Mapana - Journal of Sciences
2023, Vol. 22, Special Issue 2, 63-79
ISSN 0975-3303|https://doi.org/10.12723/mjs.sp2.4

Fuzzy Based Sentiment Classification Using Fuzzy Linguistic Hedges for Decision Making

Dr. A. Angelpreethi*

Abstract

Sentiment analysis is used to identify the attitude, opinions, and emotions of people towards a certain topic or entity. The goal of this paper is to develop a model to do sentiment classification of online product reviews using fuzzy linguistic hedges. The proposed model will be trained on a corpus of reviews, and will be able to classify reviews into a number of sentiment categories, such as positive, neutral, and negative. The proposed model will use fuzzy linguistic hedges to improve the accuracy of the sentiment analysis. The fuzzy linguistic hedges will be used to add context and nuance to the sentiment analysis, and will enable the model to better distinguish between subtle differences in sentiment. The proposed model is tested with microblog electronics dataset. The proposed model is used for making decisions.

Keywords: Fuzzy linguistic hedges, text classification, sentiment analysis, pre-processing.

* Assistant Professor, Department of Computer Science, St.Joseph's College, Trichy-2.

1. Sentiment analysis:

Sentiment analysis is a process of analysing text data and extracting subjective information from it. It is commonly used to determine the overall sentiment or opinion of a text and classify it as positive, negative or neutral. The procedure involves using natural language processing and machine learning techniques to identify and extract subjective information from text data. Sentiment investigation can be applied to reviews, conversations and other forms of text to better understand how people feel about a particular topic. Sentiment analysis can be used to assess customer satisfaction, market sentiment, and brand perception.

Fuzzy based sentiment classification is a category of sentiment analysis that habits fuzzy logic to identify the sentiment of a given text. It works by applying fuzzy set theory to determine the meaning of the text and then assigning a sentiment score to it. This approach is useful in situations where traditional sentiment analysis may not be able to accurately capture the sentiment that is being expressed. For example, if a text contains words with multiple meanings, fuzzy based sentiment analysis can help to clarify the sentiment. Additionally, fuzzy based sentiment analysis can help to capture the sentiment in a more nuanced way than traditional sentiment analysis.

2. Fuzzy Logic:

Fuzzy logic is a form of logic that allows for degrees of truth, rather than only true or false values. It can be represented mathematically by the set of fuzzy membership functions where

$$M = \{\mu(x) \mid x \in X\} \quad \text{-----------(1)}$$

X is the universe of discourse and $\mu(x)$ is a real-valued function that assigns a degree of membership to each element x in X.

3. Linguistic Hedges:

Hedges in sentiment analysis are linguistic expressions used to soften the intensity of a statement or opinion. They are used to indicate doubt, uncertainty, or a lack of commitment to a statement. Examples of hedges include words like "maybe," "perhaps," "sort of," and "kind of," as well as phrases like "I'm not sure," "It could be," and "I think." Hedges can be used to downplay the sentiment of a statement, making it appear less extreme or opinionated. They can also be used to increase the sentiment of a statement, making it appear more certain or confident. Hedges can be used to modify the sentiment of a statement and make it more palatable to the reader.

4. Review of Literature:

There have been many research studies conducted on the use of linguistic hedges in sentiment analysis. One of the earliest studies was conducted by Wallis and Dunne (1994), who explored the use of hedges in natural language processing. They found that hedges can be used to improve the accuracy of sentiment analysis by reducing the uncertainty of the sentiment expressed in a text.

Zhou et al [6] proposed a fuzzy-based sentiment classification approach that utilizes fuzzy linguistic hedges and domain-specific sentiment words. The proposed approach was evaluated on several target datasets, and the new results showed that the proposed approach outperformed other advanced approaches in terms of classification accuracy. However, this approach failed to handles the slang words in a better way.

Wang et al [7] proposed a fuzzy clustering approach with linguistic hedges and sentiment propagation for sentiment classification. The proposed approach was evaluated on two benchmark datasets, and the experimental results showed that the proposed approach outperformed other state-of-the-art approaches in terms of classification accuracy. However, this approach failed to handles the large volume of data.

Zheng et al [8] proposed a hybrid fuzzy model for sentiment classification that combines fuzzy linguistic hedges and term frequency-inverse document frequency (TF-IDF) weighting. The proposed approach was evaluated on three benchmark datasets, and the experimental results showed that the proposed approach outperformed other state-of-the-art approaches in terms of classification accuracy. Though this proposed fuzzy model failed to handle big data.

Lu et al [9] proposed a fuzzy-based sentiment classification approach that utilizes fuzzy linguistic hedges and an attention mechanism. The proposed approach was evaluated on three benchmark datasets, and the experimental results showed that the proposed approach outperformed other state-of-the-art approaches in terms of classification accuracy.

Peter D et al [1] has proposed an unsupervised learning algorithm by calculating the sematic orientation of the sentiment terms in a dataset using Pointwise Mutual Information and algorithms of information retrieval. Their method achieves 74% accuracy.

An approach for aspect-based sentiment classification was presented by Anuradha et al. in [3]. It was divided into two stages: the pre-processing stage and the categorization stage. The classifier was trained using the Naive Bayes approach during the classification phase, and fuzzy logic based on which the test set's sentences were evaluated. To give a clear description of the product, classification was improved. Sentiment classification tasks and term weighing tasks, respectively, included the incorporation of fuzzy linguistic hedges.

A supervised fuzzy inference system based on hedge functions in the presence of an adverbial modifier was created by Reshma et al. [5]. This approach was used to analyse the n-grams of adverbial modifiers. Linguistic hedges can be used to identify opinions in addition fuzzy criteria were used to increase the impact of an opinion. To describe ambiguous and imprecise information, the system generated a range of degree values, yielding encouraging outcomes.

However, the above works failed to include the slang words along with sentiment score and hedge score. Therefore, the effective model is needed to handle slang words, sentiments and hedges using fuzzy based classification for a fine-grained analysis.

5. Proposed Methodology:

Methodological diagram of the proposed work is depicted in Figure 1.

5.1 Phase I: Data Collection and Pre-processing:

5.1.1 Data Collection:

Sentiment Analysis has been performed on various electronic dataset of online microblog reviews which has collected using web crawler. The collected review consists of unwanted text to extract the knowledge from the text pre-processing techniques needs to apply for the selected datasets.

5.1.2 Pre-Processing the text:

Text pre-processing is a crucial step in natural language processing (NLP) that involves cleaning, transforming, and normalizing raw text data to prepare it for analysis [12][15]. There are various techniques used in NLP for text pre-processing, some of which are presented in Table 1.

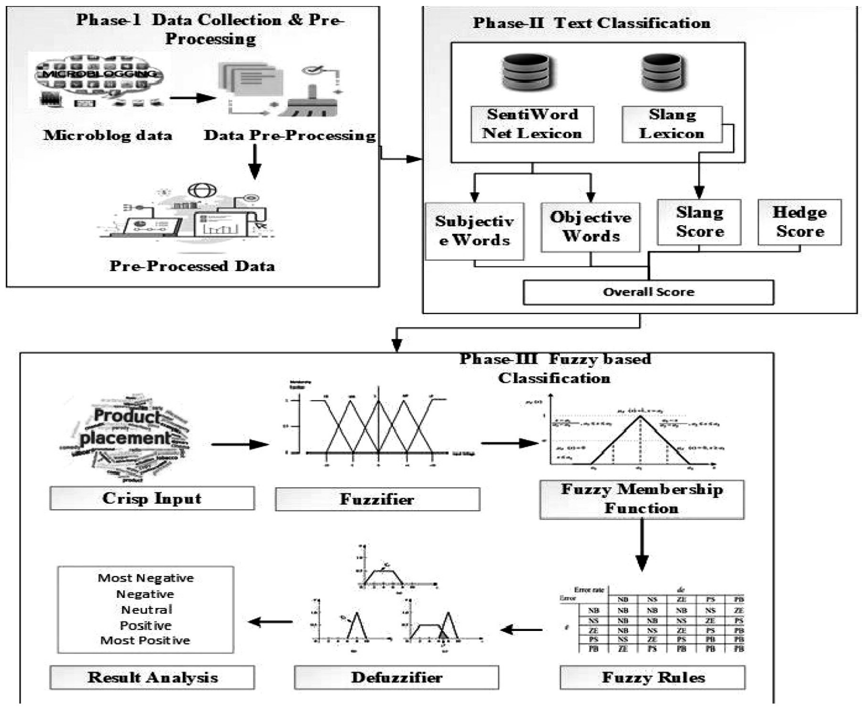

Figure 1 Sentiment classification using Fuzzy.

Table 1. Text Pre-Processing mechanism

Name of the technique	Working Mechanism	Example
Tokenization	It is the process of breaking down a sentence into smaller units such as words or sub words. Tokenization helps to create a structured form of data that can be used for further analysis.	"I love eating pizza." After tokenization, the sentence will be broken down into individual tokens as: ["I", "love", "eating", "pizza", "."]
Part-of-Speech Tagging	It is the procedure of marking each word in a text corpus with its equivalent part of speech such as adjective, noun, verb, adverb, pronoun, preposition, conjunction, interjection or other part of speech.	"I love to go to running in the park" After tagging the words will become [('I', 'PRP'), ('love', 'VBP'), ('to', 'TO'), ('go', 'VB'), ('running', 'VBG'), ('in', 'IN'), ('the', 'DT'), ('park', 'NN'), ('.', '.')]

Stop word Removal	Stop words are words that do not add any value to the analysis such as "the", "is", "a", "an", etc. These words can be removed from the text data to improve the accuracy of the analysis.	"The cat is sitting on the mat." After stop word removal, the sentence will become: "cat sitting mat."
Stemming and Lemmatization	These techniques used to reduce a word to its root form. The root form of a word is called its lemma. Stemming includes removing the suffix from the word to get its root form, whereas lemmatization using a dictionary to find the base form of the word	consider the words "running" and "ran". After stemming, the words will become "run". After lemmatization, "running" will become "run" and "ran" will remain "ran".
Lowercasing	Adapting all the text to lowercase can help to reduce the number of unique words in the text data and make it easier to analyze.	consider the sentence: "The quick BROWN fox JUMPS over the LAZY dog." After lowercase conversion, the sentence will become: "the quick brown fox jumps over the lazy dog."
Removing Punctuation	Punctuation marks such as commas, periods, exclamation marks, etc., can be removed from the text data as they do not add any value to the analysis.	"I am feeling happy!" After punctuation removal, the sentence will become: "I am feeling happy"

Overall, these techniques help to clean and standardize the text data and make it suitable for further NLP processing. Python based NLTK tool is used in this work to clean the text data.

5.2 Phase II: Text Classification using Lexicons:

Pre-processed text has been given as an input to the Phase-II. This method uses the trigram-based feature selection which consists of sentiment word, slang word and linguistic hedges. Scores are automatically allocated using the occurrence of the specific word using SentiWordNet (SWN). The manually constructed slang lexicon provides score for the slang terms.

Based on the occurrence of the lexicons the words are categorized into opinionated words (subjective) and non-opinionated words(objective).

Hedges are also called as modifiers. Modifiers either increases the sentiment value (Enhancers) or decreases the sentiment value(Reducers). Any Linguistic hedge H taken as unary operator h with the interval [0,1].

For example linguistic hedge value of enhancers and reducers are calculated using

$$h_{enh}(a) = a*a \text{ for all } a \in [0,1] \text{ ------------------(2)}$$

$$hred(a)= a^{1/2} \text{ for all } a \in [0,1] \text{ -------------------(3)}$$

Where a denotes the linguistic modifier.

The overall score of the sentence is calculated by adding the scores of sentiment words, slang words and linguistic hedges. The overall final score is given as an input to the fuzzy based classification.

5.3 Phase III: Fuzzy based Classification:

In Phase III classifies the user reviews in a fine-grained way. Fuzzy logic defines linguistic variables by assigning a range of values that are represented using natural language terms or linguistic labels. These labels can be defined using membership functions that assign a degree of membership to each value in the range for each label. The membership value of the fuzzy set is defined between the limits 0 and 1. Fuzzy membership functions are used in conjunction with fuzzy if-then rules to make decisions about the sentiment of a text, by evaluating the degree of membership of linguistic variables to different sentiment categories, and combining the results to determine the overall sentiment.

In the first stage of the phase III the sentiments, hedges and slang words score are summated and given as an input to the fuzzy based classification system. Input values are differing in their ranges. To make uniformity in their range's normalization has to be made. In the proposed system. Since each value of the words are high normalization is performed with the help of min-max normalization.

5.3.1 Min-Max normalization:

Min-max normalization is a common technique used in fuzzy-based sentiment analysis to normalize data and bring it into a common scale, so that it can be more easily compared and combined. Fuzzy logic-based sentiment analysis systems use linguistic variables and fuzzy sets to represent the degree of membership of a sentence to a particular sentiment category (such as "positive" or "negative"), and these linguistic variables often have different scales and ranges.

For example, the linguistic variable "battery life" might have a range of 0-100, while the linguistic variable "price" might have a range of 0-1000. Without normalization, these variables cannot be easily compared and combined, as they are on different scales.

Min-max normalization is used to rescale the values of the linguistic variables to a common scale (typically between 0 and 1), by subtracting the minimum value from each data point, and then dividing by the range (i.e., the maximum value minus the minimum value). This transformation ensures that all linguistic variables have the same range and scale, and that they can be more easily compared and combined using fuzzy logic operations.

$$X\_normalized = (X - X\_min) / (X\_max - X\_min) \quad \text{------------(4)}$$

where:

- X is the original value of the data point
- X_min is the minimum value of the data points in the dataset
- X_max is the maximum value of the data points in the dataset
- X_normalized is the normalized value of the data point, which ranges between [0,1]

For example, if the linguistic variable "battery life" has a range of 0-100 and a value of 80, and the linguistic variable "price" has a range of 0-1000 and a value of 600, then after min-max normalization, the values will be rescaled to 0.8 and 0.6, respectively, and can be easily compared and combined using fuzzy logic operations.

Overall, min-max normalization is a useful technique in fuzzy-based sentiment analysis as it helps to bring all variables into a common scale, enabling easy comparison and combination of linguistic variables.

5.3.2 Fuzzifier:

Fuzzification is the process of mapping a crisp (i.e., precise) input value to a degree of membership in one or more fuzzy sets. In fuzzy sentiment analysis, fuzzification is used to map input linguistic variables, such as words or phrases that express sentiment, to degrees of membership in fuzzy sets that represent sentiment categories, such as positive, neutral, and negative.

To fuzzify an input linguistic variable, we first map the input value to the appropriate fuzzy set using the corresponding membership function.

In the Proposed System uses three linguistic variables namely Design, user interface, camera quality. The word set of a linguistic variable design is defined as

W(Design)= { Stylish, modern, outdated}

W(User Interface) = { Intuitive, user-friendly , confusing}

W(camera quality) = {excellent, blurry, bad}

Table 2. linguistic variable design

linguistic variable	word set	Range
Design	Stylish	100-75
	Modern	75-50
	Outdated	Less than 50
User Interface	Intuitive	100-80
	user-friendly	80-50
	confusing	Less than 50
camera quality	excellent	100-70
	blurry	70-50
	bad	Less than 50
Customer Rating (*Output variable*)	Highly recommended	100-80
	recommended	80-60
	Average	60-40
	Satisfactory	40-20
	bad	Less than 20

5.3.3 Fuzzy Rules:

In a fuzzy logic system, the rules are typically expressed in the form of "if-then" statements, where the "if" part specifies the conditions for the rule to be applied, and the "then" part specifies the action or output that should be taken if the rule is applied.

Fuzzy logic is a mathematical framework that is designed to handle imprecision and uncertainty in data, by allowing values to be represented as fuzzy sets, which can have degrees of membership ranging from 0 to 1. If-then rules in fuzzy logic systems use these fuzzy sets to represent the relationships between inputs and outputs, and they use fuzzy reasoning to make decisions based on these rules.

The proposed system used the following rules for decision making purpose for the benefit of customers and business organizations. Some of the rules are

- If the Design is stylish, user interface is intuitive, and camera quality is excellent then the customer rating is highly recommended.
- If the Design is stylish, user interface is user-friendly, and camera quality is excellent then the customer rating is recommended.
- If the Design is modern, user interface is user-friendly, and camera quality is blurry then the customer rating is Average.
- If the Design is out-dated, user interface is user-friendly, and camera quality is blurry then the customer rating is Satisfactory.
- If the Design is out-dated, user interface is confusing, and camera quality is bad then the customer rating is bad.

5.3.4 Defuzzifier:

In fuzzy logic, defuzzification is the process of converting a fuzzy output into a crisp, numerical value that can be used for decision-making. Defuzzification is necessary because fuzzy logic operations often result in fuzzy outputs, which are represented as fuzzy sets with degrees of membership.

One common defuzzification method is the centroid method, which calculates the center of gravity of the fuzzy output set. The centroid is calculated by finding the weighted average of the values in the fuzzy set, where the weights are the degrees of membership. The formula for centroid defuzzification is:

$$output = (sum(x * \mu(x))) / sum(\mu(x)) \text{ -----------(5)}$$

where x is the value in the fuzzy set, μ(x) is the degree of membership of x, and the summation is over all values in the fuzzy set.

Here's an example of defuzzification using the centroid method. Suppose we have a fuzzy sentiment analysis system that analyzes customer reviews of a mobile phone, and outputs a fuzzy set representing the sentiment of the reviews. The fuzzy set has a triangular membership function with the parameters (0.4, 0.7, 1.0), and represents the sentiment category "positive". The degree of membership of the fuzzy set at its peak value (x=0.7) is 0.9, indicating a high degree of certainty that the sentiment is positive.

To defuzzify the output, we can use the centroid method to calculate the center of gravity of the fuzzy set:

$$\text{output} = (0.4*0 + 0.7*0.9 + 1.0*1) / (0 + 0.9 + 1) = 0.79$$

The resulting value of 0.79 represents the degree of positivity of the reviews, on a scale from 0 to 1. This crisp output can be used for decision-making, such as ranking the mobile phone among other products based on sentiment analysis.

6. Result Analysis:

The overall performance and effectiveness of the proposed approach is presented in Table 3. In addition, diagrammatically represented in Figure 2. The proposed technique is evaluated by using various evaluation metrics such as precision, recall, F-Measure and accuracy to measure its performance.

Table 3. Performance of the Proposed Approach

Evaluation Measures	Fuzzy based Approach	Accuracy in percentage
Precision	0.8102	81.02%
Recall	0.7903	79.03%
F-Measure	0.8001	80.01%
Accuracy	0.8254	82.54%

Table 3 shows the performance of the fuzzy based approach using the software metrics precision, recall , F-Measure and Accuracy.

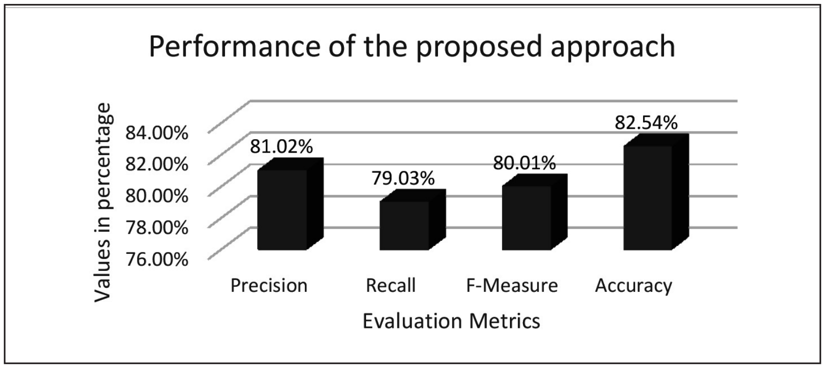

Figure 2 Results of the Proposed Approach

The proposed technique is compared with the results of the existing approaches. The comparative study of evaluation measures is depicted in Table 4 and Figure 3 respectively.

Table 4 Accuracy Comparison

S. No.	Approaches	Accuracy (%)
1	Reshma et al .,[4]	76.32
2	Masud et al., [16]	81.00
3	Saprativa et al.,[17]	68.46
4	Ayushi et al.,[18]	67.04
5	**Proposed**	**82.54**

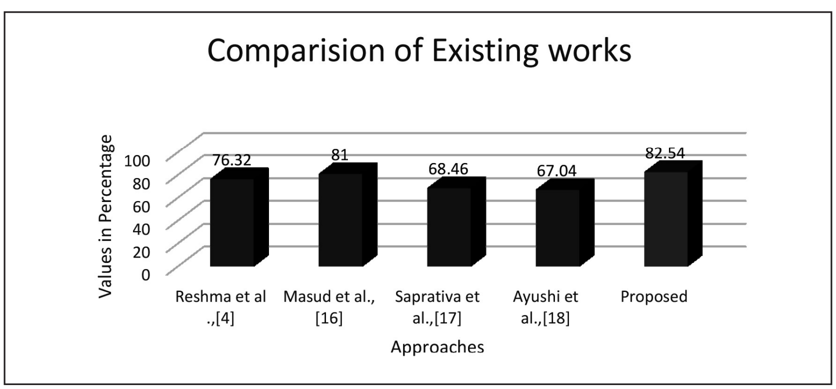

Figure 3 Comparison of Existing approaches with proposes approach

Figure 3 shows the comparison of existing work with proposed work. The X axis represents the approaches and Y axis represents the accuracy in percentage. While comparing existing work, the proposed work acquires better accuracy.

7. Conclusion:

The proposed fuzzy based model performs fine-grained sentiment classification of online product reviews using fuzzy linguistic hedges. The model presents the fuzzy based rules to classify the user sentiments. The methodological diagram of the proposed system is presented. The crisp inputs, membership functions, linguistic variables and output variables are defined. This proposed system which help the internet users for making decisions. The proposed model handles sentiments, linguistic hedges, slang words using fuzzy based sentiment classification. 82.54% Accuracy achieved and the results are compared with existing approaches. This model is simple but it outperforms well.

References:

1. Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 417-424, july 2002
2. Yan Dang, Yulei Zhang, and Hsinchun Chen, "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," Journal IEEE Intelligent Systems, Volume 25 Issue 4, Pages 46-53, July 2010.
3. Anuradha, G., Joel, D.J.: Fuzzy based summarization of product reviews for better analysis.
   Indian J. Sci. Technol. 9(31), 1–9 (2016)

4. V. Reshma, A. John, Aspect based summarization of reviews using naïve Bayesian classifier and fuzzy logic, in 2015 International Conference on Control Communication & Computing India (ICCC), IEEE, Trivandrum, India, 2015, pp. 617–621.
5. R. Katarya, O.P. Verma, An effective web page recommender system with fuzzy c-mean clustering, Multimed. Tools Appl. 76 (2017), 21481–21496.
6. Zhou, J., He, J., & Zhao, Y. (2021). Fuzzy-Based Sentiment Classification Using Fuzzy Linguistic Hedges and Domain-Specific Sentiment Words. IEEE Access, 9, 81315-81328.
7. Wang, H., & Zhang, L. (2021). Fuzzy clustering with linguistic hedges and sentiment propagation for sentiment classification. Journal of Intelligent & Fuzzy Systems, 41(1), 67-79.
8. Zheng, Y., Liu, H., & Zhang, J. (2021). A hybrid fuzzy model for sentiment classification based on text reviews. Neural Computing and Applications, 33(7), 2945-2960.
9. Lu, Y., Chen, X., Hu, J., & Zhang, J. (2020). Sentiment classification based on linguistic hedges and attention mechanism. Expert Systems with Applications, 139, 112844.
10. Karimi, S., Mohammadi, M. R., & Sadeghi, H. (2021). A fuzzy-rough hybrid approach to feature selection for sentiment classification. Soft Computing, 25(2), 1345-1358.
11. Wang, M., Li, Y., & Zhang, Y. (2020). Sentiment classification of product reviews using fuzzy comprehensive evaluation method. Soft Computing, 24(23), 17425-17434.
12. A.Angelpreethi, Dr.S.Britto Ramesh Kumar (2018)," Dictionary based approach to improve the accuracy of opinion mining on big data", International Journal of

scientific research in computer science and management studies, volume 7, issue 5 (sep 2018).

13. Dr. P. Bastin thiyagaraj, Dr..A. Aloysius (2023), "Data Reduction Based Truth Discovery Analysis by Resolving the Conflicts in Big Data using Continuous Data", International Journal for Innovative Engineering and Management Research, volume 12, issue 02, ISSN 2456-5083.
14. Supriya, B.N., Kallimani, V., Prakash, S., Akki, C.B(2016).: Twitter sentiment analysis using binary classification technique. In: Vinh, P.C., Barolli, L. (eds.) ICTCC 2016. LNICST, volume. 168, pp. 391–396. Springer, Cham.
15. A. Angelpreethi and S. B. R. Kumar (2017), "An Enhanced Architecture for Feature Based Opinion Mining from Product Reviews," in 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, Tamil Nadu, India, pp. 89–92.
16. Fazal Masud Kundi, Aurangzeb Khan, Shakeel Ahmed and Muhammad Zubair Asghar, "Lexicon-Based Sentiment Analysis in the Social Web", Journal of Basic and Applied Scientific Research, Vol. 4, Issue 6, 2014, pp. 238-348.
17. Saprativa Bhattacharjee, Anirban Das, Ujjwal Bhattacharjee, Swanpan K. Parui and Sudipta Roy, "Sentiment Analysis using Cosine Similarity Measure", 2nd International Conference on Recent Trends in Information Systems, IEEE, 2015, pp. 27-32.
18. Ayushi Dalmia, Manish Gupta, and Vasudeva Varma, "IIIT-H at Sem Eval 2015: Twitter Sentiment Analysis The good, the bad and the neutral" Proceedings of the 9th International Workshop on Semantic Evaluation, 2015, pp. 520-526.

****