International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-6, April 2021

Intelligent Estimation of Social Media Sentimental Features using Deep Learning with Natural Language Processing Strategies

Sarojini Yarramsetti, Anvar Shathik J, Renisha P. S

Abstract*: In this digital world, experience sharing, knowledge exploration, taught posting and other related social exploitations are common to every individual as well as social media/network such as FaceBook, Twitter, etc plays a vital role in such kinds of activities. In general, many social network based sentimental feature extraction details and logics are available as well as many researchers work on that domain for last few years. But all those research specification are narrowed in the sense of building a way for estimating the opinions and sentiments with respect to the tweets and posts the user raised on the social network or any other related web interfacing medium. Many social network schemes provides an ability to the users to push the voice tweets and voice messages, so that the voice messages may contain some harmful as well as normal and important contents. In this paper, a new methodology is designed called Intensive Deep Learning based Voice Estimation Principle (IDLVEP), in which it is used to identify the voice message content and extract the features based on the Natural Language Processing (NLP) logic. The association of such Deep Learning and Natural Language Processing provides an efficient approach to build the powerful data processing model to identify the sentimental features from the social networking medium. This hybrid logic provides support for both text based and voice based tweet sentimental feature estimations. The Natural Language Processing principles assists the proposed approach of IDLVEP to extracts the voice content from the input message and provides a raw text content, based on that the deep learning principles classify the messages with respect to the estimation of harmful or normal tweets. The tweets raised by the user are initially sub-divided into two categories such as voice tweets and text tweets. The voice tweets will be taken care by the NLP principles and the text enabled tweets will be handled by means of deep learning principles, in which the voice tweets are also extracted and taken care by the deep learning principle only. The social network has two different faces such as provides support to developments as well as the same it provides a way to access that for harmful things. So, that this approach of IDLVEP identifies the harmful contents from the user tweets and remove that in an intelligent manner by using the proposed approach classification strategies. This paper concentrates on identifying the sentimental features from the user tweets and provides the harm free social network environment to the society.

Keywords: Deep Learning, Natural Language Processing, Sentimental Features, Opinion Mining.

Manuscript received on April 20, 2021.
Revised Manuscript received on April 23, 2021.
Manuscript published on April 30, 2021.

*****Correspondence Author

Sarojini Yarramsetti*, Department of Computer Science and Engineering, Srinivas University College of Engineering and Technology, Mangalore, Karnataka, India.Email: ys10676@gmail.com

Anvar Shathik J, Department of Computer Science and Engineering, Srinivas University College of Engineering and Technology, Mangalore, Karnataka, India. Email: anvarshathik@gmail.com

Renisha.P.S, Department of Computer Science and Engineering, Srinivas University College of Engineering and Technology, Mangalore, Karnataka, India. Email: renisha.ps@gmail.com

I. INTRODUCTION

Sentiments are essential to aspects of human activity and every emotion characterizes the human and the respective opinion mining is the topic of research, in which it evaluates the views and feelings of people. Sentimental feature progression shapes the fundamentals of feeling focus on moment. This could assist individuals to have a deeper awareness on the part and feelings suggested in online content. Existing work focuses mainly on the categorization of feelings, while the examination of how the perspective of a subject was inspired by certain subjects or the potential differences between subjects from the facet of sentiment was ignored. This paper is intended to build a novel approach to processes the model for sentimental complexity estimations and interrelations in social networking areas like Twitter. The proposed approach utilizes the advancements of both Natural Language Processing and the Deep Learning Strategies are used to shape statistical analysis of the user sentiments and define a new methodology to learn the associations between the opinions shared on social networking environment. The proposed framework itself uses Deep Learning procedures to characterize the feeling at a specified period, based on the mentioned issues in the past. A series of researches were conducted on a real life dataset with millions of tweets scampered from Twitter. The research proved a case method to examine the sentimental composition of various topics and the increase of opinion rich consumer behavior, including such websites such as trip advisor, blogs and micro - blogs (for example, Twitter, FaceBook and so on), has stimulated the attention of sentimental feature estimations [1][2][3]. Opinion Mining and the associated investigations has proven useful in a number of commercial advanced analytics which include production/product management, product innovations and company growth management [1][2].

Nearly every day significant quantity of instant messages including such Twitter and Weibo are compiled over social networking sites. Users reach wider their true ideas and experiences on such platforms. Hash tags, beginning with a symbol in front of words or phrases, are commonly used throughout tweets as coarse grained subjects [4] and Twitter hash tags.

For instance, the hash tag #Covid was recently a long-term trend subject matter on Twitter. The results of sentence and tweet sentiment analyses could offers extremely beneficial data[4], with most of the existing tasks focusing on feeling analysis [5][6]. Along with an instance that use NLP approaches including certain Deep convolutional Machine Learning approaches [7] to anticipate if the sentimental feature polarity estimation of the tweet is Normal, Abnormal or Moderate. The general social networking medium is influenced with comments, posts, tweets and other related taught sharing methodologies to share their opinions and perceptions to the globe. The major problem need to consider in this case is the false messages and negative content/post/tweet sharing. Some people trying to spread the false and harmful contents via social media by hiding their identity, so that many followers think like the information posted in the social network environment is correct as well as act accordingly. This kind of situation cause a severe destruction in many ways, so that a new mechanism is required to solve the issue of identifying the tweets posted in the social networking environment. Now-a-days the social networking medium allows the user to post the voice messages in their port, so that the proposed approach needs to identify the voice tweets as well. For this concern Natural Language Processing (NLP) logic is associated with the proposed approach of deep learning to provide an efficient solution in this mentioned issue. The polarity of the tweets are defined under three constrains such as Normal, Moderate and Abnormal. The following figure, Fig-1 illustrates the text estimation principles of the deep learning strategy as well as the text mining principles are explored by means of analyzing the words presented into the sentences/tweets. In which the same logic is suitable for voice assisted messages as well but the pre-processing such as data extraction from voice is required for doing such process.

The tweets in which it contains the normal and regular content means that will be considered as a normal tweet, the tweet in which it contains the negative and false content means it will be treated as an abnormal tweet and finally the tweet is not matched with both these constraints will be treated as a moderate tweet. Based on these deviations the proposed approach classification process differentiates the content and provides the results accordingly. This kind of sentimental estimation process identifies the human mindset and activities happening in social networking area as well as it will be helpful for the cyber-crime department to deal with such issues in an intelligent manner.

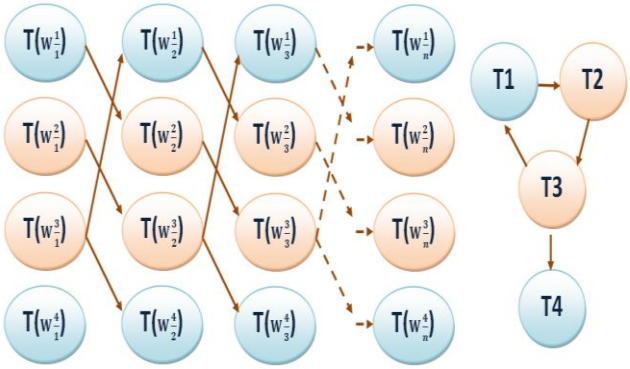

Fig.1 Text Classification Model (T) with respect to Words (W). This figure shows the directed data flow model on the corner from text ratio of T1 to T4.

The major contributions of this paper are summarized as follows:

(i) To develop a framework to provide an efficient support to the cyber-crime department to analyze the tweets in an intelligent manner.

(ii) This framework differentiates the tweets based on two categories such as voice based tweets and the text based tweets.

(iii) The voice based tweets are handled with respect to the features of Natural Language Processing and the extracted textual features are forwarded to the next process of classification.

(iv) The deep learning principles are applied to the textual features of tweet processing and sentimental estimations. As well as the voice textual features also estimated by the same logic of deep learning enabled classification.

(v) The polarity of the tweets is easily identified by means of the novel implementations and accordingly the tweets are decided to post into the social networking medium. The abnormal tweets are rejected to post into the social medium.

The rest of this paper describe regarding Related Study over section 2, further section of Section 3 illustrates the proposed system methodologies in detail and Section 4 illustrates the Results and Discussion portion of the paper and the final section, Section 5 illustrates the concept of Conclusion and Future Scope of the proposed paper. These all will be explained in detail over the further section summaries.

II. RELATED STUDY

In the year of 2020, the authors "Huizhi Liang' et al., [8]" have proposed a paper related to sentimental feature estimation based on topic analysis with respect to the dynamical Bayesian networking logic. In this paper [8], the authors described such as sentimental feature classification is among the critical priorities of understanding natural-language processing scheme and the s sentimental feature exploration shapes the composition of feeling focus on period as well as the sentimental feature exploration shapes the composition of feeling focus on particular period. This approach [8] assists many people to believe in social network and provides a way to understand the social networking nature with tweet/post raising concepts. Several literatures focus on the same sentimental features in past before [8], but all are influenced by several other dependencies such as privacy, opinion theft and so on. These all issues are resolved by using this sentimental extraction logic presented in this paper [8] as well as the in this paper, a new Gaussian Process Dynamic Bayesian Network model is introduced to analyze the sentiment features with respect to classification logic. The inter-relations between the tweets are clearly analyzed with the help of this proposed approach as well as the empirical process cross-validates around 9.72 million twitter data over this period of investigation and provides an appropriate result in outcome. In the year of 2020, the authors "Tunga Gungor' et al., [9]" have proposed a paper related to sentiment analysis over social networking medium based on aspect ratio analysis scheme.

In this paper [9], the authors described such as sentimental feature estimation findings also use repetitive or multiplicative machine learning architectures and it is mentioned in several literatures. Current generations obtain the effect and broadcast messages on feeling labels all across the lexical item through a review. On either side, recurring frameworks extricate semantic meaning first from messages and use the feeling relevant data all through instruction. There are now only some very studies which implement all these designs into some kind of deep learning model for the categorization of sentimental features. In this paper [9], a new neural network structure is proposed, in which it incorporates recurrent and multiplicative neural replication for aspect assisted text and opinion mining process. By using electoral districts and dependence compilers, initially split so every assessment into sub reviews that contain sentiment related information pertaining to the respective aspect conditions. Upon accumulating and mentoring the recidivist neural trees made from sub-views, we add everyone outcome to the repetitive model. This approach [9] is estimated to quartet on two data sets of various styles in native language and have achieved cutting-edge results and have exceeded the preliminary research with a considerable margin in each of these areas.

In the year of 2020, the authors "Manuel.J.Cobo' et al., [10]" have proposed a paper related to mining the user opinions with respect to machine learning concepts based on emotions recognition principles. In this paper [10], the authors described such as due to recent advances in social network models, large incentive to invest in machine learning and a high rise in consumer expertise, the adverts industry has gone through a quantum leap over the last decade. In this context, text mining, sentiment analysis and the understanding of emotions eventually lead to these most sought-after advertising goals: offering relevant ads on a scale. In recent years the significance of analyses on recommender systems, sentiment classification and moral reasoning has increased exponentially. The highlight of this new situation was the research community's interest in studying the relation between certain technologies and the diffusion of intelligent publicity and context. This article examines the relationships among sentiment classification, opinion mining and the understanding of emotions in publicity. The prime purpose is to confirm the present incarnation of all these research, to explore questions, strategies, conclusions, themes and gaps and to determine their importance in the existing collaboration promotional process. In order to achieve those objectives, 919 research papers years between 2010 and 2019 have been analyzed, retrieved and investigated based on results from the web of science.

III. METHODOLOGIES

In this paper, a novel deep learning methodology is introduced called Intensive Deep Learning based Voice Estimation Principle (IDLVEP), in which it is used to solve the issues presented into the social network taught sharing principles as well as the proposed approach of IDLVEP provides an efficient voice tweet sentimental analysis feature by using the Natural Language Processing (NLP) procedures. This paper combines both the logic of deep learning strategy and the Natural Language Processing to produce a new hybrid algorithm called IDLVEP, in which it extracts the content from the social network such as twitter, FaceBook and so on. The extracted contents are processed based on the two deviations such as text based tweet estimation and the voice based tweet estimation. The text based contents are extracted from the user tweet and apply the pre-processing logic to the data to eliminate the Stopwords in the content and process the content with respect to deep learning principles. The machine is already trained by using the proposed approach based on the dataset gathered by the social network such as Twitter. The dataset contains the summary of Positive and Negative tweet contents with the proper polarity ratio, in which the new training model is created, based on this polarity and maintain that into the server for cross-validating the test cases occurred on real world social networking environment. The same scenario is valid for voice assisted tweets, in which the pre-processing is different in case of voice messages that extracts the voice content based on Natural Language Processing logic from the user tweets and provide the extracted textual data to the deep learning principles for further process. The following table, Table-1 shows the dataset samples with proper polarity values.

Table-1: Dataset Attributes with Polarity and Weight

Attribute	Polarity	Weight
Acclaim	Normal	52%
Acclaimed	Normal	36%
Acclamation	Normal	83%
Accolade	Normal	92%
Accolades	Normal	72%
Abnormal	Abnormal	95%
Abolish	Abnormal	67%
Abominable	Abnormal	90%
Abominably	Abnormal	81%
Worthy	Normal	94%
Wow	Normal	86%
Wowed	Normal	69%
Wrought	Abnormal	79%
Yawn	Abnormal	58%
Zap	Abnormal	37%
Zapped	Abnormal	64%
Wowing	Normal	96%
Wows	Normal	96%
Zaps	Abnormal	37%
Wrong	Abnormal	97%
Yay	Normal	61%
Villainously	Abnormal	72%
Youthful	Normal	23%
Vilify	Abnormal	29%
Zeal	Normal	41%
Vindictive	Abnormal	70%

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

The following equation is used to estimate the weight factors of the tweets, in which the weight ratio is normalized in the form of Wm, 1 → 1/M. In this scenario M is the text length constant that will be defined by means of 1 to n.

$$W(1 \ldots M) = \frac{T(Wi(1-n) \pm \sqrt{x^2 - M}}{T \sqrt[2]{n}} \quad (1)$$

Where W indicates the weight factor, T indicates the text range with the words W_i(1-to-n) and x^2 indicates the time factor for estimating the weight with respect to text length M. Based on these weight metric the polarity ratio is deviated and the tweets are classified accordingly based on such logic. The following equation is used to estimating the scaling level of the tweets and the polarity ratio of the tweet participates in high impact with word density (D) of the text phrase (T).

$$T(D) \leftarrow \sum \left(1 + \frac{W(1)}{\sqrt[1]{n}} + \frac{W(2)}{\sqrt[2]{n}} + \frac{W(3)}{\sqrt[3]{n}} + \cdots + \frac{W(n)}{\sqrt[n]{n}} \right) \quad (2)$$

Where the W(1) to W(n) indicates the estimated weight factor, acquired from equation (1) and the T(D) indicates the text density estimation object.

The following algorithm illustrates the logic of proposed approach Intensive Deep Learning based Voice Estimation Principle (IDLVEP), in which the process flow of the algorithm is illustrated in clear manner with proper Pseudocode.

Algorithm: Intensive Deep Learning based Voice Estimation Principle

Input: *Time Series, Input Data*

Output: *Polarity Estimated Value with estimation accuracy*

1. Gather the input from the user end via social network medium.

2. Analyze the received input is analog or digital.

3. The analog form of input records is considered as the voice message and the digital form of input is considered as a regular text based tweet.

4. The analog tweets are converted into text phrases by using Google Voice Conversion principles.

5. This API is powered by Google Voice to Text Conversion logic.

6. The converted text is cross-validated with Natural Language Processing principles to analyze the language pattern.

7. The same logic is repeated for total phrase presented into the voice message.

Pseudocode:

```
do{
incr i;
find T(M);
i=T(M).length;
Categorize Words from 1 to n;
var x=Word(i).pattern;
x=text.Convert(x);
}while(i<=0);
```

8. Accumulate all the textual features from 1 to n-1.

9. Obtain the trained dataset model into the procedure.

10. Cross validate the accumulated textual characteristics to the model trained.

11. Identify the polarity of each words specified into T(M) based on the tweet classification logic.

12. Based on the overall weight ratio W(i), the final polarity of the tweet is defined.

13. Return the final polarity value for accuracy estimation.

Pseudocode:

```
return Pol(T(M));
```

14. Estimate the accuracy of the polarity returned.

Pseudocode:

```
var str, acc;
str=Pol(T(M));
acc=Calc.Accracy(str);
return acc;
```

IV. RESULTS AND DISCUSSION

This paper provides a solution to identify the sentimental features based on the tweets generated by the users on social medium. Throughout this section, a concept of how to perform real-time data sets and the associated experimentations to identify the sentimental features from the user tweets will be discussed in clear manner. Input from the approach proposed is the feeling time series data for subjects, during which the process begins with the collection of hash tags and tweets from the social network medium over a given period of time. Then detect each hash tag's sentiment dynamics using existing classification approaches and discuss the construction of the proposed approach on the basis of the time series of input feelings and the analysis of feelings results of the constructed sentiment development procedures. This paper associates the two powerful approaches in hybrid manner such as Deep Learning and the Natural Language Processing (NLP) to provide high efficient results in prediction with proper polarity ratio. The proposed hybrid logic is named as Intensive Deep Learning based Voice Estimation Principle (IDLVEP), in which it assures the resulting performance based on model creation, accuracy on predictions, reduced error rates and the time efficiency. The following figure, Fig-2 illustrates the proposed approach training efficiency with respect to specific time period, in which the x-axis shows the time period and the y-axis shows the number of data processed on that specific time period.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

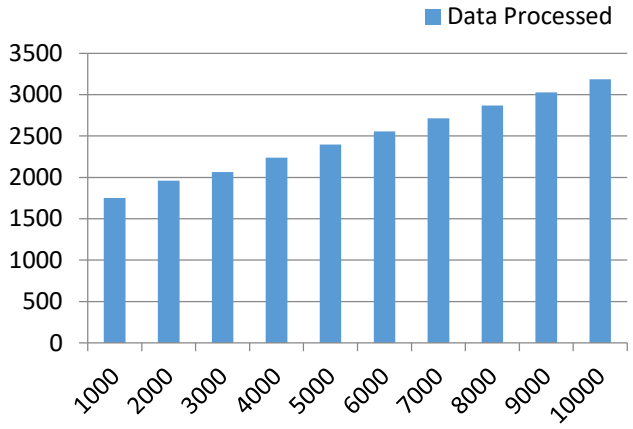

Fig.2 Data Processing Efficiency w.r.t Time

The following figure, Fig-3 illustrates the error estimation scenario of the proposed approach IDLVEP, in which the data training freuquency is highly dependednt on the error frequyency because in the testing phase the chances of error occurrence is low. The training phase error estimation results are visualized in grpahical manner with the following figure in clear manner. This figure shows the outcome of training model data processing frequency with respect to error rates, in which the x-axis portrays the nmber of data processing while train the system by using proposed deep learning principle as well as the y-axis indicates the error ratio.

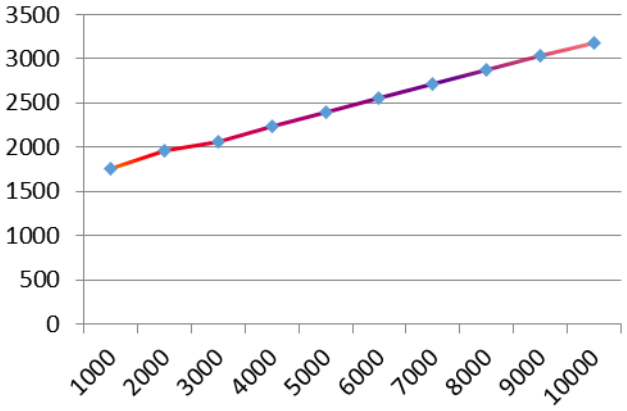

Fig.3 Error Ratio Estimation

The following figure, Fig-4 illustrates the tweet categorization efficiency, in which the overall sentimental features and the associated data is acqired form the social networking medium. The proposed approach of Intensive Deep Learning based Voice Estimantion Principle identifies the text related tweets and the voice related tweets in an intelligent manner by means of text categorization logic, in which it deviates the analog and digital data in separate manner.

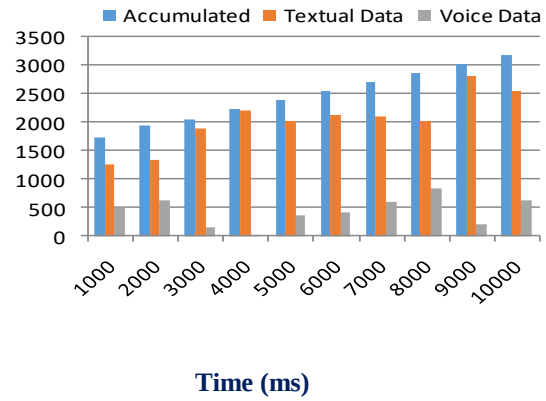

Fig.4 Data Categorization Efficiency

The following figure, Fig-5 illustrates the proposed approach confusion matrix diagram, in which the confusion matric is plotted based on the normal and abnormal data ranges with proper polarity specifications. The graph plot is based on the weight ratio of the polarity specified into the table, Table-1.

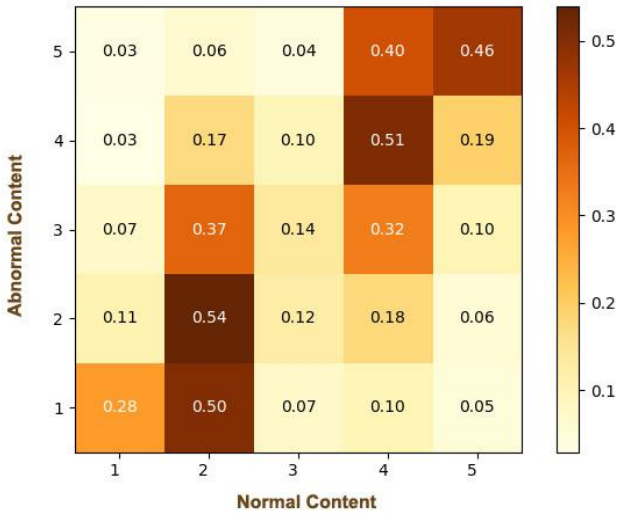

Fig.5 Conusion Matrix

V. CONCLUSION AND FUTURE SCOPE

In this paper, a new social network analysis mechanism is designed in order to identify the sentimental features from the tweets raised on the social network with respect to identify the negative and false news spread over the medium. This kind of analysis is illustrated on several classical approaches but all are strucked under certain issues such as time efficiency and error ratio. But in this paper, a new hybrid logic is introduced called Intensive Deep Learning based Voice Estimation Principle (IDLVEP), in which it is derived by using two powerful technologies called Deep Learning and Natural Language Processing. These two features are combined together to provide an efficient data processing logic as well as the proposed mechanism is suitable for identifying the text based tweets and voice based tweets in an efficient manner.

Retrieval Number: 100.1/ijitee.F88250410621
DOI: 10.35940/ijitee.F8825.0410621
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

The resulting figure, Fig-2 shows the training process efficiency of the proposed approach with respect to time constraint. The same data are acquired from the figure, Fig-2 and define an error ratio estimation, in which it is portrayed clearly in figure, Fig-3. The proposed approach data categorization accuracy is portrayed in clear manner over the resulting figure, Fig-5 with respect to time complexities. For all the proposed logic is suitable for analyzing the user tweets and provide a flaw free social networking environment to the users.

In future, the proposed work can further be extended by means of adding some cryptographic principles to the data maintained into the server for processing. These kind of cryptographic principles are really useful in terms of two constrains such as the crypto contents maintained into the server end indicates the safety measures and at the same time it ensures the size reduction in server end.

REFERENCES

1. B. Liu, Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). San Rafael, CA, USA: Morgan & Claypool, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
2. B. Pang and L. Lee, ``Opinion mining and sentiment analysis, Found. Trends Inf. Retr., vol. 2, nos. 12, pp. 1135, 2008, doi: 10.1561/1500000011.
3. Y. Zhao, B. Qin, T. Liu, and D. Tang, ``Social sentiment sensor: Avisualization system for topic detection and topic sentiment analysis on microblog," Multimedia Tools Appl., vol. 75, no. 15, pp. 88438860, Aug. 2014.
4. X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, ``Topic sentiment analysis in Twitter: A graph-based hashtag sentiment classication approach, in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), New York, NY, USA, 2011, pp. 10311040, doi: 10.1145/2063576.2063726.
5. K. Macropol, P. Bogdanov, A. K. Singh, L. Petzold, and X. Yan, I act, therefore I judge: Network sentiment dynamics based on user activity change," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), 2013, pp. 396402.
6. U. Kursuncu, M. Gaur, U. Lokala, K. Thirunarayan, A. Sheth, and I. B. Arpinar, ``Predictive analysis on Twitter: Techniques and applications, in Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. Cham, Switzerland: Springer, 2019, pp. 67104, doi: 10.1007/978-3-319-94105-9_4.
7. Y. Kim, ``Convolutional neural networks for sentence classication, in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2014, pp. 17461751.
8. Huizhi Liang, Umarani Ganeshbabu and Thomas Thorne, "A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution", IEEE Access, 2020.
9. Cem Rifki Aydin and Tunga Güngör, "Combination of Recursive and Recurrent Neural Networks for Aspect-Based Sentiment Analysis Using Inter-Aspect Relations", IEEE Access, 2020.
10. Pablo Sánchez-Núñez, Manuel J. Cobo, Carlos De Las Heras-Pedrosa, et al., "Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis", IEEE Access, 2020

AUTHORS PROFILE

Dr. Sarojini Yarramsetti, working Professor in the Department of Computer Science Engineering, Srinivas University College of Engineering and Technology, Mangalore, India. Dr. Sarojini received her Bachelor degree in Information Technology from Bannari Amman Institute of Technology, Sathyamangalam. She did her M.E in Computer Science Engg from M.I.E.T Engg. College, Tiruchirappalli. She did her Ph.D. in Computer Science Engineering from Anna University, Chennai. Dr. Sarojini's research area of interest includes Software Engineering. She has published research papers in peer reviewed national and international journals of repute.

Mr. Anvar Shathik heads the Department of Cloud Technology & Data Science (Senior Faculty, iNurture Education Solutions, Bangalore), Srinivas University College of Engineering and Technology, Mangalore, India. He is having 15 years of teaching experience in various Engineering colleges. Anvar Shathik received her Bachelor degree in Computer Science and Engineering from Vels Srinivasa College of Engineering & Technology, Anna University, Chennai, India. He pursed his M.E in Computer Science Engineering from Madha Engineering College Anna University, Chennai, India. He pursuing Ph.D. in Natural Language Processing using Deep Learning in Srinivas University, Mangalore, and Karnataka. His research area of interest includes Machine Learning, Deep Learning and Data Science. He published more than 10 research papers in a reputed journal.

Renisha.P.S is an Assistant Professor of Cloud Technology and Data Science department at Srinivas University with 9years of experience. She received his Bachelor Degree in IT from National College of Engineering, Tirunelveli and Master degree in Network Engineering from Anna University, Coimbatore. She has presented papers at conferences and published papers in various journals. Her publication interest include Natural Language processing and Deep Learning. She is a member of I2OR ED-Tech community and International Association of Engineers.