

Sentiment Analysis of Social Network Comments for Identifying Opinion Leaders Using Machine Learning

Ripunjay Kumar

Research Scholar Jabalpur, India

Article received: 21/04/2026, Article Accepted: 28/05/2026, Article Published: 30/06/2026

Doi: - <https://doi.org/10.55640/ijidml-v03i06-04>

© 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the [Creative Commons Attribution License 4.0 \(CC-BY\)](#), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

ABSTRACT

The rate of social media growth is the greatest determinant in the spread of information. In order to understand and change opinion of the majority, it is necessary to find Key Opinion Leaders. In this research, a technique is provided for detecting opinion leaders in social media communications by analyzing their sentiment using Sentiment140. Following a long pre-processing sequence that includes text cleaning, tokenization, stemming, normalization, and Bag-of-Words feature extraction, the class imbalance problem is solved by using the SMOTE approach. The suggested method find geographical patterns and causal linkages in the text by using deep learning models like Bidirectional Long Short-Term Memory (BiLSTM) networks and Convolutional Neural Networks (CNNs). Compared to other machine learning models, such as Naive Bayes, Adaboost, and Logistic Regression, the suggested models performed better in the empirical data. The F1-score (F1), accuracy (ACC), and recall (REC) of these models are all above average, reaching at 98%. The results indicated that the deep learning algorithms can perform very well in the field of identifying opinion leaders based on sentiment analysis.

KEYWORDS

Text Classification, Social Network, Machine Learning, Opinion Leader, Social Media Analytics Sentiment Analysis, Short Text Classification, Deep Learning.

I. INTRODUCTION

The proliferation of social media websites has changed the way people communicate, share their opinions and discuss their thoughts [1]. There is a wealth of user-generated textual data available on social media sites like Reddit, Facebook, and Twitter that might provide insight into audience attitudes [2][3]. Opinion mining, or sentiment analysis, has recently been interested in the subjective information included in this unstructured data [4]. This capability is useful for researchers, businesses, and lawmakers since it can determine if language is neutral, positive, or negative. They can learn about public opinion and habits of conduct through this categorization [5]. Social media data poses problems to analyze because of its linguistic ambiguity, sarcasm, informal expressions and varying writing styles.

The information revolution, as an outcome of the Internet, has changed the way that content is produced, used and shared [6]. Social networking websites are

living ecosystems where millions of people interact, share ideas, and have an influence on one another's opinions [7]. In these kinds of digital spaces, these informal online communities are a crucial part of public discourse, particularly in socio-political spaces [8][9]. People engage in conversations, share their opinions about politicians and help to shape the group's opinion [10].

The majority of research on opinion leaders is based on network structure and user activity and does not address the qualitative nature of influence such as positivity, expertise, and credibility [11]. Such techniques can show controversial or very active users instead of the truly influential and professional opinion makers [12]. In order to overcome this constraint, a sentiment analysis-based approach that combines topic classification to detect the influencers that have a positive sentiment for the topic of discussion is introduced [13]. Using sentiment and thematic analysis of user-generated content can identify people who are adding value and contributing to

discussion in a positive and constructive way, providing a more accurate view of opinion leaders[14].

As social data grows in volume and complexity, traditional lexicon-based and keyword-driven approaches fail to capture the nuances, sarcasm and changing language patterns of social data [15]. The area of sentiment analysis has seen a rise in the use of DL and ML approaches [16][17]. These advanced methods enhance ACC accuracy and sentiment classification flexibility by utilizing massive amounts of data and complex models [18]. Language-independent and language-aware sentiment analysis systems using ML and DL can monitor in real time, process large amounts of information, and identify opinion leaders more accurately, as they are able to cope with linguistic variations and context dependency more effectively [19].

A. Motivation and Contribution

The necessity to assess user mood and identify prominent opinion leaders has arisen due to the social media's growing influence on public opinion. Typical methods do not reflect nuances and effective impact in context and give unreliable answers. Advancement in the field of sentiment analysis Utilizing state-of-the-art ML algorithms drives ACC and effectively identifies positive opinion leaders. As below, the study makes important strides in a number of critical areas of research:

- Proposed a sentiment-based opinion leaders identification model for the Sentiment140 dataset using social network data.
- Extensive pre-processing pipeline adopted including Text Cleaning, Tokenization, Stemming, Normalization, Feature Extraction (Bag of Words).
- Using the SMOTE method to improve the model learning process, successfully dealt with the class imbalance crisis.
- Created and used CNN and BiLSTM DL models to mine text databases for geographical information and contextual relationships.
- The suggested model's classification capabilities were thoroughly evaluated using a range of performance indicators, including REC, ACC, PRE, and F1.

This study is motivated by the increasing demand for identifying opinion leaders in massive social media data to grasp public opinion accurately. The novelty of this work is the integration of the advanced DL models into a balanced well-preprocessed Sentiment140 dataset with SMOTE. Higher ACC and more resilience makes this combination significantly outperform typical ML algorithms in sentiment categorization.

B. Organization of the Paper

The paper is organised as follows: Section II presents the literature search for finding influential users using their content on social media, The dataset, its preprocessing, and the model's execution are detailed in Section III. The study's results and comparisons are presented in Section

IV. Finally, recommendations for future research are presented in Section V.

II. LITERATURE REVIEW

The development of this study is made simpler by doing a thorough evaluation of previous research on opinion leader identification in social networks. The present studies, models, data, major discoveries, and issues in the field are summarized in Table I.

S. Borhade et al. (2026) Processed data is managed in the HDFS and queried via Hive to derive aggregated insights. The proposed system is effective in analyzing large-scale user feedback, as it is shown to be effective on Reddit forums related to the iPhone 17, with an emotion detection ACC of 88.7% and a sentiment classification ACC of 92.4% [20].

V. S. Abeynayake et al. (2026) examined the efficacy of the SentimentAnalysis model in Romanized Sinhala-English sentiment analysis by mining a sizable Sinhala sentiment corpus for pre-training the Sinhala model SinLlama, which is a derivative of the Llama-3-8B model. A combined Romanized and coded text's semantic subtleties may be better understood by the improved SinLlama model, according to the testing findings, which demonstrate an ACC of 74.2%, surpassing the classic ML baselines [21].

X. Zhang et al. (2025) developed a novel model for social media sentiment analysis and prediction using GNN, FP-Growth frequent itemset, and LDA topic model. Achieving an ACC of 91.30% when training and testing the model on popular Weibo conversation topics validates the efficiency of adding user connection networks and multimodal characteristics [22].

S. Prabha et al. (2025) proposed to all six techniques SVM has outperformed for classification of toxic and non-toxic comments and normal text classification in general. Results from SVM are the best with an ACC of 93.61% and an F1 of 93% Proposed technique has improvised ACC score as compared to existing techniques [23].

M. K. Singh and S. Kumar (2024) created to predict the following five unique emotional states: "Happy," "Angry," "Depression," "Sad," and "Fear.". Unbelievable in comparison to previous systems used for human emotion identification, the suggested method has attained an ACC and F1 higher than 93% [24].

K. Rai and P. Kumar (2023) DL models, DTs, RF, support vector machines, LR with TF-IDF, and count vector-based models are among the ML approaches that have been studied and implemented. With a success rate of 92%, the DL model surpassed all other ML models [25].

TABLE I. RECENT STUDIES ON SOCIAL NETWORK COMMENTS USING MACHINE LEARNING TECHNIQUES

Author	Proposed Work	Results	Key Findings	Limitations & Future Work
Borhade et al. (2026)	Developed a large-scale social media sentiment and emotion analysis framework using HDFS and Hive for processing Reddit discussions related to iPhone 17	92.4% sentiment classification accuracy and 88.7% emotion detection accuracy	Demonstrated effective large-scale user feedback analysis using distributed big data technologies	Future work can focus on multilingual sentiment analysis and real-time streaming data processing
Abeynayake et al. (2026)	Proposed SinLlama, a fine-tuned Llama-3-8B model trained on a large Sinhala corpus for Romanized Sinhala-English sentiment analysis	74.2% accuracy	Fine-tuned SinLlama outperformed traditional machine learning baselines in understanding code-mixed semantic nuances	Performance can be improved using larger multilingual datasets and optimized transformer architectures
Zhang et al. (2025)	Proposed a sentiment prediction framework integrating LDA topic modeling, FP-Growth frequent itemsets, and GNN on Weibo social media data	91.30% accuracy	Integration of user relationship networks and multimodal features improved sentiment prediction performance	Future work may include cross-platform validation and enhanced multimodal feature extraction
Prabha et al. (2025)	Applied six techniques with SVM for toxic and non-toxic comment classification	93.61% accuracy, 93% F1-score	SVM outperformed all models in classification tasks	Requires testing on larger and diverse datasets for generalization
Singh & Kumar (2024)	Emotion prediction across five categories (Happy, Angry, Depression, Sad, Fear)	>93% accuracy and F1-score	Model outperformed existing emotion recognition systems	Can be extended to real-time and multilingual datasets
Rai & Kumar (2023)	Compared ML (DT, RF, SVM, LR) and DL models using TF-IDF and count vectors	92% accuracy (DL)	DL models performed better than ML models	Requires optimization to reduce training time and resources

Research gaps: Although the fields of sentiment analysis and opinion leader identification have made great progress, the current literature remains insufficiently equipped to process large, noisy, and imbalanced social media data. The majority of conventional ML systems exhibit moderate performance, and most DL models fail to generalize across different datasets and languages. Thus, stronger, more scalable, and more accurate models are required that can enhance the real-time opinion leader detection in social networks.

III. RESEARCH METHODOLOGY

This approach finds sentiment analysis opinion leaders by using data pre-treatment methods as cleaning, tokenization, stemming, and normalization. To do this, it leverages the Sentiment140 dataset. After feature extraction, the dataset is first divided into a training set and a testing set using the Bag of Words approach. The dataset is then class-balanced using SMOTE. Following their analysis of the processed data, the CNN and Bi-LSTM models are evaluated using ACC, PRE, REC, and F1. The Social Network Comments offer a ML-based flowchart (Fig. 1).

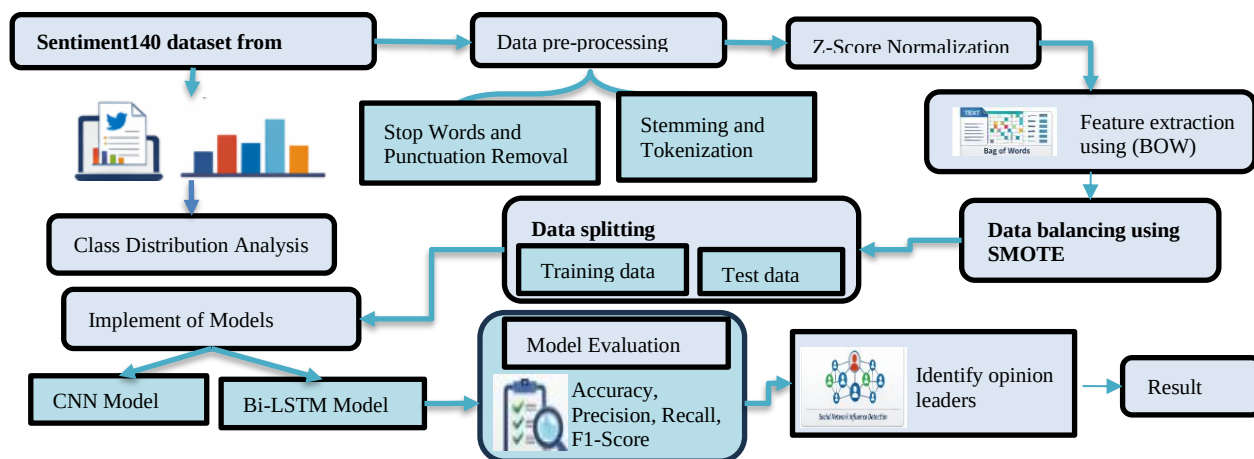


Fig. 1. Proposed flowchart for Social Network Comments for Identifying Opinion Leaders using machine learning

The next section lays out the suggested technique step-by-step:

C. Data Gathering and Analysis

A little over 1.5 million tweets that were tagged as good, negative, or neutral make up the Sentiment140 dataset.

Using the presence of emoticons, sentiment labels are automatically assigned to each tweet, offering a massive resource for training and testing sentiment analysis algorithms. The distribution of attacks, correlation of features etc. are explored in data visualization below:

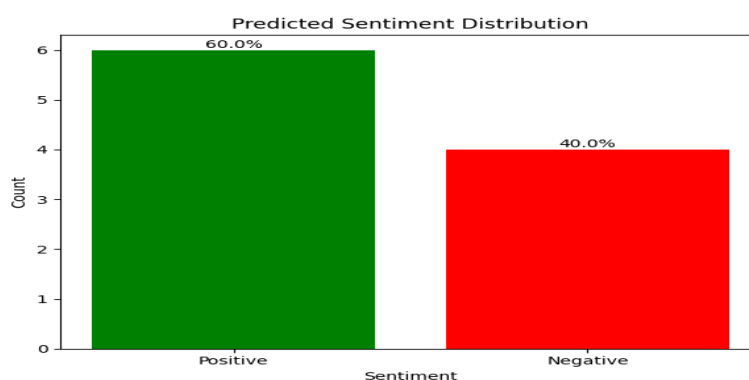


Fig. 2. Bar graph of class distribution of Sentiment140 dataset

The predicted sentiment distribution bar chart shown in Fig. 2 shows a minor imbalance of classes, with more

instances in the positive class (6 instances - 60%) than in the negative class (4 instances - 40%). This is equivalent to a 3:2 ratio, or more positive than negative ratio.



Fig. 3. Positive and Negative Word Clouds

The most popular positive and negative terms used in social media comments are displayed in the word clouds in Fig. 3. The abundance of the words "love," "thank," and "good" in the positive word cloud indicates a state of contentment. The negative word cloud, on the other hand, shows words such as "miss", "work", and "think", which indicate negative feelings and complaints from the users.

D. Data Pre-Processing

The Sentiment140 dataset is used in the preparation of the data where concatenating, cleaning, and feature engineering of the data took place. The preprocessing involved letters and special characters, punctuation and text cleaning, stemming, tokenization and data labeling and normalization. The primary steps of preprocessing can be summarized in the following way:

- **Elimination of Stop Words and Special Characters:** A language's most popular words are called stop words. Remove them without altering the meaning of the statement because they don't contribute anything.
- **Punctuation Removal and Text Refinement:** Textual Data cleanup. Punctuation and text refinement clean textual data by removing unnecessary symbols and noises that contribute to text analysis ACC.
- **Stemming:** Use base forms of words (running, run) to minimize dimensionality. This used to simplify analysis and improve computational performance. In achieving efficiency and scalability.

E. Tokenization

The information is broken up into separate words or phrases to make further processing easier. One way to link models with unstructured text is to sort the data. Guarantees that the support environment phrase is correctly logged in to analyze the social network; this has an immediate effect on the model's efficiency.

F. Feature normalization using Z-Score Normalization

Data normalization is a process that aims to modify or standardize data in order to make its distribution more consistent. Among the most common methods for data normalization, z-score and rescaling (min-max normalization) stand out. This study employed z-score normalization to get a set of standard deviations of 1 and

means of 0. This scaling process centers the data around the mean and gives it a standard deviation of one. Equation determines the normalization of the z-score Equation (1).

$$E' = \frac{E - \bar{M}}{\sigma_M} \quad (1)$$

For each data input, E' and E stand for the new and old values, $\bar{M}$ denotes the mean, and σ_M stands for the standard deviation.

G. Feature Extraction Using (BOW)

The purpose of feature extraction is to improve the efficiency of ML models by translating text into useful numerical vectors. The BOW method is widely used because it produces a set of new expressions using the frequencies of the words. It treats text without regard to context or word order, and transforms it into a structured set of numerical features that allow for efficient processing and analysis by models.

H. Data balancing using SMOTE

Data balancing solves the problem of class imbalance by achieving more balanced sample distribution to enhance model performance. Using Euclidean distances and KNN SMOTE technique creates synthetic samples for minority classes. New samples are generated using interpolation and inserted in the data set to increase the class balance and improve overall learning efficiency.

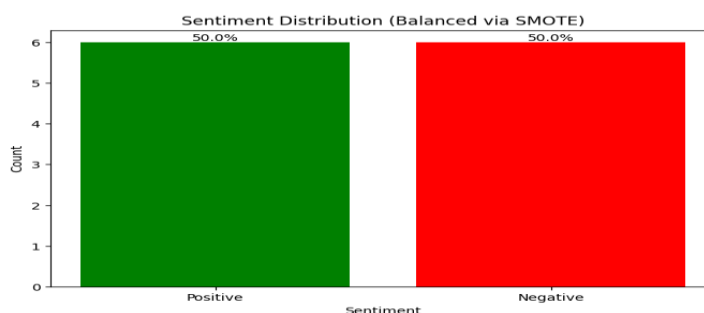


Fig. 4. Bar graph of class distribution after SMOTE of the Sentiment140 dataset

The Sentiment Distribution Bar Graph is depicted in Fig. 4 which shows effective elimination of the class imbalance, with 50% Positive class and 50% Negative class of the total number of classes. The balanced distribution ensures fair training of the model, equal learning of both sentiment classes, and better predictive fairness.

I. Data Splitting

The information consists of two parts: Before evaluating the learnt model, use 20% of the data, and use 80% to train the algorithm.

J. Implement of Models

The proposed CNN and BiLSTM models can be used to find opinion leaders in social network comments by combining feature extraction and contextual sequence

learning. This makes it easier to find sentiment and influence.

1) Proposed Convolutional Neural Network (CNN) Model

Feature learning is accomplished by CNNs through a categorization of layered layers that employ non-linear transformations. The input data is shown at the visible layer as a tensor, which is a multidimensional data array. Data in time series, for instance, may be seen as a 1D grid sampling at regular intervals; pixels in a 2D image[26]; the structure of a 3D movie; and so on all exhibit this grid-like topology. Then, many abstract characteristics are extracted using a hidden layer series. As an example, as shown in Equation (2), a 2D kernel h calculates the 2D convolution given a 2D input x .

$$(x * h)_{i,j} = x[i,j] * h[i,j] = \sum_n \sum_m x[n,m] \cdot h[i-n][j-m] \quad (2)$$

A limited input region linked to their weights and the dot product of that region.

The feature map is enhanced at the filter's output following convolution with the addition of a bias term and a point-wise nonlinearity g . The feature map h^l , where l is the l -th feature map at a certain convolutional layer, may be expressed using Equation (3). Input x , weights or coefficients W^l , and bias b_l decide the filters of the features map.

$$h_{i,j}^l = g(W^l * x)_{ij} + b_l \quad (3)$$

The 2D convolution that is defined is represented by $*$, and the activation function is $b(\cdot)$.

Equation (4) shows the rectifier activation function, which is standard in DNNs.

$$g(x) = x^+ = \max(0, x) \quad (4)$$

A CNN model with 10 training epochs contains 128 embedding dimensions and 64 batches. The Adam optimizer, the ReLU activation function, and a learning rate of 0.001 are used to reduce overfitting and increase generalization. Approximately 50% of those that enroll not remain.

2) Proposed Bidirectional Long Short-Term Memory (Bi-LSTM) Model

In Bi-LSTM, the forward hidden layer is in charge of storing and transmitting data from the past to the future, while the backward hidden layer is in charge of storing and transmitting data from the future to the past. When it comes to data representation in DL systems, Bi-LSTM performs better than ordinary LSTM. This section delves into the Bi-LSTM output, specifically covering points in Equations (5) to (7):

$$h_t^f = LSTM(x_t, h_{t-1}^f) \quad (5)$$

$$h_t^b = LSTM(x_t, h_{t-1}^b) \quad (6)$$

$$y_t = W_o h_t + b_o \quad (7)$$

The output layer's bias vector is represented as b_o , the weights from the forward layer to the output layer are W^f_{hy} , and the weights from the backward layer to the output layer are $b W^b_{hy}$. The function h_t is created by using the integrals of h^f_t and h^b_t . In order for the Bi-LSTM to learn at time t , it takes in data simultaneously from the past and the future, merging the two sets of knowledge.

The BiLSTM model trained for 10 epochs and features 128 LSTM units with a batch size of 128. Further, it employs the Adam optimizer, which allows for 0.3 dropout rate and 0.1 learning rate for textual contextual

connections.

A. Evaluation Metrics

Researchers compared the proposed model's performance across a range of feature extraction approaches using a battery of industry-standard performance indicators. In instance, REC, PRE, ACC, and the F1 were employed. To visualize ML models and compute all four metrics, a confusion matrix is a helpful tool. Within this matrix, the components that constitute the TN are FN, TP, FP, and TN. In order to measure how well a classifier does its TASK, utilize the following formulas, which are derived from Equations (8) to (11):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

The ACC rate of the trained model's predictions as a percentage of the total instances in the dataset is called its ACC rate. The ACC rate is a measure of the model's accuracy in predicting positive occurrences relative to a 100-point scale. A measure of how many positive events the model anticipated out of a total number of positive events. The F1 is a harmonic average of memory and ACC. This means that it helps to balance ACC and REC.

IV. RESULTS AND DISCUSSION

The experimental setup and performance of both models are examined in this part. These models are tested on diverse datasets and demonstrate processing efficiency and effectiveness.

A. Tools and Hardware Configuration

The implementation is done on Python 3.9 and makes use of well-known ML modules for preprocessing, handling, and visualization, including scikit-learn, NLTK, pandas, NumPy, matplotlib, and seaborn [27]. It is based on an Intel i5 12th Gen processor with 16 GB of RAM, it is a dual Windows 11 operating system and is developed using Jupyter Notebook.

B. Performance Evaluation

The models are evaluated based on critical performance metrics such REC, ACC, PRE, and F1, as shown in Table II. They were trained using the Sentiment140 dataset. These findings demonstrate that CNNs and Bidirectional LSTM models are capable of detect and predict future opinion leaders from comments made on social networks. The CNN model achieves 98% PRE, 99% REC, and 98% F1, whereas the BiLSTM model achieves 98% on all measures. These results suggest strong and dependable sentiment classification, whereby CNN model exhibits a minor difference in recall.

TABLE II. CLASSIFICATION RESULTS OF PROPOSED MODELS SOCIAL NETWORK FOR SENTIMENT140 DATASET

Matrix	CNN	BiLSTM
Accuracy	98	98
Precision	98	98
Recall	99	98
F1-score	98	98

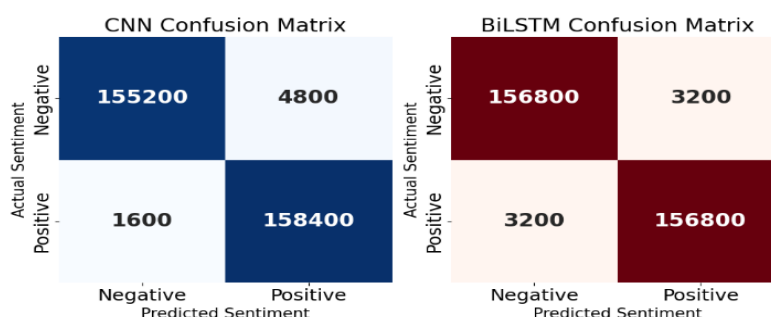


Fig. 5. Confusion Matrix Comparison of CNN and BiLSTM Models for Sentiment Classification

Fig. 5 shows the CNN and BiLSTM models' confusion matrices for sentiment categorization. The CNN model correctly classified 155,200 negative and 158,400 positive samples, with 4,800 and 1,600

misclassifications, respectively. Similarly, the BiLSTM model achieved 156,800 correct predictions for both negative and positive classes, with only 3,200 incorrect classifications, demonstrating high classification ACC for both models.

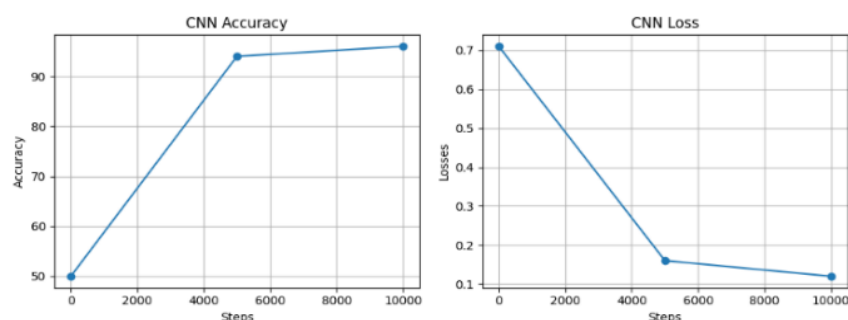


Fig. 6. Accuracy and loss Curve for the proposed CNN model

The suggested CNN model's training results, shown in Fig. 6, demonstrate a rapid and efficient learning process, a high convergence rate, and solid initial improvement. The curve of ACC increases sharply then levels off and the curve of loss always decreases during training. This trend shows successful feature learning and optimization, and the performance is stable and there are no definite indications of instability.

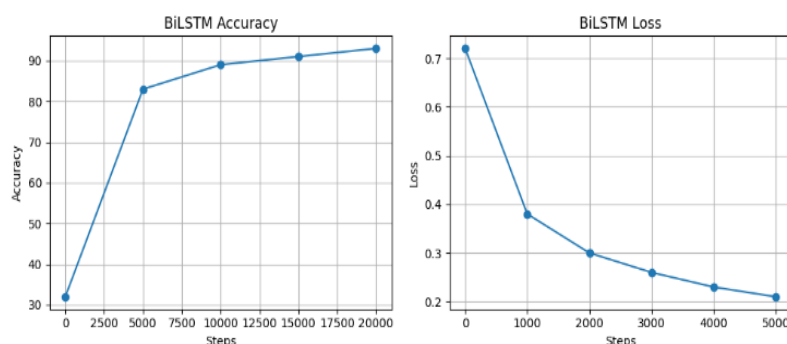


Fig. 7. Accuracy and loss Curve for the proposed BiLSTM model

Fig 7 shows the training curve of the proposed BiLSTM model, which exhibits a good learning trend with rapid convergence and stable predictive performance. The ACC curve has a very high upward trend with a tendential

stabilization and the loss curve has a steady decrease with the course of training. Since ACC is negatively correlated with loss, and can observe that the model is learning the fundamental patterns and optimizing its performance without any overfitting.

A. Comparative Analysis

The effectiveness of proposed CNN and BiLSTM models is evaluated through different methods, and the comparison between these two models is given in Table III. The performance variations amongst the models used on the Sentiment140 dataset to find opinion leaders.

Compared to conventional algorithms like BERT, Logistic regression and AdaBoost, NB has a marginal improvement but has less ACC. However, with an ACC of 98% and consistently high PRE, REC, and F1 values, the suggested CNN and BiLSTM models are the strongest and most reliable options.

TABLE III. COMPARISON OF DIFFERENT MACHINE LEARNING DEEP LEARNING MODELS FOR SOCIAL NETWORK COMMENTS

Model	Accuracy	Precision	Recall	F1-score
BERT[28]	64.50	64	65	64
LR[29]	68	69	68	68
AdaBoost [30]	74	70	82	75
NB[31]	80	80	81	80
Proposed CNN	98	98	99	98
Proposed BiLSTM	98	98	98	98

B. Discussion

The discussion highlights that the proposed CNN and BiLSTM models outperform traditional and baseline models in opinion leader identification on comments of social networks. They can extract local features and sequential context, which helps achieve better sentiment classification. The overall results are indicative of the robustness and effectiveness of the proposed models for dealing with large-scale textual data.

V. CONCLUSION AND FUTURE STUDY

Sentiment analysis of social media comments is one of the most crucial factors in assessing real-time user feedback and public opinion, and in finding suitable news and opinion-forming posts in social networks. This research presents a DL opinion leader identification strategy and compares several models that are based on DL and ML. Experimental results confirm that traditional models such as BERT, LR, AdaBoost and NB have moderate performance, while the proposed CNN and BiLSTM models have excellent performance with an ACC of 98% and can well capture the complex textual patterns and contextual relationships. Future work should focus on improving feature extraction and classification results by incorporating attention mechanisms and integrating hybrid DL architectures like CNN-LSTM and Transformer-based models. The results show that DL techniques can be effective in opinion leader detection. The system's scalability, efficiency, interpretability, and applicability may be improved with the use of explainable AI approaches and real-time processing pipelines.

REFERENCES

1. M. Zhu, X. Lin, T. Lu, and H. Wang, "Identification of Opinion Leaders in Social Networks Based on Sentiment Analysis: Evidence from an Automotive Forum," vol. 58, no. Msota, pp. 412–416, 2016, doi: 10.2991/msota-16.2016.91.
2. S. R. El-Beltagy, T. Khalil, A. Halaby, and M. Hammad, "Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis," 2018, pp. 307–319. doi: 10.1007/978-3-319-75487-1_24.
3. M. N. Paralath, "Bridging Astronomical Imagery and Text with Advanced Multi-Modal AI," in 2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India: IEEE, 2025, pp. 1–8, November. doi: 10.1109/ICCCA66364.2025.11325649.
4. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks," Soc. Netw. Anal. Min., vol. 10, no. 1, p. 53, Dec. 2020, doi: 10.1007/s13278-020-00668-1.
5. N. Alapati, R. rao Thallada, and K. Nallabothu, "Risk-Based Governance for Autonomous Decision Systems," J. Bus. Manag. Stud., vol. 8, no. 6, April, pp. 69–73, 2026, doi: 10.32996/jbms.

6. Alsayat and N. Elmitwally, "A comprehensive study for Arabic Sentiment Analysis (Challenges and Applications)," *Egypt. Informatics J.*, 2020, doi: 10.1016/j.eij.2019.06.001.
7. Y. Zahidi, Y. El Younoussi, and Y. Al-Amrani, "A powerful comparison of deep learning frameworks for Arabic sentiment analysis," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 1, p. 745, Feb. 2021, doi: 10.11591/ijece.v11i1.pp745-752.
8. Padhy, N. Seshagiri, V. Soni, A. K. Elengovan, G. Babu Thokala, and M. Kumari, "AI-Enabled Autonomous Data Preprocessing: A Scalable Architecture for Intelligent Machine Learning Pipeline Management," in *2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN)*, Jan. 2026, pp. 1318–1323. doi: 10.1109/ICIPCN67432.2026.11438503.
9. Patel, "Hybrid Deep Learning Approaches for Personalized Product Recommendation in Online Retail Systems," in *2026 6th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, FEZ, Morocco: IEEE, 2026, pp. 1–6, May. doi: 10.1109/IRASET68627.2026.11538894.
10. R. M. Duwairi and I. Qarqaz, "Arabic Sentiment Analysis Using Supervised Classification," in *2014 International Conference on Future Internet of Things and Cloud*, IEEE, Aug. 2014, pp. 579–583. doi: 10.1109/FiCloud.2014.100.
11. M. El-Masri, N. Altrabsheh, H. Mansour, and A. Ramsay, "A web-based tool for Arabic sentiment analysis," *Procedia Comput. Sci.*, vol. 117, pp. 38–45, 2017, doi: 10.1016/j.procs.2017.10.092.
12. A.-K. Al-Tamimi, A. Shatnawi, and E. Bani-Issa, "Arabic sentiment analysis of YouTube comments," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, IEEE, Oct. 2017, pp. 1–6. doi: 10.1109/AEECT.2017.8257766.
13. Alharbi, M. Kalkatawi, and M. Taileb, "Arabic Sentiment Analysis Using Deep Learning and Ensemble Methods," *Arab. J. Sci. Eng.*, vol. 46, no. 9, pp. 8913–8923, Sep. 2021, doi: 10.1007/s13369-021-05475-0.
14. R. Azmeera and J. Hyatt, "Software Errors, Product Functionality, and Organizational Cascades: Evidence from Developer-Lived Experience," 2026.
15. Guellil, F. Azouaou, and M. Mendoza, "Arabic sentiment analysis: studies, resources, and tools," *Soc. Netw. Anal. Min.*, vol. 9, no. 1, p. 56, Dec. 2019, doi: 10.1007/s13278-019-0602-x.
16. R. M. Duwairi, N. A. Ahmed, and S. Y. Al-Rifai, "Detecting sentiment embedded in Arabic social media – A lexicon-based approach," *J. Intell. Fuzzy Syst.*, vol. 29, no. 1, pp. 107–117, Sep. 2015, doi: 10.3233/IFS-151574.
17. Dixit, "Predictive Analytics in Business Intelligence for Sales Forecasting," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 03, no. 01, August, pp. 981–991, 2023, doi: 10.48175/IJAR SCT-12750G.
18. S. Dodda, H. Volikatla, and J. R. Vummadi, "Exploring the Role of AI-Enhanced Chatbots in Automating Recruitment Processes in Human Capital Management Systems," *Int. J. Emerg. Trends Comput. Sci. Inf. Technol.*, vol. 6, no. 3, July, pp. 28–36, 2025, doi: <https://doi.org/10.63282/3050-9246.IJETCSIT-V6I3P104>.
19. Alqahtani, S. B. Khan, J. Alqahtani, S. AlYami, and F. Alfayez, "Sentiment Analysis of Semantically Interoperable Social Media Platforms Using Computational Intelligence Techniques," *Appl. Sci.*, vol. 13, no. 13, Jun. 2023, doi: 10.3390/app13137599.
20. S. Borhade, S. Bhoir, A. Tumsare, H. Harkulkar, S. Mhatre, and M. Trivedi, "Reddit-Based Social Media Sentiment and Emotion Analysis for Product Improvement Using Hadoop," in *2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET)*, IEEE, Feb. 2026, pp. 816–820. doi: 10.1109/IC3ET64989.2026.11467286.
21. V. S. Abeynayake, S. A. S. Lorensuhewa, and M. A. L. Kalyani, "Fine-Tuning SinLlama for Sentiment Analysis in Romanized Sinhala-English Code-Mixed Social Media Text," in *2026 6th International Conference on Advanced Research in Computing (ICARC)*, IEEE, Feb. 2026, pp. 1–5. doi: 10.1109/ICARC68737.2026.11453630.
22. X. Zhang, J. Li, and M. Zhao, "The Research on Social Media Sentiment Analysis and Prediction Model Based on Graph Neural Networks," in *2025 5th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM)*, IEEE, Sep. 2025, pp. 376–379. doi: 10.1109/AIIM67611.2025.11232724.
23. Prabha, A. Yadav, A. Yadav, A. Kumar, and M. Tyagi, "Toxic Comments Classification using Machine Learning and Word Embedding Techniques," in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, IEEE, Feb.

- 2025, pp. 599–604. doi: 10.1109/CICTN64563.2025.10932491.
24. Singh and S. Kumar, “Stress Detection During Social Interactions with Natural Language Processing and Machine Learning,” in 2024 International Conference on Expert Clouds and Applications (ICOECA), IEEE, Apr. 2024, pp. 297–301. doi: 10.1109/ICOECA62351.2024.00060.
25. Rai and P. Kumar, “Comparative Analysis of Machine Learning and Deep Learning Techniques in Text Based Emotion Detection,” in 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), IEEE, Apr. 2023, pp. 260–263. doi: 10.1109/CISES58720.2023.10183612.
26. Singamsetty, “AI-powered Satellite Image Processing for Global Air Traffic Surveillance Techniques Using NCNN–EGSA Optimization Techniques,” in Machine Learning Based Air Traffic Surveillance System Using Image Processing, Emerald Publishing Limited, 2026, pp. 179–197. doi: 10.1108/978-1-80592-062-520251010.
27. Korat and A. Alimohammad, “Efficient Hardware Implementation of Eigen Solver,” in 2026 IEEE 16th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA: IEEE, 2026, pp. 1376–1385, January. doi: 10.1109/CCWC67433.2026.11393888.
28. Sathya, K. Thulasimani, and S. Devasurithi, “A study on twitter sentiment analysis using deep learning,” Int. Res. J. Eng. Technol., pp. 799–804, 2023.
29. S. Shah, “Twitter Sentiment Analysis Using Recurrent,” Int. J. Sci. Technoledge, vol. 10, no. 10, pp. 5–9, 2022.
30. Gaye, D. Zhang, and A. Wulamu, “A Tweet Sentiment Classification Approach Using a Hybrid Stacked Ensemble Technique,” Information, vol. 12, no. 9, p. 374, Sep. 2021, doi: 10.3390/info12090374.
31. P. Durga and T. Sudhakar, “A decision-making framework utilizing machine learning techniques, based on analysing the sentiment of tweets,” Multidiscip. Sci. J., vol. 5, p. 2023ss0102, Aug. 2023, doi: 10.31893/multiscience.2023ss0102.