Twitter sentiment analysis using adaptive neuro-fuzzy inference system with genetic algorithm*

Padmaja K.
Computer Science and Engineering
University College of Engineering Kakatiya University
Kothagudem, India
ajit2107@yahoo.com

Nagaratna P. Hegde
Computer Science and Engineering
Vasavi College of Engineering
Hyderabad, India
nagaratnaphd@gmail.com

***Abstract*—In recent decades, it is very challenging for the researchers to identify the users' sentiments from the twitter data, due to unstructured nature, misspells, abbreviations, limited size, and slangs. In order to address these problems, a new system is developed in this research study for improving the performance of twitter sentiment analysis. In this research, the proposed system comprises of three major phases; data collection, pre-processing and classification of sentiments from pre-processed tweets. Initially, the twitter sentiment analysis was carried-out by using twitter-sanders-apple 2 dataset. Generally, the raw collected tweets contain more noises by means of positive emoji's, URLs, stop words, negative emoji's, which were essentially eliminated to achieve better classification. Finally, a hybrid classifier (Adaptive Neuro-Fuzzy Inference System (ANFIS)-Genetic Algorithm (GA)) was used to classify the twitter sentiment classes as either positive class or negative class. The ANFIS classifier was the fuzzy based ontology that was designed by actualizing and the GA optimizes the fuzzy principles in the ANFIS classifier. The experimental consequence showed that the proposed methodology enhanced the accuracy in sentiment analysis up to 5.5-6% related to the existing methodology.**

Keywords—Adaptive neuro-fuzzy inference system, Genetic algorithm, Opinion mining, Twitter sentiment analysis, Pattern-based approach.

I. INTRODUCTION

In present scenario, social media websites like Tumbler, Facebook, Twitter, etc. plays a vital role in human life. In that, twitter is the emerging platform for communicating and sharing data with others. Twitter assists the users to exhibit their thinking or feelings about a few situations of real-world happenings [1]. Twitter analysis the user's opinions by using the review, blogs, and posts that enriches the business, recommender system, customer opinions, and politics [2-3]. The procedure of inspecting the polarity or intention of the twitter user messages about a topic of interest is named as sentiment analysis [4]. The information which is collected from twitter, usually analyzed for evaluating the sentiment of the information like negative or positive sentiments [5]. Generally, sentiment refers to emotions, attitude, feelings and opinion of the users that includes three levels, namely sentence, aspect, and document level [6]. The motivation behind sentiment analysis is to recognize sentences by evaluative meaning. After that, sentences are classified according to their division as either positive, negative or neutral classes [7]. In web, users regularly express their sentiments over the web through web based social networking, websites, rating and surveys. The main issues in the twitter sentiment analysis are short messages shows limited cues about sentiment, informal language, abbreviations, and acronyms that are mainly accomplished on twitter data [8].

In recent period, many classification methods are developed for the automatic classification of twitter data. Usually, raw twitter data comprises of irregular, noisy, poor structured sentence, ill formed, incomplete words, non-dictionary terms, and irregular expressions [9-10]. In addition, end-user opinions, sentiments, and emotions are used for the selection of reviews that helps the users in analyzing the positive and negative sides of a particular topic [11]. Generally, the traditional classification methodologies like naive Bayes, k-nearest neighbor, maximum entropy classifier, decision tree and support vector machine are utilized for solving the twitter sentiment classification problem in order to review the documents using natural language processing [12-14]. Usually, the naive Bayes classifier determines the posterior probability of a class that depends on the distribution of the words in the report. This classification approach works with the feature extraction that discounts the word position in the document. Similarly, k-nearest neighbor classifier is a non-parametric technique, where the learning functions are approximated locally and also all computation is deferred until classification. In order to make the proposed system simpler, GA is used for optimizing the generated rules in the ANFIS classifier in order to diminish the proposed system (ANFIS-GA) complexity. GA is defined as a global optimization method that works on the basis of survival-of-the-fittest and mechanics of natural selection. In this research, twitter sentiment analysis is implemented for analyzing the tweets by generating the fuzzy rules with the help of ANFIS classifier. The benefits of the proposed system are listed below,

- GA utilizes a simple objective function, which is based on the mathematical model.
- The ANFIS classifier combines the benefits of both artificial neural network and fuzzy logic controllers for generating the fuzzy rules.

This research paper is organized as follows. Section II describes about the survey of several recent year papers used for twitter sentiment analysis. Section III represents the challenges in the twitter sentiment analysis. The proposed system (ANFIS with GA) is explained in the section IV. Whereas, the quantitative and comparative analysis of the proposed system is detailed in the section V. Finally, the conclusion of the research work with future directions is described in the section VI.

II. LITERATURE REVIEW

Several methods have been developed by the researchers in the twitter sentiment analysis. In this sub-section, a brief evaluation of a few essential contributions to the existing literature papers is presented.

F.O. de Franca et al., [15] briefed about the twitter users during the impeachment voting event, which took place in Brazil in the year of 2016. For this purpose, a new methodology was developed for classifying the users as popular, activists and observers. The identification of these subjects on each and every group of users was used to verify the real interest of common Brazilian citizens. Hence, the developed methodology segments the user's popular activity in order to filter out the useful information. The developed approach assists other studies related to the usage of twitter during important events. Since, the experiment was performed on the collected twitter data during a small time window and the sentiment change results did not indicate shifts in opinion. The developed methodology needs more repeated words in order to summarize the particular event, which was considered as one of the major issues.

A.C. Pandey et al., [16] developed a new metaheuristic methodology that was the combination of cuckoo search and k-means algorithm for analyzing the twitter sentiment analysis. The developed methodology was utilized for identifying the cluster heads from the twitter contents. In addition, the developed methodology indiscriminate the practical suggestions to design a new system that delivers decisive reviews on any social concerns. Extensive experiments were performed and the effectiveness of developed methodology was verified by using twitter sanders apple 2 and 3 datasets. The comparative and quantitative analysis showed that the developed methodology has out-performed the existing approaches by means of classification accuracy and computation time. In some cases, the developed methodology was inefficient for large datasets and also it was not robust in content clarification of multi-media documents.

S. Rill et al., [17] developed a new system (PoliTwi) for detecting the emerging political topics using the twitter data. In this research study, in total 4,000,000 tweets were collected during the parliamentary election, 2013 in Germany from the month of April to September. Here, the new topics on twitter were detected correctly after their occurrence. At last, the developed method showed how the correctly detected topics were utilized for extending the existing knowledge bases (semantic networks or web ontologies), which were required for concept level sentiment analysis. Here, the developed system utilized the twitter hashtags during the parliamentary election. An issue in this research paper was the exclusion of controversial topics.

B. Heredia et al., [18] developed a new methodology for exploring the efficiency of social media as a resource for both prediction and polling the election result. In this research study, a new dataset was developed by using three million tweets that range from the time period of September 2016 to November 2016. Here, the polling investigation was carried-out on two levels; state and national level, and the election prediction analysis was carried-out on the only state level. In this paper, shared elector count methodology and winner-take-all methodology was used to predict the election condition. The developed methodology showed better outcomes in nation level, but the performance of developed methodology was limited in state level.

M. Bouazizi, and T. O. Ohtsuki, [19] presented a new pattern based method for detecting sarcasm on twitter data. The developed pattern based methodology generates four sets of features, which covered the dissimilar types of sarcasm. The developed methodology utilized the four sets of features for classifying the tweets as non-sarcastic and sarcastic. In addition, the developed system highlighted the position of a pattern based features in order to detect the sarcastic statements. The developed approach used part-of-speech tags for extracting the patterns to describe the level of sarcasm of tweets. In the experimental outcome, the developed approach showed better results in light of classification accuracy, precision, and recall. Besides, the patterns extracted from the current data did not cover all possible sarcastic patterns.

For addressing the above mentioned issues, a new hybrid methodology (ANFIS-GA) is developed for twitter sentiment classification.

III. CHALLENGES IN SENTIMENT ANALYSIS

In twitter application, identifying the user's sentiment is a challenging task, because user represents their feelings with respect to image, text, emoji's, etc. Researchers who endeavor to create efficient twitter sentiment analysis strategies need to confront against various difficulties that rises from the attributes of twitter. The major difficulty in the twitter sentiment analysis is the large length of the messages and different types of data (text, image, emoji's, etc.) [20]. Moreover, twitter sentiment analysis is needed to handle the advancing substances.

The most important challenges in the twitter sentiment analysis are listed below.

- **Text length:** The length of the tweets is very limited that approximately 140 characters. The shorter length of the text degrades the twitter sentiment analysis performance and also it is very difficult to recognize the users' opinions.
- **Relevance of topic:** In twitter sentiment analysis, the existing research works classify the tweets without considering the topic. Though, the existing methods randomly classify the twitter text.
- **Incorrect English:** The major issues in the twitter sentiment analysis are incorrect spellings and inaccurate English.
- **Sparsity of data:** The random communications and variations in text length influences the twitter sentiment analysis. The blogs, news, text, etc. belongs to the different kinds of languages.
- **Negation:** Recognition and identification of invalidate words from the tweets is a very challenging task.
- **Multilingual content:** The tweets are language independent and sometimes messages are similar.

IV. PROPOSED SYSTEM

The proposed work is performed in dissimilar phases that consists of twitter data collection, tokenization, data pre-processing, feature extraction and hybrid classification (combination of both ANFIS and GA). The combination of algorithms provides a better outcome for classifying the sentiments of tweets. The proposed system architecture is given in the Fig 1 and also the detailed explanation about the proposed system is given below.

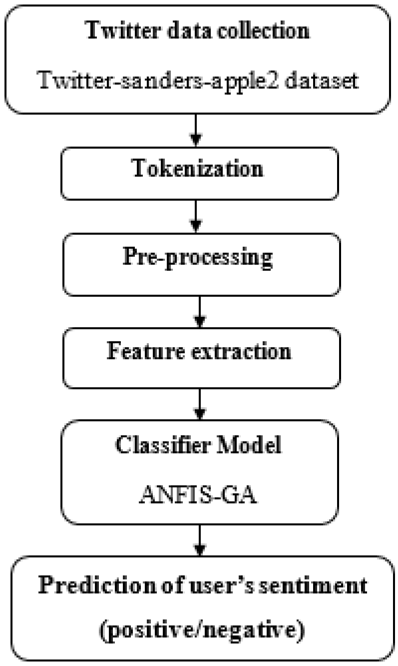

Fig. 1. Work flow of proposed methodology

A. Twitter data collection

Initially, the twitter sentiment data are collected from the dataset (twitter-sanders-apple 2) that comprises of 479 tweets, which are directly imported in the Light-SIDE. In that 479 tweets, 316 tweets are represented as negative sentiment tweets and the remaining 163 tweets are denoted as positive sentiment tweets. After collecting the twitter data, an important phase: tokenization is applied on the raw tweets for extracting the tokens, phrases and meaningful words.

B. Tokenization

The tokenization mechanism is utilized to break the tweets into symbols, tokens or phrases and meaningful words that are highly implemented for parsing process and text mining. Initially, segment the tweets into words and locate the boundaries for the segmented words in order to form tokenization. Generally, work boundary starts with one word and ending off with another word. In addition, performing tokenization is very easy, if the segmented words have more space. If there is no existence of any white space, the punctuation symbol is considered to be white space, when the tokenization process occurs. Here, the tokens notions are defined, before doing any kind of processing.

C. Data pre-processing

After tokenization, data pre-processing phase is carried out for decreasing the feature size for diminishing the computation complexity of the learning algorithm. In this research study, data-pre-processing is essential, because the raw tweets comprise of numerous features as shown in a sample tweet for pre-processing.

Data preprocessing includes the following steps,

- Eliminate the retweets, which are all starts with "RT".
- Remove the stop-words or useless words from the raw tweets.
- Convert the slangs into words with equivalent meaning.
- The external links and user name preceded with "@" are also eliminated.
- The hashtags "#" are also eliminated from the raw tweet.
- "Stemming" is carried-out for reducing every word to its root word.
- "Emoticons" are exchanged by its equivalent meaning, because it is very useful for detecting the sentiments.

D. Feature extraction

After preprocessing the raw-tweets, the data features are extracted from the pre-processed tweets. The feature extraction phase performs numerous operations; need to analyze the score value of positive and negative word, need to present the overall score of the words in the tweets, text tag count, and the frequency of negative and positive words are also analyzed. In this scenario, numerous features are given as the input for learning classifiers. Here, the word count is determined as the total number of words present in each tweet after the preprocessing is completed.

The negative word count is denoted as the number of negative words exists in each and every tweet. Tag count is represented as the number "@" tags utilized in every tweet. Successively, positive word count is denoted as the number positive words exists in each and every tweet. Here, positive score is the score that achieved after totaling the positive adjectives. Respectively, negative score is the score that attained after totaling the negative adjectives. Generally, score is determined as the final result that subtract negative score from the positive score in each tweets.

E. Sentiment Classification

The twitter sentiment classification depends on ANFIS classifier that relies upon the fuzzy rational. In addition, GA algorithm is utilized to optimize the fuzzy rules that are being produced for generating ontology.

1) Adaptive neuro-fuzzy inference system

The ANFIS classifier accomplishes multiple targets, because it is more feasible and reliable related to the individual target. The ANFIS is a Neuro-fuzzy model that has the advantage of both neural network and fuzzy logic. At first, the learning mechanism is exploited on the selected data values $d\left(x_1'\right), d\left(x_2'\right), ...d\left(x_n'\right)$. The basic rule of ANFIS classifier is mathematically represented in the Eq. (1).

$$Rules_i = a_i d\left(x_1^{'}\right) + b_i d\left(x_2^{'}\right) + c_i d\left(x_n^{i}\right) + f_i \qquad (1)$$

Where,

$a_i, b_i, c_i \wedge f_i$ are denoted as design parameters.

Layer 1: In layer 1, every node i is a square node that includes several node functions. These node functions are selected from the bell shaped curve with minimum 0 and maximum 1 values that is mathematically represented in the Eq. (2) and (3).

$$o_{1,i} = \mu a_i d\left(x_1^{'}\right) + \mu b_i d\left(x_2^{'}\right) + \mu b_i d\left(x_n^{'}\right) \qquad (2)$$

$$o_{1,i} = \mu a_i d\left(x_1^{'}\right) = \frac{1}{1 + \left[\left(\left(x - o_i\right) / \mathrm{p}_i\right)^2\right]^{q_i}} \qquad (3)$$

Where, $o_i, p, \wedge q_i$ are represented as parameter set and μ is indicated as degree of membership functions for the fuzzy sets a_i, b_i and c_i.

Layer 2: In layer 2, each node i is a circle node $\prod$ that multiplies the incoming values and transmit the product out that is given in the Eq. (4).

$$o_{2,i} = wt_i = \mu a_i d\left(x_1^{'}\right) + \mu b_i d\left(x_2^{'}\right) + \mu c_i d\left(x_n^{'}\right), \quad i = 1,2 \qquad (4)$$

Layer 3: In layer 3, each node i is a circle node, which assesses the ratio of rules firing strength that is mathematically specified in the Eq. (5).

$$o_{3,i} = wt_i^{'} = \frac{wt_i}{\left(wt_1 + wt_2\right)}, i = 1,2 \qquad (5)$$

Layer 4:

Here, each and every node i is a square node with the node functions that is mathematically indicated in the Eq. (6).

$$o_{4,i} = wt_i^{'}.Rules_i,\ i = 1,2 \qquad (6)$$

Where, wt_i is exemplified as the output of layer 3.

Layer 5: In this layer, the incoming values are shortened and the overall output values are denoted in the Eq. (7) and (8).

$$o_{5,i} = \sum_i wt_i^{'}.Rules_i, \frac{\sum_i wt_i^{'} Rules_i}{\sum_i wt_i^{'}} \qquad (7)$$

$$o_{5,i} = wt_1^{'}.Rules_i + wt_2^{'}.Rules_i \qquad (8)$$

The fuzzy rules are generated by using ANFIS classification method. The output of ANFIS classifier is given as the input for GA for finding the optimal values for fuzzy rules.

2) Genetic algorithm

Feature optimization is a high-level process that identifies the relevant subsets of twitter data on the basis of a particular criterion. In most of the existing researches, GA is used for feature optimization for determining the relevant features. In GA, the mutual information between the features are calculated in order to identify the most optimal features that effectively decreases the computational effort. The core idea of GA is population initialization. In this application, every string comprises of a lot of fluffy standard parameters developed by ANFIS classifier. The step by step procedure of GA is listed below.

Step by step mechanism of GA

- **Initial populations are accomplished:** Chromosome lengths are generated by random binary digits, which are all inversely proportional to the dimension of initial population size. Usually, chromosomes are denoted in binary values as strings of $0s$ and $1s$.
- **Generating children for initial population:** The GA automatically picks the top two best chromosomes by using the elite-count size of two. Whereas, elite-count is less than or equal to the size of population.
- GA performs elitism, mutation, and crossover on new population.
- The fitness function of each and every chromosome is determined using nearest neighbor algorithm. The general formula of nearest neighbor distance is given in the Eq. (9).

$$fitness = \frac{k}{v} + e^{\frac{-1}{v}} \qquad (9)$$

Where, k is stated as the classification error, and v is indicated as the feature cardinality.

- **Selection mechanism:** Elite-count size: two is used in this research study, because of its speed, simplicity and efficiency.
- **Stopping criteria:** Here, stopping criteria is based on number of iterations (50), when the generation reaches its predefined value; it stops and delivers the best solution in the last generation.
- Finally, evaluate the classification accuracy of the new set of features. These relevant features are given as the input for ANFIS classifier for twitter sentiment classification.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In the experimental phase, the proposed methodology was simulated by using Python platform with 3.0 GHZ-Intel i3 processor, and 8 GB RAM. The proposed ANFIS-GA methodology performance was related to other existing methodology (Latent Dirichlet Allocation (LDA)-Possibilistic Fuzzy C-Means (PFCM) [21]) for estimating the efficiency and effectiveness of the proposed methodology. M. Trupthi, et al., [21] developed a superior topic modelling approach: LDA-PFCM in order to extract the keywords and also for finding the concerned topics. The performance of the proposed methodology was validated in light of precision, f-measure, recall and accuracy.

In proposed methodology (ANFIS-GA), the collected tweets are forwarded to text preprocessing and separate the words into tokens using tokenization. Then, extract the features by developing the polarity and scores of the tweets. By utilizing the proposed methodology (ANFIS-GA), the fuzzy rules are optimized and then estimate the efficiency and effectiveness by means of precision, f-measure, accuracy, and recall. The experimental analysis was carried-out on Python platform and the outcomes were compared with some current work in order to evaluate the effectiveness of proposed strategy.

A. Performance measure

Performance measure is defined as the regular measurement of outcomes and results that develops a reliable information about the effectiveness of proposed method (ANFIS-GA). Also, performance measure is the procedure of reporting, collecting and analyzing information about the performance of a group or individual. The mathematical equation of accuracy, and f-measure are denoted in the Eq. (10), and (11).

$$Accuracy = \frac{TN + TP}{TP + TN + FN + FP} \times 100 \tag{10}$$

$$F - measure = \frac{2TP}{(2TP + FP + FN)} \times 100 \tag{11}$$

Where, TP is signified as true positive, TN is indicated as true negative, FP is specified as false positive, and FN is indicated as false negative. In this research work, precision values are sub-categorized as positive and negative precision values that is named as precision positive (p) and precision negative (n). These two precision values are calculated by using the Eq. (12) and Eq. (13).

$$Precision(p) = \frac{TP}{FP + TP} \tag{12}$$

$$Precision(n) = \frac{TN}{FN + TN} \tag{13}$$

In the proposed system (ANFIS-GA), there are two values for recall such as, positive and negative values. Recall positive (p) and Recall negative (n) are recalling ratio, which are calculated as given in Eq. (14) and Eq. (15):

$$Recall(p) = \frac{TP}{FN + TP} \tag{14}$$

$$Recall(n) = \frac{TN}{FN + TN} \tag{15}$$

B. Quantitative analysis of proposed system

In this research study, twitter-sanders-apple2 dataset is used to evaluate the performance of existing methodology (LDA-PFCM) and the proposed methodology (ANFIS-GA). In this scenario, the f-measure, precision, accuracy, and recall value of the existing and proposed methodologies are evaluated for both negative and positive classes. In table 1, the precision, accuracy, f-measure, and recall value of existing methodology (LDA-PFCM) in the positive class is 100%, 87%, 87.71%, and 78.12%. In contrast, the proposed approach (ANFIS-GA) achieved 100%, 93%, 90.15%., and 80.10%. Correspondingly, the precision, accuracy, f-measure and recall value of existing methodology in the negative class is 72.5%, 87%, 83.88%, and 100%. Whereas, the proposed methodology (ANFIS-GA) achieved 70.4%, 92.56%, 89.32%, and 100%. Figures 2 and 3 represent the comparison of positive and negative values for both existing and proposed scheme in terms of precision, accuracy, F-measures, and recall.

TABLE I. PERFORMANCE ANALYSIS OF PROPOSED AND EXISTING SCHEME

Methods	**Accuracy (%)**		**Precision (%)**		**Recall (%)**		**F-Measure (%)**	
	Positive	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
LDA-PFCM [21]	87	87	100	72.5	78.12	100	87.71	83.88
Proposed ANFIS-GA	93	92.56	100	70.4	80.10	100	90.15	89.32

From the table 1, the graph results showed that the proposed methodology: ANFIS-GA performs significantly in all performance measurements. Here, ANFIS classifier is used for identifying the relevant fuzzy topics, and then the optimization procedure is carried-out for predicting the best outcomes. Though, the outcome is directly forecasted with-out utilizing fuzzy generation in the existing methodology (LDA-PFCM) that is considered as one of the major concerns. Finally, the performance metrics confirmed that the proposed methodology (ANFIS-GA) performs effectively in twitter sentiment analysis related to existing methodology (LDA-PFCM) in light of recall, accuracy, f-measure, and recall.

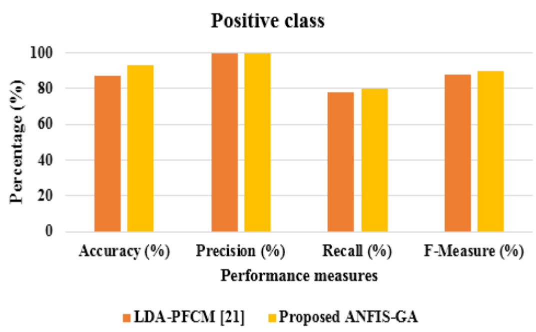

Fig. 2. Graphical comparison of proposed and existing methodology for positive class

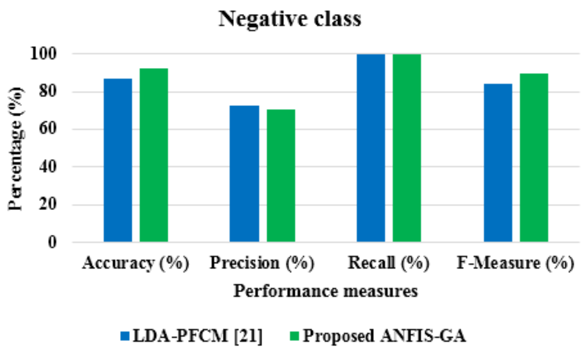

Fig. 3. Graphical comparison of proposed and existing methodology for negative class

VI. Conclusion

In present scenario, twitter sentiment analysis is one of the developing research areas to find and analyze the view-points and sentiments of the individuals. In this research study, a new system is developed for the classification of sentiment tweet classes. The combination of both ANFIS and GA achieved superior performance, while related to existing methodology. The optimization approach (GA) with multi-objective classifier (ANFIS) averagely achieved a classification accuracy of 92.78%. In addition, the proposed methodology (ANFIS-GA) is trained by utilizing two main phases; preprocessing and feature generation. This experimental analysis was verified on a publicly available database; twitter-sanders-apple2 that demonstrates the advantage of proposed methodology. Related to existing methodology (ANFIS-GA), the proposed methodology achieved a superior performance in light of accuracy, which showed 5.5-6% of enhancement in twitter sentiment analysis. In future work, a multi-objective classifier is combined with an effective feature selection approach to further enhance the classification accuracy of twitter sentiment analysis.

References

[1] A. Somani, and U. Suman, “Counter measures against evolving search engine spamming techniques,”, In proceedings of IEEE 3rd International Conference on Electronics Computer Technology (ICECT). vol. 6, pp. 214–217, 2011.

[2] P. Katta and N.P. Hegde, “A Hybrid Adaptive Neuro-Fuzzy Interface and Support Vector Machine Based Sentiment Analysis on Political Twitter Data,” International Journal of Intelligent Engineering and Systems, Vol.12, No.1, pp.165-173, 2019.

[3] E. Kušen, and M. Strembeck, “Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections,” Online Social Networks and Media, vol. 5, pp. 37-50, 2018.

[4] R. Varatharajan, G. Manogaran, M. K. Priyan, V. E. Balaş, and C. Barna, “Visual analysis of geospatial habitat suitability model based on inverse distance weighting with paired comparison analysis,” Multimedia tools and applications, vol. 77, pp. 17573-17593.

[5] P. H. Shahana, and B. Omman, “Evaluation of Features on Sentimental Analysis,” Procedia Computer Science, vol. 46, pp. 1585-1592, 2015.

[6] A. S. Al-Subaihin, and H. S. Al-Khalifa., “A system for sentiment analysis of colloquial Arabic using human computation,” The Scientific World Journal, 2014.

[7] A. Tripathy, A. Agrawal, and S. K. Rath, “Classification of Sentimental Reviews Using Machine Learning Techniques,” Procedia Computer Science, vol. 57, pp. 821-829, 2015.

[8] N. Saleena, “An Ensemble Classification System for Twitter Sentiment Analysis,” Procedia Computer Science, vol. 132, pp. 937-946, 2018

[9] E. V. Balan, M. K. Priyan, C. Gokulnath, and G. U. Devi, “Fuzzy based intrusion detection systems in MANET,” Procedia Computer Science, vol. 50, pp. 109-114.

[10] G. Manogaran, R. Varatharajan, and M. K. Priyan. “Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system,” Multimedia tools and applications, vol. 77, pp. 4379-4399, 2018.

[11] G. U. Devi, E. V. Balan, M. K. Priyan, and C. Gokulnath, “Mutual authentication scheme for IoT application,” Indian Journal of Science and Technology, vol. 8, 2015.

[12] E. Haddi, L. Xiaohui, and Y. Shi. “The role of text pre-processing in sentiment analysis”, Procedia Computer Science , vol. 17, pp. 26-32, 2013.

[13] M. S. Neethu, and R. Rajasree, “Sentiment analysis in twitter using machine learning techniques,” In proceedings of IEEE Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013.

[14] Y. Wan, and Q. Gao, “An ensemble sentiment classification system of twitter data for airline services analysis,” In proceedings of IEEE International Conference on Data Mining Workshop (ICDMW), 2015.

[15] F. O. de França, D. H. Goya, and C. L. de Camargo Penteado, “User profiling of the Twitter Social Network during the impeachment of Brazilian President”, Social Network Analysis and Mining, vol. 8, pp. 5, 2018.

[16] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method,” Information Processing & Management, vol. 53, pp. 764-779, 2017.

[17] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, “Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis,” Knowledge-Based Systems, vol. 69, pp.24-33, 2014.

[18] B. Heredia, Joseph D. Prusa, and Taghi M. Khoshgoftaar. "Social media for polling and predicting United States election outcome." Social Network Analysis and Mining, vol. 8, pp.48, 2018.

[19] M. Bouazizi, and T. O. Ohtsuki. “A pattern-based approach for sarcasm detection on Twitter,” IEEE Access, vol. 4, pp. 5477-5488, 2016.

[20] R. Varatharajan, G. Manogaran, and M. K. Priyan, “A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing,” Multimedia Tools and Applications, vol. 77, pp. 10195-10215, 2018.

[21] M. Trupthi, S. Pabboju, and G. Narsimha, “Possibilistic Fuzzy C-means Topic Modelling for Twitter Sentiment Analysis,” International Journal of Intelligent Engineering and Systems, vol. 11, 2018.