues to progress : Vasundhara Raje

Table 2. Attributes of raw data

Attribute	Value
Vocabulary size	48134
Average Length of Headline	12.295
Average Number of Characters in Headlines	74.6002
Number of Positive Instances	92382
Number of Negative Instances	108118

Table 3. Sample of news headline after pre-processing

Original text	Two Chinese companies hit roadblock with Indian investments
Lowercase	two chinese companies hit roadblock with indian investments
White Space Removal	Two chines ecompanies hit roadblock with Indian investments
Tokenization	['two', 'chinese', 'companies', 'hit', 'roadblock', 'with', 'indian', 'investments']
Stop Words Removal	['two', 'chinese', 'companies', 'hit', 'roadblock', 'indian', 'investments']
Punctuation Removal	['two', 'chinese', 'companies', 'hit', 'roadblock', 'indian', 'investments']
Lemmatization	['two', 'chinese', 'company', 'hit', 'roadblock', 'indian', 'investment']
Stemming	['two', 'chines', 'compani', 'hit', 'roadblock', 'indian', 'invest']

The study applied seven pre-processing techniques to clean and refine the text data while preserving essential information. Table 3 displays a sample of the text post-processing, illustrating how these techniques were employed to enhance data quality and prepare it for further analysis.

The study evaluates the performance of Word2Vec, Doc2Vec, TFIDF, Word2Vec-TFIDF and Doc2Vec-TFIDF feature extraction techniques using the valuation metrics- precision, recall, Fi-score and accuracy. Here, the features obtained from each feature extraction technique are fed into six different classifiers- RF, SVM, DT, GB, ET, and LR to thoroughly analyze and compare their performance in handling sentiment analysis tasks. The results of the sentiment analysis for each model are given in the following Tables. 4 - 8.

Table 4. Performance analysis of Word2Vec feature extraction technique.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.60	0.60	0.59	0.60
SVM	0.61	0.60	0.59	0.60
Decision Tree	0.54	0.54	0.54	0.54
Gradient Boosting	0.59	0.59	0.58	0.59
Extra Trees	0.60	0.61	0.59	0.61
Logistic Regression	0.60	0.60	0.59	0.60

Table 5. Performance analysis of TFIDF feature extraction technique.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.73	0.73	0.73	0.73
SVM	0.75	0.75	0.75	0.75
Decision Tree	0.66	0.65	0.66	0.66
Gradient Boosting	0.66	0.65	0.64	0.65
Extra Trees	0.74	0.74	0.74	0.74
Logistic Regression	0.73	0.73	0.73	0.73

Table 6. Performance analysis of Doc2Vec feature extraction technique.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.50	0.50	0.50	0.50
SVM	0.52	0.52	0.52	0.52
Decision Tree	0.50	0.49	0.50	0.50
Gradient Boosting	0.51	0.50	0.50	0.51
Extra Trees	0.49	0.49	0.49	0.49
Logistic Regression	0.51	0.50	0.51	0.50

Table 7. Performance analysis of hybrid Doc2Vec-TFIDF feature extraction technique

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.62	0.62	0.62	0.62
SVM	0.74	0.74	0.74	0.74
Decision Tree	0.65	0.65	0.65	0.65
Gradient Boosting	0.65	0.64	0.63	0.65
Extra Trees	0.73	0.73	0.73	0.73
Logistic Regression	0.72	0.72	0.72	0.72

Table 8. Performance analysis of hybrid Word2Vec-TFIDF feature extraction technique.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.75	0.75	0.75	0.75
SVM	0.82	0.82	0.82	0.82
Decision Tree	0.65	0.65	0.65	0.65
Gradient Boosting	0.66	0.66	0.66	0.66
Extra Trees	0.77	0.77	0.77	0.77
Logistic Regression	0.78	0.78	0.78	0.78

From Tables. 4 - 8, it is observed that the model using a hybrid feature extraction technique, Word2Vec-TFIDF, and SVM classifier outperforms all other models in terms of precision, recall, F1 score, and accuracy. This effectiveness stems from the synergy of Word2Vec and TFIDF, which together produce high-quality vectors capable of capturing both semantic relationships and important words in the data. Among the given classifiers, SVM performs with a higher accuracy for all the feature extraction techniques. This is in line with the studies [28], [29], [30], [31], where it is observed that the SVM classifier exhibits robust performance in sentiment analysis. For SVM, Word2Vec-TFIDF hybrid feature extraction displays an accuracy of 82 percent, followed by TFIDF, Doc2Vec-TFIDF, Word2Vec, and Doc2Vec feature extraction techniques. Interestingly, the feature extraction using Doc2Vec was found to be the least-performing technique, followed by Word2Vec and TFIDF. This could be due to the nature of data as news headlines typically contain fewer terms and Doc2Vec performs better when trained on larger and more diverse corpora. It can be concluded that the sentiment analysis for financial news headlines can be improved using a hybrid Word2Vec-TFIDF feature extraction technique and SVM classifier, where news headlines can be efficiently classified into positive and negative sentiments.

After obtaining sentiment scores for each news headline, a sentiment index can be constructed by aggregating the total number of positive and negative news headlines each day. Fig. 2 represents the distribution of sentiments over time. It is observed that the number of positive sentiments is higher than the negative sentiments for most days. Fig. 3 displays the word cloud for positive and negative sentiments. The word cloud is a visual representation of the frequency of terms in a given body of text. Given the dataset, the most frequent terms in both negative and positive sentiments are found to be related to the Indian market and stocks.

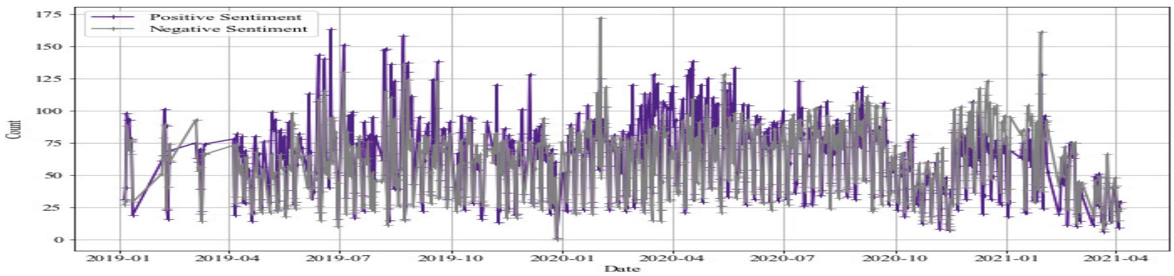

Fig. 2. Trend of positive and negative sentiments over time

Fig. 3. (a) word cloud for positive sentiments; (b) word cloud for negative sentiments

5. Conclusion

Since the past decade, sentiment analysis has been gaining wide attention in academia, especially in the field of finance. This technique captures the emotions and opinions in text data. Analyzing the sentiment of financial news helps detect emotions and control information asymmetry. Feature extraction is a crucial step in sentiment analysis, as it significantly impacts the performance of sentiment classification. Applying a hybrid feature extraction technique helps integrate the strengths of individual methods, thereby enhancing text representation. This study evaluated the performance of five different feature extraction techniques- Word2Vec, Doc2Vec, TFIDF, Word2Vec-TFIDF, and Doc2Vec-TFIDF and six different machine learning classifiers: RF, SVM, DT, GB, ET, and LR on financial news sentiments. The results show that a hybrid Word2Vec-TFIDF feature extraction technique combined with the SVM classifier outperforms all other models with an accuracy of 82 percent. For all feature extraction techniques, SVM was found to be the robust classifier with higher accuracy compared to other classifiers. Doc2Vec technique is observed to be the least performing feature extraction technique, yielding an accuracy of 50 percent when combined with SVM. This study help analyze the sentiments behind news data, which can be further utilized in various operations, including price forecasting. Future work can explore the performance of hybrid feature extraction techniques with deep learning and ensemble models. The proposed model can be further applied to other text data such as reviews, tweets, and social media comments.

References

[1] Ullah S, Md Mohsin Kabir, Talha Bin Sarwar, Safran M, Sultan Alfarhood, and Mridha MF. (2024) "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM." *Scientific reports* **14(1)**.

[2] Neha Punetha, Jain G. (2023) "Optimizing Sentiment Analysis: A Cognitive Approach with Negation Handling via Mathematical Modelling." *Cognitive Computation* **16(2):**624–40.

[3] Birjali M, Kasri M, Beni-Hssane A. (2021) "A comprehensive survey on sentiment analysis: Approaches, challenges and trends." *Knowledge-Based Systems* **226(1)**:107134.

[4] Alyoubi S, Kalkatawi M, Abukhodair F. (2023) "The Detection of Fake News in Arabic Tweets Using Deep Learning." *Applied Sciences* **13(14)**:8209.

[5] Mallik A, Kumar S. (2023) "Word2Vec and LSTM based deep learning technique for context-free fake news detection." *Multimedia Tools and Applications* **83(1)**:919–40.

[6] Dilan Lasantha, Sugandima Vidanagamachchi, Nallaperuma S. (2024) "CRIECNN: Ensemble convolutional neural network and advanced feature extraction methods for the precise forecasting of circRNA-RBP binding sites." *Computers in biology and medicine* **174**:108466–6.

[7] Mian Muhammad Danyal, Sarwar Shah Khan, Khan M, Ullah S, Muhammad Bilal Ghaffar, Khan W. (2024) "Sentiment analysis of movie reviews based on NB approaches using TF–IDF and count vectorizer." *Social network analysis and mining* **14(1)**.

[8] S. Joshua Johnson, M. Ramakrishna Murty, I. Navakanth. (2023) "A detailed review on word embedding techniques with emphasis on word2vec." *Multimedia Tools and Applications* **83(1)**: 37979–38007

[9] Nemes L, Kiss A. (2021) "Prediction of stock values changes using sentiment analysis of stock news headlines." *Journal of Information and Telecommunication* **5(3)**: 375–394.

[10]Ashtiani, Matin N., and Bijan Raahemi. (2023) "News-Based Intelligent Prediction of Financial Markets Using Text Mining and Machine Learning: A Systematic Literature Review." *Expert Systems with Applications* **217**:119509.

[11] Ren J, Dong H, Padmanabhan B, Nickerson JV. (2021) "How does social media sentiment impact mass media sentiment? A study of news in the financial markets." *Journal of the Association for Information Science and Technology* **72(9)**:1183-1197

[12] Adhikari S, Thapa S, Naseem U, Lu H, Gnana Bharathy, Prasad M. (2023) "Explainable hybrid word representations for sentiment analysis of financial news*." Neural Networks* **164**:115–23.

[13] Aakanksha Sharaff, Tushin Roy Chowdhury, Sakshi Bhandarkar.(2023) "LSTM based Sentiment Analysis of Financial News*". SN Computer Science* **4(5)**.

[14] Aluna RP, Yulita IN, Sudrajat R. (2021) "Electronic News sentiment analysis application to new normal policy during the COVID-19 pandemic using Fasttext and machine learning." In: *2021 International Conference on Artificial Intelligence and Big Data Analytics* p. 236–41.

[15] Ranathunga S, Liyanage IU. (2021) "Sentiment Analysis of Sinhala News Comments." *ACM Transactions on Asian and Low-Resource Language Information Processing* **20(4)**:1–23.

[16] Jang B, Kim I, Kim JW. (2019) "Word2vec convolutional neural networks for classification of news articles and tweets." *PLOS ONE*. **14(8):**0220976.

[17] Mishev K, Gjorgjevikj A, Stojanov R, Mishkovski I, Vodenska I, Chitkushev L,Trajanov D. (2019) "Performance evaluation of word and sentence embeddings for finance headlines sentiment analysis." *Communications in Computer and Information Science*, p.161–172.

[18] Wu S, Liu Y, Zou Z,Weng T-H. (2021) "S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis." *Connection Science* **34(1)**:44–62.

[19] Kadaparthi M. (2023) "A Deep Learning based Approach for Analyzing the Sentiments of Financial Text." In: *4th IEEE Global Conference for Advancement in Technology (GCAT)* p. 1–6.

[20] Chhabra Anusha, Arora Monika, Sharma Arpit, Singh Harsh, Verma Saurabh, Jain Rachna, Acharya Biswaranjan, Gerogiannis Vassilis C, Tzimos Dimitrios, Kanavos Andreas. (2024) "Classifying Hindi News Using Various Machine Learning and Deep Learning Techniques." *International Journal on Artificial Intelligence Tools* **33(2)**.

[21] Singh J, Singh G, Singh R. (2017) "Optimization of sentiment analysis using machine learning classifiers." *Human-centric Computing and Information Sciences.* 7(1).

[22] Shah K, Patel H, Sanghvi D, Shah M. (2020) "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification." *Augmented Human Research* **5(1).**

[23] Del-Pozo-Bueno D, Demie Kepaptsoglou, Peiró F, Sònia Estradé. (2023) "Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks." *Ultramicroscopy* **253**:113828–8.

[24] Khan Md. Hasib, Nurul Akter Towhid, Kazi Omar Faruk, Jubayer Al Mahmud, Mridha MF.(2023) "Strategies for enhancing the performance of news article classification in Bangla: Handling imbalance and interpretation." *Engineering Applications of Artificial Intelligence* **125**:106688–8.

[25] Malik A, Yash Tejas Javeri, Shah M, Ramchandra Mangrulkar. (2022) "Impact analysis of COVID-19 news headlines on global economy." *Cyber-Physical Systems*. Elsevier eBooks.189–206.

[26] Umer M, Sadiq S, karamti H, Abdulmajid Eshmawi A, Nappi M, Usman Sana M, Imran Ashra. (2022) "ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification*." Pattern Recognition Letters* **164**:224–31.

[27] Hassan SU, Ahamed J, Ahmad K. (2022) "Analytics of Machine Learning-based Algorithms for Text Classification." *Sustainable Operations and Computers* **3**:238-338

[28] Alqaryouti O, Siyam N, Abdel Monem A, Shaalan K. (2024) "Aspect-based sentiment analysis using smart government review data." *Applied Computing and Informatics* **20**:142-161.

[29] Abdel Moniem Helmy, Radwa Nassar, Nagy Ramdan. (2024) "Depression detection for twitter users using sentiment analysis in English and Arabic tweets.' *Artificial Intelligence in Medicine* **147**: 102716

[30] Dake DK, Gyimah E. (2023) "Using sentiment analysis to evaluate qualitative students' responses." *Education and Information Technologies* **28**: 4629–4647.

[31] Piryani R, Piryani B, Singh VK, Pinto D. (2020) "Sentiment analysis in Nepali: Exploring machine learning and lexicon-based approaches." *Journal of Intelligent & Fuzzy Systems* **6**;1–12.