Received: October 18, 2022. Revised: December 16, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Effective Twitter Sentiment Analysis Using Deep Belief Network with Enhanced Dragonfly Optimization Algorithm

Gouri Bompem[1*] **Nagaraju Chiluka[2]** **Dhanalakshmi Pandluri[3]**

[1]*Department of Computer Science & Engineering, KSRM College of Engineering, Kadapa, India*
[2]*Department of Computer Science and Engineering, YSR Engineering College of YVU, Proddatur, India*
[3]*Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Chittoor, India*
* Corresponding author's Email: gouri.b30@gmail.com

Abstract: In this research article, an effective model is implemented for predicting the outcome of the Indian general election 2019 and farmer's protest by utilizing the sentiment analysis of Twitter data. In the initial segment, the raw Twitter data are acquired from the Indian Political tweets 2019 and farmers protest tweets databases. Further, the denoising operations such as removal of unnecessary space, tabs, new-line, hashtag symbols, non-English characters, punctuations, numbers and special characters are used to enhance the quality of the acquired data. Then, the keyword trend analysis and topic modeling utilizing latent dirichlet allocations (LDAs) are performed for better data representation. Next, the extraction of the feature is carried out utilizing skip-gram and term frequency-document level frequency (TF-IDF) techniques, and further, the feature optimization is accomplished using enhanced dragonfly optimization algorithm (EDOA) for selecting the optimal feature vectors. In the EDOA, a Brownian motion is added for improving the probabilistic behaviors. Lastly, the selected features are given to the deep belief network (DBN) model to classify the people's sentiments into negative, neutral, and positive classes. The experimental evaluations demonstrated that the EDOA-DBN model has obtained 99.22% and 98.83% of accuracy on both Indian political tweets 2019 and farmers protest tweets databases, which are maximum related to other existing models.

Keywords: Deep belief network, Dragonfly optimization algorithm, Farmers protest, Parliament election, Sentiment analysis, Twitter data.

1. Introduction

Indian farmers' protests and general elections are the important events that occurred in the year 2019, 2020, and 2021 [1, 2]. So, it is necessary to understand the people's sentiments behind online conversations such as Twitter, Facebook, Instagram, etc., which allows us to take into consideration a broader audience that includes both in-direct and direct participants [3-5]. Currently, Twitter is one of the effective platforms for sharing informative messages related to other social media platforms [6, 7]. In recent decades, numerous machine learning models are used for people's sentiment analysis, but the majority of the existing models focused only on the textual feature vectors for constructing vector representation of the Twitter data [8, 9]. The existing models have failed in extracting the important sentiment information to obtain better classification performance [10-12]. To improve the classification performance and address the aforementioned issue, a feature optimization-based deep learning model is introduced in this manuscript. The main objectives of this article are stated as follows:

- After collecting raw tweets from the Indian Political tweets 2019 and farmers protest tweets databases, the basic data denoising operation, keyword trend analysis, and topic modeling are accomplished for better data representation.
- The discriminative feature values are excerpted from the denoised data utilizing skip-gram and

Received: October 18, 2022. Revised: December 16, 2022.

TF-IDF, which helps to improve the classification performance.

- The higher dimensional feature vectors are decreased to the lower dimensional feature vectors by proposing EDOA, which consists of Brownian motion for improving its probabilistic behavior. The dimensionality reduction process decreases the complexity of the system and computational time.
- The dimensionally reduced feature values are given to the DBN model to classify the people's sentiments: negative, neutral, and positive classes. The EDOA-DBN efficacy is validated by means of precision, accuracy, f1-score, recall, and specificity.

The article is organized as follows: manuscripts related to the Twitter sentiment evaluation are abstracted in section 2. The methodology details, simulation outcomes, and the conclusion of this work are given in sections 3, 4, and 5 respectively.

2. Related works

Joseph, [13] used a decision tree classifier for predicting the outcomes of the Indian general election in 2019 by utilizing the sentiment analysis of the Twitter data. The experimental results demonstrated that the decision tree classifier was effective in mapping the sentiments of people across several phases of polls. However, conventional machine learning models like decision trees have high variance and bias while processing the model on larger databases. Neogi [14] applied Bag of words and TF-IDF techniques to extract active feature values from the acquired twitter data. Secondly, the extracted active feature values were given to several machine learning models: Support vector machine (SVM), random forest, decision tree and Naïve Bayes for sentiment classification. In that, the random forest classifier attained higher accuracy on the farmer's protest tweets database. However, the random forest classifier was ineffective and slow in real time prediction like sentiment analysis of farmers' protest. On the other hand, Tiwari and Nagpal [15] presented knowledge-enriched attention-based hybrid transformer (KEAHT) with bidirectional encoder representation from transformer (BERT) model for social sentiment evaluation. The experimental investigation confirmed that the developed model has achieved high classification results by means of accuracy, precision, F1-score and recall, but it was computationally complex.

Shekhawat [16] introduced a new hybrid spider monkey optimizer for effective Twitter sentiment analysis. The extensive experiments conducted on the Twitter and sender-2 databases showed the efficacy of the developed model. Still, the developed model needs to be applied to the paradox and sarcastic tweets. Hassonah [17] integrated multi-verse optimizer and ReliefF method for feature optimization and the optimized feature values were given to the SVM classifier for classifying the sentiments of people as neutral, negative, and positive. However, the SVM classifier performed well in the binary classification, but it was inappropriate for multi-class classification. Kumar, and Jaiswal, [18] integrated binary moth flame and binary grey wolf optimizers for feature optimization that helps in improving the sentiment classification accuracy. The extracted feature values were given to the 5 models such as decision tree, k-nearest neighbor, naïve Bayes, SVM, and multi-layer perceptron for sentiment classification. The experiment conducted on the SemEval 2016 and 2017 databases demonstrated the efficacy of the developed model, but the computational time was maximum in this literature study.

Wang, and Hu, [19] implemented an attentional graph neural network for effective Twitter sentiment analysis. The experiments performed on the real-time databases confirmed the efficacy of the implemented deep learning model. However, the neural network required a larger amount of data to obtain high classification results, which was computationally costly. Additionally, Pandey et al. [20] introduced a novel cuckoo search optimizer based on k-means clustering for Twitter sentiment analysis. The statistical analysis and the experimental results showed that the developed model outperformed the existing models on the benchmark twitter databases; still, the developed model needs to be focused on opinion classification. To address the aforementioned concerns, an EDOA-DBN model is implemented for effective sentiment analysis using Twitter data. Where, the proposed EDOA is effective related to other conventional algorithms: stochastic komodo algorithm [21], multi leader optimizer [22], three influential members based optimizer [23], random selected leader based optimizer [24], and squirrel search optimizer [25].

3. Methodology

In the Twitter sentiment analysis, the proposed study comprises six phases such as data acquisition: Indian political tweets 2019 and farmers protest tweets databases, data denoising, exploratory analysis: keyword trend analysis and topic modeling

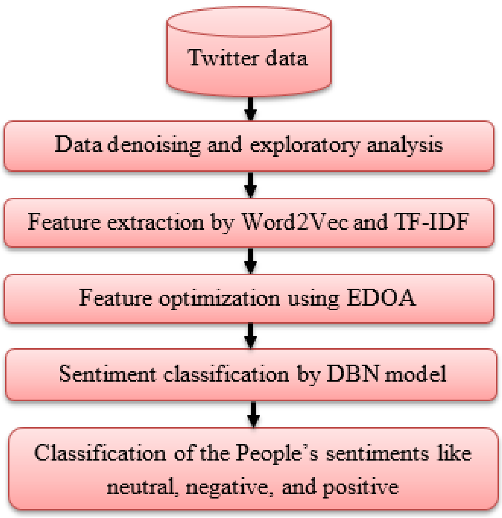

Figure. 1 Flowchart of the proposed study

by using LDA technique, extraction of features: Word2Vec and TF-IDF, feature optimization: EDOA, and sentiment analysis: DBN model. The flowchart of the proposed study is depicted in Fig. 1.

3.1 Data acquisition and denoising

The proposed EDOA-DBN's effectiveness is validated on two databases such as Indian political tweets 2019 and farmers protest tweets databases. Firstly, the Indian political tweets 2019 database is collected from 14th February to 16th May, which summed up to 1.4 TB. Each day, the most important 5000 tweets of both ruling and opposition parties are extracted which are used for experimental analysis. In addition, the farmers' protest tweets database consists of 18,000 tweets, which are acquired over 4 months. In this database, the tweets are acquired from 9061 users and stored in a Comma Separated Values (CSV) format and this database comprises four attributes such as Tweets, Username, Tweet ID, and Data/time.

After acquiring the tweets from the Indian political tweets 2019 and farmers protest tweets databases, the following denoising operations are performed.

1st step: Remove unnecessary spaces, tabs, and new lines from the acquired tweets.

2nd step: Remove hashtag symbols such as #India, #central government, #protest, etc., @users symbols, and uniform-resources-locators in the acquired tweets where it does not contribute to analysing the messages.

3rd step: Remove non-English characters, where the EDOA-DBN model focused only on the information analysis, especially related to English language characters.

4th step: Convert repeated words like "soooooo ambitious" to "so ambitious".

5th step: Convert emojis into short textual representations by utilizing the python emoji2 library.

6th step: Remove special characters, punctuations, and numbers from the databases, where it does not contribute to enhancing the sentiment classification performance.

3.2 Exploratory analysis

After denoising, the acquired data, the exploratory analysis is performed to achieve a more comprehensive representation of the databases. In this study, the exploratory analysis consists of two phases keyword trend analysis and topic modeling by LDA technique.

3.2.1. Keyword trend analysis

Initially, the keyword trend analysis is performed on the denoised databases to identify the frequently utilized words. According to both databases, the Indian people talk about Modi, media, Punjab, agriculture, policy, demonetization, petrol, crises, pandemic, covid-19, tractors, farmers, etc. In this study, the top 10 commonly used keywords on both databases are denoted in Table 1. Further, the common keywords in positive, negative, and neutral classes are stated in Fig. 2.

Table 1. Top 10 commonly utilized keywords

Key-words	Counts
Punjab	2900
Policy	1812
Modi	1786
Farmer	1445
Covid-19	1396
Crises	1208
Pandemic	1192
Demonetization	999
Petrol	982
Agriculture	894

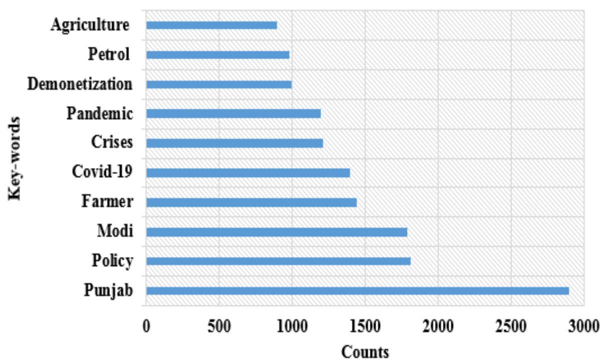

Figure. 2 Top 10 common keywords utilized in positive, negative, and neutral classes

Table 2. Top 10 keywords in the generated six topics

Topic 1	Modi	Crises	Petrol	India	Farmer	Agriculture	Policy	Punjab	Protest	Covid-19
Topic 2	Petrol	Diesel	Money	Modi	Corona	Protest	North	Policy	India	Agriculture
Topic 3	Leader	Modi	Media	Protest	Tractor	Policy	Stop	Farmer	Delhi	Punjab
Topic 4	Crises	Pandemic	Modi	Useless	Punjab	Government	Media	Policy	India	Agriculture
Topic 5	Protest	Covid-19	Punjab	Policy	Farmer	Agriculture	Delhi	India	Modi	Crises
Topic 6	Modi	Policy	India	Hike	GST	Punjab	Media	Farmer	Delhi	Foreign

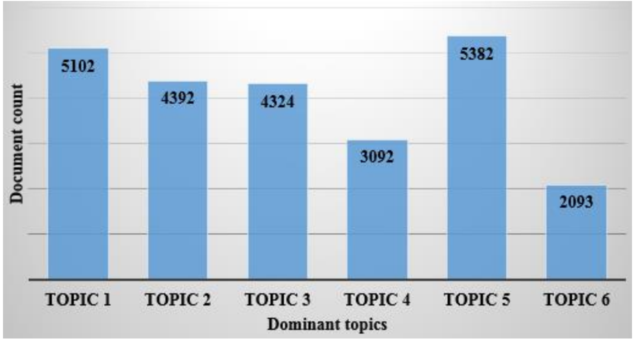

Figure. 3 Dominant topic's distributions

3.2.2. Topic modeling using the LDA technique

In this scenario, the topic distribution is performed utilizing the LDA technique to analyze the topics in the acquired databases. The LDA is an effective topic modeling technique and it selects the appropriate topics from the acquired databases. Further, each tweet is classified as either negative, neutral, or positive classes according to the concept. First, the LDA generates the topics for the relevant tweets or words as the Dirichlet distribution, and then, every tweet is described with the probability-distributions-functions pr, which is defined in Eqs. (1-3).

$$pr(\aleph|\pi) = \frac{\Gamma(\sum_{i=1}^{k}\pi_i)}{\prod_{i=1}^{k}\Gamma(\pi_i)} \aleph_1^{\pi_1-1} \ldots\ldots \aleph_k^{\pi_k-1} \quad (1)$$

$$pr(\aleph, x, y|\pi, \mu) = pr(\aleph|\pi) \prod_{n=1}^{N} pr(x_n|\aleph)\, pr(y_n|x_n, \beta) \quad (2)$$

$$pr(D|\pi, \mu) = \prod_{d=1}^{M} \int pr(\aleph_d|\pi) \times \left(\prod_{n=1}^{N_d} \sum_{x_{dn}} pr(x_{dn}|\aleph_d) pr(y_{dn}|x_{dn}, \mu)\right) d\aleph_d \quad (3)$$

Whereas y represents observed texts, $\aleph$ indicates document-level-topic-vectors, D represents Dirichlet distributions, π denotes Dirichlet parameters, Γ indicates gamma functions, N states the number of tweets, M indicates text reviews, x states topic assignments up to k^{th} texts and μ denote the number of topics which is six in this study and it is represented as word distributions. The top 10 keywords in the generated six topics are represented in table 2, and the dominant topic's distributions are indicated in Fig. 3 [26].

3.3 Feature extraction

After performing exploratory analysis, the extraction of features is performed utilizing TF-IDF and Word2Vec. In this study, the TF-IDF is a text vectorization method where it extracts discriminative features from the denoised tweets. The TF-IDF computes how regularly a term t appears in the tweets, which is mathematically stated in Eqs. (4-5).

$$TF = \frac{Number\ of\ times\ a\ term\ (t)\ appers\ in\ a\ tweet}{Total\ number\ of\ terms\ in\ a\ tweet} \quad (4)$$

$$IDF = log\frac{Total\ number\ of\ tweets}{Number\ of\ tweets\ with\ term\ (t)} \quad (5)$$

In addition, the Word2Vec technique is applied to effectively learn the relationship between the words in a large text corpus. Hence, the synonymous words are detected once the network is trained. In this research manuscript, the Word2Vec technique (Skip-gram) is implemented for predicting the contextual words for the target word. In the Skip-gram technique, the target word is input and the contextual words are output. The Skip-gram technique includes 2 major benefits in this research (i) requires limited memory compared to other Word2Vec techniques and (ii) an effective unsupervised learning model, which works well on any raw text data. Here, 5824 and 4928 feature vectors are extracted using TF-IDF and Skip-gram techniques on both Indian political tweets 2019 and farmers protest tweets databases. Then, the multidimensional feature vectors are fed to the EDOA for dimensionality reduction, which significantly decreases the computational time and system complexity of the proposed model [27-28].

3.4 Feature optimization

After extracting active feature vectors, the feature optimization is performed by proposing EDOA, which is an effective metaheuristic optimizer, and it mimics both static and dynamic behaviors of a dragonfly. The developed EDOA includes two major steps such as exploration and exploitation, which are modeled dynamically and statically to avoid enemy attacks and to better food search. Generally, the swarms have three behaviors like separation, alignment, and cohesion. In the developed EDOA, two more behaviors are included moving towards food and avoiding the enemy. These two behaviors increase the survival time of the swarms, and the dragonfly's location is updated in the search space based on two vectors such as step and position. In the developed EDOA, the step vector is considered as speed which represents the dragonfly's direction. In addition, the position vector is updated after calculating the step vector [29].

In the developed EDOA, the coefficients such as separation, inertia coefficient, cohesion, iteration number, enemy factor, food factor, and alignment are employed for performing both exploratory and exploitative behaviors. In the exploration process, the alignment coefficients are higher and the cohesion coefficients are limited. In the exploitation process, the alignment coefficients are lower and the cohesion coefficients are higher. In the existing DOA, the levy fight process is performed to improve the randomness, probabilistic behaviors, and discovery of dragonflies. However, the levy flight process enhances the DOA's effectiveness to a certain extent. Where, the step size control is contrary to the levy flight process, and the agent needs to go outside the space while considering a long step. To address the above-mentioned issue, a Brownian motion P_g is added to the traditional DOA to improve its probabilistic behaviors, the discovery of dragonflies and randomness. The Brownian motion P_g is defined in Eqs. (6) and (7).

$$P_g = \frac{1}{s\sqrt{2\pi}} exp\,(-\frac{(dimension-agents)^2}{2s^2}) \quad (6)$$

$$s = \sqrt{\frac{m_t}{m_s}}, \text{ and } m_s = 100 \times m_t \quad (7)$$

Where, m_s indicates the number of sudden motions, and $m_t = 0.01$ represents the motion time of an agent. The hyper-parameters setting of the EDOA are stated as follows: the number of iterations is 100, dimensions are the same as the extracted feature vectors, number of search agents is 50, and the search domain is [0 1]. The proposed EDOA chooses 2801 and 2730 feature vectors from the extracted feature vectors of Indian political tweets 2019 and farmers protest tweets databases, which are given to the DBN model for sentiment analysis.

3.5 Sentiment analysis

After optimal feature optimization, the sentiment classification is performed using the DBN model, which consists of restricted Boltzmann machines (RBMs) for classification. Though, the learned activation units of the 1st RBM are considered as the input for succeeding RBMs. Additionally, the DBN is an undirected graphical model where the hidden units are integrated with the visible variables by utilizing undirected weights. Related to other models, the DBNs are constrained, because there is no connection between the hidden units and visible variables. Whereas, the probability distribution p_d of hidden units h, visible variables m, and the energy function $E(m,h;\theta)$ are defined in Eq. (8).

$$-\log p_d(m,h)\,\alpha\, E(m,h;\theta) = -\sum_{i=1}^{|V|}\sum_{j=1}^{|Q|} w_{ij} m_i h_j - \sum_{i=1}^{|V|} b_i\, m_i - \sum_{j=1}^{|Q|} a_j h_j \quad (8)$$

Where, α indicates learning rate, a_j and b_i are bias, $\theta = (w,b,a)$ specifies parameter set, and w_{ij} represents symmetric weight between the visual variables of m . Additionally, $|Q|$ represents the number of hidden layers and $|V|$ denotes the number of visible layers. In the DBN model, the hidden units h and conditional probability distribution of visible variables m are stated in Eqs. (9-10).

$$p_d(h_j|m;\theta) = sigm \sum_{i=1}^{|V|} w_{ij} m_i + a_j \quad (9)$$

$$p_d(m_i|h;\theta) = sigm \sum_{j=1}^{|Q|} w_{ij} h_i + b_j \quad (10)$$

Where θ is a parameter, and it is learned by contrastive divergence and $sigm(M) = (\frac{1}{1+e^{-m}})$ represents sigmoid-activation-function. The parameter θ is obtained using RBM in the DBN model, which is determined as $p_d(h|\theta)$ and $p_d(m|h,\theta)$ [30]. The probability of generating a new visible variable is mathematically denoted in Eq. (11).

$$p_d(m) = \sum_h p_d(h|\theta) p_d(m|h,\theta) \quad (11)$$

Table 3. Simulation outcome of the EDOA-DBN on the Indian political tweets 2019 database

Optimizers	Classifiers	Precision (%)	Accuracy (%)	F1-score (%)	Recall (%)	Specificity (%)
GOA	Autoencoder	88.10	88.90	90.80	88.44	88.80
FOA		89.28	90.18	91.22	90.70	90.10
DOA		90.38	91.32	91.70	92.80	91.04
EDOA		91.28	92.10	92.92	92.98	92.88
GOA	LSTM	89.22	90.02	91.18	92.10	92.45
FOA		92.28	91.98	93.20	93.45	92.62
DOA		93.42	92.80	93.92	93.80	93.94
EDOA		94.84	93.94	94.78	94.50	94.60
GOA	Bi-LSTM	92.44	92.66	93.04	94.90	93.90
FOA		94.84	95.06	94.98	95.88	94.10
DOA		95.98	95.80	95.16	97.98	95.40
EDOA		97.97	96.72	96.88	98.90	96.68
GOA	DBN	93.45	94.38	94.30	96.62	96.04
FOA		95.40	97.54	96.50	98.60	97.76
DOA		97.68	98.92	98.76	99.03	98.88
EDOA		99.02	99.22	99.04	99.12	99.30

After learning θ from an RBM, the term $p_d(m|h,\theta)$ is kept, and further, the term $p_d(h|\theta)$ is exchanged using consecutive RBMs, which treat the prior RBM hidden layer as a visible variable. The hyper-parameters setting of the DBN model is determined as follows: transfer function is the sigmoid activation function, the learning rate is 0.01, the drop-out rate is 0.1, the batch size is 0.5, the initial momentum is 0.5, maximum iteration is 100, and final momentum is 0.9. Whereas, the experimental outcomes of the developed EDOA-DBN model are stated in the next section.

4. Simulation results

The developed EDOA-DBN model is simulated by using a python software environment with a machine learning library named sci-kit learn. In this scenario, the EDOA-DBN model is analyzed on a system with a windows 10 (64-bit) operating system, Intel core i9 processor, 4 TB hard disk, and 128 GB RAM. Further, the proposed model's effectiveness is validated by utilizing different evaluation metrics like precision, accuracy, F1-score, recall, and specificity on the Indian political tweets 2019 and farmers' protest tweets databases. In this scenario, the accuracy is determined as the ratio of correctly classified results to the total results and the f1-score is determined as the harmonic mean of recall and precision value. Further, the precision is specified as the ratio of correctly classified results to the total classified positive results and the recall is stated as the ratio of correctly classified results to the actual total positive results. In addition, specificity is defined as the ratio of correctly predicted negative results to the actual total negative results. The undertaken evaluation metrics are mathematically indicated in Eqs. (12-16).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{12}$$

$$Recall = \frac{TP}{TP+FN} \times 100 \tag{13}$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \tag{14}$$

$$Precision = \frac{TP}{TP+FP} \times 100 \tag{15}$$

$$F1-score = \frac{2TP}{FP+2TP+FN} \times 100 \tag{16}$$

Where true negative (TN) states actual negative predictions, false negative (FN) represents incorrect negative predictions, true positive (TP) represents actual positive predictions, and false positive (FP) states incorrect positive predictions.

4.1 Performance analysis of the Indian political tweets 2019 database

In this segment, the efficacy of the developed EDOA-DBN methods is analyzed on the Indian political tweets 2019 database in terms of precision, accuracy, f1-score, recall and specificity. As specified in Table 3, the experimental examination is performed with numerous classifiers: autoencoder, long short term memory (LSTM), Bidirectional LSTM (Bi-LSTM) and DBN, and optimizers such as DOA, EDOA, grasshopper optimization algorithm (GOA), and fire-fly optimization algorithm (FOA). By investigating Table 3, the combination: EDOA-DBNs achieved maximum classification results in the

Received: October 18, 2022. Revised: December 16, 2022.

Table 4. The experimental outcome of the EDOA-DBN model on the farmers' protest tweets database

Optimizers	Classifiers	Precision (%)	Accuracy (%)	F1-score (%)	Recall (%)	Specificity (%)
GOA	Autoencoder	90.12	90.60	90.18	90.40	90.70
FOA		90.64	91.12	92.28	92.10	92.18
DOA		91.70	92.30	93.70	93.22	93.30
EDOA		92.30	92.80	94.02	93.98	93.86
GOA	LSTM	90.24	90.01	90.18	91.18	92.40
FOA		91.90	92.03	93.83	93.15	93.68
DOA		92.40	93.25	94.42	93.90	93.99
EDOA		93.66	94.80	94.78	94.20	94.40
GOA	Bi-LSTM	93.80	93.67	94.09	93.98	92.92
FOA		94.04	96.06	96.82	95.18	94.08
DOA		95.12	96.82	96.96	96.68	95.46
EDOA		96.80	97.78	97.80	97.56	96.70
GOA	DBN	95.40	95.44	97.20	94.68	95.96
FOA		97.12	96.58	97.56	95.66	97.36
DOA		98.14	98.12	98.16	97.99	98.16
EDOA		98.72	98.83	98.94	98.54	98.80

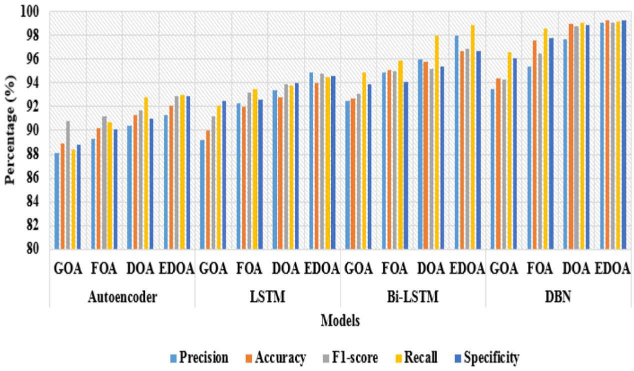

Figure. 4 Graphical presentation of the EDOA-DBN model on the Indian political tweets 2019 database

Twitter sentiment analysis compared to other combinations. As represented in Table 3, the developed EDOA-DBN model obtained 99.02% of precision, 99.22% of accuracy, 99.04% of F1-score, 99.12% of recall value, and 99.30% of specificity in the Indian political tweets 2019 database. The graphical presentation of the EDOA-DBN model on the Indian political tweets 2019 database is denoted in Fig. 4. The multiple- layers in DBN allow models to become effective to perform intensive computational tasks, and learning the complex feature vectors on the larger unstructured databases.

4.2 Performance analysis of the farmer's protest tweets database

Similar to Table 3, the experimental outcome of the EDOA-DBN model on the farmers' protest tweets database is denoted in Table 4, and it is validated in light of precision, accuracy, F1-score, recall, and specificity. By inspecting Table 4, the EDOA-DBN model has achieved 98.72% of precision, 98.83% of accuracy, 98.94% of F1-score, 98.54% of recall, and 98.80% of specificity in the farmers' protest tweets database. Though, the achieved simulation results are maximumly related to the comparative optimizers and classifiers like Autoencoder, LSTM, Bi-LSTM, DOA, GOA, and FOA. A graphical representation of the EDOA-DBN model on the farmers' protests tweets database is indicated in Fig. 5. In this study, the EDOA is developed for selecting the optimum feature values from the extracted feature values. This process helps in reducing the model's complexity to linear, and it is computed based on the order of magnitude and input size. Additionally, the computational time of the EDOA-DBN model is 45.21, and 33.36 seconds for Indian political tweets 2019 and farmers protest tweets databases, which are minimum related to other combinations, and the system complexity is linear, due to effective feature optimization.

On the other hand, the proposed EDOA-DBN model's efficacy is analyzed with numerous k-fold cross-validations such as 10-folds, 5-folds and 3-folds. By viewing table 5, the EDOA-DBN model has achieved higher classification results in 5-folds (80:20% training and testing of data) compared to other cross-fold validations on both databases. The cross-fold validation helps in better usage of acquired data, and also delivers more information about the proposed models' performance. A graphical presentation of the EDOA-DBN model with different k-fold cross-validations is indicated in Fig. 6.

4.3 Comparative analysis

The comparative results of the prior models and the proposed EDOA-DBN model are stated in table 6. Joseph, [13] used a decision tree classifier for

Table 5. The experimental outcome of the EDOA-DBN model with different k-fold cross validations

Indian political tweets 2019 database					
Cross-folds	**Precision (%)**	**Accuracy (%)**	**F1-score (%)**	**Recall (%)**	**Specificity (%)**
3-folds	98.20	97.64	98.02	98.34	98.70
5-folds	99.02	99.22	99.04	99.12	99.30
10-folds	97.43	92.20	94.37	96.05	98.88
Farmers protest tweets database					
Cross-folds	Precision (%)	Accuracy (%)	F1-score (%)	Recall (%)	Specificity (%)
3-folds	95.40	96.78	97.47	98.03	96.82
5-folds	98.72	98.83	98.94	98.54	98.80
10-folds	93.20	93.92	94.30	95.40	95.38

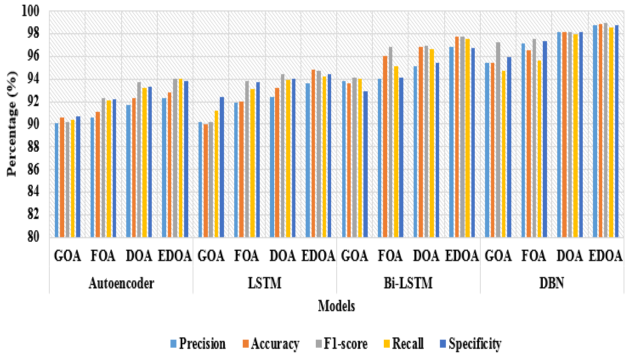

Figure. 5 Graphical presentation of the EDOA-DBN model on the farmers' protest tweets database

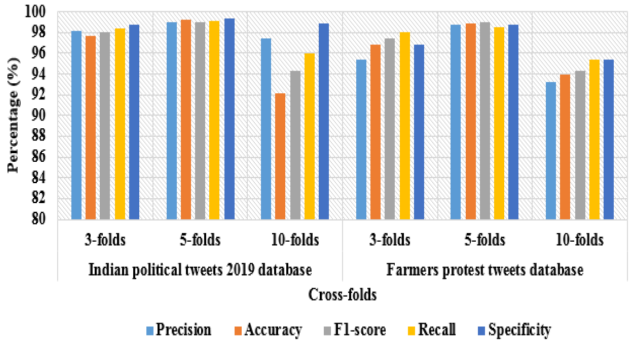

Figure. 6 Graphical presentation of the EDOA-DBN model with different k-fold cross-validations

Table 6. Comparative results of the prior models and the proposed EDOA-DBN model

Models	Database	Accuracy (%)
Decision tree [13]	Indian political tweets 2019	97
Random forest [14]	Farmers protest tweets	96.60
EDOA-DBN	Indian political tweets 2019	99.22
	Farmers protest tweets	98.83

predicting sentiments of the India general election 2019. The experimental evaluations demonstrated that the implemented model attained 97% of classification accuracy on the Indian political tweets 2019 database [31]. In addition, Neogi et al. [14] integrated Bag of words and TF-IDF techniques to extract features from the input data. The extracted features were classified by utilizing many machine learning classifiers such as SVM, decision tree, Naïve Bayes and random forest. As indicated in the resulting phase, the random forest has achieved a higher classification accuracy of 96.6% on the farmers' protest tweets database.

Related to these existing papers, the proposed EDOA-DBN model achieved higher classification results on both databases, as mentioned in Table 6. In addition, the EDOA-DBN model has limited computational time and system complexity, which are the major problems addressed in the literature section.

5. Conclusion

In this manuscript, the experiments are performed to find the general opinions (sentiments) of the Indian people in the events like farmers' protests and the 2019 parliament election. Firstly, the Indian people's tweets are acquired from Indian political tweets 2019 and farmers protest tweets databases and further, the basic data denoising operations are accomplished for improving the acquired twitter data quality. Next, the exploratory analysis is performed like keyword-trend analysis and topic-modeling by using the LDA technique. In addition, the discriminative feature vectors are extracted from the denoised data utilizing Skip-Gram and TF-IDF techniques, and then, the dimensions of the extracted feature vectors are diminished by proposing EDOA that helps in reducing the computational time and system complexity. Finally, the selected feature values are given to the DBN model to classify the sentiments of the people such as negative, positive, and neutral. The experimental evaluations demonstrated that the EDOA-DBN model has attained significant performance in sentiment classification related to other classifiers and optimizers in light of precision, accuracy, F1-score, recall, and specificity. As depicted in the resulting phase, the EDOA-DBN model has obtained 99.22% and 98.83% of accuracy on the Indian political tweets 2019 and farmers protest tweets databases. As a future extension, an ensemble deep learning model can be incorporated with the EDOA-DBN to further improve classification performance on other larger unstructured databases.

Parameter	Notations
pr	Probability-distributions-functions
y	Observed texts
$\aleph$	Document-level-topic-vectors
D	Dirichlet distributions
π	Dirichlet parameters
Γ	Gamma functions
N	Number of tweets
M	Text reviews
x	Topic assignments up to k^{th} texts
μ	Number of topics
m_s	Number of sudden motions
m_t	Motion time of an agent
P_g	Brownian motion
p_d	Probability distribution
h	Hidden units
α	Learning rate
a_j and b_i	Bias
w_{ij}	Symmetric weight between the visual variables
$\|Q\|$	Number of hidden layers
$\|V\|$	Number of visible layers
$sigm(M)$	Sigmoid-activation-function
TP	True positive
TN	True negative
FP	False positive
FN	False negative

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper background work, conceptualization, methodology, dataset collection, implementation, result analysis and comparison, preparing and editing draft, visualization have been done by first author. The supervision, review of work and project administration, have been done by second and third author.

References

[1] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers", *Future Generation Computer Systems*, Vol. 106, pp. 92-104, 2020.

[2] A. Sharma and U. Ghose, "Sentimental analysis of twitter data with respect to general elections in India", Proce*dia Computer Scien*ce, Vol. 173, pp. 325-334, 2020.

[3] P. Chauhan, N. Sharma and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 2, pp. 2601-2627, 2021.

[4] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts", *Mathematical and Computational Applications*, Vol. 23, No. 1, p. 11, 2018.

[5] B. Bansal, and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets", *Procedia Computer Science*, Vol. 135, pp. 346-353, 2018.

[6] T. A. Neyazi, and R. Schroeder, "Was the 2019 Indian Election won by Digital media?", *The Communication Review*, Vol. 24, No. 2, pp. 87-106, 2021.

[7] S. Datta, and S. Chakrabarti, "Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer", *Sādhanā*, Vol. 46, No. 2, pp. 1-23, 2021.

[8] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM model for document-level sentiment analysis", *Machine Learning and Knowledge Extraction*, Vol. 1, No. 3, pp. 832-847, 2019.

[9] K. Shazly, M. Eid, and H. Salem, "An Efficient Hybrid Approach for Twitter Sentiment Analysis based on Bidirectional Recurrent Neural Networks", *International Journal of Computer Applications*, Vol. 175, No. 17, pp. 32-36, 2020.

[10] H. Naz, S. Ahuja, and D. Kumar, "DT-FNN based effective hybrid classification scheme for twitter sentiment analysis", *Multimedia Tools and Applications*, Vol. 80, No. 8, pp. 11443-11458, 2021.

[11] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. S. Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model", *Computational Intelligence*, Vol. 37, No. 1, pp. 409-434, 2021.

[12] S. Tam, R. B. Said, and Ö. Ö. Tanriöver, "A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification", *IEEE Access*, Vol. 9, pp. 41283-41293, 2021.

[13] F. J. J. Joseph, "Twitter based outcome predictions of 2019 Indian general elections using decision tree", In: *Proc. of 2019 4th International Conference on Information Technology (InCIT)*, pp. 50-53, 2019.

[14] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data", *International Journal of Information Management Data Insights*, Vol. 1, No. 2, p. 100019, 2021.

[15] D. Tiwari and B. Nagpal, "KEAHT: A Knowledge-Enriched Attention-Based Hybrid Transformer Model for Social Sentiment Analysis", *New Generation Computing*, pp. 1-38, 2022.

[16] S. S. Shekhawat, S. Shringi, and H. Sharma, "Twitter sentiment analysis using hybrid Spider Monkey optimization method", *Evolutionary Intelligence*, Vol. 14, No. 3, pp. 1307-1316, 2021.

[17] M. A. Hassonah, R. A. Sayyed, A. Rodan, A. Z. Ala'M, I. Aljarah, and H. Faris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter", *Knowledge-Based Systems*, Vol. 192, p. 105353, 2020.

[18] A. Kumar and A. Jaiswal, "Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter", *Multimedia Tools and Applications*, Vol. 78, No. 20, pp. 29529-29553, 2019.

[19] M. Wang and G. Hu, "A Novel Method for Twitter Sentiment Analysis Based on Attentional-Graph Neural Network", *Information*, Vol. 11, No. 2, p. 92, 2020.

[20] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method", *Information Processing & Management*, Vol. 53, No. 4, pp. 764-779, 2017.

[21] P.D. Kusuma and M. Kallista, "Stochastic Komodo Algorithm", *International Journal of Intelligence Engineering and Systems*, Vol. 15, No. 4, pp. 156-166, 2022, doi: 10.22266/ijies2022.0831.15.

[22] M. Dehghani, Z. Montazeri, A. Dehghani, R. A. R. Mendoza, H. Samet, J. M. Guerrero, and G. Dhiman, "MLO: Multi leader optimizer", *International Journal of Intelligence Engineering and Systems*, Vol. 13, pp. 364-373, 2020, doi: 10.22266/ijies2020.1231.32.

[23] F. A. Zeidabadi, M. Dehghani, and O. P. Malik, "TIMBO: Three influential members based optimizer", *International Journal of Intelligent Engineering and Systems*, Vol.14, No. 5, pp. 121-128, 2021, doi: 10.22266/ijies2021.1031.12.

[24] F. A. Zeidabadi, M. Dehghani, and O. P. Malik, "RSLBO: Random Selected Leader Based Optimizer", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 529-538, 2021, doi: 10.22266/ijies2021.1031.46.

[25] M. Sumanl, V. P. Sakthivel, and P. D. Sathya, "Squirrel search optimizer: nature inspired metaheuristic strategy for solving disparate economic dispatch problems", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 5, pp. 111-121, 2020, doi: 10.22266/ijies2020.1031.11.

[26] M. Pavlinek and V. Podgorelec, "Text classification method based on self-training and LDA topic models", *Expert Systems with Applications*, Vol. 80, pp. 83-93, 2017.

[27] S. W. Kim, and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes", *Human-centric Computing and Information Sciences*, Vol. 9, No. 1, pp. 1-21, 2019.

[28] P. T. Hung and K. Yamanishi, "Word2vec skip-gram dimensionality selection via sequential normalized maximum likelihood", *Entropy*, Vol. 23, No. 8, p. 997, 2021.

[29] M. Mafarja, I. Aljarah, A. A. Heidari, H. Faris, P. F. Viger, X. Li, and S. Mirjalili, "Binary dragonfly optimization for feature selection using time-varying transfer functions", *Knowledge-Based Systems*, Vol. 161, pp. 185-204, 2018.

[30] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network", *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 10, pp. 7319-7327, 2020.

[31] Database links:
https://www.kaggle.com/datasets/codesagar/indian-political-tweets-2019-feb-to-may-sample
https://www.kaggle.com/datasets/prathamsharma123/farmers-protest-tweets-dataset-csv