IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11241

A Topographical Sentimental Analysis of COVID-19 Tweets

Dr. Sasikala. P[1]*, Shilpa K[2]

[1]Assistant professor, Department of Computer science, Nrupathunga University, India

[2]Assistant professor, Department of Computer science, Nrupathunga University, India

Abstract: The spread of Covid-19 has resulted in worldwide health concerns. Social media is increasingly used to share news and opinions about it. Therefore, it is the need of the hour to implement different measures to safeguard the countries by demystifying the pertinent facts and information. This paper focused on sentiment analysis on Twitter data related to Covid-19 using RNN. Originally, the input data collected from the dataset are pre-processed and applied to the sentiment lexicon dictionary for estimating sentiment values. The output data with the sentiment values are clustered using the Probability-based K- medoid algorithm to find whether the tweet caries any sentiment or not. Then from clustered results, the important is extracted and tagged using Part of Speech (POS). Then the Optima POS words are selected by using the Uniform Distribution based Cat Swarm Optimization (UD-CSO) algorithm and are subjected to Xavior initialization and Linear Logistic based Recurrent Neural Network (XL2-RNN) for classification. The experimental results indicated the effectiveness of the proposed sentiment analysis model.

Keywords: Probability-based K- medoid algorithm, Uniform Distribution based Cat Swarm Optimization (UD-CSO), Part of Speech (POS), Xavior initialization and Linear Logistic based Recurrent Neural Network (XL2-RNN), Covid-19 tweets, Sentiment Analysis

1. INTRODUCTION

The COVID-19 outbreak has been declared a pandemic by the World Health Organization (WHO), because of its high spreading and severity, which can cause severe pneumonia, respiratory failure, and death [3]. The COVID-19 outbreak has forced people to change their regular routine lives and practice social distancing. Such a sudden change can drastically increase people's stress levels and lead to other mental health issues [21]. However, many mental symptoms like worry, fear, frustration, depression, and anxiety could occur and cause serious mental health issues to people due to the long-time social activity restriction during the pandemic period [3-i]. Multiple studies have investigated the economic and social impacts of COVID-19, but what mental impact the drastic life changes bring to people and how to quantify it at the population level are yet to be studied [1].

Nowadays, people use social media like Facebook, Twitter, Weibo, and WhatsApp to express their opinion [23]. Thus, people have used the Internet as a way of escape, increasing virtual communication to compensate for the decrease and overall worsening of human interactions at the social level [6]. Given the developing situation with the pandemic, social media allows people to inform themselves and get updates from official sources [4-i]. Social Media can facilitate pre-diagnosis of a clinical mental health condition related to anxiety, depression, or anxious depression in active extroverts who verbalize and share their internal restlessness [17]. On the other hand, it was witnessed increased usage of online social media Twitter, during the lockdown [19]. According to Internet Live Statistics meanwhile 2013, the number of tweets sent every day has stretched 500 million [4]. With the sentiment analysis of covid 19 data from Twitter the mental health of people can be seen in the world, whether they still feel safe or are they in the excessive worry stage. Therefore, in this paper, an efficient framework of XL2-RNN based sentiment analysis model using Covid-19 Twitter data is proposed.

The remainder of the paper is structured as follows: Section 2 surveys the existing sentiment analysis models. Section 3 elaborates the framework of the proposed system. Section 4 discusses the experimental results and summarizes the findings, and Section 5 concludes the paper.

2. RELATED WORK

Mohammad Ehsan Basiri et al. [11] presented a novel method based on the fusion of four deep learning and one classical supervised machine learning model for sentiment analysis of coronavirus-related tweets from eight countries. Also, we analyzed coronavirus-related searches using Google Trends to better understand the change in the sentiment pattern at different times and places. Our findings reveal that the coronavirus attracted the attention of people from different countries at different times in varying intensities.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⋇ Vol. 11, Issue 2, February 2022
DOI: 10.17148/IJARCCE.2022.11241

Harleen Kaur et al. [12] analysed Twitter data through the R programming language. The Twitter data was collected based on hashtag keywords, including COVID-19, coronavirus, deaths, new cases, recovered. An algorithm called Hybrid Heterogeneous Support Vector Machine (H-SVM) was designed and performed the sentiment classification and classified them positive, negative, and neutral sentiment scores. Also, the performance of the algorithm was compared on certain parameters like precision, recall, F1 score, and accuracy with Recurrent Neural Network (RNN) and Support Vector Machine (SVM).

Md. Shahriare Satu et al. [13] aimed to design an intelligent clustering-based classification and topic extracting model named TClustVID that analyzes COVID-19-related public tweets to extract significant sentiments with high accuracy. COVID-19 Twitter datasets were collected and employed a range of data preprocessing methods to clean the raw data, then applied tokenization and produced a word-to-index dictionary. Thereafter, different classifications were employed on those datasets which enabled the exploration of the performance of traditional classification and TClustVID. The analysis found that TClustVID showed higher performance compared to traditional methodologies. Finally, significant topics were extracted from the clusters, split them into positive, neutral, and negative sentiments, and identified the most frequent topics using the model.

3. PROPOSED SENTIMENT ANALYSIS SYSTEM

Social media has great importance in individuals' life and connects people to the rest of the world. It is now considered a new source of stress, depression, and anxiety for people due to ambiguous information circulated over social media. This offers an opportunity for researchers and data scientists to access the data for academic and research use for analysing the mental health of the person. This paper proposed an efficient framework of a sentiment analysis system that analyses the Covid-19 Tweets using XL2-RNN. The proposed system first clusters the tweets as subjective and objective and then classifies the subjective tweets as it expresses positive, negative, or neutral sentiment. The block diagram of the proposed methodology is shown in below figure 1.

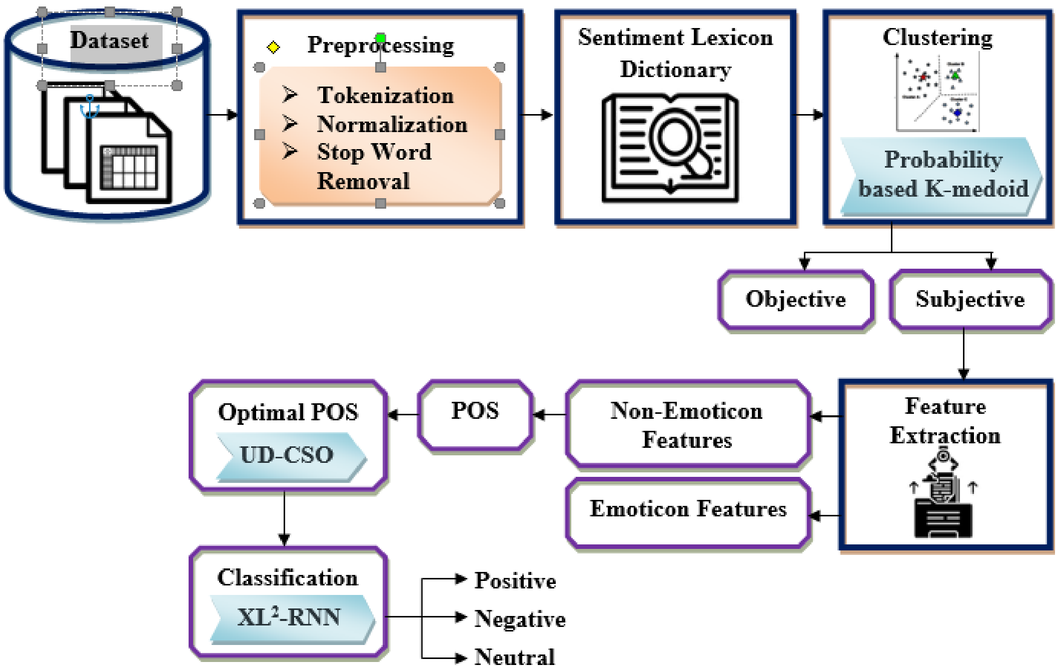

FIGURE 1: Block diagram of the proposed methodology

3.1 Pre-processing

Initially, the tweet data collected from the dataset are pre-processed to make it applicable to the model. The pre-processing helps to get rid of the unhelpful parts such as stop words, punctuation, numbers, single characters, and multiple spaces. The pre-processing steps used in the proposed model is expressed as follows,

Tokenization: It is the process of breaking the data into units called tokens. In the proposed model each word is converted into a token which can be a word, number, or punctuation.

Normalization: This step is for spelling improvement or to convert the unusual text to its standard form.

Stop Word Removal: This step eliminates the words that do not carry meanings to reduce the data dimensions.

Hence the pre-processed data are expressed as,

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁒ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11241

$$TD_{pre(i)} = \{TD_{pre(1)}, TD_{pre(2)}, \ldots\ldots, TD_{pre(N)}\} \tag{1}$$

Where, $TD_{pre(i)}$ denotes the pre-processed data.

3.2 Sentiment Lexicon Dictionary

The sentiment lexicon dictionary is used to compute the overall sentiment of the sentence based on individual words. Here, the pre-processed tweets $TD_{pre(i)}$ are scanned word by word, and the word that contains sentiment was found out. It is used to find whether the tweet has any sentiment or not. For this approach, a dictionary of positive and negative words is required, in which the sentiment values from the dictionary are assigned to all positive and negative words in the tweet. Based on the sentiment values the tweets are clustered in the following phase.

3.3 Clustering

With the sentiment values of the input $TD_{sen(i)}$, the tweets are clustered into subjective and objective. An objective sentence is a sentence that expresses some factual information about the world, while a subjective sentence expresses some personal feelings or beliefs. For clustering probability-based K-medoid clustering technique is used. K-medoid is an unsupervised partitioning technique that divides the data into the desired number of clusters. However the distance metric used in the existing algorithm doesn't offers accurate results, in the proposed method the conditional probability is estimated for each word instead of distance calculation.

Step 1: Select r random points from the input data. r is the number of clusters to be formed. As the proposed method forms only two clusters, $r = 2$ point are randomly selected. The selected points are considered as the medoids.

Step 2: Compute cost for each data point from r medoids. The cost is computed based on conditional probability. It can be expressed as,

$$\left.\begin{array}{l} \phi_p(+ve \mid TD_{sen(i)}) \rightarrow positive\ \ TD_{sen(i)} \\ \phi_p(-ve \mid TD_{sen(i)}) \rightarrow negative\ \ TD_{sen(i)} \end{array}\right\} \tag{2}$$

Where, ϕ_p denotes the probability estimated for each positive $+ve$ word and positive $-ve$ word.

Step 3: If the word in the tweet carries the probabilities of the positive or negative word, the tweet is assigned to the first medoid. Else, it is assigned for another medoid. This step forms the required number of clusters. The number of clusters formed ς_r can be expressed as,

$$\varsigma_r = \{\varsigma_1, \varsigma_2\} \tag{3}$$

Step 4: Compute the mean of all data points in each cluster.

Step 5: The steps are repeated until the same data points are assigned to each cluster.

The pseudo-code for the proposed probability-based K-medoid algorithm is shown in below figure 2,

Input: *preprocessed data with sentiment values* $TD_{sen(i)}$

Output: *number of clusters* ς_r

Begin

  Initialize input data $TD_{sen(i)}$

  Determine number of clusters

  Randomly select number of medoids

    $TD_{med(i)} = \{TD_{med(1)}, TD_{med(2)}\}$

  For each remaining data

    Estimate conditional probability

      $\left.\begin{array}{l} \phi_p(+ve \mid TD_{sen(i)}) \rightarrow positive\ \ TD_{sen(i)} \\ \phi_p(-ve \mid TD_{sen(i)}) \rightarrow negative\ \ TD_{sen(i)} \end{array}\right\}$

    Assign data point to the nearest medoid

  End for

  Return number of clusters ς_r

End

FIGURE 2: pseudo-code of the proposed probability-based K-medoid algorithm

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⁑ Vol. 11, Issue 2, February 2022
DOI: 10.17148/IJARCCE.2022.11241

As a result, two types of tweets are obtained, subjective $\varsigma_{1(sub)}$ and objective $\varsigma_{2(obj)}$. From these clusters, the subjective tweets $\varsigma_{1(sub)}$ are utilized for further processing.

3.4 Feature Extraction

After clustering the important features are extracted from the subjective tweets $\varsigma_{1(sub)}$. Two kinds of features are extracted namely, emoticons and non-emoticons features. The expressive, non-verbal components of the written language are known as emoticon features. Emoticon features boost the sentiment analysis as they give positive or negative sense to the sentence by a visual expression. The simple texts without any kind of symbols or expressive components are known as non-emoticon features. The extracted features are expressed as,

$$\chi_{fea(n)} = \{{}^{\varsigma_{1(sub)}}\chi_{EF(n)}, {}^{\varsigma_{1(sub)}}\chi_{NEF(n)}\} \tag{4}$$

$${}^{\varsigma_{1(sub)}}\chi_{EF(n)} = \{\chi_{EF(1)}, \chi_{EF(n)}, \dots, \chi_{EF(N)} \tag{5}$$

$${}^{\varsigma_{1(sub)}}\chi_{NEF(n)} = \{\chi_{NEF(1)}, \chi_{NEF(2)}, \dots, \chi_{NEF(N)}\} \tag{6}$$

Where, $\chi_{fea(n)}$ denotes the total features extracted containing emoticon ${}^{\varsigma_{1(sub)}}\chi_{EF(n)}$ and non-emoticon features ${}^{\varsigma_{1(sub)}}\chi_{NEF(n)}$. Then the emoticon features are assigned with some score values based on their polarity and are denoted as, ${}^{S}\chi_{EF(n)}$

3.5 Part of Speech

The non-emoticon features extracted ${}^{\varsigma_{1(sub)}}\chi_{NEF(n)}$ are then applied for Part of Speech (POS). In POS the features are labelled with grammatical descriptions, such as **Noun, Adjective, Adverb.** Using POS each word is tagged and it helps to get the context in which the word is used.
The features with the POS tags are expressed as,

$${}^{\chi}POS_{NEF(n)} = \{POS_{NEF(1)}, POS_{NEF(2)}, \dots, POS_{NEF(N)}\} \tag{7}$$

Where, ${}^{\chi}POS_{NEF(n)}$ denotes the features with the POS tags.

3.6 Optimization

From the POS features ${}^{\chi}POS_{NEF(n)}$ only the parts optimal for classification are filtered by using the Uniform Distribution based Cat Swarm Optimization (UD-CSO) algorithm. CSO is a metaheuristic algorithm inspired by the resting and tracking behaviours of cats. The algorithm contains two modes namely, seeking and tracing. Seeking mode considers the resting behaviour of the cats where the cats spend more time resting but still pay attention to their target. On the other hand, the tracing mode expresses the hunting capabilities of cats. The conventional CSO algorithm suffers from poor convergence for large dimension problem. In order to solve this random number in the solution search part of seeking mode is generated by using uniform distribution which balances the global and local searches. The algorithm steps are as follows,

Step 1: The number of N cats (here the cats are considered as the words obtained from POS ${}^{\chi}POS_{NEF(n)}$) are initialized. Each cat has its position with N dimensions, velocity values, fitness function, and a flag.
Step 2: The cats are randomly positioned in the N dimensional search space where each cat has the velocity value smaller than the predefined range of maximum velocity value. For each cat, the fitness is evaluated.
Step 3: The cats are distributed into the seeking and tracing modes based on the mixture ratio. Fitness is evaluated and the cat with the best fitness value is selected and kept in the memory. Here the fitness value is evaluated based on the classifier accuracy.
Step 4: Again the cats go through the seeking and tracing modes for the next iteration.
Step 5: The above process is continued until the solution with the worst fitness value is found.
The seeking and tracing modes of the above process are detailed as follows,
Seeking Mode
This mode reflects the resting behaviour of cats, in which four important roles such as Seeking Memory Pool (SMP), Seeking Range of selected Dimension (SRD), Counts of Dimension to Change (CDC), and Self-Position Considering (SPC) are carried out.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⵋ Vol. 11, Issue 2, February 2022
DOI: 10.17148/IJARCCE.2022.11241

- SMP signifies the number of position movements of the cat.
- CDC selects the number of dimensions to be modified.
- SRD defines the amount of mutation and modifications for the dimensions selected by CDC.
- SPC determines whether the present position of the cat is a candidate position for the next iteration or not.

During the seeking mode $l = SMP$ number of copies is created for the current position of the cat. If SPC is true then the current position is retained as one of the candidate and $l = SMP - 1$. For each copy, CDC dimensions are selected randomly and SRD percent of present values is added or subtracted randomly based on CDC. The new position can be updated as,

$$^{new}\xi_{pos(^{\chi}POS_{NEF(n)})} = (1+\lambda * SRD) *^{old}\xi_{pos(^{\chi}POS_{NEF(n)})} \tag{8}$$

Where, $^{new}\xi_{pos(^{\chi}POS_{NEF(n)})}$ is next position replaces the current position $^{old}\xi_{pos(^{\chi}POS_{NEF(n)})}$, $\lambda = \dfrac{1}{\alpha - \beta}$ is the factor uniformly distributed in which α, β are the locating and scaling parameters.

In this way, the old positions are replaced for all copies. Then the fitness is evaluated and the position with the best fitness value is selected to be the next position for the cat. If the fitness value is not equal then the selecting probability is computed as,

$$p = \frac{|\psi_i - \psi_s|}{\psi_{\max} - \psi_{\min}} \tag{9}$$

When the goal is to minimization $\psi_s = \psi_{\min}$ otherwise, $\psi_s = \psi_{\max}$. If all fitness values are equal, then all the selecting probability of each candidate point is set to be 1.

Tracing Mode

This mode replicates the tracing behaviour of cats for finding food. In this mode, the velocity values for each dimension of the cat are updated. The velocity values are updated as,

$$^{new}\varpi_{vel} = a * \varpi_{vel} + \beta * h(^{best}\xi_{pos(^{\chi}POS_{NEF(n)})} - \xi_{pos(^{\chi}POS_{NEF(n)})}) \tag{10}$$

Where, $^{best}\xi_{pos(^{\chi}POS_{NEF(n)})}$ denotes the best position obtained, $^{new}\varpi_{vel}$, ϖ_{vel} denotes the new and old velocity values, $a \in (0,1)$ is the weight vector, $\xi_{pos(^{\chi}POS_{NEF(n)})}$ denotes the current position, β is the user-defined parameter, h is the random value.

Then the updated values are checked where if the velocity is larger than the predefined value then it is set to be in the range. After that, the position of the cat is updated as,

$$^{new}\xi_{pos(^{\chi}POS_{NEF(n)})} = \xi^{N}_{pos(^{\chi}POS_{NEF(n)})} + \varpi^{N}_{vel} \tag{11}$$

Where, $^{new}\xi_{pos(^{\chi}POS_{NEF(n)})}$ is the newly updated position, $\xi^{N}_{pos(^{\chi}POS_{NEF(n)})}$, ϖ^{N}_{vel} are the current position and old velocity values of the cat in the N^{th} dimension. In this way, the optimal features $^{opt(\chi)}POS_{NEF(n)}$ are selected by using the proposed UD-CSO algorithm.

3.7 Classification

Finally, the optimal features $^{opt(\chi)}POS_{NEF(n)}$ and the emoticon features $^{S}\chi_{EF(n)}$ extracted are rendered as the input to the Xavior initialization and Linear Logistic based Recurrent Neural Network (XL^2-RNN) for classification. RNN works similarly to the traditional neural network except that it can memorize the input. This is because RNN has internal memory by which it remembers its input and allows the network to be very precise in predicting what's coming next. RNN considers the current input and what it has calculated for the previously received input for producing the output. Hence the information cycles through a loop. The random weight initialization in the existing RNN causes increased loss of the network and the higher training time. To alleviate these issues, in the proposed network the weights are initialized by using the Xaviour Initialization technique and the linear logistic activation function is used. The algorithm steps are as follows,

At any given time i, the current input is the combination of input at current state J_i and the previous state J_{i-1}. The formula for calculating current state is given as follow,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11241

$$hd_i = \Re_i(hd_{i-1}, J_i) \tag{12}$$

$$\Re_i = \frac{(hd_{i-1}, J_i)}{1+\exp(-(hd_{i-1}, J_i))} \tag{13}$$

Where, hd_i is the Current state which holds the information about the previous state hd_{i-1}, $J_i \in ({}^{opt(\chi)}POS_{NEF(n)}, {}^{S}\chi_{EF(n)})$ is the input state, $\Re_i$ is the linear logistic activation function. The current state hd_i is activated as,

$$hd_i = \Re_i(Q_{hd}\, hd_{i-1} + Q_{ip,hd} J_i) \tag{14}$$

Where, Q_{hd} , Q_{ip} are the weights of the recurrent neuron and the weight of the input neuron.

Thus, the current hd_i becomes hd_{i-1} for the next time step. After all the time steps are completed, the final current state is used to calculate the output as,

$$op_i = Q_{hd,op}\, hd_i \tag{15}$$

Where, op_i denotes the output, $Q_{hd,op}$ denotes the weight of the output layer.

The output is then compared to the actual output i.e., the target output and the error are generated. The error is then backpropagated to the network to update the weights and hence the network is trained. During back propagation the weights are updated as,

$$Q_{ip,hd,op} = X\left(\left(\frac{-\sqrt{6}}{a+b}, \frac{\sqrt{6}}{a+b}\right)\right)$$

Where, a b are the number of inputs and outputs of the network.

Finally, the classifier output contains three classes where the subjective tweets are classified as positive, negative, and neutral.

4. RESULT AND DISCUSSION

In this section, the performance analysis and comparative analysis are carried out to evaluate the effectiveness of the proposed model. The proposed sentiment analysis model on Covid-19 tweets is implemented in the working platform of JAVA. The analyses are done by using Covid-19 Twitter Dataset.

4.1 Performance analysis

In this section performance of the proposed classification, clustering, and optimal POS selection techniques are compared with the existing techniques. The proposed XL^2-RNN is compared with the existing Recurrent Neural Network (RNN), Convolution Neural Network (CNN), and Artificial Neural Network (ANN) in respect of sensitivity, specificity, accuracy, and F-measure. Then the clustering time of the proposed probability-based K-medoid technique is contrasted against the existing K-medoid, K-means, Partition Around Medoids (PAM), and Fuzzy C-means (FCM) methods. Followed by the fitness vs iteration analysis for the proposed UD-CSO and existing Cat Swarm Optimization (CSO), Sandpiper Algorithm (SPA), Chimp Optimization Algorithm (ChOA), and Particle Swarm Optimization (PSO) takes place.

Table 1: Performance Analysis of classifiers

Methods	**Sensitivity**	**Specificity**	**Accuracy**	**F-measure**
Proposed XL^2-RNN	97.4785	97.5234	98.7548	96.5234
RNN	94.7458	95.7426	96.5478	94.8424
CNN	92.7534	93.1421	94.6154	93.2335
ANN	91.1524	92.1534	92.1245	91.4524

The performance analysis of the proposed and existing classifiers in terms of sensitivity, specificity, accuracy, and F-measure are shown in above table 1. The higher value of sensitivity, specificity, accuracy, and F-measure signifies the improved performance of the classifier. From the above table, it is evident that the proposed method has higher performance for all metrics. The proposed XL^2-RNN classifies the subjective tweets with an accuracy of 97.4785 whereas,

IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⋇ Vol. 11, Issue 2, February 2022
DOI: 10.17148/IJARCCE.2022.11241

the existing RNN (94.7458), CNN (92.7534), and ANN (91.1524) methods attained lower accuracies. Thus the analysis concludes that the proposed RNN has higher performance than the existing classifiers.

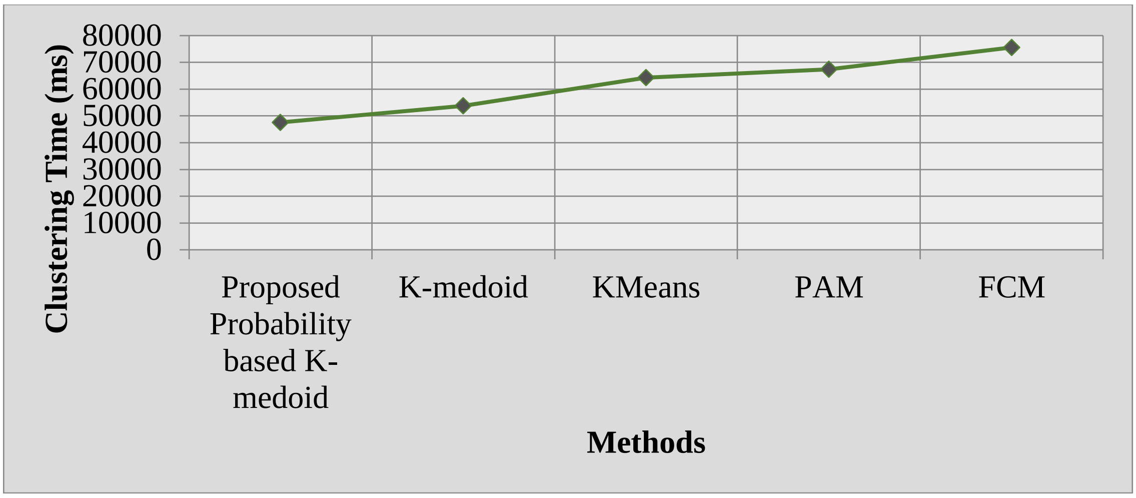

FIGURE 3: Clustering Time Analysis

Above figure 3 demonstrates the clustering time of the proposed and existing K-medoid, K-means, PAM, and FCM clustering techniques. The clustering time of the K-medoid, K-means, PAM, and FCM are 53751 ms, 64264 ms, 67437 ms, and 75574 ms which are higher than the proposed method. The proposed probability-based K-medoid method takes much lesser time clustering which is 47574 ms. Hence, as a result, the proposed clustering techniques have better performance when compared to the existing methods.

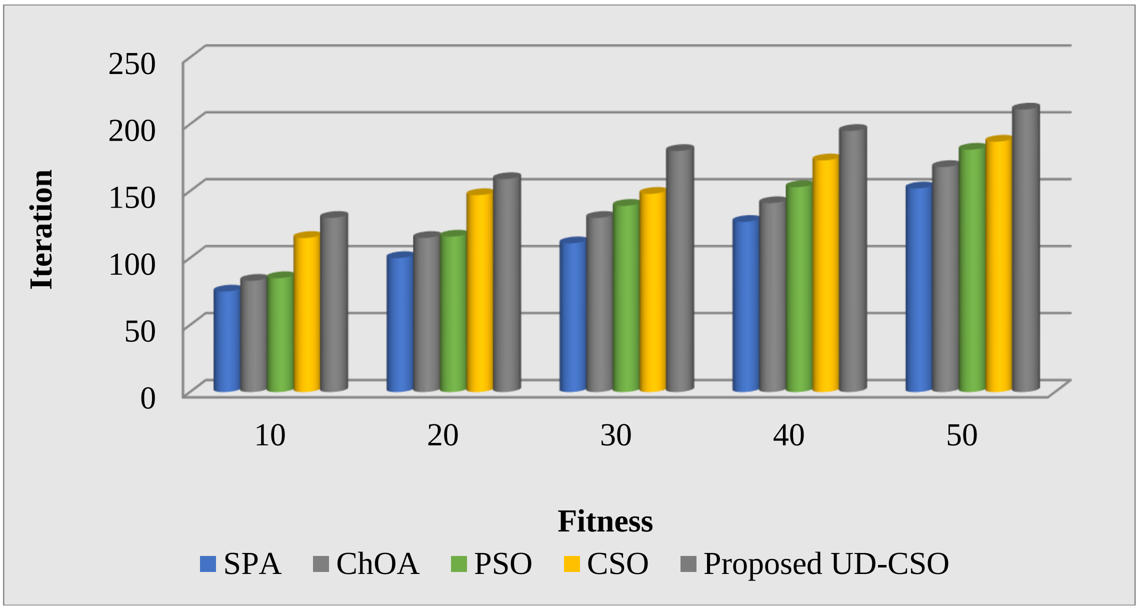

FIGURE 4: Fitness vs Iteration Analysis

The fitness vs iteration analysis of the proposed and existing method is shown in above figure 4. The fitness is evaluated for the number of iterations ranging from 10 to 50. For the minimum number of 10 iterations, the fitness attained by the proposed method is 130. But the existing SPA, ChOA, PSO, and CSO methods have the fitness of 75, 83, 85, and 115 lower than the proposed method. In this way, for the remaining number of iterations also the proposed method achieved fitness value better than the existing methods. Thus, the analysis delivers that the performance of the proposed method is superior to existing methods.

DOI: 10.17148/IJARCCE.2022.11241

5. CONCLUSION

This paper proposes an efficient model for sentiment analysis on Covid-19 tweets. The model includes several phases, where the input collected from the dataset is preprocessed and applied to the sentiment lexicon dictionary. Then the subjective tweets are identified through clustering and important features are extracted. Then XL^2-RNN is employed to define the polarity of the tweet. To determine the effectiveness of the proposed technique, the performance of XL^2-RNN is compared with three existing techniques. The experimental results prove that the proposed XL^2-RNN proffered an excellent performance than existing techniques. The proposed RNN yielded a higher accuracy of 97.4785%. The performance analysis proves that the proposed scheme's accuracy and its efficiency are better than the existing techniques.

AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the pa-per was free of plagiarism.

REFERENCES:

[1] Klaifer Garcia and Lilian Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA", Applied Soft Computing Journal, vol. 101, pp. 1-15, 2021. doi: 10.1016/j.asoc.2020.107057
[2] Diya Li, Harshita Chaudhary and Zhe Zhang, "Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining", International Journal of Environmental Research and Public Health, vol. 17, no. 14, pp. 1-22, 2020. https://doi.org/10.3390/ijerph17144988
[3] Hui Yin, Shuiqiao Yang and Jianxin Li, "Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media", 2020, https://doi.org/10.48550/arXiv.2007.02304.
[4] Yipeng Zhang, Hanjia Lyu, Yubao Liu, Xiyang Zhang, Yu Wang and Jiebo Luo, "Monitoring depression trend on twitter during the COVID-19 pandemic", 2020, https://doi.org/10.48550/arXiv.2007.00228.
[5] Shreya Ghosh and Tarique Anwar, "Depression intensity estimation via social media a deep learning approach", IEEE Transactions on Computational Social Systems, vol. 8, no. 6, pp. 1465-1474, 2021. **DOI:** 10.1109/TCSS.2021.3084154
[6] Javier Cabezas, Daniela Moctezuma, Alberto Fernandez-Isabel and Isaac Martin de Diego, "Detecting emotional evolution on twitter during the COVID-19 pandemic using text analysis", International Journal of Environmental Research and Public Health, vol. 18, no. 13, pp. 1-20, 2021. https://doi.org/10.3390/ijerph18136981
[7] Jean Marie Tshimula, Belkacem Chikhaoui and Shengrui Wang, "COVID-19 detecting depression signals during stay-at-home period", 2021, https://doi.org/10.48550/arXiv.2103.00597.
[8] Akshi Kumar, Aditi Sharma and Anshika Arora, "Anxious depression prediction in real-time social data", International Conference on Advanced Engineering, Science, Management and Technology, 14-15 March 2019, Dehradun, Uttarakhand, India, 2019.
[9] Jianlong Zhou, Hamad Zogan, Shuiqiao Yang, Shoaib Jameel, Guandong Xu and Fang Chen, "Detecting community depression dynamics due to COVID-19 pandemic in australia", IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 982-991, 2021. DOI: 10.1109/TCSS.2020.3047604
[10] Imamah and Fika Hastarita Rachman, "Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regresion", Information Technology International Seminar (ITIS), 14-16 October 2020, Surabaya, Indonesia, 2020. **DOI:** 10.1109/ITIS50118.2020.9320958
[11] Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Somayeh Asadi and Rajendra Acharrya U, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets", Knowledge-Based Systems, vol. 228, pp. 1-21, 2021. 10.1016/j.knosys.2021.107242
[12] Harleen Kaur, Shafqat Ul Ahsaan, Bhavya Alankar and Victor Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets", Information Systems Frontiers, vol. 23, no. 2, pp. 1417-1429, 2021. **https://doi.org/10.1007/s10796-021-10135-7**
[13] Shahriare Satu Md, Imran Khan Md, Mufti Mahmud, Shahadat Uddin, Matthew A. Summers, Julian M. W Quinn and Mohammad Ali Moni, "TClustVID a novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets", Knowledge-Based Systems, vol. 226, pp. 1-16, 2021. DOI: 10.1016/j.knosys.2021.107126