Journal of Scientific & Industrial Research
Vol. 82, January 2023, pp. 93-100
DOI: 10.56042/jsir.v82i1.69943

Using a Novel Hybrid Krill Herd and Bat based Recurrent Replica to Estimate the Sentiment Values of Twitter based Political Data

I Lakshmi Manikyamba[1]* & A Krishna Mohan[2]

[1]Department of Computer Science and Engineering, JNTUK, Kakinada, Andhra Pradesh, 533 003 India
[2]Department of Computer Science and Engineering, UCEK, JNTUK, Kakinada, Andhra Pradesh, 533 003 India

Received 20 May 2022; revised 08 September 2022; accepted 07 October 2022

Big data is an essential part of the world since it is directly applicable to many functions. Twitter is an essential social network or big data replicating political information. However, big data sentiment analysis in opinion mining is challenging for complex information. In this approach, the Twitter-based political datasets are taken as input. Furthermore, the sentiment analysis of twitter-based political multilingual datasets like Hindi and English is not easy because of the complicated data. Therefore, this paper introduces a novel Hybrid Krill Herd and Bat-based Recurrent Replica (HKHBRR) to evaluate the sentiment values of twitter-based political data. Here, the fitness functions of the krill herd and bat optimization model are initialized in the dense layer to enhance the accuracy, precision, etc., and also reduce the error rate. Initially, Twitter-based political datasets are taken as input, and these collected datasets are also trained to this proposed approach. Moreover, the proposed deep learning technique is implemented in the Python framework. Thus, the outcomes of the developed model are compared with existing techniques and have attained the finest results of 98.68% accuracy and 0.5% error.

Keywords: Big data, Multilingual datasets, Opinion mining, Sentiment analysis, Text summarization

Introduction

In today's scenario, expressing ideas and thoughts on social networks has become a trending topic.[1] That social network is termed "big data." Moreover, all information is arranged in either a structured or unstructured format.[2] It makes the feature extraction function and the manual text valuation a difficult.[3] So, Natural Language Processing (NLP) has been introduced in the big data industry; hence, the NLP paradigm offers machine-human interaction communication without human risk.[4] Furthermore, on social sites, Twitter is the most popular commutation field for sharing crucial info as short messages that are described as tweets.[5] If anyone is attracted to a particular tweet, they will continuously follow the specific user's tweets.[6]

Moreover, the tweets are received from worldwide users. So, the NLP approach can evaluate the opinion value in all languages.[7] In India, the tweets are mostly in English and Hindi, so categorizing the sentiment value of each tweet is more difficult because of the different language corpus.[8] Henceforth, to understand different languages, a multilingual framework should be designed in the NLP frame[9], which means the developed ML or DL can answer or classify multilingual text.[10] Aside from these, social networking sites assist the public in effectively analyzing social events with users worldwide.[11]

Furthermore, in NLP, sentiment, or opinion, emotion value estimation is the key paradigm to validate the emotion of each text.[12] Also, Machine learning (ML) and deep learning (DL) techniques are used to verify the sentiment analysis.[13] Moreover, estimating customer reviews in an online business is the key to improving the online business.[14] Because, based on customer reviews, only online business is managed.[15] In online marketing, every product review is more critical because only the sold rating is increased based on the product reviews.[2] The buyer often analyses the product reviews before buying a particular product.[16] Conversely, an optimization framework is implemented in the ML module to maximize emotion categorization accuracy.[17] Here, The heuristic frame model's objective function is used as input to the ML classification model. As a result, the precision of the opinion specification has increased.[18] The text-driven-based opinion estimation has been elaborately described for both unstructured and structured benchmark tweet datasets. In other cases, visual images are utilized as the training datasets. In the

Author for Correspondence
E-mail: lakshmi.isit@gmail.com

online application, introducing the sentiment scheme is a much-needed model to understand people's thoughts and improve the service based on them. It helps me earn more money from my online business.

Numerous recent ML and DL models, such as Bayesian networks[19], transform replicas[20], etc., have been introduced to end this opinion classification issue. However, the specification is still challenging because of the wide range of data. So, the present research paper focused on ending this opinion classification problem by designing a hybrid optimized deep learning framework. Moreover, the twitter-based political dataset as an English-Hindi corpus is taken to estimate the proposed replica.

Related Work

Few recent associated works of sentiment analysis are detailed as follows:

Sentiment value estimation for tweets dataset is the finest topic in the big data industry. So, Ruz *et al.*[19] developed a Bayesian network frame to calculate the opinion measure in Twitter datasets. It has pertained to a better emotional estimation accuracy score. Moreover, its proficient value is determined by developing the parameter assessment frame with other old replicas. However, the projected approach is not suitable for multilingual text.

A transformer replica estimate Twitter data's opinion measure.[20] Moreover, this bidirectional layer is designed as a memory frame to store the data. Here, the key focus of this transform model is to explain the deep emotions that are present in the tweet data. But it takes more time to execute the functions.

In other cases, this sentiment value estimation is effectively utilized in the stock market for selling the product which Elaborated sentiment classification research for stock market data.[21] In this case, the stock market-based Twitter statistics are trained into the system. The current sentiment value in Twitter datasets is then specified using a support vector-based regression model. Finally, its parameters are validated with other approaches and have gained the best outcome. However, it is suitable for multilingual languages.

The multilingual dataset becomes challenging in most NLP problems because of complicated training samples. A convolution neural frame context model[22] helps to evaluate the sentiment value with a high exactness measure to overcome this dataset issue. Subsequently, the projected frame has obtained 1.3% of sentiment classification accuracy. However, it required more time for process execution.

For this purpose, the prediction model utilizes a recurrent and neural convolution model to infer the sentiment of tweets.[23] Initially, the entire dataset is filtered using the preprocessing module. Consequently, the prediction of emotion is processed in the dense frame of the recurrent model. It has also gained a high sentiment specification accuracy of 93.78%. However, the developed paradigm is not suitable for the multilingual module.

A hybrid strategy called a cat swarm optimization-based long short-term memory neural network (CSO-LSTMNN) improves the sentiment classification rate.[24] As a result, it has gained the highest exactness rate of sentiment classification in the multilingual database, but it has taken more time than conventional optimization.

The ensemble learning replica categorizes the sentiment value in the social media database. Several methods are used to classify the opinions from the dataset, such as support vector classification, optimization, etc.[25] Hence, it has gained 88.2% accuracy. But it is complex in design.

System Model and Problem Statement

Sentiment Analysis (SA) is the process of evaluating the sentiment score on every tweet attitude at the back of a sequence of languages that is otherwise called opinion mining. In SA, initially, the information is collected from the big data based on political datasets, opinions, etc. Henceforth, the preprocessing stage can remove all unwanted information present in the collected tweets, such as punctuation, repeated words, symbols, numbers, and URLs, to train the information.

Then calculate the sentiment analysis value based on three factors, such as positive, negative, and neutral. Usually, opinion specification is a challenging task in big data because of complicated data. Furthermore, the sentiment analysis of twitter-based political multilingual datasets like Hindi and English is not at all an easy task because of the complicated data. Also, the tweets are very short messages with many meanings and emotions. So, in rare cases, The ML model could not foresee the tweet's underlying emotional state. A deep learning strategy is introduced, though the DL model takes more time to execute the task. The main motive of the sentiment classification in Twitter data is that short sentences are called "tweets." So, from the tweets, identifying the sentiment value is difficult. Researchers have already done much work to end these issues, but it is still no end. Also, the error is often raised in sentiment classification because of large, complex, noisy data.

Researchers today are trying to find a scientific way to improve political data analytics by using sentiment analysis in Natural Language Processing to cut down on all kinds of problems. So, hybrid optimization is suggested to increase the rate of correct classification and decrease the rate of wrong classification.

Proposed HKHBRR Methodology

An HKHBRR is planned for this current work to design and analyze the sentiment value in political tweets. Here, the combined corpus Hindi-English dataset validates the proposed design performance. Initially, the system is used to train the dataset, after which a brand new HKHBRR is developed. To classify the tone of each tweet, we update the fitness function of krill heard and Bat in a recurrent dense layer. The proposed layout is illustrated in Fig. 1 in greater detail.

Data Collection

Customers commonly share their emotions in public forums like discussion boards, polls, and item surveys on their private blogs on social networking sites like Facebook and Twitter. The material is vast and complicated because sentiments and feelings are expressed utilizing a variety of jargon, composition settings, short structures, and slang. The manual study of opinion data reveals that it is essentially unintelligible.

Preprocessing

The initial step of sentiment analysis is preprocessing. Many approaches are applied in datasets to diminish unwanted data and measurement methodologies and support the development of the organization's efficiency. Moreover, the preprocessing stage contains,

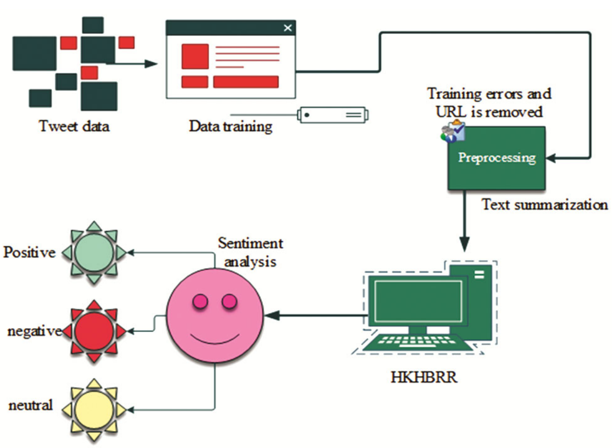

Fig. 1 — Proposed methodology

- Training
- Conversion
- Removal
- Tokenization

In this preprocessing layer, the datasets are tokenized and summarized. Moreover, tokenization is the process of exchanging paragraphs into a sentence and also removes unwanted information. Furthermore, before broadcasting the vector, the converted text is changed into the token.

Text Summarization

At this stage, there are three main steps for summarising the text which are mentioned as recognition, elimination, and summary creation. Initially, error-free content is recognized from the collected dataset. The text is changed to the sentence, and that sentence is denoted as a vector form that is f1, f2, etc. The f1 and f2 are likewise titles, location, the first line of the sentence, sentence length, etc. Then the summarised data is created in the text summarization stage.

Sentiment Detection

At this point, the subjectivity of each sentence in the audit and assessment is checked. Sentences containing emotional articulations are gathered, and the sentence is then forwarded to the intended. Sentiment analysis is carried out at several levels using conventional computer techniques. SA can be broadly categorized into positive and negative elements. Every sentence opinion gathered during this stage of the SA process is identified and categorized as positive, negative, great, and dreadful. Transforming unstructured input into significant data is the central concept of assumption examination. Following the investigation's conclusion, the findings are displayed on diagrams such as pie charts, bar graphs, and line charts.

Process of HKHBRR

Initially, the collected Twitter-based political datasets are trained on the system, which has a python environment. The proposed approach removes unwanted information, such as dots, commas, etc., present in the collected Twitter-based political datasets. The error-removing layer is detailed in Eq. (1)

$$E_r = \sum_{n=1}^{k}[L_{n,i} = L_{n,i}]P_s / \sum_{n=1}^{k}[L_{n',i} = L_{n,i}] \qquad \ldots (1)$$

where, L is the total content taken from the Twitter datasets, is the particular content, and k represents every word present in the collected dataset. Moreover, the opinion specification procedure is more straightforward, while errors are eliminated. After removing the unwanted data, a new solution is determined by Eq. (2).

$$D_n = [L_n^j / L_n^k] + (R-1) \quad \ldots (2)$$

where, D_n represents new datasets, L_n^j is denoted as a new content after removing the error, L_n^k means a new selected specified sentence, and R is the duplicated or unwanted sentence.

The proposed HKHBRR approach is the predicting mechanism and then calculates the sentiment analysis value after removing the error. Sentiment analysis is determined by Eq. (3).

$$S_A = \{E_R[La^*] + D_n[(L_n^j / L_n^k) + (R-1)] \quad \ldots (3)$$

where, S_A represents sentiment analysis D_n represents new datasets represents error removing dataset. FS represents Feature Selection and TNFS represents Total Number of Feature Selections. FS mainly reduces the input variable numbers in predictive model development. In that case, the input variable is decreased to lower the cost of computation and improve the performance of the model.

The Twitter-based political datasets are trained on the system as shown in Fig. 2. Primarily, the dataset's unwanted information and repeated words are removed by the proposed technique. Therefore, the filtered error-free data is updated to the HKHBRR fitness function, which is imported to the classification layer. The proposed HKHBRR approach is detailed in Algorithm 1 to evaluate the sentiment values of political-based Twitter data.

Algorithm 1 — Proposed HKHBRR Approach

Start

Train the datasets — *//twitter-based political multilingual dataset (Hindi and English) is taken as input*

Initialize the parameter — *// P_s, k, L represents particular content, all collected datasets, total contents*

//Text summarization
// removal of Unwanted information

If E_R =[$L_{n,i} = L_{n,i}$] P_s — *// occurrence of unwanted information is removed using the eqn. (1)*

Analyze and remove the unwanted data

Else

Return to estimation step

// Eliminate the repeated words

If($D_n = 0$***)*** — *// Repeated words are eliminated using the eqn. (2)*

Repeated words are removed

Else

Return to calculation step

End if

// Calculate the sentiment analysis of the given filtered dataset

Update the Parameters — *// P, N, N_l, D_p is denoted as Positive, negative, neutral*

// filtered data is trained in the HKHBRR approach — *// twitter-based political datasets are trained into the*

Randomly update the function using the HKHBRR equation

$S_A = \{E_R[La^*] + D_n[$ — *// estimate the sentiment analysis value*

//Update the D_p ***as Twitter-based political datasets***

// identify the sentiment analysis (positive, negative, and neutral) — *//twitter-based political datasets*

If (S_A***=1*** $\rightarrow$ ***P)then*** every sentence positive — *Positive*

Else

Return the finest classification

If(S_A***=0*** $\rightarrow$ ***N), then*** every sentence negative — *negative*

Else

Return the finest classification

If (S_A***=-1*** $\rightarrow$ N_l***), then*** every sentence neutral — $N_l \rightarrow$ Neutral

Else

Return the finest classification

End if

Best possible solution

Stop

Initially, the Twitter-based political multilingual (English and Hindi) datasets are taken as input, and these datasets are also trained into the proposed HKHBRR approach. Then this trained dataset is entered into the preprocessing stage, and the unwanted information and repeated words are removed based on the fitness function. Hereafter, the sentiment score is estimated from the filtered dataset. The flow chart of the proposed HKHBRR mechanism is shown in Fig. 3.

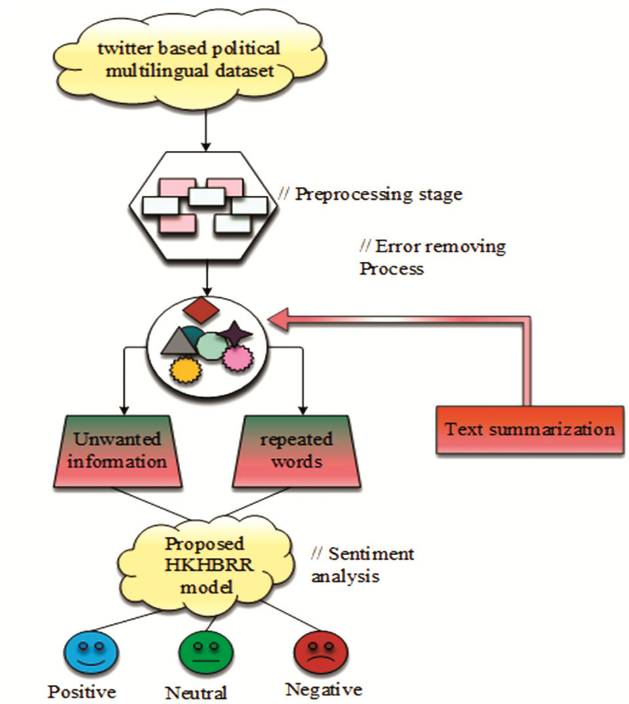

Fig. 2 — Process of proposed HKHBRR approach

Results and Discussion

Big data needs NLP, especially for sentiment analysis. The success of the proposed replica is predicted in this study using the political tweet datasets, which are made up of various political-related data. Additionally, the Python framework incorporates the proposed HKHBRR. The effective method that has been suggested is utilized to improve the classification of sentiment into positive, negative, and neutral. The proposed replica also performs better and is implemented in the Python framework. Based on the fact that neither positive nor negative words are present in the sentence, neutral sentences are categorized here. Positive aspect terms categorize positive sentences, whereas negative aspect terms classify negative statements.

Case Study

Political datasets based on Twitter are used to assess how well the replica has worked. Table 1 displays the gathered data sets.

Let us consider 30 contents taken from the Twitter-based political dataset with its word arrangement is

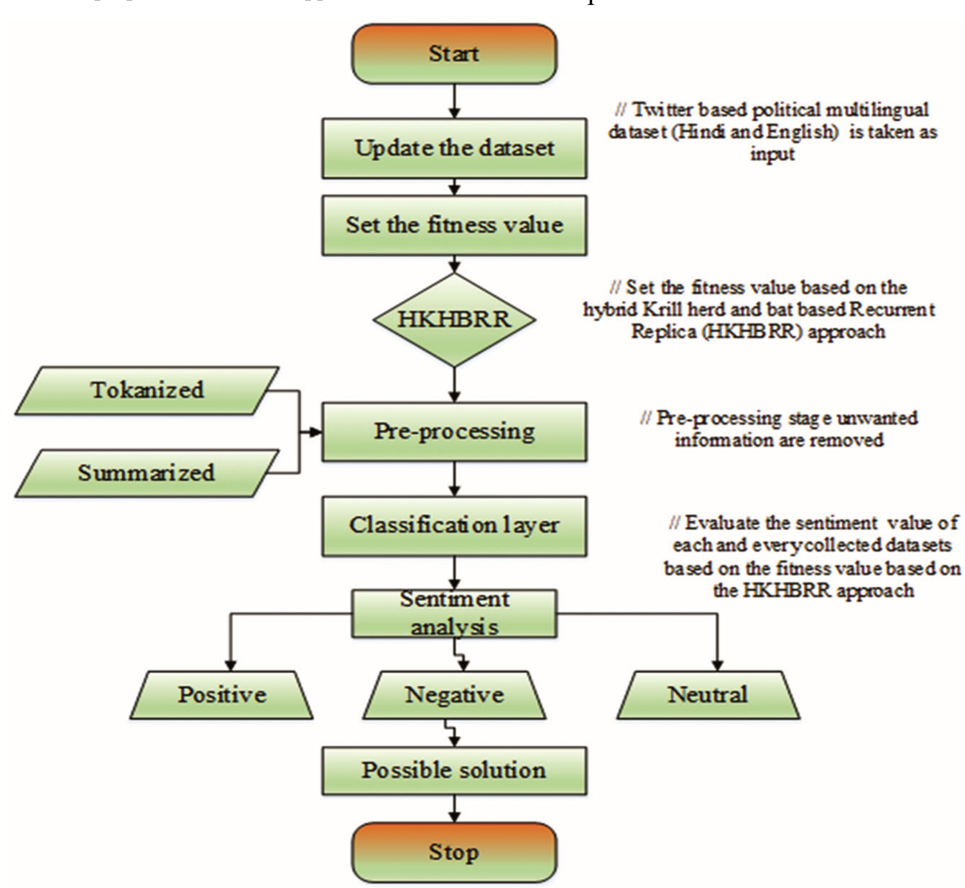

Fig. 3 — Flowchart for the proposed approach

Table 1 — Twitter-based political multilingual dataset

Sl. No	Twitter-based political dataset English	Hindi	Positive	Negative	Neutral
1	Dr. Jason Johnson, do you know why most of the stories never made much sense to me? because Joe Biden was in the White house.	डॉ.जैसन जॉनसन आपजानतेहैं कि ज्यादातर कहानियों ने मुझे कभी ज्यादा प्रभावित क्यों नहीं किया क्योंकि जो बिडेन व्हाइट हाउस में थे	—	—	−1
2	To provide the vote for Joe Bidden and Puerto Rico should never disregard American president Donald Trump's inactions that lead to the defeat	जोबिडेन और प्यूर्टोरिको के लिए वोट प्रदान करने के लिए अमेरिकी राष्ट्रपति डोनाल्ड ट्रम्प की निष्क्रियता की कभी हार नहीं माननी चाहिए	—	0	—
3	The US Presidential Elections are planned for the third of November and with not exactly seven days left	अमेरिकी राष्ट्रपति चुनावों की योजना नवंबर के तीसरे और ठीक सात दिन शेषन हीं है	1	—	—
4	The present president of Republicans was Donald Trump and Democratic contestant Joe Biden.	रिपब्लिकन के वर्तमान राष्ट्रपति डोनाल्डट्रम्प और डेमोक्रेटिक प्रतियोगी जो बिडेन थे।	1	—	—
5	The selection process of the president will be held at the pass this as last Tuesday, which is one week before the elections are to be held.	राष्ट्रपति की चयन प्रक्रिया पिछले मंगलवार को पारित की जाएगी, जो चुनाव होने से एक सप्ताह पहले होगी।	—	—	−1

Table 2 — Comparison of performance measures

Technique	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	Error Rate
CSO-LSTMNN	96.89	0.73	0.72	0.75	2.54
HELM	88	0.87	0.82	0.84	1.43
ACO-PSO	96	0.79	0.78	0.85	0.90
PSO	89	0.57	0.65	0.87	0.73
Proposed approach	98.68	0.90	0.90	0.90	0.50

300. L=30, n=300, and P_s=2. Apply these values in Eq. (1) and attained Eq. (4),

$$E_r = \sum_{1=1}^{k}[30_{2,300} = 30_{2,300}]\,2 / \sum_{1-1}^{k}[30_{2.300} = 30_{2.300}] \quad \ldots (4)$$

Then calculate the error-free new dataset based on the repeated words from Twitter-based political data. Consider L_n^j=2, L_n^k=300, and R=3, these values substitute in Eq. (2) and attained Eq. (5),

$$D_n = [2/300] + (3 - 1) = 2.05 \quad \ldots (5)$$

From this procedure, the sentence length is improved. Then calculate the sentiment analysis on the trained data.

Here, $E_r = 1, D_n = 2$ these values are substituted in Eq. (3) and obtained Eq. (6),

$$S_A = \{1[\,La^*\,] + 2.05[2/300] + (3 - 1)\} \quad \ldots (6)$$

The sentiment values are denoted as S_A

$$= \begin{cases} 0 & \rightarrow negative \\ 1 & \rightarrow positive \\ -1 & \rightarrow neutral \end{cases}$$

The sentiment analysis result is a negative one. Its value is zero, and it is considered to be neutral. Finally, if the value equals 1, it is said to be positive. Otherwise, it is said to be negative.

Performance Evaluation

The intended strategy is developed in Python, and its efficiency is measured in terms of accuracy, precision, recall, F-measure, and error rate in comparison to other methods. ACO-PSO algorithms, CSO-LSTMNN, Hybrid Ensemble Learning Model (HELM), Particle Swarm Optimization (PSO), and HELM are some other existing methodologies with which the efficiency of the suggested replica is also compared.[24–27]

Accuracy (A)

The major three classifiers used to assess the accuracy of sentiment analysis are positive, negative, and neutral. Moreover, the accuracy of the sentiment value can be determined using the Eq. (7).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad \ldots (7)$$

where, TP stands for "True Positive," TN for "True Negative," FP for "False Positive," and FN for "False Negative," respectively.

The CSO-LSTMNN achieved accuracy measure is 96.89%, HELM attained 88% accuracy, the ACO-PSO algorithm earned a 96% accuracy measure, and the PSO predicted accuracy rate is 89%. As a result, the proposed method worked 98.68% of the time, and values are shown in Table 2.

Precision (P)

More fake positives are implied by lower precision, while fewer bogus positives are implied by higher accuracy. This is frequently a possibility with the review because reducing recall is a straightforward way to increase precision. The precision rate of sentiment analysis is calculated using an Eq. (8).

$$Precision = \frac{TP}{TP + FP} \quad \ldots (8)$$

A comparison of precision measurement with existing approaches is presented in Table 2. The precision rate of the CSO-LSTMNN approach is 0.73, HELM's achieved precision measure is 0.87, the attained precision rate ACO-PSO is 0.79, PSO predicts 0.57 of the proposed replica's precision rate, and the proposed replica has achieved 0.90 precision measures.

Recall (R_l)

The recall is verified by the full amount of content and the total number of precise sentiment specifications. Moreover, the recall measure is determined by the Eq. (9).

$$\text{Recall} = \frac{FP}{FP+FN} \quad \ldots (9)$$

The developed approach achieved a recall measure of 0.90 in Table 2. As a result, the recall measure for the current replica CSO-LSTMNN reached 0.72, the recall measure for the HELM method was 0.82, and the recall measure for ACO-PSO increased to 0.78. PSO accomplished a recall measure of 0.65.

F-measure (F_m)

The F-measure is determined by Eq. (10). Moreover, it is defined as the medium mean between recall and precision. Consequently, F-measure is compared with four existing techniques: CSO-LSTMNN, HELM, ACO-PSO, and PSO. The comparison is presented in Table 2.

$$F_m = 2 \times \frac{P \times R_l}{P + R_l} \quad \ldots (10)$$

The suggested strategy aims to achieve an F-measure of 0.90. As a result, the HELM approach achieved 0.84 F-measure, the PSO technique attained 0.87 F-measure, the ACO-PSO technique predicted 0.85 F-measure, and the previous techniques of CSO-LSTMNN achieved 0.75 F-measure.

Error Rate

Misclassification often happens because of large amounts of complex data, and the proposed approach attains a lower error rate than other techniques. Also, sentiment classification based on multilingual datasets makes it possible to analyze Hindi and English well are given in Table 2 shows how well the two languages compare.

The suggested approach also produced an error rate of 0.5. The HELM strategy achieved a 1.43 error rate, the ACO-PSO technique forecasted a 0.9 error rate, and the PSO technique's error rate reached 0.73 using currently available CSO-LSTMNN techniques. As a result, when compared to current techniques, the suggested strategy is quite effective.

A multilingual dataset is used to perform sentiment analysis using the suggested HKHBRR model. The accuracy of this model is 98.68%, and the comparative results demonstrate the effectiveness of the suggested strategy.

Conclusions

This study introduces a brand-new sentiment analysis model called the Hybrid Krill Herd and Bat-based Recurrent Replica (HKHBRR). The bilingual (Hindi and English) twitter-based political dataset's text summarising and sentiment analysis was also covered in detail in this study. A political dataset based on Twitter was used to assess the sentiment value. Additionally, three factors—such as positive, negative, and neutral—must be present in the dataset that was gathered. This examination focuses mostly on a study regarding opinion mining with categorization and assesses the sentiment levels of political datasets derived from Twitter. Additionally, the innovative proposed framework is implemented in the Python framework, and the suggested replica's effectiveness is evaluated compared to other methods currently in use. Finally, the proposed method has achieved 98.68% accuracy, 0.90 recall, and 0.5% error.

References

1 Rajput A, Natural language processing, sentiment analysis and clinical analytics, *arXiv*. https://doi.org/10.48550/arXiv.1902.00679.

2 Mishev K, Gjorgjevikj A, Vodenska I, Chitkushev L T & Trajanov D, Evaluation of sentiment analysis in finance: From lexicons to transformers, *IEEE Access*, **8** (2020) 131662–131682.

3 Basiri M E, Nemati S, Abdar M, Cambria E & Acharya U R, ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis, *Future Gener Comput Syst*, **115** (2021) 279–294.

4 Sankar H, Subramaniyaswamy V, Vijayakumar V, Arun Kumar S, Logesh R & Umamakeswari A, Intelligent sentiment analysis approach using edge computing-based deep learning technique, *Softw Pract Exp*, **50(5)** (2020) 645–657.

5 Maryame N, Najima D & Rachida A, Spatio-temporal context for improving sentiment analysis accuracy, in *2018 6th Int Conf Multimed Comput Syst* (ICMCS) 2018, 1–6.

6 Lappeman J, Clark R, Evans J, Sierra-Rubia L & Gordon P, Studying social media sentiment using human validated analysis, *MethodsX* **7** (2020) 100867.

7 Goel P, Goel V, Gupta A K, Multilingual data analysis to classify sentiment analysis for tweets using NLP and classification algorithm, LNNS-Adv Data Info Sci Lect Notes Netw Syst, 94 (2020) 271–280.

8 Kauffmann E, Peral J, Gil D, Ferrández A, Sellers R & Mora H, A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making, *Ind Mark Manag* (2019) 1–15.

9 Gridach M, A Framework based on (Probabilistic) soft logic and neural network for NLP, *Appl Soft Comput*, **93** (2020) 106232.

10 Watkins J, Fabielli M & Mahmud M, SENSE: A student performance quantifier using sentiment analysis, *Int Joint Conf Neural Networks (IJCNN)* (Glasgow, United Kingdom) 2020, 1–6, doi: 10.1109/IJCNN48605.2020.9207721.

11 García-Díaz J A, Cánovas-García M & Valencia-García R, Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America, *Future Gener Comput Syst*, **112** (2020) 641–657.

12 Tiwari D & Nagpal B, Ensemble of methods sentiment analysis: A survey, in *7th Int Conf Comput Sustain Global Dev (INDIACom)* (New Delhi, India) 2020, 150–155.

13 Shaukat Z, Zulfiqar A A, Xiao C, Azeem M & Mahmood T, Sentiment analysis on IMDB using lexicon and neural networks, *SN Appl Sci*, **2(2)** (2020) 1–10.

14 Jelodar H, Wang Y, Rabbani M, Ahmadi S B B, Boukela L, Zhao R & Larik R S A, A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments, *Multimed Tools Appl*, **80(2)** (2021) 4155–4181.

15 Cobos R, Jurado F & Blazquez-Herranz A, A content analysis system that supports sentiment analysis for subjectivity and polarity detection in online courses, *Rev Iberoam de Tecnol del Aprendiz*, **14(4)** (2019) 177–187.

16 Amin A, Hossain I, Akther A & Alam Bengali K M, VADER: a sentiment analysis approach using modified vader, *Int Conf Comput Electr Electron Commun Eng (ECCE)* (MGIMO University, Moscow) 2019, 1–6.

17 GaoZ, Feng A, Song X & Wu X, Target-dependent sentiment classification with BERT, *IEEE Access*, **7** (2019) 154290–154299.

18 Dey U K & Noor A, For sentiment analysis using NLP, *Int Conf Comput Commun Inf (ICCCI)* (Hendaye France) 2019, 1–6.

19 Ruz G A, Henríquez P A & Mascareño A, Sentiment analysis of twitter data during critical events through bayesian networks classifiers, *Future Gener Comput Syst*, **106** (2020) 92–104.

20 Naseem U, Razzak I, Musial K & Imran M, Transformer based deep intelligent contextual embedding for twitter sentiment analysis, *Future Gener Comput Syst*, **113** (2020) 58–69.

21 Pai P F & Liu C H, Predicting vehicle sales by sentiment analysis of twitter data and stock market values, *IEEE Access*, **6** (2018) 57655–57662.

22 Kumar A & Garg G, Sentiment analysis of multimodal twitter data, *Multimed Tools Appl*, **78** (2019) 24103–24119, https://doi.org/10.1007/s11042-019-7390-1

23 Feizollah A, Ainin S, Anuar N B, Abdullah N A B & Hazim M, Halal products on twitter: Data extraction and sentiment analysis using stack of deep learning algorithms, *IEEE Access*, **7** (2019) 83354–83362.

24 Alarifi A, Tolba A, Al-Makhadmeh Z & Said W, A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks, *J Supercomput*, **76(6)** (2020) 4414–4429.

25 Sharma S & Jain A, Hybrid ensemble learning with feature selection for sentiment classification in social media, *Int J Inf Retr Res*, **10(2)** (2020) 40–58.

26 Krithiga R & Ilavarasan E, A reliable modified whale optimization algorithm based approach for feature selection to classify twitter spam profiles, *Microprocess Microsyst*, (2020) 103451.

27 Dubey A K, Kumar A & Agrawal R, An efficient ACO-PSO-based framework for data classification and preprocessing in big data, *Evol Intell*, **14(2)** (2021) 909–922.