International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

ROLE OF ARTIFICIAL NEURAL NETWORK IN OPINION CLASSIFICATION

[1]Dr. Helen Josephine V L, [2]Gomathi Thiyagarajan, [3]Dr. V. S. Anita Sofia

***ABSTRACT - ****In the present world the internet users enlarged rapidly, consumers are more involved in searching and selecting best product. E-commerce organizations are also spent a lot of time, efforts and cost to investigate the comments and feedback about their products. This scrutiny would assists the corporations to enhance their services and products at low cost, which in turn would facilitate them to extend and flourish in the industry. To extract knowledge from vast unstructured data which is in the form of customer reviews finds it very difficult to analyze and evaluate. This research paper analyze the role of neural network in opining mining classification and proposed Gain Ratio feed forward neural network (GR_FFNN) algorithm to improve classification accuracy. This paper also compares the GR_FFNN algorithm with the existing neural network classifier.*

Keywords *– Classification, Machine Learning, Neural Network, Opinion Mining*

I. INTRODUCTION

An enormous amount of information's is obtainable from the Internet in the form of online product reviews, feedback, micro blogs and comments on the website and in social media like Facebook, Orkut, and Twitter. Consumers give more attention to Google, Amazon, Yelp and Facebook online reviews. Both consumers and providers are forced to analyze and investigate the reviews to decide the right product and potential customers. The reviews are valuable only if it is honest, communicating the true feelings and sentiment towards the product, the service, the politician, the global event or any object. The opinion mining and sentiment analysis has become significant and vital area of web content mining and text mining and many researchers paying attention over the past decade. Recent studies suggest that artificial neural network based algorithms provide better performance in opinion classification. This research paper is organized in five sections. The categories of Sources containing rich opinions are explained section 2. Section 3 describes the motivation and literature survey. Section 4 explains the role of artificial neural network and proposed gain ratio feed forward neural network classification algorithm. Results and discussions are examined in Section 5. Section 6 concludes this article.

II. SOURCES CONTAINING RICH OPINIONS

A. Online Reviews

[1] Associate Professor, Department of Computer Applications, CMR Institute of Technology, Bengaluru, helenjose.cbe@gmail.com
[2] Assistant Professor, Department of Computer Applications, CMR Institute of Technology, Bengaluru, gomathi.t@cmrit.ac.in
[3] Associates Professor, Department of Computer Technology, Sri Krishna Arts and Science College, Coimbatore, anitasofiav@skasc.ac.in

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

In this decade, e-commerce has exponential growth and offers an excellent platform for the customers and vendors for the product or services. Online reviews, surveys, and other related substances are the potential driving force for the customers to choose the product. These reviews create a strong effect on consumers' choice by through assisting them to choose for the product smartly. An online review also helps business by allowing them to enhance their products and provide assistance to their business [1]. Organizations regularly analyze the reviews that are posted by their clients. Opinion Mining identifies and extracts the opinions in customer review document and categorizes the opinions of clients in terms of either positive, negative or neutral. These outputs are assisted in different fields like merchandising, recommender systems, contextual advertising, etc.

B. Social Media

In the 21st century, social network media is one of the extensive exchanges of information using technology. Easy access to micro-blogging platform where there is no restriction in message formats provides comfortable feeling to the Internet user to express their views and experiences in micro-blog than blogs and emails. Twitter is one of biggest social network and micro-blogging website and it contains short and quick messages called "Tweets". Each tweet has a maximum of 140 characters [2]. Every minute 350,000+ tweets are tweeted on Twitter and on average around 200 Billon tweets every year. Twitter has 336 million monthly active clients and it contains a good source of information. Facebook, Google+ are the other social networking websites where the people carryout user frequently talks and share about current affairs and their opinions. In recent times, Twitter has become popular around the world. Numbers of dynamic users are increasing every year and therefore it has been an important source to mine opinions.

C. Blogs

There are a large number of blogs posted on the internet every day. They are an essential source of knowledge. The blog is also known as a weblog, it an online informational website or journal in which a person can share his/her knowledge and opinions in regular intervals. Opinion mining of the viewpoints might be more important for business organizations where they investigate the opinion of the products.

III. LITERATURE SURVEY

Artificial Neural Networks (ANN) as a soft computing technology is uncommonly been studied in the academic journal of opinion classification. But in recent years, extensively advanced in artificial neural network methodology, similar to high-speed training algorithm for deep neural networks has emerged. The Neural Network with suitable network structure is able to identify the relationship or correlation among input variables[3]. However very few studies are available in opinion classification using neural networks, the literature does not provide more work in opinion classification by means of the probabilistic neural network[4].

Sentiment classification task consists of two stages: (i) identifying and extracting useful features (ii) classifying the reviews with the help of available supervised learning methods, Support Vector Machines (SVM) and Naı¨ve Bayes (NB)[5]. SVM have been widely used to learn the opinion. At the same time, Artificial Neural Networks

(ANN) have hardly ever been used to analyze opinion in comparative statement. Morae.R. et al. have presented a back propagation neural network based document level opinion classification for movie review dataset[6].

Ghiassi.M. et.al. have introduced an approach for feature dimension reduction based on *n*-grams and have used statistical analysis to extend a Twitter-specific lexicon for opinion mining[7]. This Twitter-specific lexicon was added with the brand-specific terms. Dynamic neural network has been used to classify the sentiment. This research reveals that the compact lexicon set decreases the model complexity, covers the entire corpus and achieves enhanced opinion sentiment classification accuracy. The Twitter-specific lexicon was compared with the conventional sentiment lexicon. The experimental results show that diverse Twitter-specific lexicon is extensively more efficient in terms of classification accuracy and other performance metrics. Opinion classification model was implemented with the help of Twitter-specific lexicon and machine learning approach namely DAN2 The results reveals that DAN2 provides more perfect opinion classification results than SVM with the Twitter-specific lexicon.

Anuj et al., (2012) have recommended opinion classification model with the help of back-propagation artificial neural networks (BPANN) which is one of the most popular neural networks[8]. He has utilized three different famous opinion lexicons and concepts of Information Gain for identifying and extracting opinion. These extracted are aspects further used to train the BPANN. This technique combines the advantages of BPANN in order to get better classification accuracy along with inherent subjectivity understanding which is obtainable in the opinion lexicons. The experiments were conducted on the movie and hotel review corpora. The results revealed that the recommended method facilitates to reduce dimension of the feature set. Moreover, it provides precise results for opinion classification.

N. Mohd Naw et al., introduced modification in back propagation algorithm along with conjugate gradient optimization algorithm which results new algorithm, Fletcher-Reeves update with adaptive gain called CGFR/AG[9]. This technique contains three different stages. 1) Adopting gain variation term of the activation function which modifies on standard back propagation algorithm. 2) Evaluate the error based on gradient decent with the corresponding to the weights and gain values 3) Obtain latest new search direction by utilize the information evaluated through the stage (2) along with the farmer search direction. This recommended method enhanced the efficiency in training back propagation algorithm by incorporating the updated initial search direction[10].

Huawang Shi, (2009) took initiative to provide optimized neural network through properly handling the genetic operators[11]. Back propagation algorithm suffered with drawback of slow convergence that would be solved by using momentum and problem of premature convergence also eliminated.

Norhamreeza et al., (2011) recommended a technique to enhance back propagation from getting trapped with local minima problem which results delay in optimal solution convergence and neuron dispersion in the hidden layer[12]. Asha Gowda Karegowda et al., explained in detail about the feature selection and evaluation method namely, Correlation-based.[13] The problems were identified with Information gain. It was biased towards attributes with many values. Gain parameters did not work for new data. Gain ratio is a relation between information gain to the essential information. Compared to IG, Gain ratio overcomes the flaws of information gain Gongde Guo [14].

IV. Artificial Neural Network

In Late 1950, Cornell, Frant Rosenblatt constructed various naturally motivated learning apparatus and simulated them on a digital computer[15]. He called these designs "perceptrons" to highlight their perceptive, rather than logical, abilities. The plan of many of his devices was to be trained through the example to differentiate whether an input was a member of each class of inputs, either class ‘A’, or class ‘B’. It was attained through examples of members of each class along together with the right classification. The class members, any finite number being allowed, were represented by (n-1) vectors where x_i^j denotes the element -‘i’ of the vector ‘j; Given a vector, the simplest perceptron would compute.

$$\sum_{i=1}^{n-1} w_i x_i^j - t \tag{1}$$

where w_i is denoted as weights, updatable during learning and analogous in the role to synapses in the human brain, applied to the input elements x_i^j, and t, also modifiable was called a threshold. Let net^j denotes the amount by which the weighted input exceeds the threshold t (where net^j can be negative), the output unit (artificial neuron) would output either 1, or -1 if net^j exceeded 0. The Threshold is linear unit mathematically by the input output relation for case j.

$$o^j = f(net^j) = f(\textstyle\sum_{i=1}^{n-1} w_i x_i^j - t) \tag{2}$$

Then Rosenblatt and others found the perceptron convergence theorem that emphasizes that if there exist any weight that provides the desired output for all training set, one such set will be discovered in a finite number of steps. The initial value of the weight may be assigned to zero and in each iteration, in the training dataset, the weight is modified by the rule.

$$w_i(new) = w_i(old) + \tfrac{\eta}{2}(d_j - o_j) x_i^j \tag{3}$$

The disadvantage of the single perceptron is that it classifies the dataset into two classes and the dataset must be linearly separable. These disadvantages had been overcome through multilayer perceptron that would classify multiple classes and would work even for nonlinear the dataset.

A standard neural network (NN) consists of many simple and connected processors which are called neurons. Each neuron generates a sequence of real-valued activation. Input neurons get activated through sensors perceiving the surroundings; other neurons get activated through weighted connections from previously active neurons. The basic unit of artificial neural networks, namely, the artificial neurons which are assembled and interconnected in layers are simulates the basic functions of natural neurons. Compare to biological neural artificial neurons are much simpler. ANNs are consists of multiple nodes, which imitate biological neurons of the human brain. The following are the assumption made in the artificial neuron.

Information is processed in the artificial neurons –Signals are exchange the information between neurons through conventional links

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

- Each neuron is interconnected between the neurons by links with a weight; they interact with each other neuron by exchanging signals between neurons which is multiplied by the weights. The weights determine the strengths of the interconnections between two neurons.
- Each neuron in the neural network concerns an activation function to its input signals so as to standardize its output signal [16].
- The result of activation function is passed on to other neurons in the next layer. The output at each node is entitled as node value.

Each input x_i is multiplied by the respective weight w_i. Then, these products are simply summed, fed through activation function. The outcome of the activation function will be considered as inputs to the next layer or it will generate the result in the output.

These ANN has been effectively applied to real-world classification problem in the various field of research namely image and pattern recognition, text classification, speech recognition, fault detection, and medical diagnosis and language translation. This section explains the concepts of existing artificial neural network based on classification algorithm.

One of the major advantages of neural networks is their capacity to learn through training dataset[17]. Training dataset helps to train the neural network; it consists of furnishing the neural network set of training pairs, each of which contains set of input features and the corresponding target values [18].

A. Elman Neural *Network*

Jeffrey Elman proposed a neural network that was recurrent in nature and constructed to learn sequential and time-varying patterns. To this extent, the algorithms would identify and predict trained set of values or events. An Elman network consists of three units namely, input units, hidden units and, output units along with the addition of context units.[19]

The input units and context units trigger the hidden units and subsequently, hidden layer neurons feed forward to active the output layer neurons. At the same time, hidden layer neurons also provide the feedback to the context units. This represents the forward activation in the neural network. Based on the task, the learning phase may or may not happen in this time cycle. The hidden layer is linked to these context units through fixed weight At every time step, the input is fed-forward and a learning rule is applied [16].

B. Feed Forward Neural Network

There are two categories in the feedforward neural network 1) Fully connected feedforward neural network - Every single neuron in the each layer are connected with every other neuron in the previous and next layer of the neural network 2) Partially connected feedforward neural network - some neuron in input layer or hidden layer is not connected with the hidden layer or output layer

In the first step network will be initialized through fixing variables in the range +1 to -1.

1. x_i denotes 'i'th input node value.
2. $w_{j,i}$ denotes the link value or weight from the 'i'th input node to the 'j'th hidden layer node.
3. the net input to the 'j'th node in the hidden layer is calculated by the following formula.

$$\sum_{i=0}^{n} w_{j,i} x_i \tag{4}$$

The output of ‘j’th node in the hidden layer is calculated by any one of the activation function.

1. Sigmoid or Logistic

$$f(x) = \frac{\mathbf{1}}{\mathbf{1+e^{-x}}} \tag{5}$$

This is one of the popular activation functions. The important reason is the output values within the range 0 to 1.

2. Tanh—Hyperbolic tangent

The range of the tanh function is from (-1 to 1).

$$f(x) = \frac{\mathbf{1-e^{-2x}}}{\mathbf{1+e^{-2x}}} \tag{6}$$

3. ReLu -Rectified linear units

R(x) is zero when ‘x’ is less than zero and R(x) is equal to ‘x’ when ‘x’ is above or equal to zero.

R(x) = max(0,x) i.e if x < 0 , R(x) = 0 and if x >= 0 , R(x) = x.

the output of the ‘j’th node in the hidden layer - *Activation* $\left\{ \sum_{i=0}^{n} \boldsymbol{w_{j,i} x_i} \right\}$

C. Gain Ratio - Forward Neural Network(GR_FFNN)

Information Gain (IG) evaluates the total amount of information in bits concerning the class prediction, if the only information available is the presence of a feature and the corresponding class distribution [20]. In decision tree algorithm each consecutive recursion decides the best attribute to separate the data at the current node based on the values of the attribute. The best attribute is selected according to a function that seeks to reduce the impurity after partitioning. In further words, it increases the level of purity. Therefore, the main criteria in decision tree learning are the option of the impurity method. Information gain and Gain ratio are the most famous impurity methods utilized for decision tree learning.

Information gain calculates how much information a particular feature gives about the class. Features that absolutely partition should give more information. But the distinct and unrelated features will not give more information. In other words, it calculates the reduction in entropy. Entropy is the asses of impurity, uncertainty or disorder in a bunch of examples. The information gain measure depends on the entropy function from information theory[21].

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

1. Initialize all weights W1 through Wn with a random number between (0, 1).
2. Set Learning Rate to 0.1, No. of epochs =500
3. Set Minimum threshold =0.01
4. Set the values of all input nodes.
5. For epo=1 to 500

(Hidden Layer)

6. Calculate hidden node values

Summation function

$H_i = B + \Sigma w_i x_i$ where i=1 to n

$H_i = B + w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n$

Activation function (sigmoid)

$H_A = f(H_i)$ where $f(x) = \frac{1}{1+e^{-x}}$

7. Calculate additional inputs from input layer by using Gain Ratio formula

$$Gain\ Ratio\ (S, A) = \frac{Information\ Gain\ (S, A)}{SplitInformation(S, A)}$$

$$Information\ Gain(S, A) = E(S) - \sum_{v \in value\ S(A)} \frac{|S_v|}{|S|} E(S_v)$$

$$E(s) = -\sum_{i=1}^{m} p_{i \log p_i}$$

$$SplitInformation(S, A) = -\sum_{v \in value\ S(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|}$$

8. Select top 5 positive and negative features feed into the output layer

Perform Summation function

$y_k = \Sigma w_i x_i$ where i=1 to 10

$y_k = w_1x_1 + w_2x_2 + \ldots\ldots + w_{10}x_{10}$

Activation function (sigmoid)

$H_A = f(y_k)$ where $f(x) = \frac{1}{1+e^{-x}}$

9. Calculate error value using cross entropy

$$\delta_k = -y \log \hat{y} - (1-y)\log(1-\hat{y})$$

10. if ((error(δ_k)<0.01 or (epo<500))

Estimate new weight w= old weight + δ_k * learning rate + momentum

Continue step 5 to 9

Else

Goto step 12

End if

11. Next epo
12. End

Information Gain(S,A) = expected reduction in entropy due to branching on attribute A

= original entropy−entropy after branching

Entropy – Entropy: (im)purity in an random gathering of data. The calculation of entropy E(S) for m values in each feature is

$$E(s) = -\sum_{i=1}^{m} p_{i \log p_i} \tag{7}$$

Where 'p' is the proportion of positive and negative, 'S' is the sample dataset. Entropy value for each class label is calculated[22]. Then, Entropy is calculated for every value in a feature with its corresponding class label.

Information gain works very well for most cases, but for few variables which have a huge number of classes (or values. Gain ratio conquers the disadvantage of information gain by taking into account the number of classes would outcome prior to make the split. Gain Ratio removes the disadvantage of information gain by taking the essential data of a split into consideration. The information gain should be maximized with small number of partitions. Attributes need to be selected based on

$$Gain\ Ratio\ (A) = \frac{Information\ Gain(S,A) \quad \leftarrow \text{need to be Maximized}}{SplitInfo\ (S,A) \quad \leftarrow \text{need to be Minimized}}$$

The split information value denotes the potential information created by dividing the training dataset ‘D’ into ‘v’ partitions, succeeding to ‘v’ outcomes on attribute ‘A’.

V. DISCUSSION

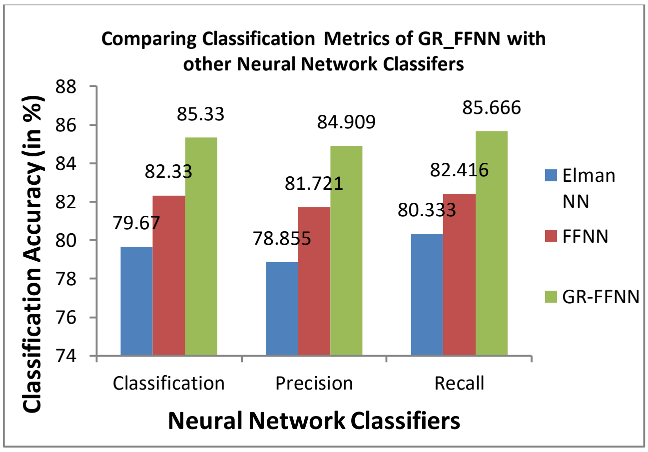
In this study, 300 learning app reviews (positive, negative and neutral) were collected from the Android Market Website for scrutiny. Mobile learning app review dataset which is the outcome of SWSF (Significant Word, SVD, and Filtering) pre-processed algorithm is fed in GR_FFNN for opinion classification [23]. The classification accuracy achieved from the experiment is 85.33%. The experiments were conducted with the existing ANN classifiers Elman neural network, and Feed Forward neural network. The classification accuracy and other classification metrics namely precision and recall have been evaluated and the values were compared. The Proposed GR-FFNN method achieved the classification accuracy of 85.33%, whereas the Elman neural network gained 79.67% and FFNN gained 82.33%. The classification accuracies and other metrics precision and recall are tabulated in Table-5.1.

The proposed algorithm GR_FFNN is validated against the existing ANN classifiers with the classification accuracy measures Precision and Recall. This result reveals that the proposed algorithm yields a better result than the existing ANN algorithm. Fig 5.1 illustrates that the affirmed methodology provides more classification accuracy. The comparative analysis it is evident to note that the proposed GR-FFNN classifier supersedes the existing ANN classifiers.

Table 5.1 : Comparing Classification metrics of GR_FFNN with existing Neural Network Classifiers

Algorithm	Elman NN	FFNN	GR-FFNN
Classification	79.67	82.33	85.33
Precision	78.855	81.721	84.909
Recall	80.333	82.416	85.666

Fig 5.1 Comparing classification metrics of GR_FFNN with other Neural Network Classifiers



VI. CONCLUSION

The main focus of this article is to analyze the role of the neural network in opinion mining classification. The proposed neural network algorithm enhances classification metrics namely classification accuracy, precision and recall. The proposed algorithm GR_FFNN is compared with existing neural network classifier Elman NN and Feed Forward NN. The comparative analysis it is evident to note that the proposed GR_FFNN classifier supersedes the existing ANN classifiers.

REFERENCES

1. X.Hu and B.Wu, 2009, Classification and Summarization of Pros and Cons for Customer Reviews, Web Intelligence and Intelligent Agent Technologies, Vol. 3, pp. 73 – 76.
2. Go A., Bhayani R., and Huang L.,(2009), Twitter sentiment classification using distant supervision‖ CS224N Natural Language Processing in Spring, pp.1– 6.
3. Lina L. Dhande and Dr. Prof. Girish K. Patnaik , " Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier " , International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 4, July - August 2014 , pp. 313-320 , ISSN 2278-6856.
4. Cambria. E., Mazzocco.T., & Hussain. A. (2013), Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining, Biologically Inspired Cognitive Architectures, vol. 4, pp. 41-53.
5. V. Dhanalakshmi,Dhivya Bino,A. M. Saravanan,2016, Opinion mining from student feedback data using supervised learning algorithms,, 3rd MEC International Conference on Big Data and Smart City (ICBDSC)
6. Moraes. R., Valiati. J. F. & GaviãO Neto. W. P. (2013), Document-level sentiment classification: An empirical comparison between SVM and ANN, Expert Systems with Applications, vol. 40, no.2, pp. 621-633.
7. Ghiassi. M., Skinner.J., & Zimbrqa. D. (2013), Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network, Expert Systems with Applications: An International Journal, vol. 40, no. 16, pp. 6266-6282

8. Anuj Sharma Shubhamoy Dey, (2012), A Document-Level Sentiment Analysis Approach Using Artificial Neural Network and Sentiment Lexicons, ACM SIGAPP Applied Computing Review December 2012.
9. N. Mohd Nawi, R.S. Ransing and M.S. Ransing, (2008), An Improved Conjugate Gradient Based Learning Algorithm for Back Propagation Neural Networks, International Journal of Information and Mathematical Sciences. Vol. 4, no.1.
10. Ozan Irsoy and Claire Cardie, (2014), Opinion Mining with Deep Recurrent Neural Networks, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp 720–728.
11. Huawang Shi, (2009), Evolving Artificial Neural Networks Using GA and Momentum, Second International Symposium on Electronic Commerce and Security, Nanchang, pp. 475-478.
12. Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh, (2011), Solving Local Minima Problem in Back Propagation Algorithm Using Adaptive Gain, Adaptive Momentum and Adaptive Learning Rate on Classification Problems, International Conference Mathematical and Computational Biology, pp. 448-455.
13. Asha Gowda Karegowda, M.A. Jayaram, A.S. Manjunath , “Feature Feature Subset Selection Problem using Wrapper Approach in Supervised Learning”, .International Journal on Computer Applications(IJCA) Volume 1 , 2010.
14. Gongde Guo, Daniel Neagu1 and Mark Cronin, (2001), A Study on Feature Selection for Toxicity Prediction EPSRC Project: PYTHIA – Predictive Toxicology Knowledge representation and Processing Tool based on a Hybrid Intelligent Systems Approach.
15. Rosenblatt, Frant, (1958), The Perceptron: A probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, vol 65, no. 6, pp. 386-408
16. Fausett, Laurene. (1994), Fundamentals of neural networks: Architectures, algorithms, and applications. New Jersey: Prentice Hall.
17. Jack V.Tu et.al, (1996), Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, Journal of Clinical Epidemiology, vol. 49, no.11, pp. 1225-1231
18. Jiang Tengjiao, Zhong Minjuan, Liao Shumei, and Luo Siwen, (2016), Mining Opinion Word (2016), Mining Opinion Word from Customer Review, International Journal of Database and Theory and Application, vol.9, no.2, pp. 129-136.
19. Elman, Jeffrey L. (1990). Finding structure in time. Cognitive Science, vol. 14, pp. 179-211.
20. Isik Yilmaz, Nazan Yalcin Erik, Oguz Kaynar,(2011), Different types of learning algorithms of artificial neural network (ANN) models for prediction of gross calorific value (GCV) of coals, Scientific Research and Essays, vol. 5, no.16, pp. 2242-2249.
21. Huang, Z., D.D. Zeng, and H. Chen.(2007), Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommender Systems. Management Science, vol.53, no.7, pp. 1146-1164.
22. Shikha Chourasia, (2013), Survey paper on improved methods of ID3 decision tree classification, International Journal of Scientific and Research Publications, vol. 3, no. 12
23. H. Josephine, S. Duraisamy, "Novel pre-processing framework to improve classification accuracy in opinion mining", *International journal of computing*, vol. 17, no. 4, pp. 234-242, 2018.