

Ensemble Approach for Twitter Sentiment Analysis

Dimple Tiwari

Ambedkar Institute of Advanced Communication Technologies and Research Govt. of NCT of Delhi, India
E-mail: dimple.tiwari88@gmail.com

Nanhay Singh

Ambedkar Institute of Advanced Communication Technologies and Research Govt. of NCT of Delhi, India
E-mail: nsingh1973@gmail.com

Received: 30 March 2019; Accepted: 07 June 2019; Published: 08 August 2019

Abstract—Due to enlargement of social network and online marketing websites. The Blogs and reviews of the user are acquired from these websites. And these become useful for analysis and Decision making for various types of products, marketing and movie etc. with the extent of the usefulness of social Reviews. It is to be needed carefully analysis of that data. There are various techniques and methods are available that can accurately analyses the social information and provides greater accuracy for the analysis. But one of the major issues available with the social media data is that data is unstructured and noisy. It is to be required to solve this problem. So here in this paper a framework is proposed that includes latest data preprocessing techniques instead of noise removal like stemming, Lemmatization and Tokenization. After Pre-Processing of data ensemble methods is applied that increase the accuracy of previous classification algorithms. This method is inherent from bagging concept. First apply Decision Tree, KNeighbor and Naive Bayes classifier that not provide batter accuracy after that boosting concept is applied with the help of AdaBoost method that improves the accuracy of previous classical classifiers. At last our proposed ensemble method ExtraTree classifier is applied that inherent from bagging concept. Here we use the Extra Tree classifier that take the various sample are taken from training set and various random trees are created. It is also called as extremely randomized tree that provides extreme refined view. So that, it is to be conveying that The ExtraTree classifier of bagging ensemble method outperforms than all other techniques that are previously applied in this paper. with using some novel pre-processing techniques data that produced is more refined and that provides clean and pure base for the implementation of ensemble techniques. And also contributes in improving the accuracy of the applied methods.

Index Terms—Stemming, Lemmatization, Tokenization, Naive Bayes, Decision Tree, KNeighbors, AdaBoost, ExtraTree classifier and sentiment analysis.

I. INTRODUCTION

In the modern era each and every user and discoverer use internet or various types of social sites and mass media to search latest topics of nation and world. News information is an extensive outlook of the social network massive information on the website. On the daily basis more than the thousands of people share new information regarding the statistics. Each day exclusive data is added on the internet that creates new records of information. Browsing and talk through the events and news become standard of standard routine lifestyle. That's why high-priced or you can say important information of government and corporation can found easily on the internet. The analyzing sentiment from the information that is provided by news is one of the major responsibilities.

For accessing meaningful information from people review, lots of machine learning algorithms are applied by the scholars. Sentiment analysis can be fetch on the basis of three different levels. Such as Aspect level, sentence level and document level [1]. Aspect level technique found sentiment on the basis of all expression of sentiment shows within a document and aspect by which it related. Sentence level shows weather the sentence expresses Positive, Negative or Neutral review. And the final one Document level presents the sentiment on the basis of whole document classification either Positive, Negative or Neutral. In the literary texts, most of the approaches and models of aspect-based analysis depends on a unique classifier that is to be trained on annotated data for recognizing polarity of an opinion on the basis of positive, negative or Neutral direction [2]. In the Document level sentiment analysis sentiment is analyses on the opinion of author that expressed in the document. It is the simplest way to found sentiment [3]. A sentence level is a deep form of sentiment analysis that aimed to obtain more detailed view of the each and every opinion present in the sentence of document [4]. There are major two types of machine learning techniques

available for sentiment analysis one is: Supervised and second is: Un-supervised. In the Supervised technique data is labeled and trained to get appropriate output which help to take correct and efficient decision making [5]. But in the case of unsupervised technique there is no need of labeled data. For solving the problem of unlabeled data clustering technique is used [6].

Here in this proposed Framework supervised approach is used. Various classification algorithms are used for sentiment analysis. But from that we apply only three best algorithms are Decision Tree, KNeighbor and Naive Bayes. Instead of it two ensemble methods AdaBoost and ExtraTree classifier are also used. That maximizes the accuracy with the help of Bagging and Boosting. Some Pre-Processing task like Tokenization, Lemmatization and stemming is used. From that Tokenization process is used for dividing textual information into a single word. For implementing this concept various open source tools are available [7]. Stemming algorithms are used for prerequisite task of any text mining application for searching the root words from various semantically same kinds of morphed words. In the lemmatization lemmas are used for improving the text similarity matrix.

The major contribution of this work is stated below:

- Dataset is collected from the Kaggle.com. From there US-Airline data is selected for performing the sentiment analysis. This data contains around 14,604 tweets.
- In the Pre-Processing of data with the noise removal three advance concepts Stemming, Lemmatization and Tokenization is added for providing the better accuracy.
- Three classification algorithms decision Tree, Kneighbor and Naive Bayes are applied on the tweets. After that boosting concept AdaBoost is applied.
- Finally, with Boosting concept our novel approach ExtraTree classifier is applied. And it is to be sated that ExtraTree ensemble method outperforms then others.

The structure of this paper is described as: section 2, presents literature review. Section 3, indicates the detailed view of applied algorithms. Section 4, proposed Framework is explained with results. Section 5, describes the comparison between all applied algorithms 6, and shows the conclusion and future work.

II. RELATED WORK

The development of social network especially twitters is growing very fast now days. Twitter is mostly used to comment on a person, product or movie. Twitter is mostly used to comment on a person, product or movie. By using twitter, it is to be very easy to provide user sentiments on any topic. That can be analysis by various machine learning techniques or by combination of lexicon approach. That provides very much good accuracy [8]. Here the new quality model is presented as

‘Sentiment Analysis as a service’ (SAaas) to evaluate multihull social information services. Spatio-temporal properties of social users are focused to identify the locations of disease outbreaks [9]. Online social network users are to be considered as social sensors that provides entrusting and required information. ‘Select Additive Learning’ (SAL) procedure can improves the trained neural network generalization for multimodal sentiment analysis. It also shows SAL approach outperforms in all three (verbal, visual and acoustic) modeling with their fusion [10].

Both Feature and decision-level fusion methods work effectively together as information sources. That provides effective multiple modalities. YouTube dataset achieve an accuracy of around 80% that outperforms state-of-the-art system more than 20% [11]. Images are also used for presenting the sentiment analysis. ‘Temporal Convolutional neural Network’ (CNN) is presented in that each pair of the images at particular time t and $t+1$ are combined into single image. In the deep studies of CNN, every hidden layer is convolving a weighted matrix with the activation of matrix at the layer below that is trained [12]. Twenty-one methods for sentence-level sentiment analysis are available for English language that compare with two language specific methods. Based on nine specific-language datasets, it shows multilingual sentence-level approach is also supported [13]. Naive Bayes and Levantine algorithm are used to evaluate the emotion into various categories from given social network sites. This method provides excellent performance for the news or Real time data that available on social sites [14]. Unsupervised Neural language model is used for providing the training to initial word embeddings that further tuned method of deep-learning model on the basis of different supervised corpus. Here letters are trained in supervised training. Pre-trained parameters are used to start the model that outperforms for sentiment analysis [15].

Two phases approach consisting online and offline processes can used for sentiment analysis. In offline technique, datasets of tweets are pre-processed for extracting useful information that will be used as train classifier. After providing training to classification model it stores in secondary storage for loading and using it in online phase [16]. Natural language processing and machine learning techniques are gaining inside in the field of sentiment analysis. Two-step method is used for sentiment classification. First lexicon method is used for scoring positive, negative and neutral polarity. Second tweets with low polarity strength are passed by support vector machine (SVM) classifier and this two-step method outperforms [17]. Netflix and Stanford model is widely cited model of sentiment analysis. Human trafficking in web data is identified by ensemble sentiment analysis method. DARPA and MEMEX are two human trafficking related web pages are available on that binary sentiment like positive/negative and categories or multi-class sentiment like love, neutral etc. are concluded [18]. Twitter sentiment analysis is tricky and handy according to another data source of sentiment

analysis due to misspelling repeated and slang words. It is to be known that minimum length of tweet is 140 characters. That is to be requiring finding correct sentiment from each word. For this purpose, highly accurate model is presented consisting feature vector and classifier such as SVM and Naive Bayes [19].

Natural language processing especially semantics and word sense disambiguation increase the accuracy of classification. This ensemble classifier combines the effect of various individual classifiers. That outperforms than previous classifier by 3-5%. Combined approach of rule-based classifier and supervised learning can use for sentiment analysis. And the SVM is trained on dependency and the feature of sentiment lexicon [21]. Phonetic-Based framework is presented for normalizing micro text to plain text in English. Hence it improves the classification accuracy by >4% in terms of detecting polarity after the normalization process [22].

III. DETAILED STRUCTURE OF INCLUDED CLASSIFICATION ALGORITHMS

This section is presented various techniques in detail that are used in this paper for sentiment analysis. Here five algorithms are used for showing the results that provides different accuracy. KNeighbors, Decision Tree, Naive Bayes are classical classifiers and AdaBoost, ExtraTree classifiers are the ensemble methods.

A. KNeighbors Algorithm

The Neighbors based classification algorithm is a kind of instance based or non-generalized learning. General internal model is not constructed in this technique. It is just computed by simple majority vote of the neighbors that are classes to each other. Nearest neighbors of the point are assigned as query point.

The K-Neighbors classification is one of the types of Neighbors classification. It is major used technique. The value K's best choice is highly data dependent. Generally, it is to be saying that K shows the noise effect, but makes the less distinct the boundaries of classification. KNeighbors classifier is implements on the basis of learning K closest Neighbors. Here K represents the integer value that is declared by the users.

B. Decision Tree Algorithm

This classifier is worked on the basis of dividing the working area repetitively into multiple sub parts. Decision Tree can possibly produce important insight into interactions between variables [23]. Because, there are two different regions of same class. Class is either divided on the basis of purity in classification or impurity in classification. Clear separation of work classes is done easily. But when there is impurity in classes. It is to be solved by the concept of Entropy. Entropy works on the basis degree of how particular element is occurred randomly. Mathematically it is calculated using probability of element as:

$$H = -\sum p(x) \log p(x) \quad (1)$$

P(x) is a probability of element x.

Secondly Decision Tree continues the division of classes on the basis of Information Gain. This term stated that select the division that has less impurity. Information Gain is based on decreases the entropy after the dataset is split. For constructing Decision Tree, it is to be requiring finding the attribute that return highest Information Gain [23].

$$GainRatio = \frac{\Delta \text{inf } o}{\text{split inf } o} \quad (2)$$

It is a fact that Largest Information Gain = Smallest Entropy.

C. Naive Bayes

This classifier is work by grouping the various classification algorithms that are based on Bayes Theorem. Naive Bayes is a probabilistic classifier that calculates group of probabilities [24]. It is not a single algorithm but a group of various algorithms that share common principle. It works on the basis of strong assumption. Major advantage of this algorithm is that it just needs very small training data to finds the parameters. In spite of its strong assumption and simplicity it provides better results in many domains. Bayes theorem implements on the basis of formula as follows:

$$P(C | X) = P(X | C).P(X) / P(C) \quad (3)$$

Where X denotes point and C denotes class. After that by doing the assumption for that data point $X = \{X_1, X_2, \dots, X_n\}$ the probability of every attribute that arise in this class is independent. The probability of X is as follows:

$$P(C / X) = P(C). \prod P(X; C) \quad (4)$$

D. AdaBoost Algorithm:

It was the first actual successful boosting algorithm that is developed for binary classification. This algorithm supports the Boosting concept that is ensemble method to create strong classifier from many weak classifiers. AdaBoost is a great method that enhances the performance of Decision Tree on binary classification problem. AdaBoost can increase the performance of any machine learning algorithm. It is best used with weak learners. This classifier combines weak classifier algorithm to make a strong classifier. An object can poorly classify by single algorithm. But by combining multiple classifiers with training set selection at every iteration and provides occurrence weight in final voting. It is to be score better accuracy. Here are two concepts are arising one is how to select training subset that done by every weak classifier is trained using a random subset

from overall training group. Second is how to assign a weight. The classifier that has 50% accuracy gives zero weight. And negative weight is given to classifiers that have less than 50% accuracy.

$$H(X) = \text{SIGN}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (5)$$

$h_t(x)$ is the output of weak classifier t for input x
 α_t is weight assigned to classifier.
 α_t is calculated as follows:
 $\alpha_t = 0.5 * \ln\left(\frac{1-E}{E}\right)$; weight of classifier is straight forward, it is based on the error rate E .
 AdaBoost algorithm provides great score because it is dependent on many weak classifiers for final decision.

IV. PROPOSED FRAMEWORK

This partition is presented the proposed framework and its working. Various new perspectives are added in this framework for providing new scheme of sentiment analysis using machine learning algorithms. Various pre-processing techniques are added in this framework that is helpful for boosting the accuracy of machine learning algorithms. This proposed framework is presented below:

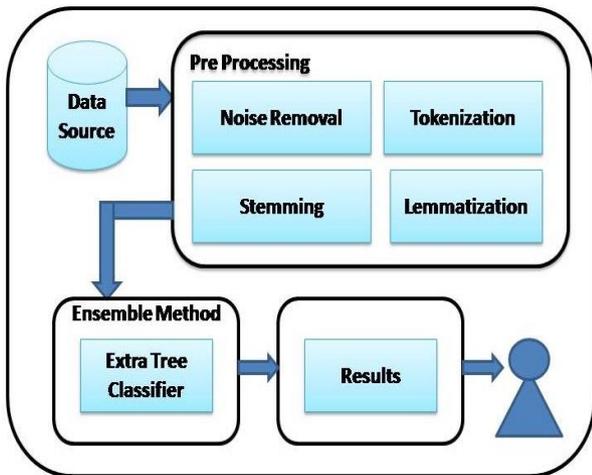


Fig.1. Proposed Framework of Sentiment Analysis

A. Data Collection

Data is collected from the “Kaggle.com”. From here US Airline data is extracted that included data of five regions (American, Delta, Southwest united, USAirways, Vergin America). This data is classified on the basis of three polarities (Positive, Negative and Neutral). Positive Sentiment is denoted by 1, negative sentiment is denoted by 0 and neutral sentiment is denoted by 2. Total number of negative tweets is 9187, positive are 3099 and neutral are 2363. So, it is to be calculated that total tweets related to US Airline is 14640. Divisions of tweets are given below table as follows:

Table 1. Tweets Count

Polarity	Count
Negative	9178
Positive	2363
Neutral	3099

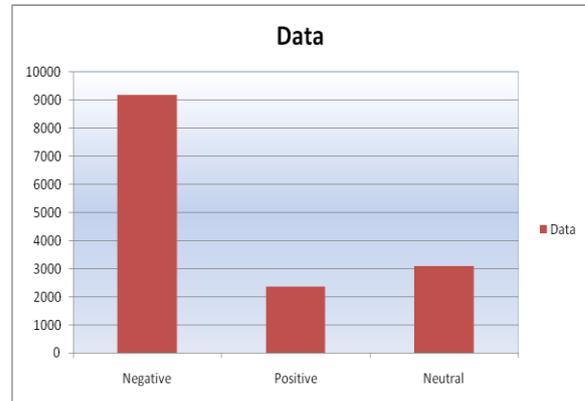


Fig.2. Tweets Count Graph

This bar chart represents the tweets division on the basis of polarity and their total count. 0, 1, 2 are stated as Negative, Positive and Neutral polarity. Which are presented by different colors. Here we also presented some sample data and their information.

Table 2. Data before preprocessing

No	Id	Text	Sentiment
0	5.703010e+17	@VirginAmerica Plus you'v....	1
1	5.703010e+17	@VirginAmerica yes, Nearly.....	1
2	5.703000e+17	@virginAmerica well, I didn't...	1
3	5.702950e+17	@VirginAmerica it was amazi..	1
4	5.702000e+17	@VirginAmerica I & It;3 prett...	1

There is GetInfo () method is used for getting the information related to data. This information presents that there are three columns in the data first is Id that data type is float, second is Text that is object type and third column is sentiment that holds integer type of data. Total memory usage is 343.2 kb.

Table 3. Data Information

Type	Information
Class	Pandas.core.frame.DataFrame
Range Index	14660 entries, 0 to 14639
Data Columns	Total 3 columns
Id	14640 non-null float64
Text	14640 non-null object
Sentiment	14640 non-null int64
DTypes	Float64(1), int64(1), object(1)
Memory Usage	343.2+KB

B. Data Pre-Processing

In this step data is pre-processed before applying the machine learning algorithms. Before applying advance processing, Noise is removed from data like remove white spaces, remove punctuations, Remove StopWords, remove URLs, remove numbers and convert to lower. It is a basic step that has to perform on tweets for calculating a sentiment. After this classical work three new concepts are added that are very useful for improving the accuracy of machine learning algorithms on text data. These concepts are added in proposed work their detailed study is given below as follows:

Stemming: This algorithm is applied as pre-processing task for any text classification application for extracting the root words or stemmed words that are semantically similar. A word that belongs to single root carries same semantic meaning in a particular text. Basically, this technique is helpful for retrieving basic information from any un-structured text data. Various algorithms are used for stemming are Lovin's stemmer, Porter's stemmer, Paice stemmer, N-Gram stemming and successor variety stemming [25]. These algorithms work on the basis of affixes to suffixes and prefixes. Stemming is very much helpful for the sentiment analysis. It also retrieves and search many more forms of words that gives better results.

Lemmatization: This works is same as like stemming. But the processing of Lemmetizer is different from stemming. It works on the basis of morphological analysis of words. For this purpose, it contains detailed dictionaries from which algorithm search link from back to its lemma. Deep learning is very much required for creating the deep dictionaries that provide searching to algorithms for proper form of the word. After the completion of Lemmatization process noise is reduced and results are produced more accurate. Finding a correct lemma for particular word in context is a typical and complicated task in fact that word can inflected form or more than one lexeme and each have different lemmas [26].

Tokenization: This is a process of replacement of sensitive data with unique symbols that are identifiable. This returns all the necessary information about the text without breaking its security. In the pre-processing it breaks up a sequence of strings into various parts like words, phrases or keywords and those elements are called tokens. These tokens become an input for different processes. Usually tokenization is work as trivial task for lots of languages, mainly for those that use separation between words [27].

After the completion of all pre-processing task data is ready for the implementation of machine learning algorithms. In this proposed work all the algorithms of machine learning that applied are belongs to supervised learning. So, for implementing supervised algorithms data is first trained and after the training of data actual

testing is done. For this data is split into two parts one is training data and second is testing data their division ratio is shown in below table is as follows:

Table 4. Data Partition

Division	Amount
Training	9808
Testing	4832
Total	14630

C. Ensemble Method

Bagging is a concept that is very much helpful for increasing the accuracy of algorithms. Bagging is an ensemble meta-algorithm of machine learning that provide the accuracy and stability to the previous classical algorithms. Here novel method in sentiment analysis Extra Tree Classifier stands for (extremely randomized trees) is applied. This is different from classical Decision Tree in the way they built. Its main objective is to further randomizing a tree building for numerical input feature. For splitting the nodes random split is used for each of the Max_Feature randomly selected feature and split the best from the chosen. This class implements a Meta estimator that fits various randomized Decision Trees on different sub-samples of dataset and uses averaging from improving the predictive accuracy. This method is more productive in the way of characterizing by a large number of numerical features differ more or less continuously. It provides smoothness and also reduces computational burdens that are linked to standard trees. This algorithm belongs to bagging concept. That provides the stability to the accuracy that is calculated.

D. Result

ExtraTree Ensemble method provides 76% accuracy that is much greater than previous classification algorithms accuracy. After the calculation of all algorithms proposed work is to be stated that Extra Tree Ensemble Method outperforms than all of the algorithms that implemented in this work. This Novel Method works very efficiently on the text data for sentiment analysis.

V. COMPARISON

In this paper work we apply four algorithms before implementing our novel approach. Three algorithms Naive Bayes, KNeighbor and Decision Tree are classical algorithms and one AdaBoost is Boosting concept. Our comparison shows on five algorithms that are applied on US-Airlines data. All implemented algorithms provide different scores. Ensemble methods provide more accuracy than classical algorithms as (Decision Tree 63%, KNeighbors 67%, Naive Bayes 69%, and AdaBoost 74%). And finally, our novel method outperforms than all of others and provides 76% accuracy.

Table 5. Algorithms Accuracy

S-No	Algorithm	Working	Accuracy
1.	Decision Tree	This classifier is worked on the basis of dividing the working area repetitively into Multiple sub parts.	63%
2.	KNeighbors	It is just computed by simple majority vote Of the neighbors that are classes to each other.	67%
3.	Naive Bayes	Bayes is a probabilistic classifier that calculate Group of probabilities.	69%
4.	AdaBoost	This algorithm supports the Boosting conceptthat is ensemble method to create strong Classifier from many weak classifiers.	74%
5.	ExtraTree	ExtraTree further randomizing a tree Building for numerical input feature.	76%

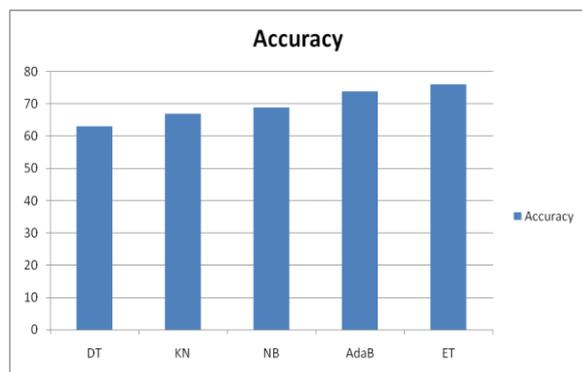


Fig. 3. Comparison Results

VI. CONCLUSION AND FUTURE WORK

This work presents ensemble method for twitter sentiment analysis. That includes set of relevant features for sentiment analysis. It includes some set of advance features for pre-process of text data and also applied some ensemble method for sentiment analysis that support Bagging and Boosting Concept. Vast information related to US-Airline is generated as tweets on twitter. For analyzing a sentiment of US-Airline three classical classification algorithms are applied but that are not perform efficiently and provides low accuracy. So, for improving the accuracy of these algorithms two ensemble methods AdaBoost and Extra Tree are applied. Finally, it is to be concluded that Extra Tree classifier outperforms than all of the others applied algorithms of this work.

In future Deep Learning concept will apply on the social media data for sentiment analysis instead of machine learning. Deep learning provides more accurate results on huge data rather machine learning. It is more powerful and flexible approach.

REFERENCES

- [1] Ortigosa A, Mart  JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. *Computers in human behavior*. 2014 Feb 1;31:527-41.
- [2] Perikos I, Hatzilygeroudis I. Aspect based sentiment analysis in social media with classifier ensembles. In2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) 2017 May 24 (pp. 273-278). IEEE.
- [3] Ghiassi M, Skinner J, Zimbra D. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*. 2013 Nov 15;40(16):6266-82.
- [4] Zainuddin N, Selamat A, Ibrahim R. Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Applied Intelligence*. 2018 May 1:1-5.
- [5] Gautam G, Yadav D. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In2014 Seventh International Conference on Contemporary Computing (IC3) 2014 Aug 7 (pp. 437-442). IEEE.
- [6] Hastie T, Tibshirani R, Friedman J. Additive models, trees, and related methods. In*The elements of statistical learning 2009* (pp. 295-336). Springer, New York, NY.
- [7] Vijayarani S, Janani MR. Text mining: open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACIJ)*. 2016 Jan;3(1):37-47.
- [8] Sabariah MK, Effendy V. Sentiment analysis on Twitter using the combination of lexicon-based and support vector machine for assessing the performance of a television program. In2015 3rd International Conference on Information and Communication Technology (ICoICT) 2015 May 27 (pp. 386-390). IEEE.
- [9] Ali K, Dong H, Bouguettaya A, Erradi A, Hadjidj R. Sentiment analysis as a service: a social media based sentiment analysis framework. In2017 IEEE International Conference on Web Services (ICWS) 2017 Jun 25 (pp. 660-667). IEEE.
- [10] Wang H, Meghawat A, Morency LP, Xing EP. Select-additive learning: Improving generalization in multimodal sentiment analysis. In2017 IEEE International Conference on Multimedia and Expo (ICME) 2017 Jul 10 (pp. 949-954). IEEE.
- [11] Poria S, Cambria E, Howard N, Huang GB, Hussain A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 2016 Jan 22;174:50-9.
- [12] Poria S, Chaturvedi I, Cambria E, Hussain A. Convolutional MKL based multimodal emotion recognition and sentiment analysis. In2016 IEEE 16th international conference on data mining (ICDM) 2016 Dec 12 (pp. 439-448). IEEE.
- [13] Araujo M, Reis J, Pereira A, Benevenuto F. An evaluation of machine translation for multilingual sentence-level sentiment analysis. In*Proceedings of the 31st Annual ACM Symposium on Applied Computing 2016 Apr 4* (pp. 1140-1145). ACM.
- [14] Shahare FF. Sentiment analysis for the news data based on the social media. In2017 International Conference on

- Intelligent Computing and Control Systems (ICICCS) 2017 Jun 15 (pp. 1365-1370). IEEE.
- [15] Severyn A, Moschitti A. Twitter sentiment analysis with deep convolutional neural networks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval 2015 Aug 9 (pp. 959-962). ACM.
- [16] Karanasou M, Ampla A, Doukeridis C, Halkidi M. Scalable and real-time sentiment analysis of twitter data. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) 2016 Dec 12 (pp. 944-951). IEEE.
- [17] Bindal N, Chatterjee N. A Two-Step Method for Sentiment Analysis of Tweets. In 2016 International Conference on Information Technology (ICIT) 2016 Dec 22 (pp. 218-224). IEEE.
- [18] Mensikova A, Mattmann CA. Ensemble sentiment analysis to identify human trafficking in web data.
- [19] Amolik A, Jivane N, Bhandari M, Venkatesan M. Twitter sentiment analysis of movie reviews using machine learning techniques. *International Journal of Engineering and Technology*. 2016;7(6):1-7.
- [20] Kanakaraj M, Guddeti RM. Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015) 2015 Feb 7 (pp. 169-170). IEEE.
- [21] Chikersal P, Poria S, Cambria E. SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) 2015 (pp. 647-651).
- [22] Satapathy R, Guerreiro C, Chaturvedi I, Cambria E. Phonetic-based microtext normalization for twitter sentiment analysis. In 2017 IEEE International Conference on Data Mining Workshops (ICDMW) 2017 Nov 18 (pp. 407-413). IEEE.
- [23] Bifet A, Frank E. Sentiment knowledge discovery in twitter streaming data. In International conference on discovery science 2010 Oct 6 (pp. 1-15). Springer, Berlin, Heidelberg.
- [24] Patil TR, Sherekar SS. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications*. 2013 Apr;6(2):256-61.
- [25] Gupta R, Jivani AG. Analyzing the Stemming Paradigm. In International Conference on Information and Communication Technology for Intelligent Systems 2017 Mar 25 (pp. 333-342). Springer, Cham.
- [26] Aker A, Petrak J, Sabbah F. An extensible multilingual open source lemmatizer. In Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017 2017 Nov 17 (pp. 40-45). ACL.
- [27] Straka M, Hajic J, Strakov áJ. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In LREC 2016 May.

Authors' Profiles



published many papers of sentiment analysis in various conferences.



has published various research papers in reputed journals and conferences.

How to cite this paper: Dimple Tiwari, Nanhay Singh, "Ensemble Approach for Twitter Sentiment Analysis", *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.11, No.8, pp.20-26, 2019. DOI: 10.5815/ijitcs.2019.08.03