Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23
© 2021 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DETECTING MALICIOUS ACTIVITIES ON TWITTER DATA FOR SENTIMENT ANALYSIS USING A NOVEL OPTIMIZED MACHINE LEARNING APPROACH

V. LAXMI NARASAMMA[1], DR. M. SREEDEVI[2]

Department Of Computer Science And Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur, Andhra Pradesh, India-522502

E-mail: [1]laxmi8866@gmail.com, [2]msreedevi_27@kluniversity.in

ABSTRACT

In Natural Language Processing (NLP), Twitter data is used for sentiment analysis and it is most prevalent theme in recent era. However, the security attacks on the Twitter data have been increased by hackers which reduced the performance of the sentiment analysis. Thus to detect the malicious activities in the Twitter data, a novel Spider Monkey based Generalized Intelligent (SMbGI) framework is developed in this paper. This model utilizes Twitter-based data about the coronavirus disease 2019 (COVID-19) to detect the malware activities for improving the classification of sentiments. Moreover, this model imposed a malicious attack on the data for recognizing the developed SMbGI model efficiencies. Thus, the proposed SMbGI approach has been effectually detecting malicious functions and enhances sentiment classification. Moreover, Python tool is used for sentiment analysis, and it computed the parameters like accuracy, recall, precision, F-measure, and error rate. Lastly, the attained outcomes are compared with recent existing works to identify the performance of the SMbGI approach.

Keywords: *Sentiment Analysis, Natural Language Processing, Spider Monkey Optimization, Twitter Data, Malicious Activities*

1. INTRODUCTION

Nowadays, social media is playing an important part in human life and users are posting a lot of information on social media sites like Twitter, YouTube, Facebook, and Instagram [1]. In addition, users get a lot of regretful situations as they upload their personal information and details on social media [2]. The Twitter data involves the opinion of the user, daily activities, ideas, personal details, and experience [3]. Furthermore, the sentimental inquiry based on Twitter's statistics has been cited as a semantic parts of the tweets' scientific study [4]. Sentiment analysis was the technique of obtaining information from multiple social media sources like Twitter; on the basis of user opinions it was classified [5]. In general, tweets reflect public opinion based on specific information about topic or product. Opinion from public is generally classified as neutral, negative and positive tweets [6]. Additionally, if someone continues to follow your messages or tweets may be liked or impressed by a specific person and it may create many security issues for the users [7]. Normally, the opinion of the users is classified into various types like positive sentences, negative sentences, and neutral sentences [9] based on the features of the sentences [8]. Nowadays, the classification of sentiments using Machine Learning (ML) [10] methods by the Twitter data is the most trending part, which is used for easily identifying the originality of the sentences or tweets [11].

Additionally, the existing works analyzed the sentiments using Twitter data about natural disasters, critical events, social movements, tourism, diseases, etc [12]. Also, the developed methods have analyzed the tweets for identifying the polarity of the sentences [13]. Subsequently, emotions are the main thing for classifying the polarity of the sentences in an easy manner [14]. The developed models analyze the trained details like emotions and aspect terms for detecting sentiments [15]. But, the classification of these types in the tweets is difficult because of utilizing a large quantity of data [16]. Additionally, the classification of sentiments may be affected by various kinds of malicious attacks, which creates errors while doing the classification process [17]. Additionally, many approaches were developed to perform opinion classification based on ML and Deep Learning (DL) approaches [18].

Moreover, the existing methods like Bayesian network [21], Artificial Neural Network (ANN) [22], and Recurrent Neural Network (RNN) [25] approaches are utilized for sentiment analysis, but these methods have attained some limitations like high False Positive Rate (FPR), high complexity, long time duration, and less security [19]. Moreover, the existing methods were easily affected by malware that increased the error and reduced the performance of the model [20]. Thus to overcome these issues, a novel optimized ML approach was developed in this research for performing sentiment analysis while detecting malicious activities.

The process of this work is described as follows. The recent works are analyzed and detailed in Section 2 and the problem statement of the work is mentioned in Section 3. Also, Section 4 elaborates on the developed method and the obtained outcomes are discussed in Section 5. Moreover, the research paper was concluded in Section 6.

2. RELATED WORKS

The recent works based on NLP by Twitter data are detailed as below,

Nowadays, sentiment analysis utilizing Twitter data was the most common part that is used for solving problems during serious events like social movements and natural disasters. Here, Gonzalo A. Ruz *et al* [21] introduced a classifier based on the Bayesian network for performing the sentiment analysis. This model has utilized two categories of datasets in Spanish based on Catalan independence and the Chilean earthquake. Additionally, this model used the Baye's factor for providing better classification. Thus, this model has attained better outcomes in prediction while comparing Random Forest (RF) and Support Vector Machines (SVM) techniques. However, this model takes more time to complete the whole process because of large datasets.

Gaurav Kumar *et al* [22] used the ANN for identifying the malware consumer nodes in the data while performing the sentiment analysis. This network analyzed the conversation through chart and remark situation in the social network sites like Twitter, Facebook, and Blog. Also, it performs on the data about the Neet, Demonetization, and Jallikattu. Additionally, the introduced ANN model has effectively differentiated the malicious consumer nodes to increase the security of the sentiments. But, the efficiency of the model is very low because of utilizing various social network sites, which increased the complexity.

The usage of social media like Facebook and Twitter has increased day by day so the security of public opinion is reduced because of malicious users. Here, Zidong Jiang *et al* [23] introduced the troll detection model through sentiment analysis and the activity information about the user on Sina Weibo. Subsequently, this detection model performs the segmentation process on the Chinese sentences, word embedding process and calculates the sentiment score. Thus, the model effectively performed the troll detection while the person browses Sina Weibo.However, the FPR rate of this model is high than other ML approaches.

Xiang Sun *et al* [24] introduced the Event Detection approach using Scoring as well as the Word Embedding model (ED-SWE) for determining the important events in the Twitter data. Also, this model has generated the event summary based on the utilized keywords in a large number of tweets. Here, the word embedding mechanism is employed for identifying the original value of the vector on Twitter data. Additionally, this model detected the live events with high-performance output but the real-time data has increased the difficulty and time duration.

To classify the tweets from Twitter, R. Geetha et al [25] introduced a model based on RNN. This model used a large number of tweets using 23 numbers cyber-keywords that can classify the insensitive and sensitive tweets. Here, the textual features are extracted with the use of word embedding models and auto-encoders. Thus, the developed RNN approach has attained a 75% accuracy value for categorizing the tweets as personally sensitive or insensitive. However, this approach did not analyze all categories of tweets, it used only personal tweets. The summary of the investigated papers is detailed in Table 1.

Author	Year	Methods	Benefits	Demerits
Gonzalo A. Ruz et al [21]	2020	Bayesian network with Bay's Factor (BF)	It effectively classifies the sentiments.	This model utilized more time to complete the entire process.
Gaurav Kumar et al [22]	2020	ANN	It improves the security of the sentiments to classify the malicious nodes.	The complexity of this model is high.
Zidong Jiang et al [23]	2021	troll detection model	It performs troll detection on real-time data.	The FPR value is high.
Xiang Sun et al [24]	2021	ED-SWE	This approach effectively detected the live events and achieved high performance.	This model was utilized a large time duration.
R. Geetha et al [25]	2020	RNN	This approach has effectively classified sensitive tweets or insensitive tweets.	It not possible for processing all categories of Twitter data.

The important steps of the proposed model are detailed as below,

- The user opinions about the COVID-19 tweets are gathered from Twitter that is utilized for processing.
- Subsequently, a novel Spider Monkey based Generalized Intelligent (SMbGI) approach is developed to classify the sentiments and to detect malicious activities.
  - Develop the malicious node in the sentiment analysis to check the efficiency of the proposed SMbGI model.
  - The fitness function of the spider monkey is updated in the classification layer of the Generalized Intelligent framework for monitoring the malicious activities in the sentiments.
  - Hence, the proposed SMbGI has effectively detected malicious functions like DDoS attacks and the sentiments were classified as neutral, negative and positive.
  - Subsequently, the model implementation is done on the Python tool and the calculated metrics are compared with other approaches.

3. PROBLEM DEFINITION AND SYSTEM MODEL

The process of sentiment analysis or subjectivity analysis in NLP has been carried out through the big data attained from social media networking. While considering social media data like Facebook, Instagram, Twitter, etc, the sentiment analysis is mostly done on the Twitter dataset. Moreover, the opinion classification in the tweets is a difficult task because of their complexity. Additionally, these data are affected by various malicious activities like phishing attacks [26].The general process of opinion classification is detailed in fig. 1 that involves the processes of data collection, pre-processing, sentiment classification and the issues when the malware is present in the data. Here, malware detection is necessary to improve the sentiment classification process. Thus, a novel ML approach is developed in this work to detect malware attacks and make the classification process easier.

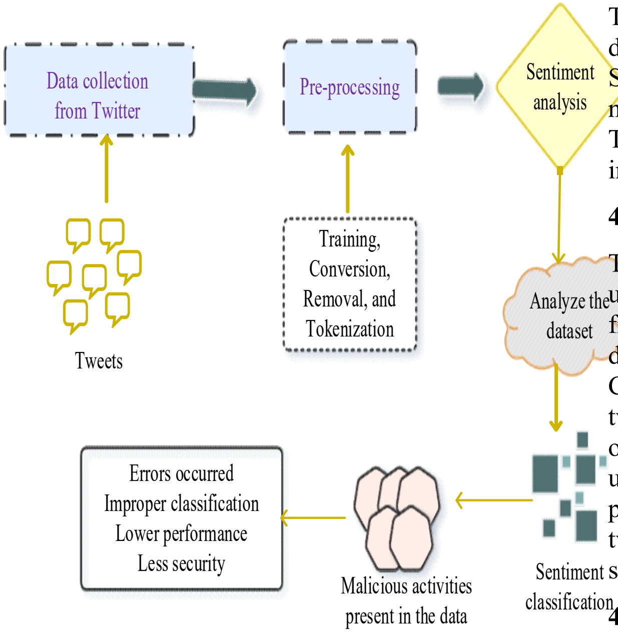

Fig. 1. System Model

4. PROPOSED SMbGI METHODOLOGY

The categorization of sentiments and the security enhancement of tweets from Twitter data is a difficult task. To improve the classification and security of the sentiments, an innovative Spider Monkey based Generalized Intelligent (SMbGI) framework is proposed in this research. In this, the Spider monkey fitness function is utilized in the classification layer of the Generalized Intelligent model to monitor malicious activities. Moreover, the local leader and global leader fitness function of the spider monkey is utilized for monitoring and detecting the malicious activities on the data.

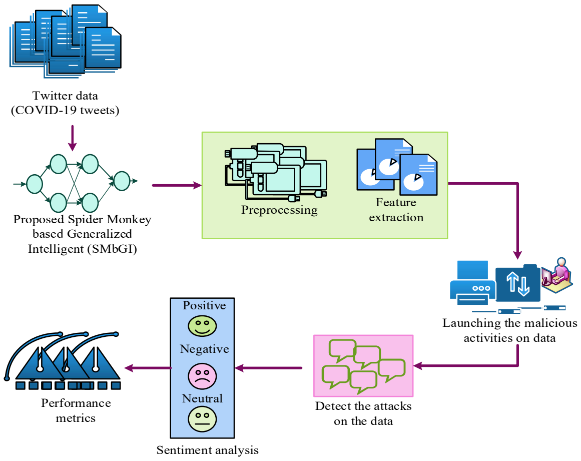
Twitter data
(COVID-19 tweets)
Proposed Spider Monkey based Generalized Intelligent (SMbGI)
Preprocessing
Feature extraction
Launching the malicious activities on data
Detect the attacks on the data
Positive
Negative
Neutral
Sentiment analysis
Performance metrics

Fig-2 Process Of Smbgi Methodology

Thus, the proposed SMbGI model has effectively detected the malware and classifies the sentiments. Subsequently, a malicious attack is imposed in the network to check the proposed model efficiency. The proposed SMbGI overall procedure is detailed in fig.2.

4.1 DATASET DESCRIPTION

This research performs the sentiment analysis using tweets about the COVID-19 disease vaccine from Twitter that are collected from Kaggle. This dataset involves the user opinion about the COVID-19, their name, location, ID, hash tags, tweet date, friends, and followers. Also, the tweets of 38460 numbers of users are in the dataset that is used in this work for analyzing the sentiments. The proposed SMbGI model utilized the collected tweets about COVID-19 for performing the sentiment analysis.

4.2 PROCESS OF SMbGI

In this research, the introduced SMbGI approach is utilized for performing the sentiment analysis and enhancing security. Here, the dataset based on the Twitter data about COVID-19 is used for processing. Also, the proposed technique performs the feature extraction, pre-processing, and malware detection functions and classification, which is detailed in fig. 3. Primarily, the collected dataset is trained to the proposed SMbGIs input layer that is given in eqn. (1)

$$S = \phi(T)_N ; N = 1,2,3..... \qquad (1)$$

where, S denotes the dataset initialization parameter, $\phi(T)_N$ denotes tweets in the dataset, and N represents the quantity of tweets.

- ***Pre-processing and feature extraction***

This step is utilized for diminishing the undesirable information from the collected tweets. Here, the proposed model is used to reduce emojis, special characters, and URLs from the tweets.After the dataset training, preprocessing process and feature extraction are performed on the dataset for removing unnecessary things in the collected tweets. In the feature extraction process, the aspect terms of the dataset are extracted for easily identifying the category of opinion. Thus the preprocessing process is done by eqn. (2),

$$P(T)=(\phi T_N-\delta)(Ap+An) \qquad (2)$$

where, $P(T)$ is the function of performing pre-processing, δ represents the factor of noise, special characters, emojis, and URLs that are removed from the dataset, and Ap, Np is the positive and negative aspect terms, which are extracted for sentiment analysis. Subsequently, the polarity score of each tweet is calculated by the proposed model.

- **Malware detection**

Subsequently, the fitness function of Spider Monkey Optimization (SMO)[27] is initiated in the Action Selection Network (ASN) of the Generalized Intelligent framework to monitor the malicious behaviors in the sentiments.The need of SMO of this model is to monitor and detect malicious activities like DDoS attacks in the tweets while performing sentiment analysis. Initially, the IP address of the input data is stored in the system by the proposed model. Here, DDoS attacks create the fault IP address to hack the data. Subsequently, the proposed method continuously monitors the data, if it is any different attack is present then the neglecting factor remove the attack. If the proposed model identifies the different IP address (attack IP address) on the data then neglects the attack. Here, the population of the spider monkey is considered as the number of tweets. The initialization process of the SM model in the classification layer is given in eqn. (3),

$$K=P(T)*sm_N \qquad (3)$$

Where, $P(T)$ is the preprocessed output and sm_N represents the SMO fitness function, which monitors the sentiments.

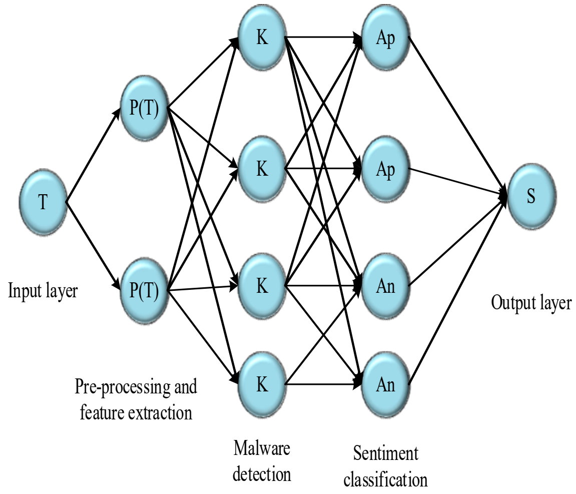

Fig-3 Structure Of Smbgi

The local leader fitness function of SMO is given in eqn.(4), which is used for identifying malicious activities.

$$sm_N = smT+\tilde{U}(0,1)(LL_N)\lambda+\tilde{U}(-1,1)(smT-m_n) \qquad (4)$$

where, LL_N represents the IP address of the malicious tweets, λ is the malware detection and neglecting parameter, m_n is the malicious activities, smT denotes the total quantity of tweets and $\tilde{U}(0,1)$, $\tilde{U}(-1,1)$ denotes the equally distributed arbitrary number for classifying the sentiments.

Algorithm:1 Process ofSMbGI for sentiment analysis

```
Start
{
Initialization ()
    Initialize the dataset $S$ // COVID-19 Tweets from Twitter data
For all $S$ do
Preprocessing()
    Remove noise, repeated words, errors, numbers, URLs, special characters, and stop words
Feature extraction()
    Extract the words features
    Define the Ap and An//positive and negative aspect terms
    Extract the aspect terms and calculate the polarity score
Malware detection()
    Initiate the fitness function of SM
    Store the IP address $(IP_N)$ of the data
    Launch the attack on data
      $m_n \rightarrow fault\_IPaddress(IP_m)$//
      $m_n$ -malicious activity
     Monitor the data using local leader fitness
     If $(IP_N \neq IP_m)$
     then
     Initiate $\lambda$ //detection and neglecting
     parameter
     Remove the malware
     End if
Sentiment Classification()
    Analyze the tweets using Aspect terms Ap and
    An
    If $T_w \rightarrow Ap > An$ then //
    $T_w$ -words in tweets
     $Sentiment \leftarrow P$ //(+1) positive tweet
    Else if $T_w \rightarrow An > Ap$ then
     $Sentiment \leftarrow N$ //(-1) negative tweet
    Otherwise
     $Sentiment \leftarrow N_l$ // (0) neutral tweet
    End if
Secure classification
Calculate performance metrics
}
Stop
```

Subsequently, the malicious activities are imposed on the model to check the efficiency of the proposed approach. Here, the global leader of SMO is utilized for imposing the malicious activities in the data that is given in eqn. (5),

$$(sm_N+1)=smT+\widetilde{U}(0,1)(GL_N)\lambda+\widetilde{U}(-1,1)(smT-m_n) \quad 5$$

Where, GL_N represents the attack launching parameter, λ is the malware detection and neglecting parameter, m_n represents the malicious activities.

If the malware is identified in the location then it is neglected by the proposed model. Moreover, the proposed model analyzes the data using aspect terms to identify the category of tweets. If the tweets have a high number of positive aspect terms then it is considered a positive tweet. Also, if the aspect terms of the tweet are negative more than positive then it is a negative tweet. The remaining tweets are considered neutral tweets. Hence, the sentiments of the entire dataset have been securely classified by the proposed SMBGI model. The entire process of the SMbGI model is explained in algorithm 1 and flow chart is represented in fig.4.

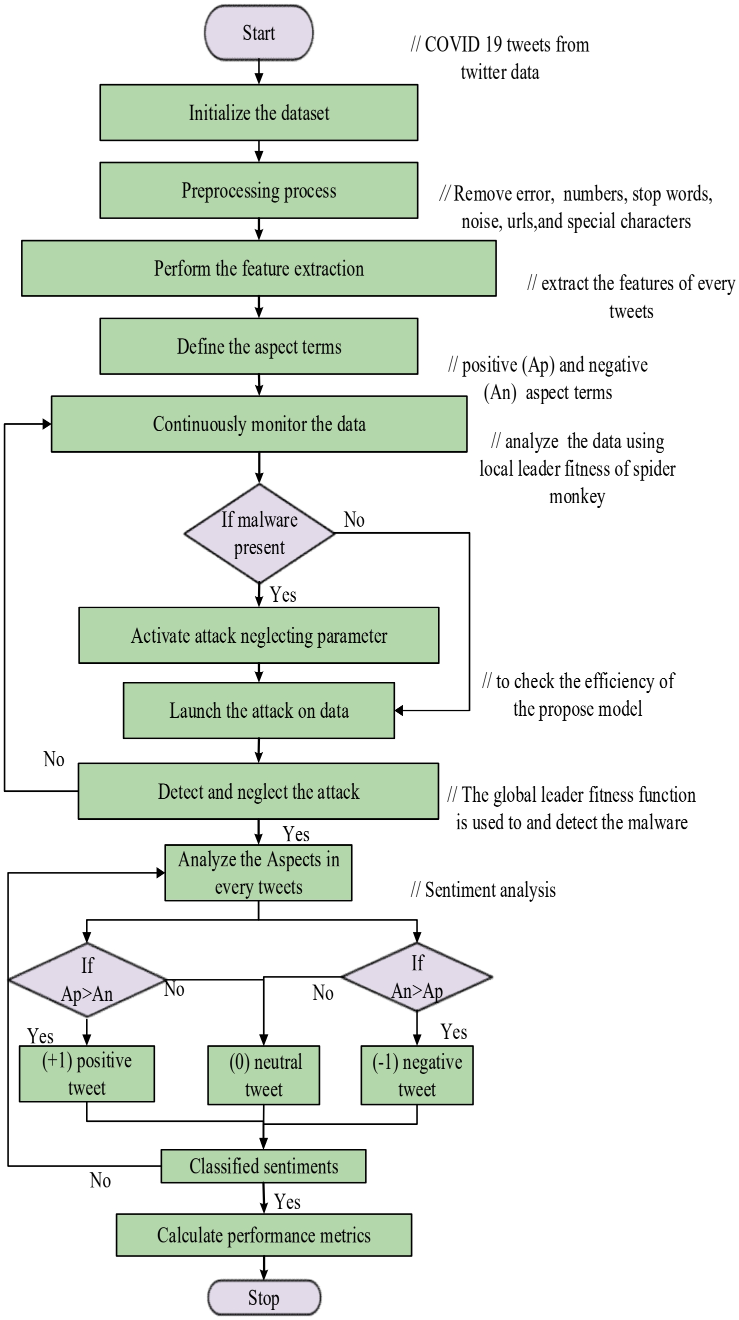
Start
// COVID 19 tweets from twitter data
Initialize the dataset
Preprocessing process
// Remove error, numbers, stop words, noise, urls,and special characters
Perform the feature extraction
// extract the features of every tweets
Define the aspect terms
// positive (Ap) and negative (An) aspect terms
Continuously monitor the data
// analyze the data using local leader fitness of spider monkey
If malware present
No
Yes
Activate attack neglecting parameter
Launch the attack on data
// to check the efficiency of the propose model
No
Detect and neglect the attack
// The global leader fitness function is used to and detect the malware
Yes
Analyze the Aspects in every tweets
// Sentiment analysis
If Ap>An
No
No
If An>Ap
Yes
Yes
(+1) positive tweet
(0) neutral tweet
(-1) negative tweet
Classified sentiments
No
Yes
Calculate performance metrics
Stop

Fig.4 Flow Chart Of The Proposed Model

5. RESULTS AND DISCUSSION

The simulation of this work is done with the use of the Python tool and the performance metrics are calculated. Moreover, the effectiveness of the proposed strategy has been calculated by prevailing works with regards of precision, F-measure, accuracy, recall, and error rate. The proposed SMbGI model has effectively detected the malware activities in the sentiment analysis.

5.1 Case study

The opinion classification and security enhancement using Twitter data are done by the proposed SMbGI model. The tweets about the disease COVID-19 vaccine are possessed to perform sentiment classification. Consider 1000 numbers of tweets about COVID-19 were poised and instructed to the system. The dataset initialization process is done by eqn. (1) and is mentioned in eqn. (6).

$$S = \phi(T)_N ; N = 1,2,3.......,1000 \tag{6}$$

Here, the quantity of tweets is considered as 1000 that are trained to the system, which is done in the initial layer of the generalized intelligence framework.The next layer of the proposed model performs the preprocessing function to remove the errors, noise, stop words and repeated words in the tweets.Subsequently, the aspect terms are defined and extracted in the process of feature extraction. The defined positive aspect terms (Ap) are immunity, vaccine, protect, healthy, etc., and the negative aspect terms (An) are death, side effects, sick, spread, etc. Moreover, the working process of the presented SMbGI technique is graphically represented in fig. 5.

The fitness function of the spider monkey local leader is used for monitoring the data to identify whether the malware is present or not. If the malware is present in the tweet then it provides errors and inaccurate classification of the opinions. Thus, the global leader fitness function is initiated to detect and neglect the attack. Finally, the sentiments are classified in an accurate manner and the proposed model continuously monitors the data if the malware is entered then it is automatically neglected. Thus, the developed SMbGI approach has attained better outcomes for performing malware detection and opinion classification.

5.2 Calculation of performance metrics

In this research, the developed SMbGI approach is used for classifying and securing the sentiments; using Python that are implemented. Furthermore, the interpretation measurements are evaluated and compared with existing methods to identify the effectiveness of the presented method. Therefore, parameters such as precision, error rate and accuracy of the presented approach are checked with prevailing techniques like BF [21], ANN [22], and Convolution Neural Network (CNN) based combined classifier (CF) [28].

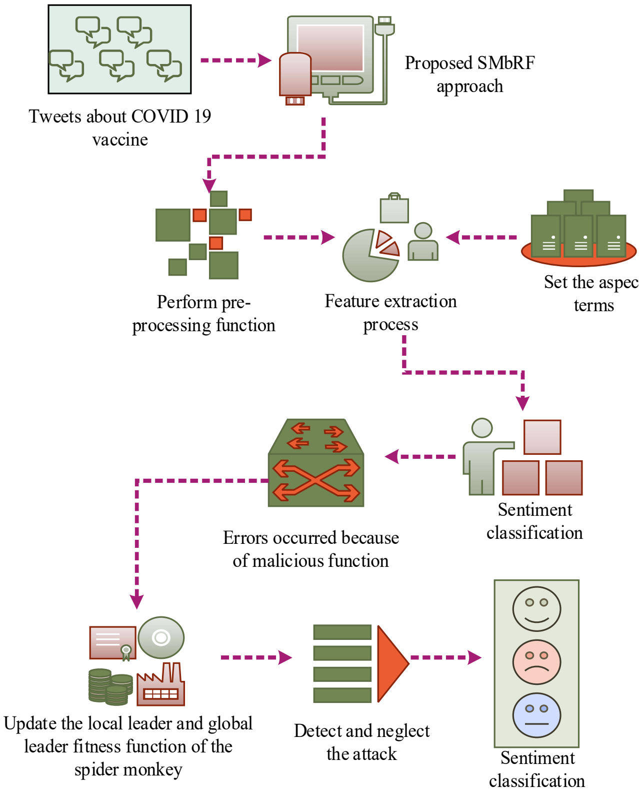
Tweets about COVID 19 vaccine
Proposed SMbRF approach
Perform pre-processing function
Feature extraction process
Set the aspect terms
Errors occurred because of malicious function
Sentiment classification
Update the local leader and global leader fitness function of the spider monkey
Detect and neglect the attack
Sentiment classification

Fig. 5 The Working Process Of The Smbgi Model

5.2.1 Accuracy measurement

For recognizing the effectiveness of the SMbGI while detecting the malicious activities on the data, accuracy calculation is used. The accuracy measurement is calculated based on the classified tweets and the formula is mentioned in eqn. (7),

$$A = \left(\frac{T_P + T_N}{T_P + F_P + F_N + T_N} \right) \tag{7}$$

where, T_P denotes the true positive value that denotes the total quantity of positive tweets which are accurately classified, the true negative value is denoted as T_N which denotes the total amount of negative tweets which are appropriately classified, the false positive value is represented as F_P which describes the total quantity of positive tweets which is incorrectly classified and F_N denotes the false negative value that is the total amount of negative tweets which is incorrectly classified.

Table.2 Comparative Analysis Of Accuracy

No. of tweets	Accuracy (%)			
	BF	ANN	CNN-CF	SMbGI [propos d]
200	85.8	80.19	93.38	99.5
400	82.9	78.08	92.53	98.64
600	80.8	77.63	91.46	98.07
800	80	76.34	89.87	97.86
1000	78	74.76	88.56	97.28

The accuracy calculation of this research is compared with other methods like BF, ANN, and CNN-CF for identifying the performance of the SMbGI model, which is given in Table.2. The existing approaches like BF, ANN, and CNN-CF models were achieved 85%, 80%, and 93% accuracy respectively.

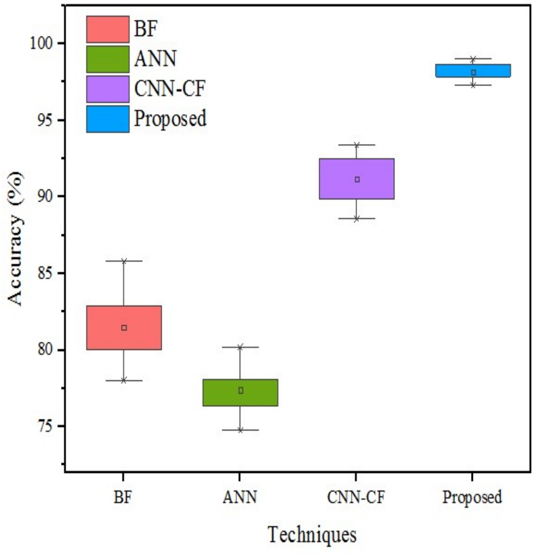

Fig.6 Comparison Of Accuracy Measurement

The proposed approach has obtained an almost high accuracy value of 99% than the other methods, which proves the efficiency of the SMbGI model. The accuracy value comparison is shown in fig. 6.

5.2.2 Precision measurement

The precision value evaluation is used to identify the performance of the presented SMbGI approach while detecting malware functions. Furthermore, the proposed model's precision value is computed using eqn. (8),

$$P = \left(\frac{T_P}{T_P + F_P} \right) \tag{8}$$

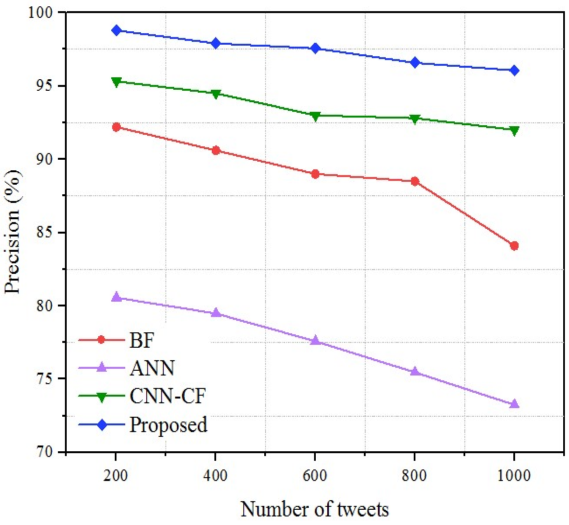

Fig. 7 Precision Measurement Comparison

The calculated precision of the proposed SMbGI methodology is compared with other existing approaches like BF, ANN, and CNN-CF for identifying the performance of the SMbGI model, which is given in Table 3.

Table 3 Comparative analysis of precision

No.of tweets	Precision (%)			
	BF	ANN	CNN-CF	SMbGI [proposed]
200	92.2	80.56	95.31	99.68
400	90.6	79.48	94.5	97.9
600	89	77.59	93	97.56
800	88.5	75.47	92.8	96.58
1000	84.1	73.25	92	96.06

The existing approaches like BF, ANN, and CNN-CF models achieved 92%, 80%, and 95% precision respectively. The proposed approach has acquired high precision value of 98.79% than the other methods and the comparison is illustrated in fig. 7

5.2.3 Recall measurement

The measurement of recall is employed to recognize the sensitivity of the proposed SMbGI approach. It is the ratio of true positive value to the addition of true positive and false negative values, which is calculated using eqn. (9),

$$R = \left(\frac{T_P}{T_P + F_N} \right) \qquad (9)$$

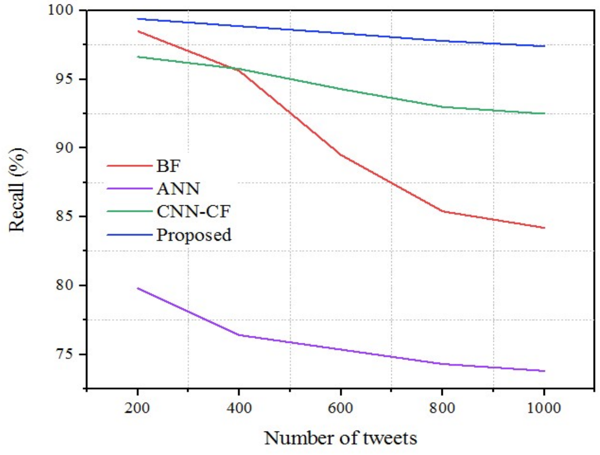

Fig. 8 Recall Measurement Comparison

The calculation of the recall value using the proposed SMbGI methodology is compared with other existent approaches like BF, ANN, and CNN-CF that is given in Table 4. The performance metrics of the proposed model are calculated based on the quantity of tweets.

Table 4 Comparative Analysis Of Recall

No. of tweets	Recall (%)			
	BF	ANN	CNN-CF	SMbGI [proposed]
200	98.5	79.8	96.64	99.79
400	95.6	76.4	95.76	98.87
600	89.5	75.35	94.3	98.36
800	85.4	74.3	93	97.8
1000	84.2	73.8	92.5	97.42

Moreover, the approach BF attained almost 98.5% recall value, and the ANN model attained the lowest recall value of 79.8% for classifying sentiments. Additionally, the existing CNN-CF approach has achieved a 96.64% recall value and the proposed SMbGI approach has obtained a 99.4% high recall value than other methods.Thus, the recall comparison is represented in fig. 8.

5.2.4 F1-measure measurement

The F1 score calculation is defined as the weighted average and recall value of the calculated precision value. Therefore, this computational score takes into account both false positives and false negatives, which is calculated using eqn. (10).

$$F1 - score = \left(2 \frac{P * R}{P + R} \right) \qquad (10)$$

Table 5 Comparative Analysis Of F1-Measure

No. of tweets	F1-measure (%)			
	BF	ANN	CNN-CF	SMbGI [proposed]
200	90.8	78.15	93.12	99.74
400	89.5	76.37	93	98.58
600	88.46	75.38	92.58	97.95
800	87.9	74	92.07	97.18
1000	86.8	73.6	91	96.73

The calculation of the F1-measure detailed the relation among the precision value and recall value that is compared with other existing methods like BF, ANN, and CNN-CF, which is mentioned Table 5.

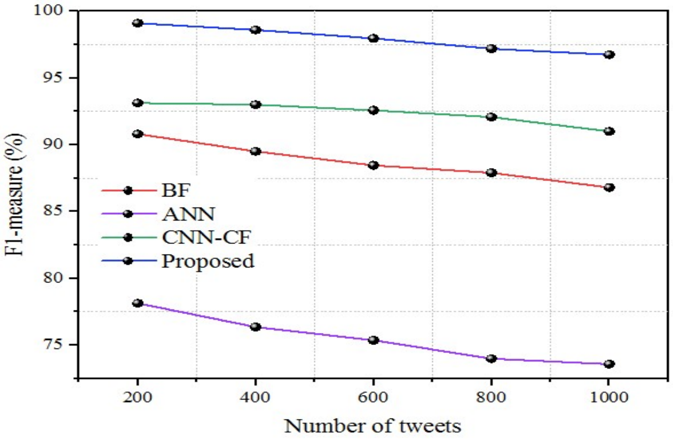

Fig. 9 F1-Measure Calculation Comparison

Moreover, the approach CNN-CF has attained almost 93.12% F1-measure value, and the ANN model has attained the lowest F1-measure value as 78.14% for classifying sentiments. Additionally, the existing BF approach has achieved almost 90% F1-measure value and the proposed SMbGI approach has obtained 99.09% high F1-measure value than other methods, which are shown in fig. 9.

5.2.5 Error rate measurement

The erroneous nature of the predicted results is called the error of the developed SMbGI method.

The error is expressed as an error rate when the target values are definite, which is calculated using eqn.(11).

$$Error\_rate = \left(\frac{F_P + F_N}{T_P + T_N + F_P + F_N} \right) \quad (11)$$

Table.6 Comparative Analysis Of Error Rate

No. of tweets	Error rate (%)			
	BF	ANN	CNN-CF	SMbGI [proposed]
200	0.008	0.06	0.007	0.0025
400	0.0095	0.085	0.01	0.0029
600	0.035	0.094	0.058	0.003
800	0.075	0.15	0.075	0.0035
1000	0.09	0.25	0.08	0.0049

The calculation of the error rate value of the presented SMbGI methodology is compared with other approaches like BF, ANN, and CNN-CF that is given in Table 6 and represented in fig.10. The calculation of error rate is utilized for identifying the efficiency of the model because a high error rate represents lower performance.

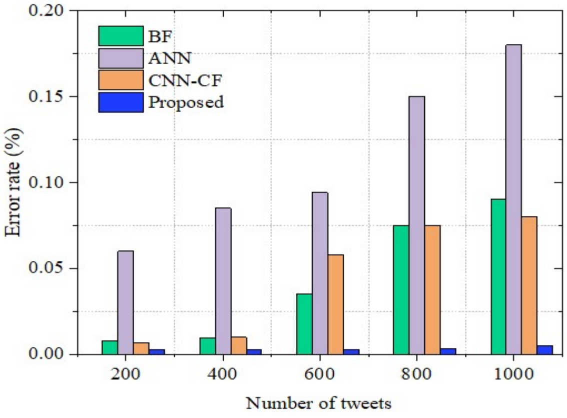

Fig.10 Comparison Of The Error Rate Calculation

6. CONCLUSION

In this research, a novel Spider Monkey-based Generalized Intelligence mechanism is developed for detecting malicious functions while performing sentiment analysis. This approach utilized the tweets about the COVID-19 vaccine collected from Twitter. This model performs preprocessing function to remove the unwanted errors, URLs, and stop words on the tweets. Also, the process of feature extraction is performed to eliminate the aspect terms. Subsequently, the proposed SMbGI effectively analyzed and detected the malicious activities in the data. To check the effectiveness of the proposed SMbGI model, a malware function is launched to the system, which is detected and neglected by the SMbGI approach. Finally, the sentiment classification is done on the tweets that are categorized as negative, neutral and positive tweets. Therefore, the proposed approach has attained a high accuracy value of 99% with a less error rate of 0.002% than other existent models.

REFERENCES

[1] Bairavel, S., and M. Krishnamurthy. "Novel OGBEE-based feature selection and feature-level fusion with MLP neural network for social media multimodal sentiment analysis." Soft Computing 24 (2020): 18431-18445.

[2] Valle-Cruz, David, et al. "Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods." Cognitive Computation (2021): 1-16.

[3] Balakrishnan, Vimala, Shahzaib Khan, and Hamid R. Arabnia. "Improving cyberbullying detection using Twitter users' psychological features and machine learning." Computers & Security 90 (2020): 101710.

[4] Naseem, Usman, et al. "Transformer based deep intelligent contextual embedding for twitter sentiment analysis." Future Generation Computer Systems 113 (2020): 58-69.

[5] Hassonah, Mohammad A., et al. "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter." Knowledge-Based Systems 192 (2020): 105353.

[6] Hassan, Saeed-Ul, et al. "Predicting literature's early impact with sentiment analysis in Twitter." Knowledge-Based Systems 192 (2020): 105383.

[7] Zhai, Wei, Zhong-Ren Peng, and Faxi Yuan. "Examine the effects of neighborhood equity on disaster situational awareness: Harness machine learning and geotagged Twitter data." International Journal of Disaster Risk Reduction 48 (2020): 101611.

[8] Alharbi, Ahmed Sulaiman M., and Elise de Doncker. "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information." Cognitive Systems Research 54 (2019): 50-61.

[9] Sailunaz, Kashfia, and Reda Alhajj. "Emotion and sentiment analysis from Twitter text." Journal of Computational Science 36 (2019): 101003.

10.Bhaumik, Ujjayanta, and Dharmveer Kumar Yadav. "Sentiment Analysis Using Twitter." Computational Intelligence and Machine Learning. Springer, Singapore, 2021. 59-66.

11.Nagarajan, Senthil Murugan, and Usha Devi Gandhi. "Classifying streaming of Twitter data based on sentiment analysis using hybridization." Neural Computing and Applications 31.5 (2019): 1425-1433.

[10] Dashtipour, Kia, et al. "An ensemble based classification approach for persian sentiment analysis." Progresses in Artificial Intelligence and Neural Systems. Springer, Singapore, 2021. 207-215.

[11] Bisht, Akanksha, et al. "Detection of hate speech and offensive language in twitter data using lstm model." Recent Trends in Image and Signal Processing in Computer Vision. Springer, Singapore, 2020. 243-264.

[12] Behera, Ranjan Kumar, et al. "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data." Information Processing & Management 58.1 (2021): 102435.

[13] Praveen, S. V., Rajesh Ittamalla, and Gerard Deepak. "Analyzing Indian general public's perspective on anxiety, stress and trauma during Covid-19-A machine learning study of 840,000 tweets." Diabetes & Metabolic Syndrome: Clinical Research & Reviews (2021).

[14] Vashishtha, Srishti, and Seba Susan. "Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis." Expert Systems with Applications 169 (2021): 114323.

[15] Hew, Khe Foon, et al. "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach." Computers & Education 145 (2020): 103724.

[16] Kumar, Akshi, et al. "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data." Information Processing & Management 57.1 (2020): 102141.

[17] Greco, Francesca, and Alessandro Polli. "Security perception and people well-being." Social Indicators Research (2020): 1-18.

[18] Basha, C. Bagath, and S. Rajaprakash. "Enhancing the security using SRB18 method of Embedding Computing." Microprocessors and Microsystems 77 (2020): 103125.

[19] Ruz, Gonzalo A., Pablo A. Henríquez, and Aldo Mascareño. "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers." Future Generation Computer Systems 106 (2020): 92-104.

[20] Kumar, Gaurav, and V. Rishiwal. "Malicious User Nodes Detection by Web Mining Based Artificial Intelligence Technique." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 28.01 (2020): 1-24.

[21] Jiang, Zidong, Fabio Di Troia, and Mark Stamp. "Sentiment Analysis for Troll Detection on Weibo." Malware Analysis Using Artificial Intelligence and Deep Learning. Springer, Cham, 2021. 555-579.

[22] Sun, Xiang, et al. "ED-SWE: event detection based on scoring and word embedding in online social networks for the internet of people." Digital Communications and Networks (2021).

[23] Geetha, R., S. Karthika, and S. Mohanavalli. "Tweet Classification Using Deep Learning Approach to Predict Sensitive Personal Data." Advances in Electrical and Computer

Technologies. Springer, Singapore, 2020. 171-180.

[24] Djaballah, Kamel Ahsene, et al. "A new approach for the detection and analysis of phishing in social networks: the case of Twitter." 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2020.

[25] Shekhawat, Sayar Singh, Sakshi Shringi, and Harish Sharma. "Twitter sentiment analysis using hybrid Spider Monkey optimization method." Evolutionary Intelligence (2020): 1-10.

[26] Alom, Zulfikar, Barbara Carminati, and Elena Ferrari. "A deep learning model for Twitter spam detection." Online Social Networks and Media 18 (2020): 100079.