

A Novel Ant Colony Based DBN Framework to Analyze the Drug Reviews

Nazia Tazeen^{1*}, K. Sandhya Rani²

¹Department of Computer Science & Engineering, SPMVV Tirupati, India

E-mail: *naziabaseer@gmail.com

²Department of Computer Science, SPMVV Tirupati, India.

E-mail: sandhyaranikasireddy@yahoo.co.in

Received: 27 September 2021; Revised: 25 October 2021; Accepted: 03 November 2021; Published: 08 December 2021

Abstract: Nowadays, big data is directing the entire advanced world with its function and applications. Moreover, to make better decisions from the ever emerging big data belonging to the respective organizations, deep learning (DL) models are required. DL is also widely used in the sentiment classification tasks considering data from social networks. Furthermore, sentiment classification signifies the best way to analyze the big data and make decisions accordingly. Analyzing the sentiments from big data applications is quite challenging task and also requires more time for the execution process. Therefore, to analyze and classify big data emerging from social networks in a better way, DL models are utilized. DL techniques are being used among the researchers to get high end results. A novel Ant Colony-based Deep Belief Neural Network (AC-DBN) framework is proposed in this research. Drug review tweets are opted to perform sentiment classification by using the proposed framework in python environment. A model fitness function is initiated in the DL framework and is observed that it is attaining high accuracy with low computation time. Additionally, the obtained results attained from the proposed framework are validated with existing methods for evaluating the efficiency of the proposed AC-DBN approach.

Index Terms: Ant colony optimization, opinion specification, big data, sentiment classification, deep learning, Deep Belief Neural framework.

1. Introduction

In the present data age, communicating thoughts and reviews in informal ways has led to huge production of data from big organizations [1]. Vast amount of data named as big data emerged from almost all the organizations due to increase in the usage of computers with internet and as everything moved to online these days and is found in unstructured form [2]. To convert the big data from unstructured form to structured form in order to make best use of big data, it requires keen data preprocessing [3]. Natural Language Processing (NLP) is carried out to process the big data and grab the sentiments from the natural language used by the people of certain organizations or offices or institutions [4,5]. Sentiment is an opinion or attitude that is influenced through emotion, excitement, etc. Moreover, sentiment analysis is the process of categorizing the polarity of the collected dataset in two ways that are positive and negative [6]. In addition, sentiment datasets are derived from many terms like political reviews, customer reviews, drug reviews, movie reviews, etc. the investigators have tuned the review dataset based on the negative, positive and neutral. Here, the polarity score is automatically labeled based on the sentiment score. [7]. Thus, the NLP approach can assess the assessment esteem on the whole languages. Machine learning (ML) is a part of Artificial Intelligence, that enables a machine to naturally learn from the given data and further make decisions accordingly. ML techniques improve from the past models without being changed explicitly [8]. The data collection process to perform sentiment analysis is demonstrated in fig.1.

Data obtaining from social networks has to be cleaned well before performing sentiment analysis [9]. Many researchers have demonstrated diverse ways to collect the data emerging from social networks and perform proper preprocessing in order to analyze the sentiments and make proper decision accordingly by classifying the sentiments based on polarity scores [10]. To accomplish sentiment classification tasks, suitable machine learning techniques are employed. One of the research works exposed the extremist's sentiment classification based on some reviews from social media [11]. With the headway in ML approaches, researchers need to use proper methods to get desired results [12]. To admire the feeling of every content NLP, sentiment, or opinion, either feeling esteem assessment is used [13]. Likewise, the sentiment assessment is done with the utilization of ML or DL approaches [14]. Also, in online business assessing the client audit is the critical interaction or acts as best feedback to improve the online business [15]. Since dependent

on the client surveys, just online business is overseen [16]. In web-based showcasing, every single item audit is more significant; on the grounds that dependent on the item surveys, just sold rating is expanded [17].

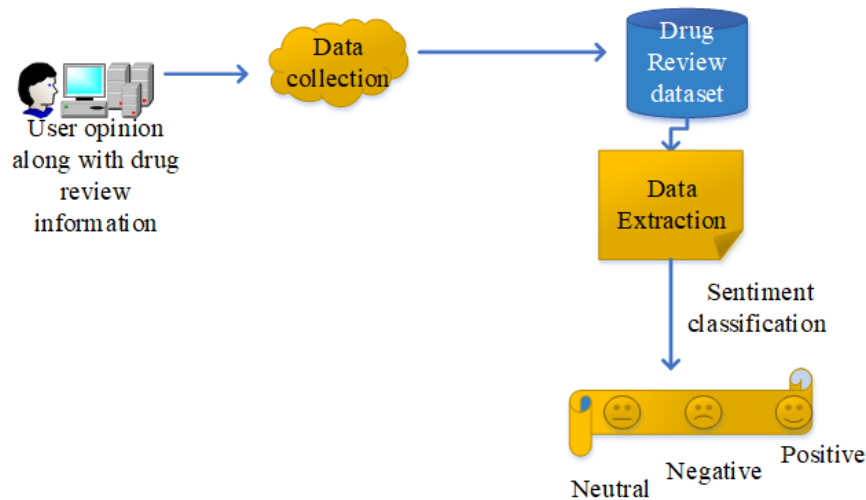


Fig.1. Sentiment analysis

Recently, numerous models [18,19] are developed to enhance the classification process of the sentiment analysis likewise drug recommendation system [20], multi-task learning model [21], and supervised scheme [22]. However, these existing methods are affected by many limitations that reduce the performance of the classification process. Moreover, the existing techniques take high computation time and are quite complex to perform the sentiment categorization on huge data. Therefore, this research work proposes a novel ant colony DBN based framework to classify the drug reviews using DL techniques. The main objective of this research is to classify the drug reviews by using the proposed framework with high accuracy. Ant colony optimization is used to enhance the accuracy in classifying drug reviews based on the fitness function of ant colony optimization algorithm. It selects the best features in order to classify sentiments from drug reviews with high accuracy. There were much research works available to classify sentiments on different data but none of the works focused on classifying drug reviews using ant colony optimization through an exclusive framework. Therefore, the proposed AC-DBN has achieved highest accuracy of 97.2%, with precision value of 95%, recall 94.8% compared to the existing works.

Key processes of current research work are explained as follows:

- Twitter-based drug review dataset is chosen for this research.
- Initially, the drug review dataset is trained by the proposed AC-DBN framework.
- The performance of the proposed method is implemented in the Python framework.
- Collected dataset is fed to the pre-processing layer; here, in this stage special characters, stop words, and repeated words are removed, and text is summarized.
- Sentiment analysis is enhanced by the fitness of Ant Colony model.
- Then, this filtered and summarized data is given to the DBN classification layer to estimate the sentiment analysis of each and every dataset.
- Consequently, the developed model is validated in terms of accuracy, precision, F-measure, recall, and opinion classification.
- Also, the attained results are compared with existing approaches and show the efficiency of the proposed AC-DBN model.

The structure of this article is designed as section 2 details associated literature; consequently, the system frame and issues definition are described in section 3, section 4 detailed a novel approach, then the progress of developed replica is explained in section 5, and the arguments are concluded in section 6.

2. Related Works

Some of the recent related work based on the aspect-based sentiment classifications is detailed as follows.

Sentiment classification and opinion mining are the most important fundamental research in social networks. Moreover, Twitter is a trendy social media that has to communicate and interact with other users in online. Sentiment classification is one of the most efficient investigations in Natural Language Processing (NLP). Drug-based reviews are provided more information to enhance pharmacovigilance coordination. Recently, several strategies are proposed to

improve the sentiment classification efficiently. Nevertheless, only few DL techniques are applied to the sentiment analysis of drug reviews. In modern days, the enhancement of innovative computer technologies has generated the textual content website. Moreover, the textual content means that health care reviews and patients wrote medical. Consequently, this textual content is very efficient and valuable in social media networks. However, these reviews cannot deal with patient treatment and doctors for their satisfaction. In order to address this issue, Basiri et al. [23] proposed two deep fusion models based on three-way decision theory to analyze the drug reviews. The first fusion model, 3-way fusion of one deep model with a traditional learning algorithm (3W1DT) developed using a deep learning method as a primary classifier and a traditional learning method as the secondary method that is used when the confidence of the deep method during classification of test samples is low. In the second proposed deep fusion model, 3-way fusion of three deep models with a traditional model (3W3DT), three deep and one traditional model are trained on the entire training data, and each classifies the test sample individually. Then, the most confident classifier is selected to classify the test drug review. Our results on the reviews based on Drugs.com dataset showed 3W1DT and 3W3DT methods outperformed the traditional and deep learning methods by 4% and the 3W3DT outperformed 3W1DT by 2% in terms of accuracy and F1-measure.

Isabel et al. [24] presented a benchmark comparison of various deep learning architectures such as Convolutional Neural Networks (CNN) and long short-term memory (LSTM) recurrent neural networks. Several combinations of these models were analyzed with effect to different pre-trained word embedding models. Bidirectional Encoder Representations from Transformers (BERT) with a Bi-LSTM for the sentiment analysis of drug reviews was introduced and the experiments show that the usage of BERT obtains the best results, but with a very high training time. On the other hand, CNN achieves acceptable results while requiring less training time.

Hossain et al. [25] made a detailed examination of recommendation systems for ML-based frameworks to analyze the sentiment classification for the drug review dataset. The drug recommender system framework is design and implements to specify the present sentiment value in general healthcare datasets. Moreover, to generate the ratings and classify the drug selection sentiment polarity is calculated. Computation time is high when compared to other approaches.

Pethalakshmi et al. [26] have proposed a novel hybrid ML-based bi-objective optimization algorithm to extract and select the features from bulk amounts of dataset. Moreover, the performance of the proposed approach is compared with existing techniques. However, the developed paradigm is not suitable for the small module. NaziaTazeen et al. [27] have proposed an effective framework for text categorization strategy and topic modeling replica to solve the diagnostic problems present in the dataset. Moreover, to attain the automated classification, the developed replica is used to enhance Dirichlet allocation for sentiment analysis and topic modeling. Also, this technique analyzes sentiment scores on the drug review dataset. The review analysis of online items will improve product quality and improves choices of the buyers [30]. Moreover, in some cases ML model failed to predict the deep emotion measure of each tweet; also, opinion specification accuracy is not improved [31].

This study was performed to predict student program completion using the Naïve Bayes classifier technique. This dataset was pre-processed, cleansed, transformed, and balanced before constructing the model. The feature selection technique was used to filter and evaluate the significance of each factor. The significant variables assessed by the feature selection technique (Weight by Correlation) were the final parameters in creating the model. The Naïve Bayes classifier was applied to predict the students' completion using the 70:30 ratios for training and testing dataset distribution. The significant attributes identified in correlation analysis spitted into 70% training data or 447 records and 30% testing data or 191 records. There were 84 out of 191 data samples, or 44% of students were predicted to complete the program. On the other hand, 107 out of 191 data samples, or 56%, were predicted as not completing the program. The accuracy values performed an 84% rating with 80.46% class precision, and 83.33% class recall in the testing dataset (n=191). The outcomes of this study have a significant impact on HEIs, particularly on college completion rates [34].

In this research, four algorithms were used to compare with the proposed novel approach. Three algorithms Naive Bayes, KNN and Decision Tree are classical algorithms and one AdaBoost is Boosting concept. The comparison shows on five algorithms that are applied on US-Airlines data. All implemented algorithms provide different scores. Ensemble methods provide more accuracy than classical algorithms as (Decision Tree 63%, KNeighbors 67%, Naive Bayes 69%, and AdaBoost 74%). And finally, our novel method outperforms than all of others and provides 76% accuracy [35].

In this study, a stacked ensemble method is proposed for sentiment analysis, which systematically combines six feature extraction methods and three classifiers. The proposed method obtained cross-validation accuracies of 89.6%, 90.7% and 67.2% on large movie, Turkish movie and SemEval-2017 datasets, respectively, outperforming the other classifiers. The accuracy improvements were shown to be statistically significant at the 99% confidence level by performing a Z-test [36].

From the literature survey, it is observed that very few attempts have been made to apply deep learning to sentiment analysis of drug reviews. Machine learning techniques need high computation to extract sentiments from huge data and are also prone to errors. ML techniques often yield good results on limited data, but fail to get high results on big data. Therefore, to overcome these issues, we thought of carrying our research by proposing a novel framework using ant colony optimization based on deep belief neural network to classify drug reviews.

Table 1. Summary of related works

Author	Method	Merits	Demerits
Basiri et al [23]	DL	It can easily identify the classification.	Processing time is high
Isabel et al [24]	CNN	Reduced the training and testing issues.	The detection of duplication error is large.
Hossain et al [25]	generation and recommendation for ML-based frameworks	Enhanced the error detection rate.	Large amount of data is not used.
Pethalakshmi et al [26]	hybrid ML approach	Reduced the detection of duplicate errors.	Only applicable for the small module.
NaziaTazeen et al [27]	text categorization strategy and topic modeling replica	More efficiency	Hybrid formation can give the better performance

3. System Model and Problem Statement

This section explains about the problem of sentiment analysis. Moreover, drug review dataset is analyzed whether it is negative, positive, or neutral opinion. For the drug review dataset conveying both negative and positive opinions, whichever is the efficient sentiment classification should be selected. Moreover, managing big data is the most important challengeable task in the NLP techniques.

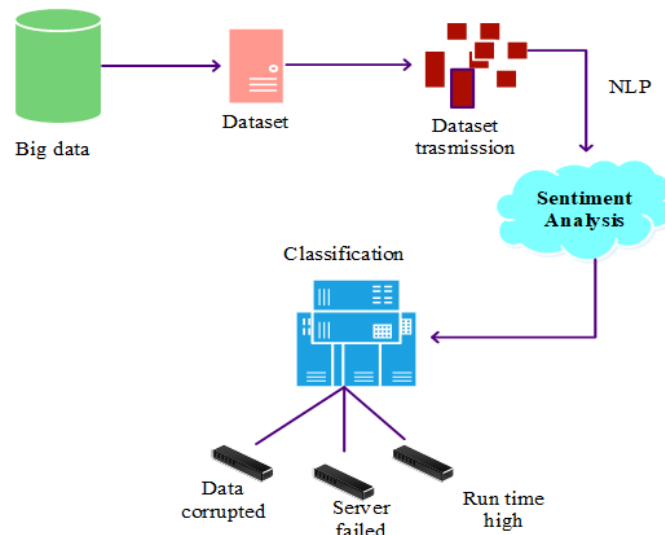


Fig.2. System model

Also, NLP approaches are used by the reviewers to get client satisfaction or customer opinion for every application based on the big data analytics. The system model and problem statement are shown in fig.2. In sentiment classification, human communication language is very complex to make the machines comprehend. So that, the proposed investigation article aimed to create a sentiment analysis framework based on the drug reviews.

3.1. Data set

Dataset is collected from Kaggle which contains drug review Twitter tweets. The valuable count is the aspect terms that embraces the importance of the amount of customers who share and review. Furthermore, dataset description is demonstrated in table.2.

Table 2. Dataset description

Sl.no	Attribute	Type of data	Explanation
1	condition	Categorical	Name of the condition
2	Ratings	Numerical arrangement	5 star ratings, 10 star ratings etc.
3	Drug name	Categorical	Name of the drug
4	Review	Text system	Patients reviews based on the drug

Here, let us consider five drug review datasets, and size of the dataset is 26.36MB [32]¹, which are elaborated in table.3.

Table 3. Drug review dataset

Sl. no	Drug review dataset	Sentiment word	Stop words	Sentiment score		
				Positive	Negative	neutral
1	In this drug has a bad smell as well as taste	bad	This, has, as, this	-	-1	-
2	Drug consumption has improved human health	improved	has	1	-	-
3	In this drug makes vomit problem	problem	This, in	-	-1	-
4	I am not satisfied with this drug	Not, am	This	-	-1	-
5	The drug is manufactured in more countries	-	is, in, the	-	-	0
6	Several dug has mixed with chemicals makes the unwanted illness	Unwanted	Has, the, with	-	-1	-

Here, the initial step of the proposed approach is preprocessing layer. It is one of the important stages to perform a particular task. Initially, the collected datasets are converted into lowercase words with the help of python library function. Moreover, the stop words are continuously present in the collected test; hence, it conveys meaningless information. Therefore, the tests are extracted with the help of the proposed approach. Consequently, preprocessing stage is demonstrated in table 4.

Table 4. Preprocessed dataset

Sl.no	Original dataset	Preprocessed dataset
1	In this drug has a bad smell as well as taste	'Drug' 'bad' 'smell' 'well' 'taste'
2	Drug consumption has improved human health	'drug' 'consumption' 'improved' 'human' 'health'
3	In this drug makes vomit problem	'drug' 'makes' 'vomit' 'problem'
4	I am not satisfied with this drug	'not' 'satisfied' 'drug'
5	The drug is manufactured in more countries	'drug' 'manufactured' 'more' 'countries'
6	Several drugs have mixed with chemicals makes the unwanted illness	'Several' 'drug' 'mixed' 'chemicals' 'makes' 'unwanted' 'illness'

Let us consider 20 types of stop words based on the drug review dataset, and their counts are demonstrated in fig.3.

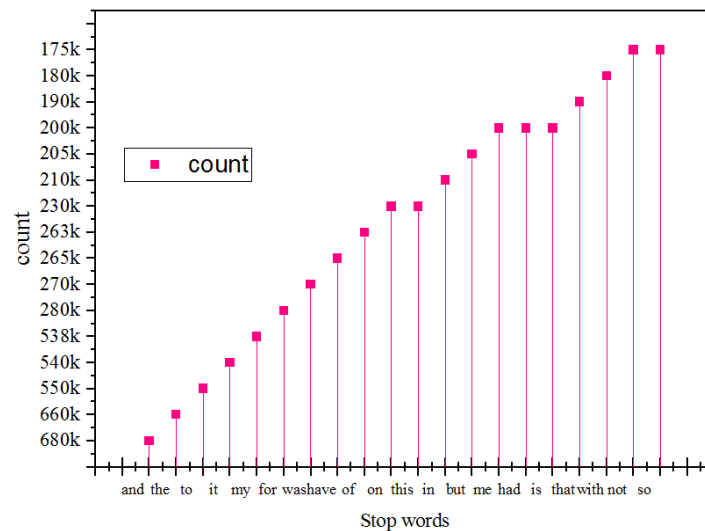


Fig.3. Dataset stop words and count representation

3.2. Deep Belief neural (DBN) framework

DBN framework is one kind of DL network that is weighted by the several layers based on the Restricted Boltzmann Machines (RBM). Moreover, the DBN framework contains two layers hidden layer and visible layer. Here,

¹ Source and size of the dataset in Kaggle link

the input datasets are trained to extract the deep hierarchical demonstration. After that RBM layer is trained the demonstration of the prior hidden layer input and next hidden layer input. Consequently, the structure of DBN framework is illustrated in fig.4.

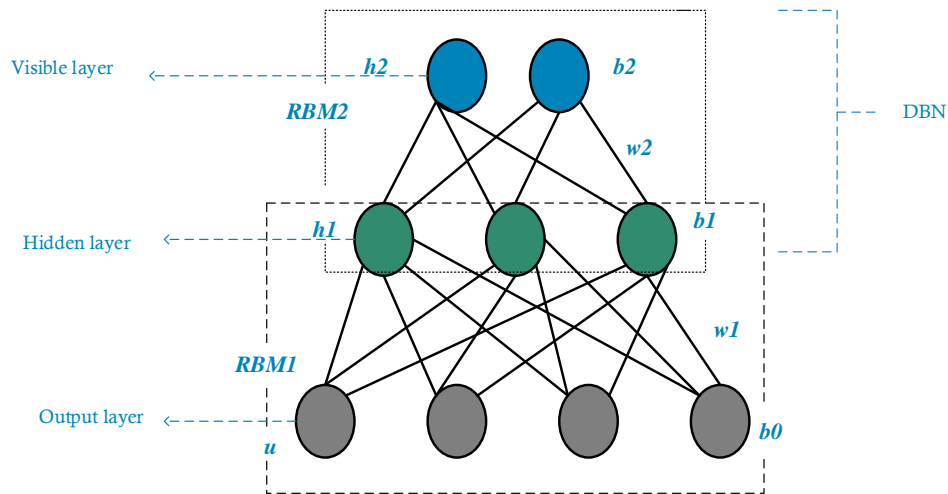


Fig.4. Basic representation of DBN framework

Here, multiple tasks cannot perform simultaneously, and real-time applications are not supported for the DBN framework. Moreover, the trained datasets are suffered from the hidden layer of the DBN framework. Also, the RBM layer is a trained dataset with the help of unsupervised manner; consequently, the hidden layer of the RBM is cannot derive the feature form of the dataset. Here, RBM is handled more information at a time; however, this stage takes more execution time. To solve this issue proposed optimization fitness function is initiated in the hidden layer.

4. Proposed Methodology

In this present work, an Ant Colony based Deep Belief neural (AC-DBN) framework is planned to design for analyzes the sentiment value in the drug review dataset. Here, the drug review dataset is used to validate the proposed design performance. Initially, a Twitter-based drug review dataset is trained to the system; consequently, a novel AC-DBN is designed. Here, the fitness function of ant colony is updated in the classification layer to categorize the sentiment value of each dataset. The proposed architecture is detailed in fig.5.

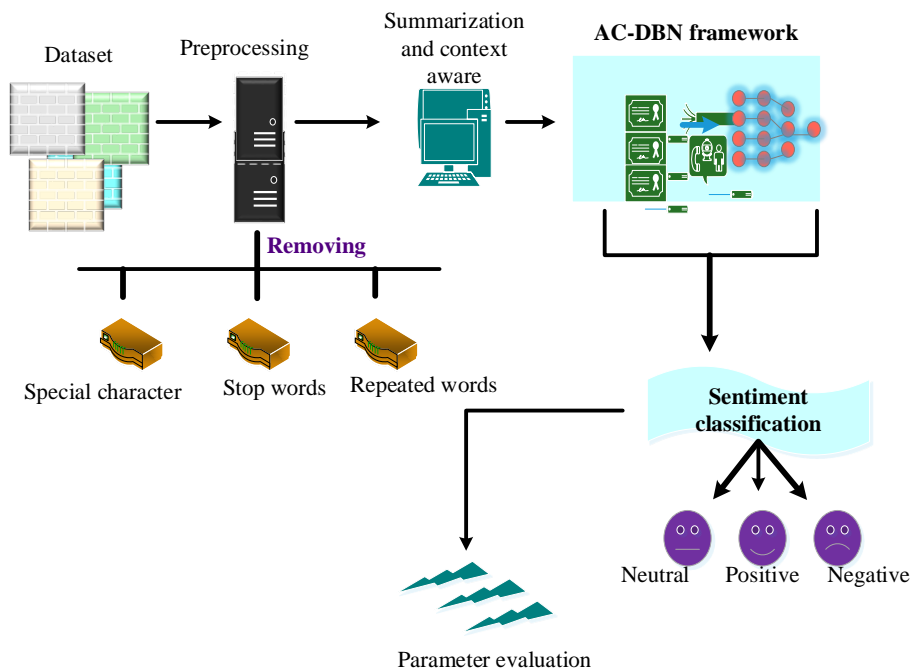


Fig.5. Proposed methodology

Here, the drug review dataset is trained to the system. Initially, the training dataset errors are removed; consequently, the error-free data is trained to the AC-DBN Classification module; the presence of ant colony fitness function in the classification model improves the opinion specification accuracy. Here, DBN is used to reduce the human efforts to pre-process the collected dataset. Also, implementation is fast and attained the highest accuracy.

➤ Preprocessing

The beginning step of sentiment analysis is preprocessing, and it is called as removing stage. If the dataset is preprocessed well, then it facilitates to enhance efficient features for extra development. Numerous techniques are developed to eliminate the insufficient information present in the collected dataset. Moreover, the preprocessing layer can deal with three main programs such as,

- Remove Special character
- Remove stop words
- Remove repeated words

Initially, remove the special character, for example, "recently proposed COVID-19 drugs, is very bad @ user's opinions" in this sentence having a lot of special characters and punctuations (, @, # ") were removed. Then, remove the stop words such as meaningful information and frequently utilized the words after the removal of repeated words in the dataset.

➤ Summarization and context-aware

Generally, drug review datasets are large in size, because the customers are expressing their opinion on social media. Moreover, context awareness is the important key for sentiment classification. For example, the word "unpredictable" has a negative classification, but the combined "unpredictable steering" has a positive classification in movie-based review. Therefore, context-aware plays a fundamental task in opinion classification. Moreover, the sentence has a reverse opinion, for example, "the product is excellent," and other "the product is not excellent". Here, the first sentence contains "good," so it is termed as positive specification, and the second sentence contains the clause "not" so this sentence is termed as negative classification.

4.1. Process of AC-DBN

Initially, collect the drug review dataset in web journals, social media like Facebook and Twitter. Moreover, sentiments and emotions are communicated in different types of ways and making that information is confused. Here, Ant Colony (AC) optimization is used to predict the sentiment classification in the drug review dataset. Also, this technique is used as a Deep Belief Neural (DBN) network. The process of DBN framework is represented in algorithm.1

Process of DBN			
Start			
Step:1	Dataset source	// tsv file format	
Step:2	Verify the dataset	// based on the dug name, review, condition, ratings etc.	
Step:3	Pre-processing	remove → unwanted information	
Step:4	Feature extraction		
	Sample data 1 → train		
	Sample data 2 → train		
	Sample data n → train		
Step:5	Sentiment analyzer	// polarity score label with sentiment specification	
Step:6	Review score	//based on	1. Feature extraction
			2. Classification
			3.cross validation
Stop			

The purpose of using the DBN framework is to reduce unnecessary words efficiently. Moreover, the sentiment classification procedure is easier to remove the unwanted information present in the collected datasets. Here, preprocessing stage is also called as removing stage. Moreover, the removing stage contains some keywords to remove the words present in the collected dataset. Consequently, to remove the stop words, repeated words, and special characters using the following eqn. (1)

$$Y'_1 = Rx\{S[D]_{\max} - S[D]_{\min} + 1\} + Avg(D_{\min 1} + D_{\min 2} + \dots + D_{\min n}) \quad (1)$$

$$Y'_2 - Y'_{n-1} = \sum D$$

Where, R is denoted as a randomly selected dataset, $S[D]_{\max}$ is represented as the maximum size of dataset, $S[D]_{\min}$ is the minimum size of the dataset, and D is the content-based indexing. Here, in eqn. (2) is checked out the following condition $Y_1' < 1.0$. After that estimate, the stop words present in the collected tweets using eqn. (2),

$$T'_{t,d} = \frac{T_{t,d}}{\sum_{t'} T_{t',d}} \quad (2)$$

Where, $T'_{t,d}$ represented as occurrence stop words in the tweet and $\sum_{t'} T_{t',d}$ is denoted as the total amount of stop words present in the tweets. Consequently, the subsection of the t and d is represented as term frequency and collected tweet test, respectively. Then, compute the inverse text frequency using eqn. (3),

$$idfT'(t, d') = \log \frac{N'}{|\{t'' \in d' : t' \in d\}|} \quad (3)$$

Where, N' denoted as the total number of collected tweets and $|\{t'' \in d' : t' \in d\}|$ is the set dataset, which contains symbolized representation. Hereafter, removing the errors that mean a new solution is determined in the eqn. (4),

$$D'_n = \left\lceil \frac{Y''}{Y'_n} \right\rceil + (R_n - 1) \quad (4)$$

Where, Y'' is represented as number words present in the new dataset that means error removed dataset, Y'_n is denoted as number words removed in the previous dataset, R_n is expressed as the total number of top words are taken from this research. At next, identified sentiment classification is updated to the ant colony fitness function in the classification layer of DBN framework. Also, the opinion specification model is developed in Ant colony technique, so it is termed as AC-DBN. Initially, numerous aspect terms such as positive is denoted as '1', negative is represented as '-1', and '0' is denoted as neutral. Moreover, sentences in words are trained to the system, and the collected sentence is classified using proposed AC-DBN framework. In addition, the flow of proposed AC-DBN is demonstrated in algorithm 2.

Algorithm 2 Pseudocode for AC-DBN algorithm

Start

Initialize the parameters

Int p, n, n_l, i, D

// where, p is represented as positive words, n is denoted as negative words, n_l is the neutral sentence, i is the each and every iteration and D is the drug review dataset

Label [1, -1, 0]

Replica= AC-DBN classifier

Initialize the drug review dataset

For($i=n$; $i \geq n$; $i++$)

Find sentiment classification S_a *// using Eqn.(3)*

If ($S_a = 1$) $\rightarrow p$ *//Classification using eqn.(4)*

Here the sentence is positive

Else

Return to calculation process

Next iteration

If ($S_a = -1$) $\rightarrow n$

Here the sentence is negative

Else

Return to calculation process

Next iteration

If ($S_a = 0$) $\rightarrow n_l$

Here the sentence is neutral

Else

Return to calculation process

End for

End if

Stop

Output: Best possible solution

Moreover, the ant colony optimization [29] technique is also called as the searching mechanism, which is frequently utilized to solve optimization problems. After the preprocessing process, the collected datasets are

summarized and arranged here, the fitness function of ant colony is pheromone updating, and this fitness is initialized in the classification layer of the DBN framework. Consequently, the representation of DNN is shown in fig.6. Moreover, to identify the sentiment classification based on the following eqn. (5)

$$S_a(n+1) = [1 - a'(0)]; [1 - a'(1)]; [1 - a'(2)]; [1 - a'(3)]; \dots \dots \dots + \sum_{n=1}^a a'(L') \quad (5)$$

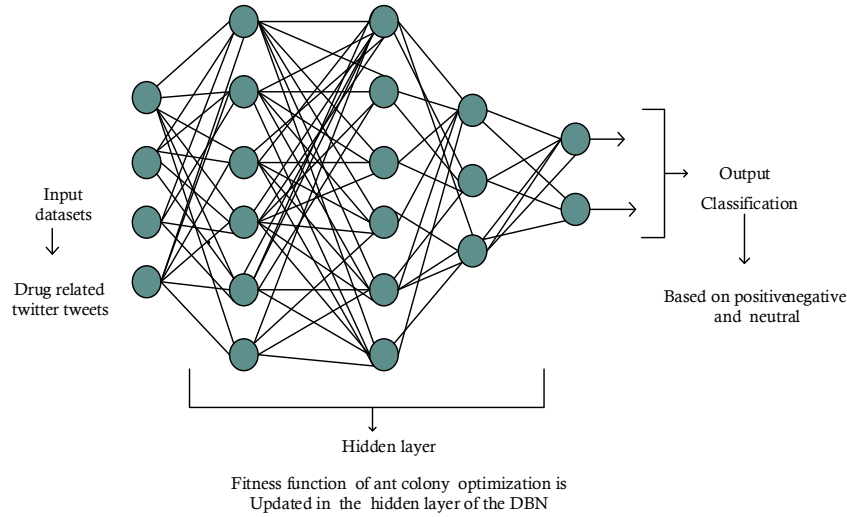


Fig.6. Representation of proposed AC-DBN

Where, a' is represented as collected sentiment word, L' is denoted as classification of the element in starting process sentence are trained in the ant colony model. Hereafter, in the deep learning model, the training sentence of opinion word is arranged in the order of 0's, 1's, and -1, which is detailed in eqn. (6),

$$\text{If the sentiment values are denoted as } S_a = \begin{cases} -1 & \rightarrow \text{negative} \\ 1 & \rightarrow \text{positive} \\ 0 & \rightarrow \text{neutral} \end{cases} \quad (6)$$

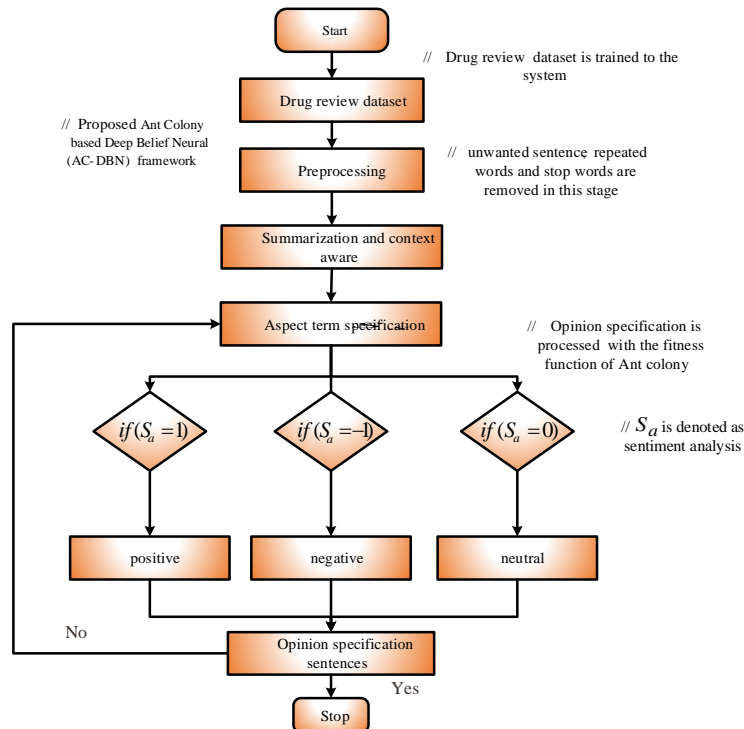


Fig.7. Working Flow of proposed method

The sentiment analysis value is equal to -1 it is said to be negative, and the value is equal to zero it is said to be neutral, and the value is equal to 1 it is said to be positive. The working flow of the proposed model is represented by the flow model in fig.7.

5. Discussion of Results

Deep learning is important in big data, especially in sentiment analysis. In this research, the drug review dataset values are predicted by this proficient of developed replica, and that dataset has three types of classes, positive, negative, and neutral. Moreover, the proposed AC-DBN framework is executed in the python platform. The proposed novel algorithm is utilized to enhance the opinion specification such as negative, positive, and neutral.

5.1. Parameter evaluation

The effectiveness of the developed technique is validated with its key metrics and achieving better outcomes by comparing other existing models like Weighted Word Representation (WWR) strategy and the Hybrid machine learning approach (H-MLA) [28], the Deep Learning (DL) model [23], Naïve bayes and KNN [32,33]. Therefore, to calculate the efficiency of the proposed model, certain metrics should be validated, such as precision, recall, F-measure, accuracy, and opinion specification. To find the successive rate of the proposed model, the existing techniques such as WWR, DL and H-MLA are implemented in the same proposed platform. Moreover, the attained results are discussed below.

A. Accuracy (A_c)

The process of accuracy calculation is classified in terms of sentiment analysis in the drug review dataset. Moreover, to estimate the accuracy of the sentiment analysis is mainly based on three aspect terms like positive, negative, and neutral. Consequently, the accuracy of the proposed model is estimated in the following eqn. (7),

$$A_c = \frac{(TN' + TP')}{(TN' + TP' + FN' + FP')} \quad (7)$$

Where, where, TN' is True Negative, TP' for true positive, FN' is false-negative and FP' is for false positive. Furthermore, comparison results are shown in fig.8.

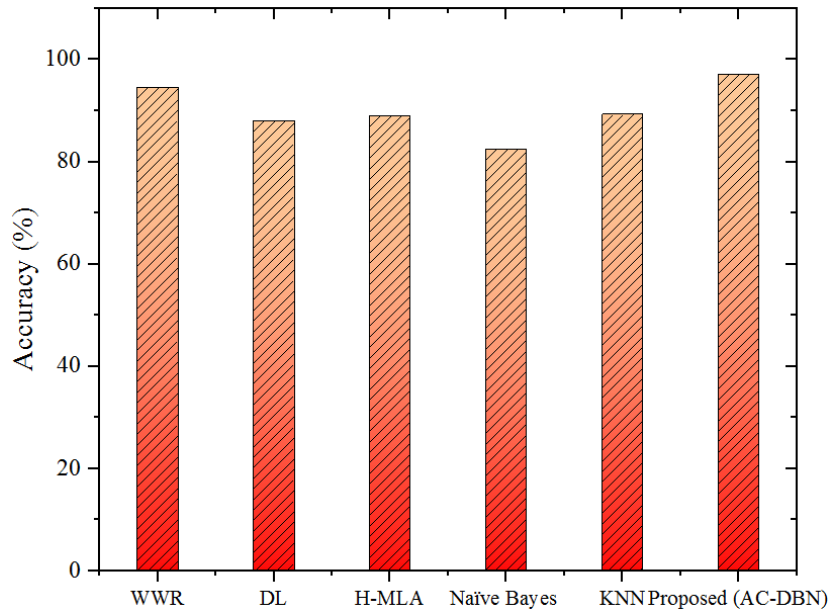


Fig.8. Comparison of accuracy

The accuracy rate of the WWR approach is 94.6%, the DL technique achieved an accuracy measure of 89%, the attained accuracy rate of H-MLA is 89%, Naïve bayes technique attained 86.6% accuracy measure, and KNN model obtained 92% of accuracy rate; consequently, the proposed AC-DBN replica achieved a 98% accuracy measure.

B. Precision (P_r)

Calculation of precision is based on the total amount of accurate classification by the total amount dataset. Therefore, the precision of the estimated amount of perfect positive calculations is divided through the number of

particular opinion sentences. Consequently, validation of precision with recent techniques is demonstrated in fig.9. The precision of the proposed approach is estimated in eqn. (8),

$$P_r = \frac{TP_t}{TP_t + FP_t} \quad (8)$$

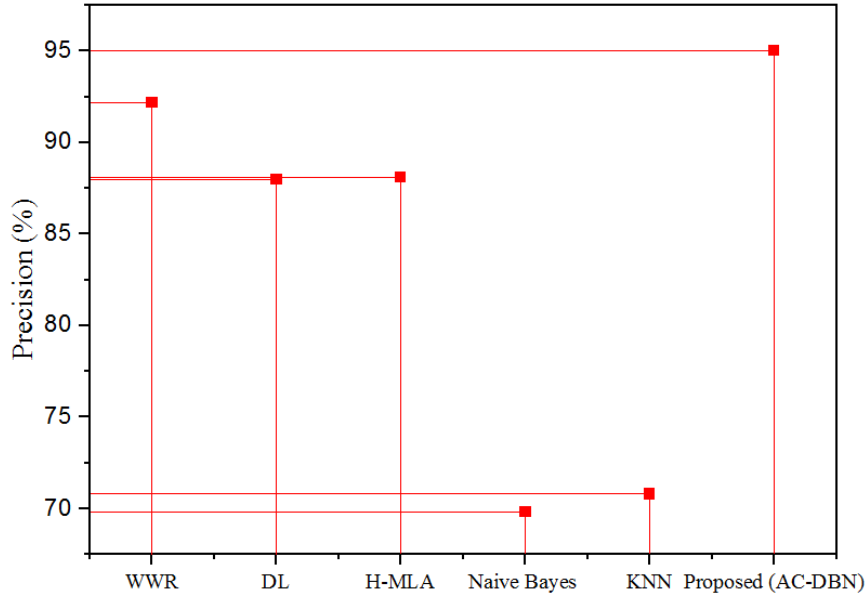


Fig.9. Comparison of precision measurement

WWR is achieved 92.2% of precision measure, DL attained 88% as the precision, H-MLA algorithm earned 88.12% precision measure, Naive Bayes technique attained 69.9% accuracy, and KNN achieve 72% accuracy measure. Consequently, the proposed AC-DBN approach attained 95% of precision rate.

C. Recall measure

The recall is defined as the ratio of properly analyzed datasets for one class to the dataset of the overall class. In addition, recall is compared with three existing techniques, namely WWR, DL, and H-MLA. Moreover, the recall measure is determined by eqn. (9),

$$R_c = \frac{FP_t}{FP_t + FN_t} \quad (9)$$

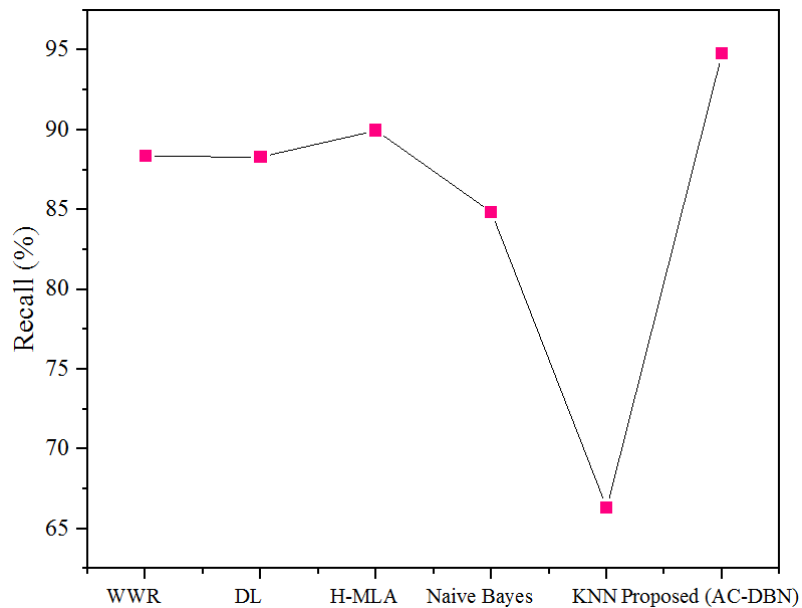


Fig.10. Recall measure

The developed method attained an 88.35% recall measure. Accordingly, the existing replica WWR attained was 0.72%, recall measure of the DL approach is 0.88%, and Naive Bayes technique reached 86.6% recall measure, KNN attains 65.5 recall measures. Finally, the H-MLA technique achieved a 0.65% recall measure, and the comparison is shown in fig 10.

D. F-measure (F_m)

F-measure is one of the most important estimation parameters in sentiment analysis. Moreover, the F-measure is measured by the precision and recall rate. Every kind of sentiment classification contains an individual F-measure to evaluate the classification. In addition, fig.11 demonstrates the statistics of F-measure. Moreover, F-measure is determined by eqn. (10).

$$F_m = 2 \times \frac{P_r \times R_c}{P_r + R_c} \quad (10)$$

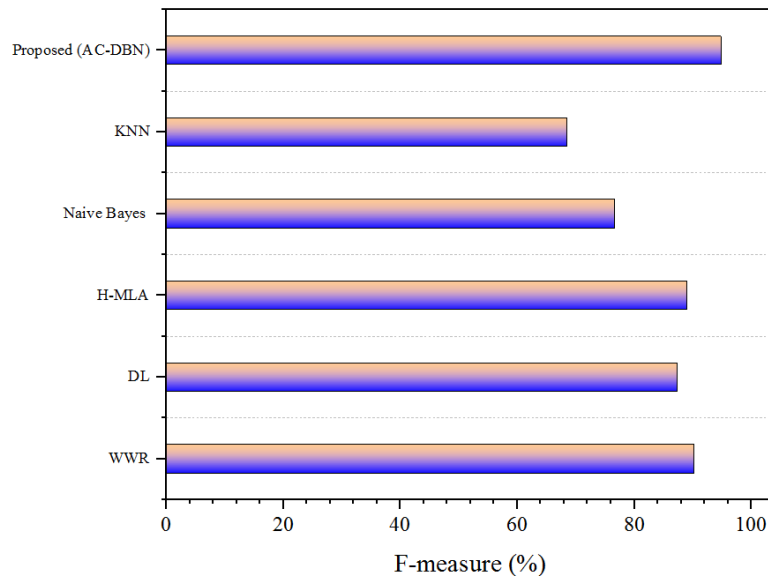


Fig.11. Comparison of F-measure

The proposed approach is attained 94.8% F-measure. Consequently, the existing techniques of WWR achieve 90.2% F-measure, the DL approach attained 87.35% F-measure, and the H-MLA technique predicted an 89.03% F-measure, Naive Bayes technique attained 76.6% F-measure, and KNN achieve 68.55 F-measure.

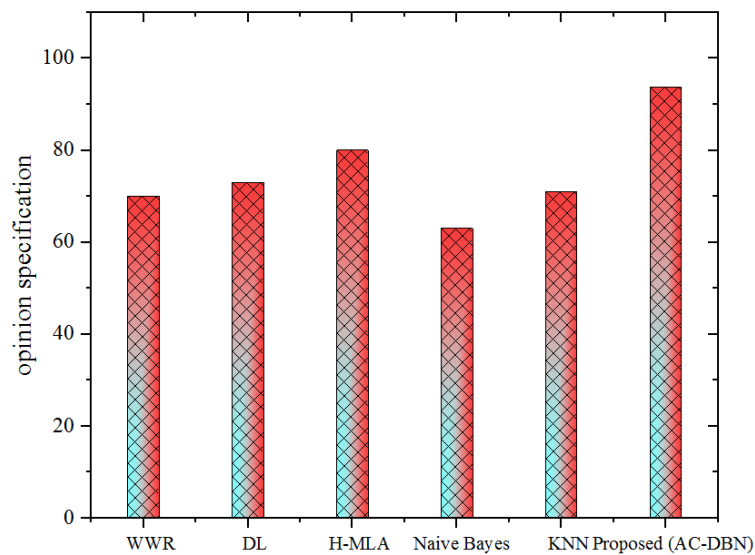


Fig.12. Comparison of opinion specification

E. Opinion specification

Opinion specification is otherwise called opinion classification. In this drug-based review, sentences are classified into three kinds of sentiment classification like positive, negative, and neutral. Moreover, the comparison of opinion specifications is demonstrated in fig.12.

By validation, the WWR attained an opinion classification rate of 70%, the DL approach attained 73% of opinion classification, H-MLA attained an opinion classification measure of 80%, Naive Bayes technique attained 68% opinion classification, and KNN achieve 75.55% opinion classification. Moreover, the proposed strategy achieved a word opinion specification of 93.8%.

5.2. Discussion

The proposed AC-DBN framework performance is compared with existing techniques such as WWR, DL, H-MLA techniques based on the sentiment classification. From all the performance evaluations, it is proved that the proposed AC-DBN performs best for sentiment classification. The overall performance of the developed model is shown in Table 5.

Table 5. Overall performance

Methods	Accuracy	Precision	Recall	F-Score
DNN	89	0.88	0.78	0.87
WWR	95	0.92	0.72	0.9
H-MLA	89	0.88	0.65	0.89
NB	87	0.7	0.86	0.76
KNN	92	0.72	0.66	0.68
AC-DBN	98	0.95	0.89	0.94

Moreover, the proposed AC-DBN model attained high performance for classifying the sentiment values in an efficient manner. Also, the performance like accuracy, error rate, f-measure, recall based on opinion classification evaluated using the AC-DBN framework.

6. Conclusion

In the modern world, analyzing sentiments from big data applications is the most challenging task. Sentiment analysis of customer reviews is carried out to predict product opinions and to improve the products further as per customer demands and sentiments. Basically, sentiments are classified as positive, negative, and neutral based on the subjectivity of a particular topic or trend. Previously, numerous ML techniques were employed to analyze sentiments on various themes and trends in different scenarios. However, it is quite difficult to predict the best technique for classifying the sentiments as it depends on the size and source of the big data emerging from. In this research, a novel AC-DBN framework is proposed for drug sentiment classification. The main aim of this research is to classify the drug review dataset. Moreover, the proposed framework includes preprocessing layer to remove the unwanted information present in the collected dataset and then the sentence is summarized with a short duration. Furthermore, the proposed AC-DBN framework outcomes are compared with the existing methods in terms of accuracy, precision, recall, F-measure, and opinion specification. Also, the proposed AC-DBN framework has attained 97.2% accuracy, 95% precision, 94.8% recall, and an F-measure rate of 94.8%. In future, we plan to classify educational sentiments using other DL techniques. This proposed framework will be useful for clinics, drug makers besides patients to get valuable sentiments from the public.

References

- [1] Xu, Guixian, et al. "Sentiment analysis of comment texts based on BiLSTM." *IEEE Access*, 7(2019): 51522-51532.
- [2] Maryame, Naji, et al. "State of the Art of Deep Learning Applications in Sentiment Analysis: Psychological Behavior Prediction." *Embedded Systems and Artificial Intelligence*. Springer, Singapore, 2020. 441-451.
- [3] Do, Hai Ha, et al. "Deep learning for aspect-based sentiment analysis: a comparative review." *Expert Systems with Applications* 118 (2019): 272-299.
- [4] Zvarevashe, Kudakwashe, and Oludayo O. Olugbara. "A framework for sentiment analysis with opinion mining of hotel reviews." *2018 Conference on information communications technology and society (ICTAS)*. IEEE, 2018.
- [5] Aggarwal, Charu C. "Opinion mining and sentiment analysis." *Machine learning for text*. Springer, Cham, 2018. 413-434.
- [6] Rathana, M., et al. "Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews." *Applied Soft Computing* 68 (2018): 765-773.
- [7] Bose, Rajesh, et al. "Analyzing political sentiment using Twitter data." *Information and communication technology for intelligent systems*. Springer, Singapore, 2019. 427-436.
- [8] Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in

commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90, 523-537.

- [9] Benlahbib, Abdessamad. "Aggregating customer review attributes for online reputation generation." *IEEE Access* (2020).
- [10] Singh, Nikhil Kumar, Deepak Singh Tomar, and Arun Kumar Sangaiah. "Sentiment analysis: a review and comparative analysis over social media." *Journal of Ambient Intelligence and Humanized Computing* 11.1 (2020): 97-117.
- [11] Ahmad, Shakeel, et al. "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques." *Human-centric Computing and Information Sciences* 9.1 (2019): 1-23.
- [12] Ramírez-Tinoco, Francisco Javier, et al. "Use of sentiment analysis techniques in healthcare domain." *Current Trends in Semantic Web Technologies: Theory and Practice*. Springer, Cham, 2019. 189-212.
- [13] Drus, Zulfadzli, and Haliyana Khalid. "Sentiment analysis in social media and its application: Systematic literature review." *Procedia Computer Science* 161 (2019): 707-714.
- [14] Vashishtha, Srishiti, and Seba Susan. "Fuzzy Interpretation of Word Polarity Scores for Unsupervised Sentiment Analysis." 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2020.
- [15] Mishra M, Barman SK, Maity D, Maiti DK (2019) Ant lion optimisation algorithm for structural damage detection using vibration data. *Journal of Civil Structural Health Monitoring* 9(1):117-136. <https://doi.org/10.1007/s13349-018-0318-z>.
- [16] Mukku SS, Oota SR, Mamidi R (2017) Tag me a label with multi-arm: Active learning for telugu sentiment analysis. *International Conference on Big Data Analytics and Knowledge Discovery*, Springer, Cham, pp 355–367. https://doi.org/10.1007/978-3-319-64283-3_26.
- [17] Nakagawa T, Inui K, Kurohashi S (2010) Dependency tree-based opinion specification using CRFs with hidden variables. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp 786–794.
- [18] Dadvar M, Hauff C, Jong F (2011) Scope of negation detection in sentiment analysis, *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, University of Amsterdam, pp 16–20.
- [19] Edgcomb JB, Zima B (2019) Machine Learning, Natural Language Processing, and the Electronic Health Record: Innovations in Mental Health Services Research. *Psychiatric Services* 70:346–349. <https://doi.org/10.1176/appi.ps.201800401>.
- [20] Garg, Satvik. "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning." 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021.
- [21] Han, Yue, Meiling Liu, and Weipeng Jing. "Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer." *IEEE Access* 8 (2020): 21314-21325.
- [22] Bhamare, Bhavana R., and JeyanthiPrabhu. "A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas." *PeerJ Computer Science* 7 (2021): e347.
- [23] Basiri, Mohammad Ehsan, et al. "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques." *Knowledge-Based Systems* 198 (2020): 105949.
- [24] Colón-Ruiz, Cristóbal, and Isabel Segura-Bedmar. "Comparing deep learning architectures for sentiment analysis on drug reviews." *Journal of Biomedical Informatics* 110 (2020): 103539.
- [25] Hossain, MdDeloar, et al. "Drugs Rating Generation and Recommendation from Sentiment Analysis of Drug Reviews using Machine Learning." 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, 2020.
- [26] Nagamanjula, R., and A. Pethalakshmi. "A novel framework based on bi-objective optimization and LAN 2 FIS for Twitter sentiment analysis." *Social Network Analysis and Mining* 10 (2020): 1-16.
- [27] NaziaTazeen and K. Sandhya Rani, "A Conceptual Data Modelling Framework for Context-Aware Text Classification" *International Journal of Advanced Computer Science and Applications (IJACSA)*,11(11),2020. <http://dx.doi.org/10.14569/IJACSA.2020.0111116>
- [28] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "A Weighted Text Representation framework for Sentiment Analysis of Medical Drug Reviews." 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM). IEEE, 2020.
- [29] Paniri, Mohsen, Mohammad BagherDowlathshahi, and HosseinNezamabadi-pour. "MLACO: A multi-label feature selection algorithm based on ant colony optimization." *Knowledge-Based Systems* 192 (2020): 105285.
- [30] Su, Chuan-Jun, and Yi Li. "Sentiment analysis and information diffusion on social media: the case of the Zika virus." *BIBE 2018; International Conference on Biological Information and Biomedical Engineering*. VDE, 2018.
- [31] Shayaa, Shahid, et al. "Sentiment analysis of big data: Methods, applications, and open challenges." *IEEE Access* 6 (2018): 37807-37827.
- [32] <https://www.kaggle.com/jessicali9530/kuc-hackathon-winter-2018>
- [33] Novendri, Risky, et al. "Sentiment analysis of YouTube movie trailer comments using naïve bayes." *Bulletin of Computer Science and Electrical Engineering* 1.1 (2020): 26-32.
- [34] Joann Galopo Perez, Eugene S. Perez, "Predicting Student Program Completion Using Naïve Bayes Classification Algorithm", *International Journal of Modern Education and Computer Science*, Vol.13, No.3, pp. 57-67, 2021.
- [35] Dimple Tiwari, Nanhay Singh, "Ensemble Approach for Twitter Sentiment Analysis", *International Journal of Information Technology and Computer Science*, Vol.11, No.8, pp.20-26, 2019.
- [36] Yasin Görmez, Yunus E. Işık, Mustafa Temiz, Zafer Aydın, "FBSEM: A Novel Feature-Based Stacked Ensemble Method for Sentiment Analysis' Comments in E-Government", *International Journal of Information Technology and Computer Science*, Vol.12, No.6, pp.11-22, 2020.

Authors' Profiles



Mrs. Nazia Tazeen, completed B. Tech, M. Tech in Computer Science and Engineering, at JNTU Hyderabad. I am pursuing PhD in Computer Science and Engineering at Sri Padmavathi Mahila Viswa Vidyalyayam, Tirupati, India. My research area is Big Data Analytics, Data Mining, Cloud computing.



Prof. K. Sandhya Rani, graduated M. Tech in Computer Science from IIT Kharagpur, in 1992. Ph.D. in Computer Science from Sri Padmavati Mahila Visvavidyalayam, Tirupathi in 2003. She is having total 33 years Work Experience in various fields. She has worked as the Rector of Sri Padmavati Mahila Visvavidyalayam, Tirupathi. Research area: Big Data Analytics, Pattern Recognition, Artificial Neural Networks and Privacy Preserving Data Mining

How to cite this paper: Nazia Tazeen, K. Sandhya Rani, "A Novel Ant Colony Based DBN Framework to Analyze the Drug Reviews", International Journal of Intelligent Systems and Applications(IJISA), Vol.13, No.6, pp.25-39, 2021. DOI: 10.5815/ijisa.2021.06.03