Yugoslav Journal of Operations Research
34 (2024), Number 4, 765-783
DOI: https://doi.org/10.2298/YJOR240315034A

AN EFFICIENT CUCCONI BASED FEATURE EXTRACTION WITH RANDOM DECISION FOREST CLASSIFICATION FOR IMPROVED SENTIMENT ANALYSIS

Dr. K. ANURADHA[*]

Research Scholar, Department of CSE, Centurion University of Technology & Management, Odisha and Assistant Professor, Dr. L. B. College of Engineering for women, Visakhapatnam, India, anukeshav76@gmail.com, kanuradha@lbce.edu.in

Dr. Banitamani MALLIK

Professor, Department of Mathematics, Centurion University of Technology & Management, Odisha, India, banita.mallik@cutm.ac.in

Dr. M. Vamsi KRISHNA

Associate Professor, Department of CSE, Aditya Engineering College, Surampalem, India, vamsimangalam@gmail.com

Received: March 2024 / Accepted: July 2024

Abstract: Sentiment analysis is a form of opinion mining technique that identifies the polarity of extracted opinions. Nowadays, opinion mining has become an important research area in recent decades to identify the polarity of the statements. Various research works have been carried out on sentiment analysis. However, the existing sentimental analysis techniques, such as time and space complexity, still have considerable limitations. To deal with these issues, this paper proposed the Cucconi Feature Extracted Random Decision Forest Classification (CFDFC) Approach. The main objective of the CFDFC approach is to provide effective sentiment analysis with improved accuracy and reduced time complexity. The proposed CFDFC framework comprisespre-processing, feature extraction, and classification. The pre-processing step eliminates stop words and stem words from user reviews. After the pre-processing step, the feature extraction process is carried out to minimize the dimensionality and time consumption for opinion classification. Cucconi's projective feature extraction process is used in this work to reduce dimensionality. Finally, the classification process is formulated using a random decision forest classifier. The random decision forest classifier uses the ID3 DT (decision tree) as a weak learner to classify the review statements. The performance

[*] Corresponding author

evaluation of the proposed approach is carried out using performance metrics such as accuracy, error rates, recall values, and time and space complexities concerning the number of review statements gathered from the dataset. The results show that the proposed CFDFC model achieves remarkable accuracy, recall, and minimal time complexity compared to existing methods.

Keywords: Sentiment analysis, review statements, weak learner, random decision forest classifier, ID3 decision tree, Cucconi projective feature extraction.

MSC: 68T50, 68T10, 68T40, 68P30.

1. INTRODUCTION

With the large development of opinionated texts on the Web, opinion mining has emerged as the hopeful one for online public opinion evaluation. The opinion word mining from online reviews is essential to identify the opinion correlations among words. A new deep learning framework was put forward in [1] for opinion mining with a temporal feature extraction layer. Bi-SRU (bidirectional simple recurrent unit) networks extract features at word and grammar levels. However, with the proposed model, the accuracy level was not improved.

Attention-based Bi-LSTM (Bidirectional Long Short-Term Memory) Networks were proposed in [2] for mining opinion aspects. The aspects from words were transformed into vectors in pre-training and subsequently used for classifying sentimentally identical and dissimilar neighbors. However, the time consumption for opinion mining was not minimized. An aspect-based opinion mining was employed in [3] to mine the latent topics. The opinion mining model was employed for topic modelling to identify the importance depending on perplexity and coherence. But, with the proposed framework, there was no reduction in the computational cost.

The model is based on CNN (Convolution Neural Networks) and LSTM networks developed in [4] to find Google Cloud’s word polarities. The solution was attained to classify opinions into positive and negative classes. However, the recall was not increased by using CNN and LSTM. A machine learning approach was employed in [5] to examine the tweets to increase the customer experience. The features were extracted from the tweets by word embedding combined with the Glove dictionary mechanism and n-gram scheme to map tweets into positive and negative classes. However, there was no improvement in recall by the designed technique.

Ideas Starbucks was employed in [6] to decide the crowdsourcing participant perception of the company. Ideas Starbucks was employed to map the users into a community format to find the ideas. But, the precision was not improved by Ideas Starbucks. An automated solution was obtained in [7] to collect the recommendations from students' feedback opinions by applying text mining and data visualization. C5.0 DT functioned better in terms of extracting recommendations from qualitative feedback. However, there was no reduction in space complexity.

The Bayesian-based model called Opinion Mine was presented in [8] to mine the opinions from Twitter Data. An imported Tweet went through processing for the untrained rule set with random variables. However, the designed framework was not capable of reducing the space complexity. A new framework was demonstrated in [9] based on product quality and customer satisfaction using diverse databases where the framework found associations between technologies and satisfied customers using structural equations and mining opinions. A partially-supervised alignment model was introduced in [10] to determine opinion relationships throughout alignments. The co-ranks of candidates were computed using a graph-based technique. Nevertheless, utilizing the intended model did not result in a reduction of complexity.

The problems found in the above literature include reduced accuracies, increased time complexities, rise in space complexities, higher computational expenses, and enhanced computational complexities. This work’s CFDFC Technique addresses the issues mentioned above. The proposed technique’s important goals and novelty are listed below:

- The key objective of the CFDFC Technique is to carry out an effective sentiment analysis yielding improved accuracy and reduced time complexity. The number of user review statements is collected from the input dataset.
- CFDFC Technique comprises pre-processes, feature extractions, and classifications. Pre-processes eliminate stop words and stem words from user reviews.
- CFDFC Technique uses cucconi projective feature extraction to extract opinion words using the review statements. Cucconi projective feature extraction is typically a dimensionality reduction technique for identifying projections in multidimensional data.
- The classification process is carried out using a random decision forest classifier. The random decision forest classifier employs theID3 DT as a weak learner to classify review statements with the extracted features. Using the voting scheme, random decision forest classification provides final classification results with higher accuracy.

The remaining parts of this article are structured into four sections, as given. In Section 2, the related works of sentiment analysis are outlined. Section 3 introduces the proposed CFDFC methodology for analyzing sentiments. Section 4 provides the various results analysis of the proposed techniques with the help of the table and graphical representation. In section 5, the conclusion of the paper is given.

2. LITERATURE REVIEW

A new topic model termed CAMEL was introduced in [11] for asymmetric sets' complementing aspect-based mining of opinions. CAMEL discovered material that was complementary to particular elements found in the collections. Instance selection for creating the reduced rule base was part of introducing the fuzzy system [12]. Using a multi-objective genetic algorithm increased accuracy while reducing the number of rules. The system's complexity level did not decrease throughout the design. The rank after clustering (RaC) technique was developed in [13] to identify influential individuals in social networks. K-means algorithm clustered social network participants to find potential opinions. Unfortunately, employing the created method did not result in a reduction in the cost of computing.

The hybrid fusion technique was employed in [14], combining features, scores and decisions of diverse modalities. The emoji features increased the classification results with text at the score level. An end-to-end system recognized the opinion list in [15] from many Twitter hashtags using a classifier trained with task-specific features. The designed system extracted the appropriate answers from relevant tweets. A consensus opinion model was designed in [16] in social communities depending on influential users and aggregation methods. The user opinions were grouped and mined to form the consensus model. Then, the opinions were propagated to reach an agreement to maintain consistency.

A novel technique was presented in [17], which extracted product characteristics and opinions from product reviews belonging to specific domains. An ontology-based semantic measure was introduced in [18] for intelligent feature selection/reduction. The proposed measure was a combination of semantic similarity and distance.

Selection issues were addressed in [19] using MPINS (Minimum-sized Positive Influential Node Set), where minimum sets of influential nodes were identified. The positive influence was exerted on each node in the network through the chosen nodes without a threshold. A new measure termed Mile stones Rank was introduced in [20] to identify the opinion leaders with influential users on specific topics.

A supervised technique, Naïve Bayes classification approach, was introduced in [21] to study the aspect ratings. The weights of aspects were determined by balancing Word frequencies and consistencies amongst reviews. Deep learning methods extracted features for opinion mining in [22]. Aspect extraction was the subtask of sentiment analysis for finding the opinion targets in opinionated text.

Saleh Naif Almuayqil et al. [23] suggested Random Minority Oversampling to improve sentiment analysis for User Tweet Review Classification. By combining the resampling of minority classes with several ordered pre-processing stages, this research suggests a paradigm for better sentiment analysis. Due to its primary emphasis on feature selection and combination, the technique's performance could differ among datasets. Many ML systems sort tweets as positive, bad, or neutral. A solution to the problem of class imbalance and performance improvement may be found by random minority oversampling.

Md. Mashiur Rahaman Mamun et al. [24] proposed the Ensemble Technique for the Classification of Textual Sentiment. Automatically classifying expressions and texts as positive, negative, or neutral is textual sentiment analysis. There has been little advancement in the creation of language processing tools for Bengali, even though it is the second-most-known Indian language and the seventh-most popular language in the world. This research suggests an ensemble-based method to divide Bengali textual emotion into positive and negative groups.

Rupali S. Patil and Satish R. Kolhe [25] recommended the Term Frequency vs. Inverse Document Frequency (TF-IDF) for sentiment analysis of Marathi tweets. The author assesses the Sentiment Analysis model's performance using the usual metrics like f1-score, accuracy, precision, and recall. Based on the trial findings, Multinomial Naïve Bayes is the best classifier for predicting the 2019 Indian State Assembly Election, with a maximum accuracy of 87.29%. The present state-of-the-art sentiment analysis of Indian text uses the suggested model, the top-ranked Naïve Bayes classifier.

AminaSamih et al. [26] introduced the improved word vector for sentiments analysis (IWVS)using the XGboost classifier. Sentiment2Vec, the suggested approach, averaged the word embeddings to create sentiment vectors. The author looked into the Polarized lexicon to categorize positive and negative emotions. The sentiment text analyzed was mapped to a feature space produced by the sentiment vectors. The selected classifier (XGboost) was fed those features. Using our technique with various sentiment datasets and machine learning models, the author could compare the F1-score of sentiment classification.

Talha Ahmed Khan et al. [27] investigated the Support Vector Machine and Random Forest for Sentiment Analysis. Techniques for pre-processing data, extracting features, training and evaluating models, and difficulties in sentiment analysis are all covered in the article. This study adds to our knowledge of sentiment analysis and sheds light on how well machine-learning methods work in this field. The results served as the basis for a classification accuracy evaluation of two machine learning methods, Random Forest and SVM. With an accuracy of 0.80394, SVM marginally beat the Random Forest method, which reached 0.78564. When it comes to the classification problem at hand, Random Forest and SVM have both shown their value by producing adequate accuracy.

3. PROPOSED METHODOLOGY

Sentiment analysis is defined as the use of NLP (natural language processing) and text analyses to recognize and perform extractions of subjective information. The primary objective of sentiment analysis is to process user reviews of specific products to generate and aggregate user opinions. However, there was no improvement in theaccuracy level, and time complexity was also not minimized using the recent opinion mining techniques. To handle these issues, the CFDFC approach is designed. The primary goal of the CFDFC approach is to increase the performance of review statement polarity identification accuracy in reduced time. The block diagram of the CFDFC approach is described in Figure 1.

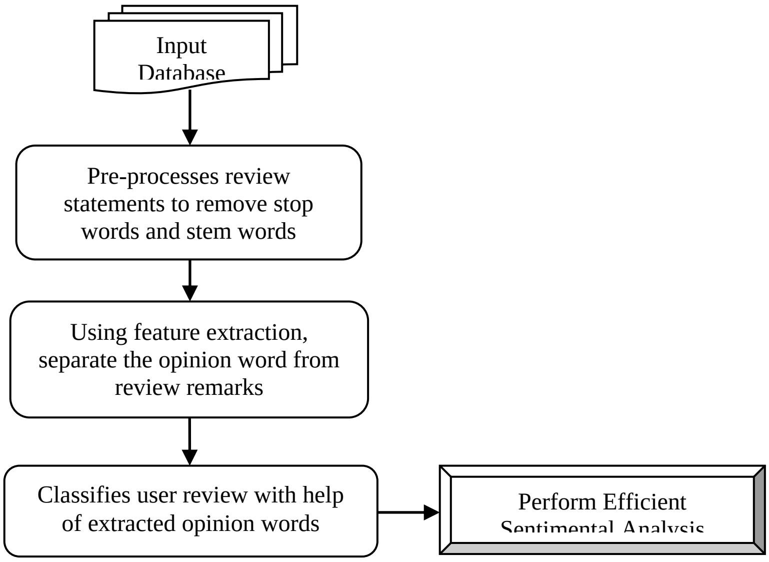
Input
Database
Pre-processes review statements to remove stop words and stem words
Using feature extraction, separate the opinion word from review remarks
Classifies user review with help of extracted opinion words
Perform Efficient
Sentimental Analysis

Figure 1: Architectural Diagram of CFERDFC Approach

The suggested CFDFC approach's architectural design, which identifies polarities from review statements, is shown in Figure 1. User reviews are initially taken from a database, and CFDFC’s Cucconi Feature Extractions subsequently identify opinions from user reviews. They also minimize review statement sizes as a part of dimensionality reduction. The reviews are then classified by the CFDFC approach's Random Decision Forests, which have greater accuracy and require less time. This is also helpful in improving sentiment analysis efficacies. The following parts briefly describe pre-processes, feature extractions, and classifications.

3.1. Data Pre-processes

Data pre-processing is crucial in the CFDFC Method for eliminating stop words from statements and reducing file sizes while maintaining improved accuracy.

3.2. Cucconi Projective Feature Extraction

The proposed CFDFC Technique performs an efficient feature extraction using the Cucconi Projective Feature Extraction process. Cucconi Projective Feature Extraction is a dimensionality reduction method used to identify projections in multidimensional data. The feature projective extraction maps the relevant features (i.e., opinion words) into the lower-dimensional space. It is formulated as,

$$fun(x) : ft_{rele} \rightarrow Subset \tag{1}$$

From (1), '$fun(x)$' symbolizes the feature projective selection to project the essential features related to the subset. The significant features are identified using the Cucconi Projective Feature Extraction Process. It is devised as,

$$\beta_{CP} = \frac{{ft_a}^2 + {ft_b}^2 - (2 * \sigma * ft_a * ft_a)}{2(1-\sigma^2)} \tag{2}$$

From (2), Cucconi Projective Index 'β_{CP}' is computed depending on the sum of the product of paired score of two features 'ft_a' and '${ft_a}^2$'. The squared score of the feature 'ft_b' is '${ft_b}^2$' and the squared score of feature 'ft_a' is '${ft_a}^2$'. Cucconi Projective Index attains the output value ranging between '0' and '1'. It is formulated as,

$$\beta_C = \{1, \qquad Opinion\ words 0, \quad Non - opinion\ words \tag{3}$$

From (3), the index with '+1' symbolizes the opinion words. The non-opinion words are indicated by an index value of "0." The usage of opinion words allows effective data categorizations. While classifying data, the CFDFC technique considers opinion terms. This work cucconi projective feature extraction algorithm can be stated as:

Algorithm 1: Cucconi Projective Feature Extraction Algorithm

```
Input: Datasets
Output: identified opinions
Begin
Step1:  Gather pre-processed user reviews with words
Step2:  for every user review
Step 3:         Measure Cucconi Index '$\beta_C$'
Step 4:         If ($\beta_C = +1$)then
Step 5:              Opinion words
Step 6:              Select opinion words
Step 7:         else
Step 8:                Non-opinion words
Step 9:              Remove non-opinion words
Step 10:        End if
Step 11:      End for
Step 12: End
```

The Cucconi projective feature extraction algorithm is explained to achieve efficient data classification. The words are projected as non-opinion or opinion terms using the Cucconi Projective Index. While categorizing the data, the CFDFC technique selects only opinions while discarding other terms. Using opinion terms helps keep opinion mining as simple as possible.

3.3. Random Decision Forest Classification

After selecting the relevant features, the classification task is performed using a random decision forest classifier. A Random Decision Forest Classifier is an ensemble classifier employed to enhance classification accuracy by constructing several weak learners. The weak learner is the base classifier that provides less accurate results. Consequently, the proposed CFDFC Technique employs the ensemble classifier for efficient sentiment analysis. The structure of the random decision forest classifier is illustrated in Figure 2.

Figure 2 describes the structure of the random decision classifier for categorizing the opinions with higher accuracy. Let us consider that training set '$\{a_i, b_i\}$' where 'a_i'represents bootstrap samples (i.e., extracted opinion words) and 'b_i'denotes the classification results. The bootstrap aggregating classifier constructs an empty set of weak learners 'wl'. The random decision forest classifier uses ID3 DT to classify the extracted opinion words. ID3 classifier algorithm constructs the DT with extracted opinion words as root nodes. The ID3 algorithm determines entropy and IG values with extracted opinion words. ID3 classifier computes the entropy with the extracted opinion words. The entropy value calculates the uncertainty amount for every opinion word. The opinion word with the smallest entropy value is employed to partition the set. IG is the difference between entropy from before and after the set is classified for every opinion word. Consequently, the information gain of opinion words'$IG(OW)$' is computed as,

$$wl_i = IG(OW) = En - \sum_{t \in T} \quad p(t)H(t) \tag{4}$$

From (4), 'En' representsa set of entropy.'$p(t)$' symbolizes the subset generated from the splitting set'$H(t)$'denotes the entropy of subset 't'. ID3 classifier determines the IG for every opinion word. By using entropy and IG, the ID3 DT is created. A DT with root and leaf nodes are generated where nodes imply opinions. Nodes have decisions that denote subsets of opinions. Leaf nodes are called terminal nodes of trees with class labels. Opinions with higher IG are selected for decisions. The path from root nodes to leaf nodes constructs classification rules from the DT. Every leaf node represents the classification result of opinion words. However, the weak learner output has training errors during the classification task. Consequently, the weak learner results are combined to achieve a strong classification output by reducing the error. The strong classification results obtained as,

$$Ot_i = \sum_{i=1}^{m} \quad wl_i(x) \tag{5}$$

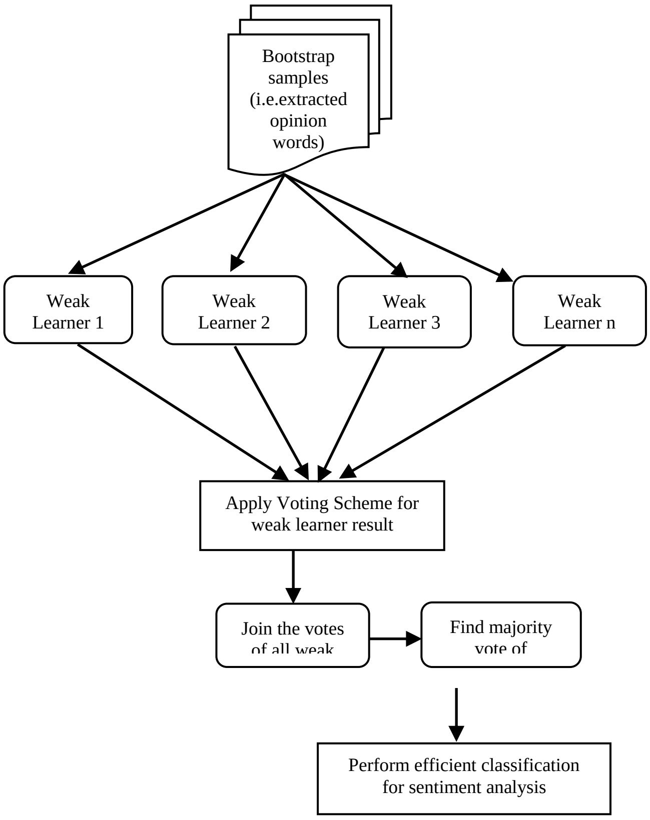
Bootstrap samples (i.e.extracted opinion words)
Weak Learner 1
Weak Learner 2
Weak Learner 3
Weak Learner n
Apply Voting Scheme for weak learner result
Join the votes of all weak
Find majority vote of
Perform efficient classification for sentiment analysis

Figure 2: Structural Diagram of Random Decision Forest Classifier

From (5), 'Ot_i'denotes the ensemble classification results. '$wl_i(x)$ denotes an output of the DT classifier. As a result, the random decision forest classifier increases the classification performance process with a minimal false positive rate. The algorithmic procedure of the CFERDFC Technique results areillustrated as,

Algorithm 2: trimmed Bootstrap Aggregating Data Classification

```
**Input:** Number of extracted opinion words '$a_i$'
**Output:** Improve classification accuracy and reduce time consumption
**Begin**
**Step1:**      **For** every extracted opinion word
**Step2:**      Construct '$m$' DTs with training data
**Step3:**      Determine information gain of opinion words
**Step4:**      Combine a set of weak learners $\sum_{i=1}^{m} wl_i(x)$
**Step5:**      Attain strong classification results.
**Step6:**      **End for**
**End**
```

Algorithm 2 explains the structure diagram of the random forest classifier. The ensemble classifier builds a set of weak classifiers with training data. The weak learner measures information gain and entropy to perform the classification. Depending on information gain, the weak learner classifies the opinion words into three classes. The weak classification results are merged to make the strong one. The ensemble classification results improve the accuracy of performance and reduce time consumption.

3.4. Preparing the Dataset

Using Java, an experimental investigation of the CFDFC Method is conducted. Tweets were gathered from the Sentiment140 dataset of Kaggle (https://www.kaggle.com/kazanova/sentiment140) [28],which had 1.6 million tweets. Using Twitter API, 1,600,000 tweets were acquired for the sentiment140 dataset. To determine the polarity of attitudes, tweets were considered as 0 – negative and 4 - positive. The dataset includes six fields: target, IDs, date, flag, user, and tweets. Tweet polarities were 0-negative, 2- neutral, and 4- positive. This dataset was unique because the training data was automatically created instead of having humans manually annotate tweets. Table 1 shows the experimental setup.

Table 1: Experimental Setup

Resources Utilized	Specifications
NVIDIA T4 x2 GPU	2560 Cuda cores, 16 GB
Intel(R)Xeon(R) CPU	x86 with a clock frequency of 2 GHz, 4 vCPU cores, 18GB
RAM	16 GB DDR4
Google TPU	8 TPU v3 cores. 128 GB
Python	version 3.10.12
Total Tweets	1,600,000 tweets
Training Set	80% of the data (1,280,000 tweets)
Testing Set	20% of the data (320,000 tweets)

4. EXPERIMENTAL ASSESSMENT

With the aid of Java, an experimental comparison is made between the proposed CFDFC Method and contemporary methods like deep learning framework [1] and attention-based position-aware Bi-LSTM [2]. Many client reviews, between 100 and 1000, are obtained for the studies. Using the data collected, the performance of the CFDFC technique was evaluated using five parameters.

- Accuracy
- Precision
- Recall
- Time complexity
- Space Complexity

A classifier's accuracy may be fine-tuned by playing about with the number of trees (n estimators). Model resilience and accuracy are improved with increased trees, yet this comes at the expense of increased computational cost and training time. By experimenting with different values for the number of trees 50, 100, 200, and 500, this study observes how the accuracy, precision, recall, and F1-score change. The model's performance may be affected by changing the value of max_features, which is the size of the feature subset. Overfitting occurs when feature subsets are too large, whereas underfitting occurs when feature subsets are too small. By experimenting with feature subset sizes such as 1000, 3000, 5000, and 10,000 features, this study may learn more about the ideal value for feature richness and model complexity.

5. RESULTS

5.1. Impact on Accuracy

Accuracy is the ratio between customer review counts, whose classifications into various classes are correct, to the overall number of customer reviews collected 'N' from the dataset. Consequently, the accuracy is devised as given below,

$$Acc = \left[\frac{T_{po}+T_{ne}}{N}\right] * 100 \tag{6}$$

From (6), 'Acc' indicates accuracy, 'T_{po}' symbolizes the true positive.'T_{ne}' denotes the number of true negatives. Accuracies are measured as percentages.

Table 2: Tabulation of Accuracy

Customer Reviews Counts	Accuracy (%)		
	New deep learning framework	Attention-based position-aware Bi-LSTM network	New deep learning framework
100	84	87	92
200	88	93	94
300	90	90	95
400	89	91	96
500	88	91	96
600	89	91	96
700	90	90	96
800	89	90	96
900	87	91	96
1000	87	90	95

Table 2 shows the accuracy performance results for customer reviews with a count between 100 and 1000. The suggested CFDFC approach, an existing new deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM Network [2] are three strategies whose accuracy is presented in the table. Compared to conventional techniques, the suggested CFDFC Methodology achieves outcomes with a better degree of precision. If there are 600 customer reviews, the accuracy attained by the CFDFC technique is 96%, while the accuracy was determined to be 89% and 91% by applying the current Attention-based position-aware Bi-LSTM and existing deep learning framework [1]. Subsequently, ten different performance results are attained for every method. Figure 3 illustrates the graphical representation of accuracy.

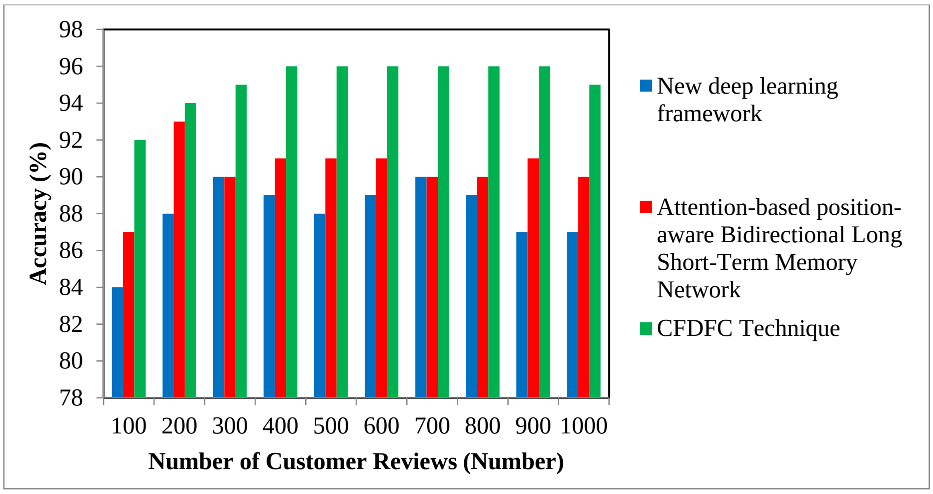

Figure 3: Measurement of Accuracy

The accuracy of the proposed CFDFC Method, the current new deep learning framework [1], and the existing Attention-based position-aware Bi-LSTM [2] are compared graphically in Figure 3. The proposed CFDFC technique's accuracy results are shown by the green colour bar, whereas the accuracy results for the current new deep learning framework [1] and the existing Attention-based position-aware Bi-LSTM Network [2] are represented by the blue and red colour bars, respectively. The graphical results are discussed that the accuracy is relatively higher when the CFDFC approach is used compared to the recent classification techniques. The improvement is credited to using the Cucconi Feature Extraction and Random Decision Forest Classification processes to extract the opinion words from the user reviews. This is useful in improving the accuracy performance during sentiment analysis. The mean of ten results shows that the accuracy of the CFDFC approach is improved by 8% compared to the new deep learning framework [1] and 5% compared to the Attention-based position-aware Bi-LSTM Network [2].

5.2. Impact on Error Rate

Error rates give proportions of customer review counts with incorrect classifications to the total number of customer reviews. Error rates can be formulated as given below:

$$Err_{Rate} = \left[\frac{F_{po}+F_{ne}}{N}\right] * 100 \tag{7}$$

From (7), 'F_{po}' denotes the false positive. 'F_{ne}' represent the false negative. Error rates are computed as percentages.

The performance results of the error rate w.r.t customer review counts taken into account in the range of 100 and 1000 are shown in Table 3. The table shows three approaches' error rates that are reviewed, including the suggested CFDFC Method, the new deep learning framework [1], and the current Attention-based position aware of Bi-LSTM [2]. The findings showed that, compared to conventional methodologies, the suggested CFDFC methodology gives outcomes with a lower error rate. If there are 200 customer reviews, the error rate achieved by the CFDFC method is 6%, whereas the error rate discovered using the new existing deep learning framework [1] and the pre-existing Attention-based position-aware Bi-LSTM Network [2] is 13% and 8%. The performance results for each method's error rate are then obtained in ten distinct ways. The graphical depiction of error rate is shown in Figure 4.

Table 3: Tabulation of Error Rate

Customer Reviews Counts	Error Rate (%)		
	New deep learning framework	Attention-based position-aware Bi-LSTM Network	New deep learning framework
100	16	13	8
200	13	8	6
300	10	10	5
400	11	9	4
500	12	9	4
600	11	9	4
700	10	10	4
800	11	10	4
900	13	9	4
1000	13	11	5

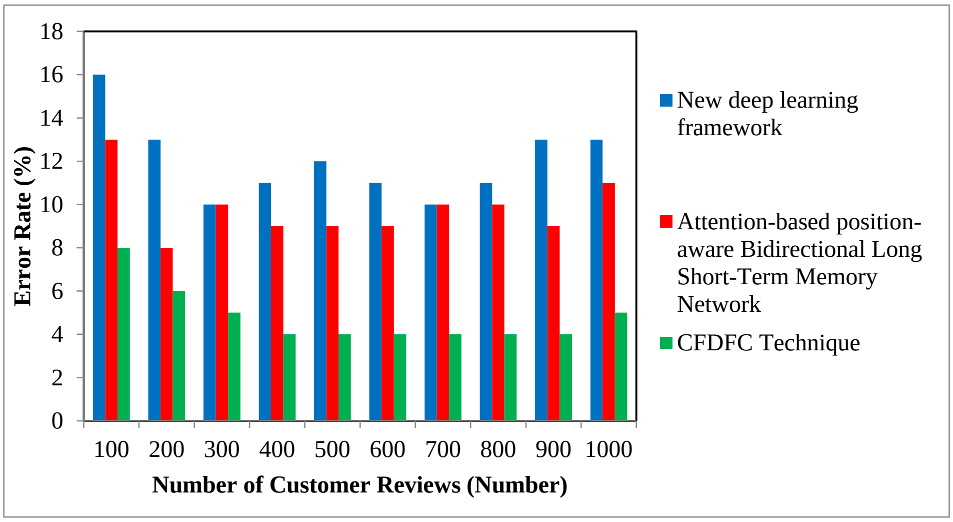

Figure 4: Measurement of Error Rate

Figure 4 compares the error rate representation using the proposed CFDFC Technique and the existing new deep learning framework [1] and Attention-based position-aware Bi-LSTM [2]. The green bar signifies the error rate results of the proposed CFDFC approach, whereas the blue and red bars denote the error rate results of the existing new deep learning framework [1] and existing Attention-based position-aware Bi-LSTM [2].

The graphical results show that the error rate is relatively lower when applying the CFDFC approach than when applying the available classification techniques. This is owing to usingthe Cucconi Feature Extraction process and Random Decision Forest Classifications to extract and classify opinions from user reviews. Meanwhile, this is useful to minimize the error rate during sentiment analysis. It is inferred from the average of ten results that the error rate of the CFDFC Technique is decreased by 60% compared to the new deep learning framework [1] and 51% compared to Attention-based position-aware Bi-LSTM [2].

5.3. Impact on Precision, Recall and F1-Score Ratio

The precision ratio has been computed using the proportion of true positive predictions relative to the total positive predictions (true positives + false positives).

$$Precision = \left[\frac{T_{po}}{T_{po}+F_{po}}\right] * 100 \quad (8)$$

It is measured by customer review counts, which are rightly classified and divided by the number of positive and negative samples. The recall is formulated as,

$$Rec = \left[\frac{T_{po}}{T_{po}+F_{ne}}\right] * 100 \quad (9)$$

From Equations (8) and (9),'Rec' denotes recall.'T_{po}' symbolizes the true positive.'F_{ne}' symbolizes the false negative and F_{po} represents the false positives. Recall values are measured as percentages.

The harmonic mean of precision and recall, which provides a balance between the two, is called F1-score.

$$F1 - score\ ratio = 2 * \frac{Precision*Recall}{Precision+Recall} * 100 \quad (10)$$

Table 4: Tabulation of Recall

Customer Reviews Counts	Recall (%)		
	New deep learning framework	Attention-based position-aware Bi-LSTM Network	New deep learning framework
100	91	94	97
200	92	96	97
300	94	95	98
400	93	96	98
500	94	95	98
600	93	95	98
700	94	95	98
800	93	95	98
900	92	95	98
1000	93	94	97

The performance results of a recall are explained in Table 4 concerning the range of customer reviews analyzed (100–1000). Three methods proposed by the CFDFC Method, an existing deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM [2]are detailed in the table below for their recollection. The findings showed that the suggested CFDFC Method achieves better recall outcomes than the conventional techniques. If there are 400 customer reviews, the recall was 93% and 96%, respectively, using the existing deep learning framework [1] and Attention-based position aware of the Bi-LSTM Network [2]. The recall was 98%, utilizing the CFDFC technique. Ten distinct recall performance results are then obtained for each strategy. The graphical depiction of recall is shown in Figure 5.

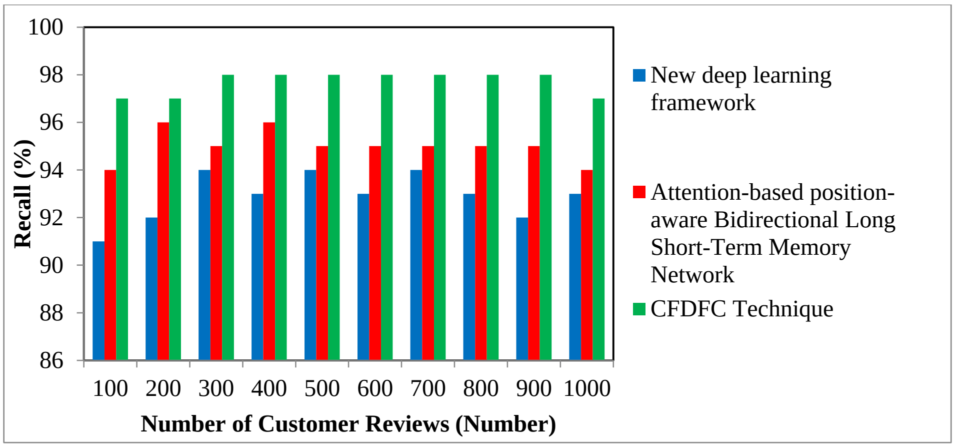

Figure 5: Measurement of Recall

The proposed CFDFC Method, an existing new deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM Network [2] are compared graphically in Figure 5 to show recall. The blue and red bars represent the recall results of the current new deep learning framework [1] and the current Attention-based position-aware Bi-LSTM Network [2, respectively], while the green bar represents the recall results of the proposed CFDFC Method. When the CFDFC strategy is employed in contrast to other accessible classification strategies, the recall is often greater, according to the provided graphical findings. This is because user reviews may be classified, and opinion words can be extracted using the Cucconi Feature Extraction and Random Decision Forest Classification processes. This aids in improving the call during sentiment analysis in the interim. According to the average of ten outcomes, the recall of the CFDFC technique is 5% higher than the new deep learning framework [1] and 3% higher than the Attention-based position-aware Bi-LSTM Network [2].

5.4. Impact on Time Complexity

Time complexity is given by the time taken to perform the opinion mining through classification. The time complexity is measured as,

$$TC = N * Time\ consumed\ for\ classifying\ one\ customer\ review \quad (11)$$

From equation (11), 'TC' symbolizes the timecomplexity.'N' symbolizes customer review counts. Time complexity is measured in units of milliseconds (ms).

Table 5: Tabulation of Time Complexity

Customer Reviews Counts	Time Complexity (ms)		
	Newdeep learning framework	Attention-based position-aware Bi-LSTM Network	New deep learning framework
100	35	28	15
200	37	30	17
300	38	32	19
400	40	35	22
500	42	37	25
600	45	40	27
700	47	42	30
800	50	44	32
900	52	45	35
1000	54	48	38

Table 5 summarizes the performance outcomes of temporal complexity in terms of the range of 100 to 1000 customer reviews that were taken into account. Three approaches, notably the suggested CFDFC Method, an existing deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM [2], are compared regarding their temporal complexity. The outcomes showed that the suggested CFDFC Method achieves results with reduced time complexity compared to the conventional techniques. Assuming there are 700 customer reviews. The time complexity acquired by the CFDFC technique is 30 ms, whereas the time complexity discovered using the current deep learning framework [1] and the existing Attention-based position-aware Bi-LSTM Network [2] is 47 ms and 42 ms. Then, ten distinct time complexity performance results are obtained for each approach. The graphical depiction of temporal complexity is shown in Figure 6.

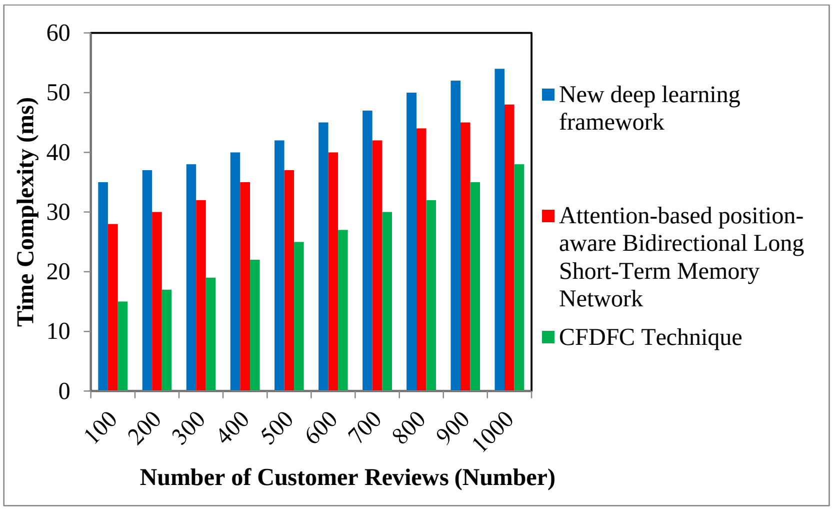

Figure 6: Measurement of Error Rate

Using the proposed CFDFC Method, a new deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM [2], the time complexity is compared graphically in Figure 6. The blue and red bars indicate the time complexity results of the existing new deep learning framework [1] and the existing Attention-based position-aware Bi-LSTM [2]. The green colour bar represents the time complexity results of the proposed CFDFC approach. The graphical findings show that using the CFDFC approach compared to more contemporary classification algorithms has considerably lower time complexity. This is because user reviews are classified, and opinion words are extracted using the Cucconi Feature Extraction and Random Decision Forest Classification procedures, respectively. Again, this helps reduce sentiment analysis's temporal complexity. The ten outcomes' average shows that the CFDFC approach's time complexity is 42% lower than that of the new deep learning framework [1] and 33% lower than that of the Attention-based position-aware Bi-LSTM [2].

5.5. Impact on Space Complexity

Space complexity is the space consumed to perform opinion mining through classification. The time complexity is measured as,

$$SC = N * Space\ consumed\ for\ classifying\ one\ customer\ review \tag{12}$$

From (12), 'SC' symbolizes the space complexity.'N' symbolizes customer review counts. Space complexity is measured in units of Megabytes (MB).

Table 6: Tabulation of Space Complexity

Customer Reviews Counts	Space Complexity(MB)		
	New deep learning framework	Attention-based position-aware Bi-LSTM Network	CFDFC Technique
100	39	32	22
200	42	34	24
300	44	37	27
400	47	40	30
500	49	42	32
600	52	45	35
700	55	48	38
800	58	50	40
900	60	53	42
1000	63	56	45

Table 6 demonstrates the performance outcomes of space complexity concerning the range of 100 to 1000 customer reviews that were considered. Three approaches, the suggested CFDFC Method, the existing deep learning framework [1], and the current Attention-based position-aware Bi-LSTM [2], are compared for their spatial complexity. The findings showed that the suggested CFDFC Method achieves outcomes with reduced space complexity compared to the existing approaches. Using the current deep learning framework [1] and Attention-based position-aware Bi-LSTM Network [2], and assuming that there are 800 customer reviews, the space complexity produced by the CFDFC technique is 30MB, the time complexity is found to be 47MB, and the time complexity is found to be 42MB. As a consequence, each technique achieves ten distinct space complexity performance results. The graphical depiction of space complexity is shown in Figure 7.

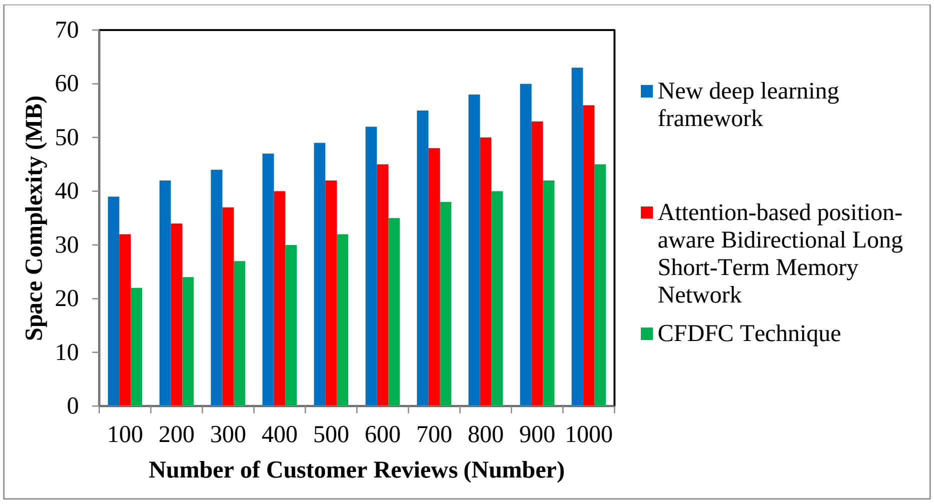

Figure 7: Measurement of Space Complexity

Using the suggested CFDFC Method, an existing new deep learning framework [1], and an existing Attention-based position-aware Bi-LSTM [2], Figure 7 demonstrates the comparison graphical depiction of space complexity. The proposed CFDFC approach's space complexity results are represented by the green colour bar, while the space complexity results of the existing new deep learning framework [1] and the existing Attention-based position-aware Bi-LSTM Network [2] are represented by the blue and red colour bars, respectively. The graphical findings show that the space complexity is substantially less difficult when the CFDFC approach is employed in contrast to the present classification techniques. This is because user reviews were classified, and opinion words were extracted using the Cucconi Feature Extraction and Random Decision Forest Classification procedures, respectively. In turn, this lessens the complexity of the space used for sentiment analysis. The arithmetic mean of 10 findings shows that the CFDFC technique has a temporal complexity of 35% lower than the new deep learning framework [1] and 24% lower than the Attention-based position-aware Bi-LSTM Network [2].

6. CONCLUSION

The CFDFC technique used in this study performs efficient sentiment analysis with improved accuracy and less time complexity. Pre-processing is done using the CFDFC Method to remove the stop and stem words from the review statements. To minimize the dimensionality of multidimensional data, the Cucconi projective feature extraction assists in extracting the opinion words from the review statements. With the help of the retrieved characteristics, the review statements are classified using an ID3 DT. The random decision forest classification uses a voting mechanism to increase the accuracy of the final classification findings. Once more, this helps to enhance sentiment analysis performance. This study tested the CFDFC Method using a dataset and compared the outcomes with two alternative baseline clustering techniques. Depending on the quantity of review statements gathered, a thorough experimental evaluation usesvarious criteria, including accuracy, error rate, recall, time complexity, and space complexity. The statistical results show that the suggested CFDFC Method outperforms contemporary approaches regarding accuracy, recall, and time complexity.

Conflict of interest statement: No conflicts of interest have been revealed by the author.

Funding. This research received no external funding.

REFERENCES

[1] T. Li, H. Xu, Z. Liu, Z. Dong, Q. Liu, J. Li, S. Fan, and X. Sun, “A spatiotemporal multi-feature extraction framework for opinion mining”, *Neurocomputing,* vol. 490, pp.337-346. 2022.

[2] A.F. Pathan, and C. Prakash, “Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory”, *Journal of King Saud University-Computer and Information Sciences,* vol. 34, no. 10, pp.8716-8726, 2022.

[3] A.F. Pathan, and C. Prakash, “Unsupervised aspect extraction algorithm for opinion mining using topic modeling”, *Global Transitions Proceedings*, vol. 2, no. 2, pp.492-499, 2021.

[4] K. Gangadharan, G.R.N. Kumari, D. Dhanasekaran, and K. Malathi, “Detection and classification of various pest attacks and infection on plants using RBPN with GA based PSO algorithm”, *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),* vol. 20, no. 3, pp.1278-1288, 2020.

[5] S. Kumar, and M. Zymbler, “A machine learning approach to analyze customer satisfaction from airline tweets”, *Journal of Big Data*, vol. 6, no. 1, pp.1-16, 2019.

[6] H. Alzahrani, P. Duverger, and N.P. Nguyen, “Contextual Polarity and Influence Mining in Online Social Networks”, In 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (pp. 1054-1061), IEEE, 2017.

[7] S. Gottipati, V. Shankararaman, and J.R. Lin, “Text analytics approach to extract course improvement suggestions from students’ feedback”, *Research and Practice in Technology Enhanced Learning,* vol. 13, pp.1-19, 2018.

[8] S. Zervoudakis, E. Marakakis, H. Kondylakis, and S. Goumas, “OpinionMine: A Bayesian-based framework for opinion mining using Twitter Data”, *Machine Learning with Applications*, vol. 3, p.100018, 2021.

[9] B. Yoon, Y. Jeong, K. Lee, and S. Lee, “A systematic approach to prioritizing R&D projects based on customer-perceived value using opinion mining”, *Technovation*, vol. 98, p. 102164, 2020.

[10] M. Arun, “Experimental investigation on energy and exergy analysis of solar water heating system using zinc oxide-based Nanofluid”, *Arabian Journal for Science and Engineering*, vol. 48, no. 3, pp.3977-3988, 2023.

[11] Y. Zuo, J. Wu, H. Zhang, D. Wang, and K. Xu, “Complementary aspect-based opinion mining”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 2, pp.249-262, 2017.

[12] C. Prajitha, K.P. Sridhar, and S. Baskar, “Noise Removal-based Thresholding framework for Arrhythmia classification”, *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 2, pp.2657-2668, 2023.

[13] B. Zhang, Y. Bai, Q. Zhang, J. Lian, and M. Li, “An opinion-leader mining method in social networks with a phased-clustering perspective”, *IEEE Access,* vol. 8, pp.31539-31550, 2020.

[14] S. Al-Azani, and E.S.M. El-Alfy, “ Early and late fusion of emojis and text to enhance opinion mining”, *IEEE Access,* vol. 9, pp.121031-121045, 2021.

[15] A. Mullick, P. Goyal, N. Ganguly, and M. Gupta, “Harnessing twitter for answering opinion list queries”, *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp.1083-1095, 2018.

[16] A. Mohammadinejad, R. Farahbakhsh, and N. Crespi, “Consensus opinion model in online social networks based on influential users”, *IEEE Access,* vol. 7, pp.28436-28451, 2019.

[17] A.D. Vo, Q.P. Nguyen, and C.Y. Ock, “Opinion–aspect relations in cognizing customer feelings via reviews”, IEEE Access, vol. 6, pp.5415-5426, 2018.

[18] S. Siddiqui, M.A. Rehman, S.M. Doudpota, and A. Waqas, “Ontology driven feature engineering for opinion mining”, *IEEE Access,* vol. 7, pp.67392-67401, 2019.

[19] J.S. He, M. Han, S. Ji, T. Du, and Z. Li, “Spreading social influence with both positive and negative opinions in online networks”, *Big Data Mining and Analytics*, vol. 2, no. 2, pp.100-117, 2019.

[20] F. Riquelme, P. Gonzalez-Cantergiani, D. Hans, R. Villarroel, and R. Munoz, “Identifying opinion leaders on social networks through milestones definition”, *IEEE Access,* vol. 7, pp.75670-75677, 2019.

[21] T. Nguyen Thi Ngoc, H. Nguyen Thi Thu, and V.A. Nguyen, “Mining aspects of customer’s review on the social network”, *Journal of Big Data,* vol. 6, pp.1-21, 2019.

[22] S. Poria, E. Cambria, and A. Gelbukh, “Aspect extraction for opinion mining with a deep convolutional neural network”, *Knowledge-Based Systems,* vol. 108, pp.42-49, 2016.

[23] S.N. Almuayqil, M. Humayun, N.Z. Jhanjhi, M.F. Almufareh, and D. Javed, “Framework for improved sentiment analysis via random minority oversampling for user tweet review classification”, *Electronics*, vol. 11, no. 19, p.3058, 2022.

[24] M.M.R. Mamun, O. Sharif, and M.M. Hoque, “Classification of textual sentiment using ensemble technique”, *SN Computer Science,* vol. 3, no. 1, p.49, 2022.

[25] R.S. Patil, and S.R. Kolhe, “Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets”, *Social Network Analysis and Mining*, vol. 12, no. 1, p.51, 2022.

[26] A. Samih, A. Ghadi, &A. Fennan, “Enhanced sentiment analysis based on improved word embeddings and XGboost”, *International Journal of Electrical & Computer Engineering* (2088-8708), vol. 13, no. 2, 2023.

[27] T.A. Khan, R. Sadiq, Z. Shahid, M.M. Alam, & M. B. M. Su'ud, “Sentiment Analysis using Support Vector Machine and Random Forest”, *Journal of Informatics and Web Engineering,* vol. 3, no. 1, 67-75, 2024.

[28] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision.”, CS224N Project Report, Stanford, vol. 1, p. 12, 2009.