

Hybrid model for sentiment analysis combination of PSO, genetic algorithm and voting classification

Garima Srivastava^{1,2}, Vaishali Singh¹, Sachin Kumar³

¹School of Engineering and Technology, Maharishi University of Information Technology, Lucknow, India

²Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Lucknow, India

³Department of Electronics Engineering, Amity School of Engineering and Technology, Amity University, Lucknow, India

Article Info

Article history:

Received Nov 8, 2023

Revised Mar 23, 2024

Accepted Apr 6, 2024

Keywords:

Ensemble model

Genetic

PSO

Sentiment analysis

Voting classification

ABSTRACT

As social network services like Weibo and Twitter have grown in popularity, natural language processing (NLP) has seen a great deal of interest in sentiment analysis of social media messages and Information mining. Social media users, whose numbers are always increasing, have the ability to exchange information on their platforms. The study of sentiment, domains and themes are closely related. Manually collecting enough labelled data from the vast array of subjects covered by large-scale social media to train sentiment classifiers across several domains would be extremely difficult. The literature review conducted concludes that models already proposed in the previous researches are not able to achieve good accuracy. This work suggests a unique model that combines of genetic algorithm and particle swarm optimization to effectively extract the features and then the voting technique is applied for the classification. Model proposed is compared with 4 ensemble datasets achieving a consistent accuracy of more than 90% for three different diversified database owing to natural selection of sequences by GA and at the same time achieves a fast convergence with PSO, the model may be employed for highly accurate recommenders demanding precision and accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Garima Srivastava

School of Engineering and Technology, Maharishi University of Information Technology

Lucknow (U.P.), India

Email: gsrivastava1@lko.amity.edu

1. INTRODUCTION

Sentiment classification, polarity classification, affect analysis, opinion mining, subjectivity analysis or Sentiment analysis is an interdisciplinary branch of research, that studies opinion of people about a various product. It combines elements of knowledge discovery, computational linguistics, natural language processing (NLP), and other fields. In contrast to actual facts, opinions and sentiments are subjective, therefore it is possible to extract them from a text using a range of computational and NLP-based approaches. As social network services like Weibo and Twitter have grown in popularity, analyzing the sentiment has attracted the NLP researchers in the last few years [1], [2]. On social networks, user opinions are often expressed in brief texts that includes a user's opinion of a certain object [3], [4]. User sentiment refers to a user's perception or attitude towards a circumstance or event (called topic). Sentiment analysis for social media texts now faces new difficulties in comparison to sentiment analysis for traditional web texts. In contrast, popular subjects on social media are continuously shifting. Sentiment analysis, themes, and domains are well known to be closely related [5]. Manually collection of labelled data from the vast array of subjects covered by large-scale social media to train sentiment classifiers across several domains would be extremely

difficult. Through text data, sentiment analysis may effectively disclose people's opinions (whether favorable, unfavorable, or neutral) about a good or service [6], [7]. The subtasks in Figure 1 of sentiment analysis are carried out by examining the user contents shared on social media.

Six components make up the system: a noise filter, a sentiment and emotion analysis engine module, a prediction analyzer, a results viewer, and a database. The "Data Collector" crawls unprocessed data from a variety of websites, including blogs, Twitter, and discussion forums. The "Noise Filter" module receives data from this module, which is a collection of scripts, and processes it, depending on whether the data sources give programmatic interface to read data [8], [9]. "Meaningless data," like ads, meaningless stuff with no comments, also noises relevant to some content are removed by the "Noise Filter/Smart Filter" [10].

In order to be used for crucial business tasks such as forecasting, monitoring, and action strategizing, Predictive analysis of events, like revenue and reputation crises, is carried out via the "Predictive Analyser". The pool of prediction algorithms and the predictor/feature set are its two main parts. The result of sentiment and emotion analysis, containing emotions like anger, grief, and anxiety as well as positive, negative, neutral, and mixed sentiments, serves as a new predictor in addition to the existing predictors and features. It will be possible to do time-series analysis, anomaly identification and customer preference analysis for sales forecasting by integrating the sentiment results and combining the prediction algorithm pool's extra insights with the emotion analysis engine [11], [12].

This research article is based on the twitter data sentiment analysis. Feature extraction, classification, and pre-processing are the three stages of the sentiment analysis process. The key step is feature extraction in which attribute set will establish relation with target set [12]-[14]. Previous research shows that author do not concentrate on the feature extraction phase which leads to low accuracy. The research work proposes a hybrid optimization model for feature extraction to increase accuracy for the sentiment analysis.

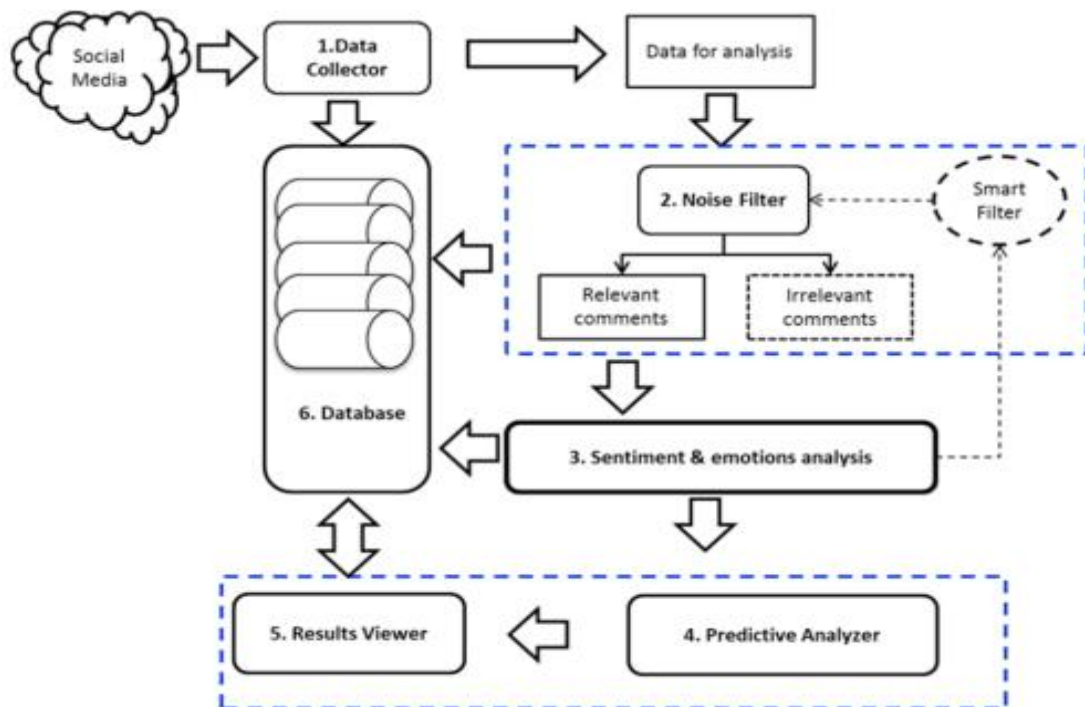


Figure 1. Social media sentiment analysis system [13]

Section 1, presents the introduction about the sentiment analysis with basic steps of social media sentiment analysis system [15]. The various models are already proposed by authors in previous years for the sentiment analysis. Section 2 of the article highlights the literature survey on sentiment analysis. Section 3 presents the proposed model with pseudo codes and flowcharts. Section 4 shows results and discussion in which detail results of proposed model is presented and compared with existing models. The last section of the paper concludes the article.

2. LITERATURE SURVEY

Hybrid algorithms can achieve faster convergence and higher enhancement. Additionally, hybrid algorithms can be more effective than thoroughbred optimization algorithms in solving engineering problems. However, hybrid algorithms can increase the difficulty of setting control parameters and may require specific design for problems, section presents a brief study on enhancing performance of opinion mining by employing diversified hybrid techniques.

Long *et al.* [15] developed an effective Bi-directional Encoder Representation from Transformers based Convolution Bi-directional Recurrent Neural Network to investigate the syntactic and semantic information for analyzing the data sentimentally and contextually. Firstly, the polarity scores were computed and assigned labels on the basis of the zero-shot classification. After that, the sentence-level semantics and contextual attributes were acquired and the embeddings were generated using a pre-trained BERT algorithm. Bi-directional long short-term memory (Bi-LSTM) sequenced the sentences.

Mahadevaswamy and Swathi [16] discussed that analysing the sentiment was a cognitive tool for extracting the emotional tone from a text. A Bi-LSTM algorithm was designed. It handled the long-term dependencies for which the memory was employed into the model to make superior predictions. The Amazon Product dataset having 104,975 was applied in the experiments. The algorithm emphasized on classifying the reviews into positive and negative classes. At last, the results exhibited that the designed algorithm yielded the accuracy around 91.4% as compared to others.

Li *et al.* [17] suggested Humor-EMOji-Slang-based (HEMOS) system to classify the sentiment for the Chinese language on the basis Deep Learning technique. The significance to identify the humor, pictograms and slang on the task related to process the social media was computed. Firstly 576 frequent Internet slang expressions were gathered as a slang lexicon. Then transformed Weibo emojis into textual attributes when a Chinese emoji lexicon was produced. The last stage employed both lexicons to AttBiLSTM (attention-based bi-directional long short-term memory recurrent neural network) and its efficacy was computed.

Pimpalkar and Raj [18] established a model based on Deep Learning and the GloVe word embedding to extract the relevant semantics of words in texts. CNN layer employed for learning the attributes. multi-layered Bi-directional long-short-term memory (MBiLSTM) deployed these portions for capturing the long-range embedded circumstances. An experiment was conducted for offering a precise answer for examining the feelings and user reviews into positive or negative. The established model yielded an accuracy of 92.05% on system test and 93.55% on validation. The supremacy of the model was proved on IMDB datasets.

Wang *et al.* [19] introduced a sentiment approach for analyzing the sentiments. At first, the optimal sentiment notion of words was presented in Microsoft Concept Graph (MCG) concerning the context of words. Secondly, the goal of this method was to extract the sentiment information associated with the words from the multi-semantics sentiment intensity lexicon. Correct representation of word semantics and sentiment was made possible by its ability to integrate sentiment information. Ultimately, a more comprehensive word representation was achieved by combining two improved word embeddings (WE) approaches. The efficacy of the strategies exhibited was documented by the outcomes across six datasets.

Tan *et al.* [20] developed a hybrid Deep learning technique called RoBERTa-LSTM which combined robustly optimized bidirectional encoder representation (BERT) with long short-term memory (LSTM) to analyse the sentiments. The initial algorithm was implemented to map the words into a compact WE space and the latter one helped in capturing the long-distance contextual semantics in an effective way. The experimental outcomes exhibited that the developed technique yielded a F1-score of 93% on IMDb dataset, 91% on Twitter US Airline dataset and 90% on Sentiment140 dataset.

He *et al.* [21] projected a fusion sentiment analysis (FSA) technique that integrated textual analysis methods with machine learning (ML) algorithms for mining the online product experience. The technique was executed in 3 stages. The sentiment dictionary was deployed as first stage for extracting the sentiment attributes. After that, the support vector machine (SVM) algorithm was put forward for recognizing sentiment polarities of reviews. Linear discriminant analysis (LDA) was utilized to extract the sentiment topics from the reviews. Generally, the weighting technique was employed for computing the sentiment contribution. In the end, the reading experiences of customers for online books available on Amazon was considered to quantify the projected technique. The experimental outcomes revealed the efficacy of the projected technique to analyse the sentiments of reviews and capture the elements which affected the reading experiences from reviews.

Ahmad *et al.* [22] introduced a technique known as attention-based multi-channel gated recurrent neural network (Att-MC-GRU) for extracting the aspects and analysing the textual reviews while classifying their sentiments. In this, a hybrid method of WE, part of speech (POS) tags, and contextual position information was put forward. This method was useful for enhancing the accuracy to recognize and predict the aspects and their associated sentiments.

Nayak *et al.* [23] suggested a modified Bayesian Boosting approach with weight-guided optimal feature selection for sentiment analysis. It included a feature selection system that weights distinct features in accordance with how crucial they are for sentiment categorization. This weight-guided feature selection method improved classification performance by identifying the most pertinent and discriminative characteristics. To maximize the boosting process, the program used a modified Bayesian boosting technique.

The availability of integrated tools and methodologies for sentiment analysis that let users experiment and evaluate various algorithms and optimisations based on personalised preferences and parameters is lacking. The work suggests a hybrid method for sentiment analysis using GA and PSO algorithms. This algorithm is helpful in choosing the optimisation features and encoding a workable solution for a problem into an individual. Each individual is viewed as an entity that supports chromosomal traits. A population is made up of several individuals [24], [25].

GA serves as an optimum search approach using natural selection and the survival of the fittest, this optimises selection of relevant word sequence from random corpus. GA operates on individuals i.e fixed size population and processes the dataset without biases that evolves over time. PSO is a fast gradient method and has a faster convergence, depending on their present position in the particle space and the associated velocity vector, the variables in PSO can have any values. Table 1 summarizes evolution of hybrid algorithms with diversified combinations for enhanced opinion mining performance, number of techniques with diversified combinations have been studied, however following gaps still exist

- a) Optimum technique for identification of best suited prominent sequence of words from random space with concept of survival of fittest for opinion mining
- b) Effective handling of bias variance trade off while classifying sequences
- c) Consistent accuracy in predictions over diversified databases

Table 1. Summary of evolution of hybrid algorithms

S.No.	Author Name	Year	Reference	Algorithm/ Technology	Metrics
1	Kaur and A. Sharma [26]	2024	24	Hybrid model combines CNN and Bi-LSTM with Elephant herd optimization	11% of users have poor experience
2	Zingade <i>et al.</i> [27]	2023	25	Hybrid TF-IDF+N-Gram model feature selection with Information Gain (IG)	78% Accuracy
3	Ezaldeen <i>et al.</i> [28]	2023	26	Hybrid method comprising DL classifier and LSTM.	F1-score of 94.46%, Precision 89%
4	Zhu <i>et al.</i> [29]	2023	27	Hybrid Multi-objective Optimization (HMO) which combines PSO with the Krill Herd Algorithm (KHA).	Accuracy: 89.1
5	Chiny <i>et al.</i> [30]	2022	28	Hybrid model assisted NLP tailored- (CNN)	F1-score of 93%,
6	Tan <i>et al.</i> [20]	2022	20	Hybrid deep learning that integrates BERT approach and Long Short-Term Memory.	Accuracy enhances
7	Gao <i>et al.</i> [31]	2021	29	Bidirectional gated recurrent unit (BiGRU) and self-attention are combined	5.91% higher Accuracy
8.	Mosa <i>et al.</i> [32]	2021	30	LSTM, rulebased lexicon and the TF -IDF method are combined in a binary classification mode	2.68% Improvement
9	API [33]	2020	31	Hybrid cat swarm optimization (HCSO) algorithm.	Precision 98%
10	Godara <i>et al.</i> [34]	2020	32	Multi-objective optimization (MOO) with Gravitational Search Algorithm (GSA) and PSO	

Hybrid algorithms can achieve faster convergence and higher enhancement by combining different types of algorithms. Hybrid algorithms can be more effective than pure optimization algorithms in solving engineering problems. However, hybrid algorithms increases the difficulty of setting control parameters and may require specific design for problems. Our goal is to integrate PSO's exploration capabilities with GA's exploitation capabilities. The proposed method initially runs PSO so that best suited particles of sequence of words from the space can be identified, GA then optimizes to estimate the most effective sequence. To extract sufficient information, the proposed algorithm classifies the combined result by voting classification, a combination of Random Forest, SVM and KNN. SVM provides an exhaustive solution for the bias variance trade-off problem. Identifiers obtained were classified using voting classifier to provide numbers that indicate the polarity values of certain words in the lexicon. Classifier forecasts an output based on the class with the highest likelihood of becoming the output.

3. METHOD

Data on social media is growing exponentially, diversified automated techniques for analysis have been proposed assisted with fuzzy systems, AI techniques, opinion mining techniques, heuristic techniques etc, work presents an algorithm combining PSO and GA for the feature extraction and classification is done using voting method. This work analyzes the sentiments on twitter data. The new mechanism has diverse stages which are defined as

3.1. Extraction of microblogs data and its pre-processing

Several clients provide information in diverse forms like the tweets for expressing their sentiments related to various topics. The twitter data sample is implemented in two categories: pessimistic and positive. Thus, the emotion scrutiny of information can be done easily in order to analyze the outcomes of several properties. But, the susceptibility for discrepancy and superfluous is found higher for unprocessed information.

3.2. Feature extraction

The crucial step in establishing the link between the attribute set and the target set is feature extraction. Hybrid optimisation algorithm combines the PSO with GA. The hybrid model is represented by the flowchart. This model may be used to choose the optimisation attributes and encode a workable solution to a problem into an individual. Each individual is viewed as an entity that supports chromosomal traits. Together, a number of individuals form a population. Prior to using GA, the main goal is to randomly create a population of chromosomes and surround it with problem variables. Next, the step focuses on analysing the generated chromosome data. The chromosomes are important for creating more chromosomes since they may clearly illustrate the best way to approach the problem. The population in this algorithm refers to the main collection of random solutions. For the purpose of coding for a solution to address the problem, each individual of the population is illustrated using a chromosome. The formula for decoding is (1).

$$X = X_{min} + \frac{X_{max}-X_{min}}{2^{Nx}-1} \sum_{n=0}^{Nx-1} b_n^X 2^n \tag{1}$$

In which, the binary representations of X's are denoted as b_0^X, \dots, b_{Nx-1}^X . The generations, or different iterations, are used to create the chromosomes. A variety of fitness indicators are used in each generation to assess the chromosomes' fitness value. A relation exists between each particle i and, the velocity vector $V_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n}]$ and the position vector $X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}]$. To carry out their search procedure, the positions of novel solutions, $x_{i,d}$ are changed at a constant rate. This method concentrates on recalling the historical position of an individual as $pbest_i$, and $gbest$ defines the current global best position that the particle swarm has found. Finding such locations causes all particle positions and velocities to be updated in accordance with the provided equations as (2) and (3).

$$v_{i,d}(t+1) = \omega \cdot v_{i,d}(t) + c_1 \cdot rand_1 \cdot (pbest_{i,d} - x_{i,d}(t)) + c_2 \cdot rand_2 \cdot (gbest_d - x_{i,d}(t)) \tag{2}$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1) \tag{3}$$

This uses t to show the t-th iteration, d to show the particle's d-th dimension, ω to indicate the inertia weight, c_1 and c_2 to show the acceleration constants, and $rand_1$ and $rand_2$ to specify the random numbers, whose distribution is done at random within the interval [0, 1]. The algorithm's efficiency is improved by reducing the inertia weight ω . Weight's definition is (4).

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) \cdot \frac{t}{t_{max}} \tag{4}$$

This indicates the maximum weight with ω_{max} , the minimum weight with ω_{min} , the number of the current iteration with t, the maximum iteration with t_{max} . Two chromosomes from the present generation are integrated using a crossover operator or a mutation operator (MO) to create the offspring. An innovative generation keeps the population size constant. Based on fitness values, certain parents and offspring are chosen, while others are disqualified for creating this generation. There are other options for optimum chromosomes. The proposed model for feature extraction is illustrated in Figure 2.

3.2.1 Feature selection algorithm

Work suggests a hybrid method for sentiment analysis using GA and PSO algorithms. This algorithm is helpful in choosing the optimisation features and encoding a workable solution for a problem into an individual. Each individual is viewed as an entity that supports chromosomal traits, mentioned below is the GA Algorithm 1 for the model:

Algorithm 1: GA – Design for model proposed.

- 1: Initiate: Generation Counter(t) = Zero
- 2: Create population p^0 - random model.
- 3: Analyse and estimate - fitness function for all entities: p^0 .
- 4: Goto step 2 - Recursive loop
- 5: Increment t as t+1.
- 6: Identify an intermediate - population p^t from previous versions p^{t-1} .
- 7: Link a numeral randomly - r within range (0,1) with p^t as each row.
- 8: Compare: $r < p_c$, if [Conditional loop] then
- 9: Register Crossover Parameter to all identified duals of p^t .
- 10: Establish updated p^t .
- 11: Stop - if [Conditional loop]
- 12: Link a numeral randomly - r1 within range (0,1) for every gene in every entity for p^t .
- 13: Compare: $r_1 < p_m$, if then
 - 14: Transmute - Gene via creating a novel merit value - Via randomly identified gene within its own domain. - Mutation operator
- 15: Establish updated p^t .
- 16: Stop - if
- 17: Analyse and estimate - fitness function for all entities p^t .
- 18: Terminate - Only if : Criteria condition = Established.

Procedure 1 (Crossover (p^1, p^2)).

1. Select: Criteria = random - choose $\lambda \in (0,1)$.
2. Offspring = 2 - $c^1 = (c_1^1, \dots, c_b^1)$ and $c^2 = (c_1^2, \dots, c_b^2)$ are generated from parents $p^1 = (p_1^1, \dots, p_b^1)$ and $p^2 = (p_1^2, \dots, p_b^2)$ where

$$c_i^1 = \lambda p_i^1 + (1 - \lambda) p_i^2,$$

$$c_i^2 = \lambda p_i^2 + (1 - \lambda) p_i^1,$$

$$i = 1, \dots, D.$$
4. Return.

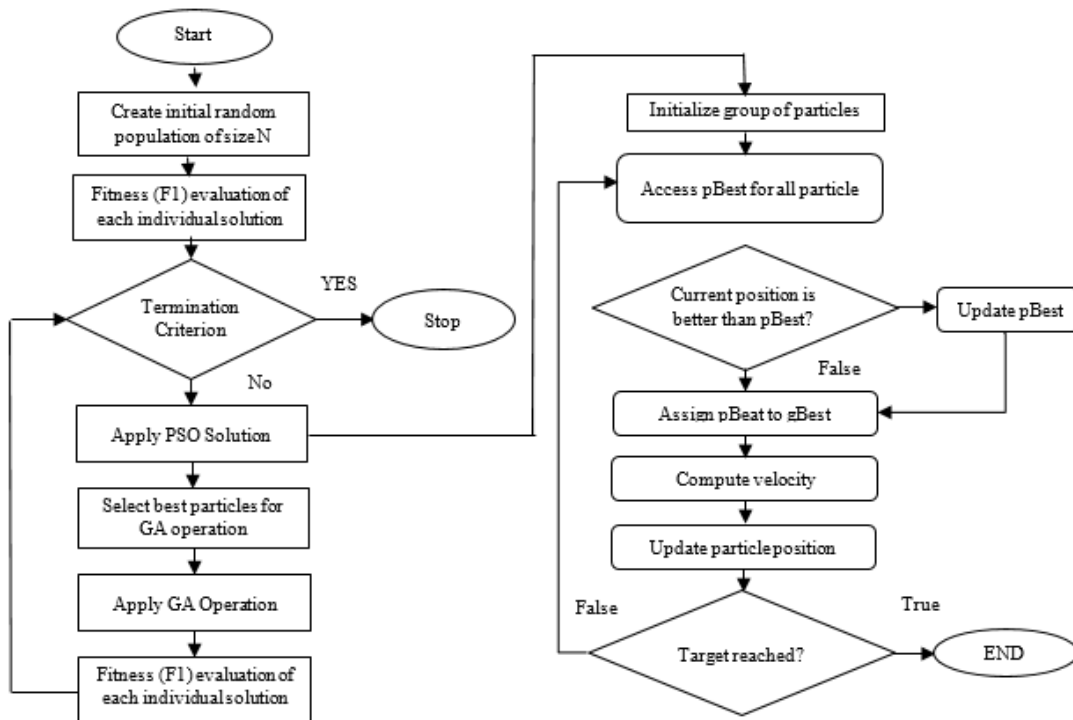


Figure 2. Proposed flowchart

Model proposed uses GA for feature selection, GA serves as an optimum search approach using natural selection and the survival of the fittest, Output from Algorithm 1: GA is processed through PSO technique, as PSO is a fast gradient method, therefore model achieves a faster convergence. Given below is the process in Algorithm 2:

Algorithm 2. Algorithm - designed.

```

1: Initialize: population size = P, Momentum constant  $c_1$  and  $c_2$ , crossover likelihood  $P_c$ , mutation likelihood  $P_m$ , partition numeral  $Part_{no}$ , total variables in every section  $v$ , total solutions in every section  $\eta$  and highest value of iterations  $Max_{itr}$ .
2: Initialize: t = Zero.
3: Loop: for ( $i = 1$  [Value set]:  $i \leq P$  [Condition]) - Continue
4: Create - Random starting population  $X_i(\vec{t})$ .
5: Analyse and estimate - Fitness function for every Agent (search):  $f(\vec{X}_i)$ 
6: Stop - loop
7: Goto step 3
8: Implement: PSO on  $X(\vec{t})$  [Entire population] as depicted in Algorithm 1
9: Implement : GA [selection operator] on  $X(\vec{t})$  [Entire population].
10: Divide population -  $X(\vec{t})$  to  $Part_{no}$  partitions,
    with every partition size as  $v \times \eta$  for all  $X'(\vec{t})$ 
11: Loop for ( $i = 1$  [Value set]:  $i \leq Part_{no}$  [Condition]) continue
12: Implement: arithmetical crossover [ depicted in Procedure 1] for every partition
made as  $X'(\vec{t})$ 
13: stop - loop
14: Implement - GA [mutation operator]:  $X(\vec{t})$  [Entire Population].
15:  $X(\vec{t})$  - update [as per results obtained].
16: Modify:  $t=t+1$ .
17: Keep modifying till ( $t \geq Max_{itr}$ ).
18: Final output - Best optimum solution.
    
```

3.3. Classification

Several classification techniques, including random forest, SVM, and KNN, used in this study for sentiment analysis. The voting classification is the combination of Random Forest, SVM and KNN. The idea of statistical learning is applied in SVM. Without addressing the more complicated problem, this method finds an indirect solution to the primary problem. The bias variance trade-off problem has a comprehensive solution provided by SVM. There are two methods to put this algorithm into practice. While the second uses kernel functions, the first involves mathematical programming. This approach uses kernel functions to split the data into P and N classes. When $y_i = +1$, class P applies, and when $y_i = -1$, class N applies.

With equal distances from each class, SVM looks for the best separation surface, or hyperplane. A combination of tree predictors makes up this method. Almost always, this algorithm produces results that are deemed appropriate. Its performance is hard to improve. This method handles a variety of data types, including nominal, binary, and numerical data. There are several trees built in RF. These trees are merged to provide precise and relevant results. Both regression and classification are handled by the RF algorithm. KNN algorithm is one type of supervised machine learning technique used to solve the prediction issues related to regression and classification.

3.3.1. Voting classifier algorithm

Proposed technique uses a hybrid optimization algorithm combining PSO and GA for feature extraction and employs voting classification, a combination of random forest, SVM and KNN for labeling. SVM provides an exhaustive solution for the bias variance trade-off problem, given below is the Algorithm 3 for voting classification:

Algorithm 3: Voting classification.

```

Input: A Sequence of duals (x, y),
       Features  $\beta \in (0,1)$ 
Output: Sequence of predicted values  $\hat{y}$  for every x
1. Set initial values of  $C_1 \dots C_n$  [experts] with weight  $\omega_i = 1/N$  for each
2. for every x in sequence
   do Collect Predictions  $C_1(x) \dots C_N(x)$ 
    
```

$$P \leftarrow \sum_i \omega_i \cdot C_i(x)$$

$$\hat{y} \leftarrow \text{Sign} \left(P - \frac{1}{2} \right)$$

```

for i ∈ 1...N
do if (Ci(x) ≠ y) then
    ωi ← β · ωi
S ← ∑i ωi
for i ∈ 1...N
do ωi ← ωi/S
    
```

4. RESULTS AND DISCUSSION

The combination of hybrid optimization algorithm combining PSO and GA for feature extraction and voting classification for classification is proposed. The suggested model is evaluated on approx. five datasets and compared with 4 ensemble datasets. 1964 instances are there in Dataset 1 as seen in Table 2, 6439 instances are there in dataset 2 as can be seen in Table 3, 1960 instances are there in dataset 3 as seen in Table 4. These datasets are gathered via the Tweepy API [33] and also gathered from the twitter. Three metrics are used to evaluate each ensemble's and the proposed classifier's performance: accuracy, precision, and recall. As illustrated in Figures 3-5 respectively, the outcomes of the proposed model are contrasted with those of four existing ensemble models on 3 different datasets. The analysis demonstrates that the proposed model achieves superior levels of accuracy, precision, and recall when compared to the other ensemble models.

Hybrid algorithms can achieve faster convergence, work suggests a hybrid model to sentiment analysis which combines GA and PSO algorithm, GA has been employed for identification of best suited prominent sequence of words from random space with concept of survival of fittest for opinion mining, PSO achieve fast convergence and SVM as combination in voting classifier for effective handling of bias variance trade off while classifying sequences. Technique proposed classifies the result by voting classification, a combination of Random Forest, SVM, and KNN.

Table 2. Results on dataset 1

Parameter	Ensemble1 [34]	Ensemble2 [34]	Ensemble3 [34]	Ensemble4 [34]	Hybrid model
Accuracy	80.12	79.87	72.35	82.45	94.45
Precision	80	79	72.34	81.91	94
Recall	81.23	80	72	80	94.12

Table 3. Results on dataset 2

Parameter	Ensemble1 [34]	Ensemble2 [34]	Ensemble3 [34]	Ensemble4 [34]	Hybrid model
Accuracy	84.45	83.87	76.78	84.56	95.56
Precision	84	83	76	84	95
Recall	84	83	76	84	95

Table 4. Results on dataset 3

Parameter	Ensemble1 [34]	Ensemble2 [34]	Ensemble3 [34]	Ensemble4 [34]	Hybrid model
Accuracy	86.12	87.16	86.78	84.46	96.89
Precision	86	87	86	84	96
Recall	86	87	86	84	96

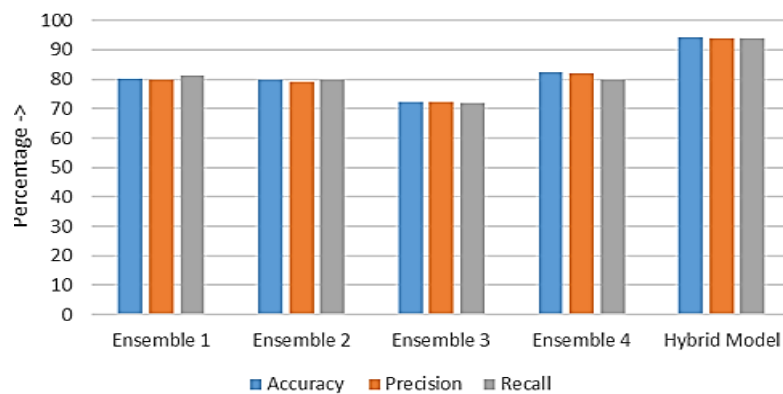


Figure 3. Results on dataset 1

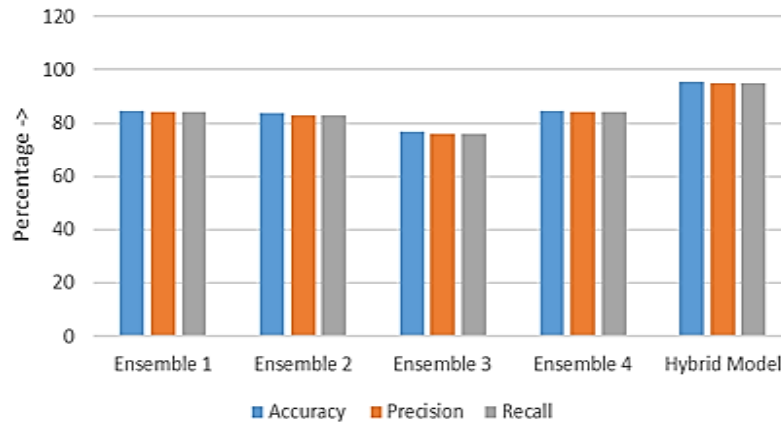


Figure 4. Results on dataset 2

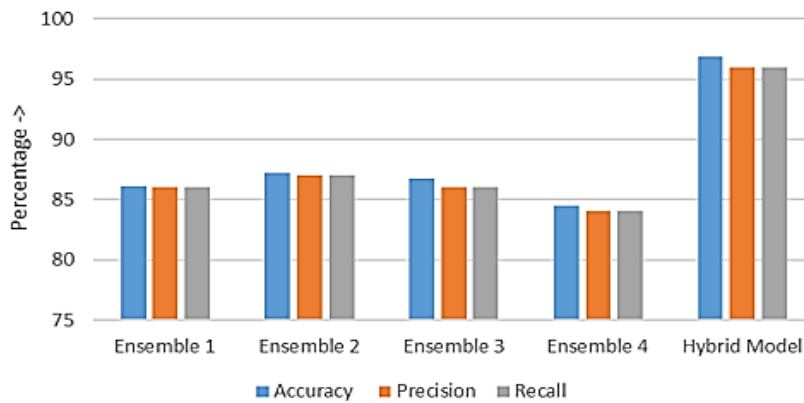


Figure 5. Results on dataset 3

Technique achieves a consistent accuracy percentage over diversified database. The proposed model is compared with 4 ensemble datasets. Dataset 1 achieves an accuracy of 94.45%, dataset 2 attains an accuracy of 95.56% and the dataset 3 attains an accuracy of 96.89%. Technique proposed achieves a consistent accuracy of more than 90% for three different diversified database owing to natural selection of sequences by GA and at the same time achieves a fast convergence with PSO, the model may be employed for highly accurate recommender systems demanding precision and accuracy.

5. CONCLUSION

Social media data can be automatically analyzed to determine the polarity of posted opinions thanks to sentiment analysis technologies. In recent years, these systems have been expanded to examine additional factors, such as a user's attitude towards a subject or their emotions, and have even combined text analytics with other inputs, like multimedia analysis or analysis of social media. It is observed that existing models are unable to achieve high accuracy because those algorithms are unable to extract features properly. In this research work for feature extraction a hybrid algorithm which is the combination of PSO and GA is proposed then for the classification a voting method is applied. The suggested model is evaluated using three datasets with varying numbers of instances and compared with four existing ensemble models to test reliability of the suggested model. It is analyzed from results that suggested model achieve maximum recall, precision and accuracy. The suggested novel model is compared with four ensemble models which are ensemble 1,2,3, and 4. The ensemble models are also the hybrid model for the sentiment analysis. The accuracy of ensemble models is between 75 to 85% approx. on all three datasets and on the other side accuracy of proposed model is between 90% to 95%. These results illustrate that proposed model has approx. 8% to 10% high accuracy as compared to ensemble models for sentiment analysis.




REFERENCES

- [1] E. Daniati, R. Firliana, A. S. Wardani, and A. C. Zarkasi, "Evaluation framework for decision making based on sentiment analysis in social media," in *2021 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation, ICAMIMIA 2021 - Proceeding*, Dec. 2021, pp. 47–51, doi: 10.1109/ICAMIMIA54022.2021.9807790.
- [2] S. Singh, K. Kumar, and B. Kumar, "Sentiment analysis of twitter data using TF-IDF and machine learning techniques," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, COM-IT-CON 2022*, May 2022, pp. 252–255, doi: 10.1109/COM-IT-CON54601.2022.9850477.
- [3] H. C. Soong, N. B. A. Jalil, R. Kumar Ayyasamy, and R. Akbar, "The essential of sentiment analysis and opinion mining in social media : Introduction and survey of the recent approaches and techniques," in *ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, Apr. 2019, pp. 272–277, doi: 10.1109/ISCAIE.2019.8743799.
- [4] J. Xu, Z. Li, F. Huang, C. Li, and P. S. Yu, "Visual sentiment analysis with social relations-guided multiattention networks," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4472–4484, Jun. 2022, doi: 10.1109/TCYB.2020.3027766.
- [5] S. Urolagin, J. Nayak, and U. R. Acharya, "Gabor CNN based intelligent system for visual sentiment analysis of social media data on cloud environment," *IEEE Access*, vol. 10, pp. 132455–132471, 2022, doi: 10.1109/ACCESS.2022.3228263.
- [6] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese text sentiment analysis based on extended sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019, doi: 10.1109/ACCESS.2019.2907772.
- [7] K. Zhang, Y. Zhu, W. Zhang, W. Zhang, and Y. Zhu, "Transfer correlation between textual content to images for sentiment analysis," *IEEE Access*, vol. 8, pp. 35276–35289, 2020, doi: 10.1109/ACCESS.2020.2975036.
- [8] J. Sangeetha and U. Kumaran, "A hybrid optimization algorithm using BiLSTM structure for sentiment analysis," *Measurement: Sensors*, vol. 25, p. 100619, Feb. 2023, doi: 10.1016/j.measen.2022.100619.
- [9] S. Al-Azani and E. S. M. El-Alfy, "Enhanced video analytics for sentiment analysis based on fusing textual, auditory and visual information," *IEEE Access*, vol. 8, pp. 136843–136857, 2020, doi: 10.1109/ACCESS.2020.3011977.
- [10] F. Iqbal *et al.*, "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction," *IEEE Access*, vol. 7, pp. 14637–14652, 2019, doi: 10.1109/ACCESS.2019.2892852.
- [11] R. Cai *et al.*, "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE Access*, vol. 8, pp. 171408–171415, 2020, doi: 10.1109/ACCESS.2020.3024750.
- [12] A. Rodriguez, Y. L. Chen, and C. Argueta, "FADOHS: framework for detection and integration of unstructured data of hate speech on facebook using sentiment and emotion analysis," *IEEE Access*, vol. 10, pp. 22400–22419, 2022, doi: 10.1109/ACCESS.2022.3151098.
- [13] B. Amina and T. Azim, "SCANPEPLENS: a framework for automatic lexicon generation and sentiment analysis of micro blogging data on China Pakistan economic corridor," *IEEE Access*, vol. 7, pp. 133876–133887, 2019, doi: 10.1109/ACCESS.2019.2940528.
- [14] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, p. 100157, Jul. 2022, doi: 10.1016/j.array.2022.100157.
- [15] F. Long, K. Zhou, and W. Ou, "Sentiment analysis of text based on bidirectional LSTM with multi-head attention," *IEEE Access*, vol. 7, pp. 141960–141969, 2019, doi: 10.1109/ACCESS.2019.2942614.
- [16] U. B. Mahadevaswamy and P. Swathi, "Sentiment analysis using bidirectional LSTM network," *Procedia Computer Science*, vol. 218, pp. 45–56, 2022, doi: 10.1016/j.procs.2022.12.400.
- [17] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media," *Information Processing and Management*, vol. 57, no. 6, p. 102290, Nov. 2020, doi: 10.1016/j.ipm.2020.102290.
- [18] A. Pimpalkar and J. R. Raj R, "MBiLSTMGloVe: Embedding GloVe knowledge into the corpus using multi-layer BiLSTM deep learning model for social media sentiment analysis," *Expert Systems with Applications*, vol. 203, p. 117581, Oct. 2022, doi: 10.1016/j.eswa.2022.117581.
- [19] Y. Wang, G. Huang, J. Li, H. Li, Y. Zhou, and H. Jiang, "Refined global word embeddings based on sentiment concept for sentiment analysis," *IEEE Access*, vol. 9, pp. 37075–37085, 2021, doi: 10.1109/ACCESS.2021.3062654.
- [20] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A hybrid model for sentiment analysis with transformer and recurrent neural network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.
- [21] H. He, G. Zhou, and S. Zhao, "Exploring E-commerce product experience based on fusion sentiment analysis method," *IEEE Access*, vol. 10, pp. 110248–110260, 2022, doi: 10.1109/ACCESS.2022.3214752.
- [22] W. Ahmad, H. U. Khan, T. Iqbal, and S. Iqbal, "Attention-based multi-channel gated recurrent neural networks: a novel feature-centric approach for aspect-based sentiment classification," *IEEE Access*, vol. 11, pp. 54408–54427, 2023, doi: 10.1109/ACCESS.2023.3281889.
- [23] S. Nayak, Savita, and Y. K. Sharma, "A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis," *Decision Analytics Journal*, vol. 8, p. 100289, Sep. 2023, doi: 10.1016/j.dajour.2023.100289.
- [24] R. Vatambeti, S. V. Mantena, K. V. D. Kiran, M. Manohar, and C. Manjunath, "Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique," *Cluster Computing*, vol. 27, no. 1, pp. 655–671, Feb. 2024, doi: 10.1007/s10586-023-03970-7.
- [25] Sutriawan, Muljono, Khairunnisa, Z. Alamin, T. A. Lorosae, and S. Ramadhan, "Improving performance sentiment movie review classification using hybrid feature TFIDF, N-Gram, information gain and support vector machine," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 2, pp. 375–384, Feb. 2024, doi: 10.18280/mmep.110209.
- [26] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *Journal of Big Data*, vol. 10, no. 1, Jan. 2023, doi: 10.1186/s40537-022-00680-6.
- [27] D. S. Zingade, R. K. Deshmukh, and D. B. Kadam, "Multi-objective hybrid optimization-based feature selection for sentiment analysis," May 2023, doi: 10.1109/INCET57972.2023.10170147.
- [28] H. Ezaldeen, R. Misra, S. K. Bisoy, R. Alatrash, and R. Priyadarshini, "A hybrid E-learning recommendation integrating adaptive profiling and sentiment analysis," *Journal of Web Semantics*, vol. 72, p. 100700, Apr. 2022, doi: 10.1016/j.websem.2021.100700.
- [29] Q. Zhu, X. Jiang, and R. Ye, "Sentiment analysis of review text based on BiGRU-attention and hybrid CNN," *IEEE Access*, vol. 9, pp. 149077–149088, 2021, doi: 10.1109/ACCESS.2021.3118537.
- [30] M. Chiny, M. Chihab, Y. Chihab, and O. Bencharef, "LSTM, VADER and TF-IDF based hybrid sentiment analysis model," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, pp. 265–275, 2021, doi: 10.14569/IJACSA.2021.0120730.
- [31] X. Z. Gao, M. S. R. Nalluri, K. Kannan, and D. Sinharoy, "Multi-objective optimization of feature selection using hybrid cat swarm optimization," *Science China Technological Sciences*, vol. 64, no. 3, pp. 508–520, Jul. 2021, doi: 10.1007/s11431-019-1607-7.
- [32] M. A. Mosa, "A novel hybrid particle swarm optimization and gravitational search algorithm for multi-objective optimization of text mining," *Applied Soft Computing Journal*, vol. 90, p. 106189, May 2020, doi: 10.1016/j.asoc.2020.106189.




- [33] API — tweepy 4.14.0 documentation. 2024.[Online]. Available: <https://docs.tweepy.org/en/stable/api.html>
- [34] J. Godara, I. Batra, R. Aron, and M. Shabaz, "Ensemble classification approach for sarcasm detection," *Behavioural Neurology*, vol. 2021, pp. 1–13, Nov. 2021, doi: 10.1155/2021/9731519.

BIOGRAPHIES OF AUTHORS






Ms. Garima Srivastava    is currently working as an Assistant Professor in the Department of Computer Science and Engineering, at Amity University, Lucknow. She holds an M.Tech degree in Computer Science and Engineering from Guru Gobind Singh Indraprastha University, Delhi. She has around 15 years of teaching and research experience. Her research interest includes machine learning and NLP. She has published several research papers in National and International Conferences and Journals. She can be contacted at email: gsrivastava1@lko.amity.edu.



Dr. Vaishali Singh    is currently working as an Assistant Professor in Department of Computer Science, Maharishi University of Information Technology, Lucknow. She has received her Ph. D Degree in the year 2017 from B.B. Ambedkar University, Lucknow, India. Her research interest includes information retrieval and question answering systems. She has published many research papers in reputed International Conferences and Journals. She can be contacted at email: singh.vaishali05@gmail.com.



Dr. Sachin Kumar    a Ph.D. in Electronics from Dr R M L University, M.Tech. and B.E in Electronics & Communication from Pune University, Presently with Amity University, Lucknow Campus. He has almost 20 years of academic and research experience. His research interest includes applications of bio-medical signal processing, machine learning, artificial intelligence and applications of wireless sensor networks, He has published more than 80 Research papers, Book chapters, and articles in a number reputed Journals indexed in SCOPUS & SCI. He is an active member of Editorial Board/Reviewer in various reputed journals. He can be contacted at email: skumar3@lko.amity.edu.