International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

A Comparative Study on Sentiment Analysis Using KNN and SVM Models

J. Uma[1], Dr. K. Ramesh[2]

[1]Ph. D Research Scholar
Department of Computer Science
Periyar University, Salem, Tamil Nadu, India
umacheran83[at]gmail.com

[2]Associate Professor
PG & Research Department Computer Science and Applications
Vivekanandha Arts and Science College for Women [Autonomous]
Elayaampalayam, Tiruchengode, Tamil Nadu, India

Abstract: ***Today in this modernized era, millions and trillions of people were started sharing their day-to-day updates through the micro-blogging platforms like Twitter using where they are allowed to share their opinions and feelings within 140 characters and this restriction was done to make them concise and clear on what they are expressing through. This has turned out to be the richest source for analysis of Twitter data sentiment and belief mining by mining the sentiments both from online reviews and the social media platforms by utilizing the approach of a bag of words. Compared to other approaches, the bag of words approach is better as it accounts only for the words individually along with their count instead of neither considering the whole paragraph nor the sentence fully. Hence, this proposed paper was aimed at developing such a kind of classifier that can automatically correct and classify the unknown tweet sentiment with its label classification done by implementing few accurate techniques. Two methods are introduced in this proposed work: sentiment classification algorithm (SCA) based on support vector machine (SVM) and sentiment classification algorithm (SCA) based on k-nearest neighbour (KNN) and both methods can be evaluated for their performance depending upon the type of real tweets.***

Keywords: Twitter; Twitter data, KNN, Micro-blogging platform, social media, Sentiment, Support vector machine

1.Introduction

Today, the internet has grown tremendously with the vast population of people conveying, discussing, and exchanging their information over a topic via the internet and thus, the internet has become a tremendous piece of human existence covering a wide extent of regions, for instance, insightful information, audits, web-based media, assessments about market items, or comments about friendly issues, etc. thus making people think and decide over numerous topics. Most people constantly check out others' sentiments before taking an extreme decision. The information which is assembled is ordinarily separated to choose the notion of the information, for instance, negative suspicion or positive end. In general, notion alludes to sentiments, perspective, feelings, and appraisal. This sentimental analysis was performed to evaluate the significance rather than arranging the sentences based on their category as neutral or negative or positive. People will involve in social media activities like commenting, reviewing, and giving a rating for products online through the medium of long-range interpersonal communication [1], [2]. Usually, the analysis of sentiments of Twitter will be done in three stages as sentence stage, the perspective stage, and the document stage as mentioned in [3] and amongst all, the document stage will intend to group the report based on the type of positive type or negative type.

In the sentencing stage, sentences will be divided whereas, in the perspective stage, several parts are archived in the corpus [4], [5]. A different perspective on the same topic can be raised through Twitter and over billions and trillions of kilobytes of information are being exchanged on a single day as mentioned in [6]. Especially the product reviews are done as tweets and in the commercial market, it has been used by the cloth merchants to get product reviews progressively and at the same time, people can decide whether to buy the product or not based on those reviews of the product [7]. Analyzing the Twitter sentiments will group the sentiments as positive and negative types in which the outcome will be diminished. Usually, tweets will be irregular comments, incomplete, and un-organized and for measuring the tweets, feature selection will be done along with the pre-preparing of such words [8], [9]. Written words are called tweets which will be a shorter message of 140 characters mentioned in the language limited on the frame. Due to these limitations, it is ever difficult to understand the emotions of people and it cannot be analyzed [10], [11]. It is difficult to identify the difference between the tweets posted and hence applications like mining audit portrayal information can be utilized for assorting the highlights of the tweets with the sentimental standards for making the leadership [12], [13]. Here, two types of techniques are compared based on their accuracy to detect the accurate sentiments because outcome, it was found that compared to SVM, and Sentiment Classifier Algorithm (SCA) is performing better in all aspects.

2.Related Work

From the writing of the survey, it was seen that few studies were conducted for evaluating the classification algorithm performance in the sentiment analysis, and a comparative study over the types of classifiers was conducted by [13] by using the four types of classifiers like SVM, K-Nearest neighbors, Native Bayes, and Decision Trees and simultaneously analysis of Twitter sentiments are done along with the product reviews. Different types of sampling methods such as random type sampling method, linear type of sampling method, and the bootstrap type of sampling method are used with training samples of the product reviews. After that, the result shows that the best output can be achieved by using a bootstrap type of sampling along with SVM as it reduces the rate of errors. Unigrams and occurrences of the terms are used for classifications input but no result was observed with the influence of the input. Evaluating the performance of three types of classifiers of Passive-Aggressive (PA) Algorithm Based Classifier, Winnow classifier, and the Language Modeling (LM) Based Classifier was done and evaluated by [14] with the 100000 online product reviews of high order N-grams which is n having the values than the value of 3. Of all the others, the result proved that passive-aggressive based classifier shows better performance when they are combined and used with higher-order N-grams. Up to the feature-length of 6-grams was analyzed through the study conducted by Hang et al with no impact on classifier performance. Following this, a study was conducted by [13] with reviews of online products having 800 characters of the word and this was added as a challenge in the Twitter Sentiment analysis as the datasets from Twitter will have large sparsely. Hence, n-grams features are used by Hang‟s as the tweet may have 4 to 6 words with higher-order N-grams. Hence, a twitter-specific comparative study was required for evaluating the classification algorithms with multiple formats of the input having a direct impact on the accuracy of the classification.

2.1 Sentiment Analysis

Sentiment analysis otherwise called sentiment mining or opinion mining, or polarity mining is done to analyze the direction-based texts containing the emotions and the opinions through implementing several tasks. Of all, 4 main tasks are considered and they are class naming, target distinguishing, pre-processing the data, and the explanation granularity is considered [15]. As the words are unstructured having errors and mistakes, language pre-processing must be done in the sentiment analysis areas inclusive of checking the spelling thereby removing the punctuation marks before classification is done [16].

3.Working Procedures

Variety of features are being utilized for experiments of classifications like the word type, n- gram type, punctuation type, and the pattern type as mentioned in [9] which also included one more feature of key-based feature and these techniques can be used after the calculation of weights depending upon these features.

a)Word Feature – In word feature, binary tweet feature which will be compared with a dictionary to determine which are the stop words and which aren‟t [10] and those mentioned in [11], [12] aren‟t considered. Else, each of the word will be considered forward feature and it will be considered so if two or more punctuation symbols are encountered in the sequence [13]. Also, the word RT with meaning "retweet" will not constitute a feature as appearing inside the data set in most of the tweets. The word feature can be calculated as:

$$ws = \frac{Nf}{coun(f)}$$

In which *Nf*is denoting the tweet features present and whereas denotes the features of the whole dataset.

b)N-Gram Feature - In N-Gram Feature, a sentence will be considered as a binary n- gram feature with 2-5 consecutive words and rare words will measure highest ever than the common words ending in affecting the classification.

c) Pattern features - In Pattern Features, 3 types of words are classified as content words (CWs), high frequency words (HFWs), and the regular words (RWs) in which the word frequency is considered as *f* whereas *frf*will be representing the dataset frequency and if *frf*> *FH*, then the high frequency word will be taken into account. On the additional side, if *frf*< *FC*, the *f* is the content word with the remaining words as regular words. Using all the dataset words, frequency of a word will be calculated and estimated. All the sequences of HFWs will be considered as punctuation characters as URL, TAG, REF, and RT meta-words in detecting the pattern which will be considered as an ordered sequence of.

FC is set with upper bound of 1000 words per million and FH is set with lower bound of 10 words per million and experimentally verified that this gave better results with lower bound of FH with 100 words per million. Also, it was shown in results that there is an overlap produced between HFWs and CWs with the bounds of FH and FC. Using the simple strategy as mentioned below.

If, $f_{rf} € \left(FH, FH + \frac{FC}{2}\right)$, then the word will be identified as HFW.

If, $f_{rf}€ \left[FH + \frac{FC}{2}, FC\right)$, then the word will be identified as CW.

Patterns with 2 to 6 count of HFWs and 1 slot to 5 slots for CWs are considered but, still more patterns are required for starting and ending the HFWs thus, the minimal pattern form as (HFW) (CW slot) (HFW).

Table 1: Algorithm accuracy over 1000 tweets

Name of method	KNN with normalization (4 features)	KNN with normalization and keyword base (5 features)	Algorithm [9]	SVM with normalization (4 features)	SVM with (4 features	SVM with normalization and keyword base (5 features) with grid search	SVM with normalization and keyword base (5 features)
Tweets number	1000	1000	1000	1000	1000	1000	1000
Recall	0.69	0.72	0.83	0.68	0.68	0.82	0.59
Precision	0.72	0.55	0.53	0.63	0.75	0.65	0.57
TPR	0.79	0.73	0.72	0.76	0.71	0.75	0.79
FPR	0.29	0.27	0.57	0.59	0.34	0.48	0.43
F-Score	0.69	0.89	0.72	0.59	0.78	0.81	0.79
Accuracy	81.20%	81.56%	73.15 %	52.45 %	77.12 %	59.45 %	77.23 %

d) Punctuation Feature – Here, five types of features are classified as tweet exclamatory marks, tweet quotes, query symbols of the twitter, length of the tweet word, and the amount of capital lettering in the terms.

$$wp = (3 * Np)/(Mp * (Mw + Mng + Mpa))$$

Using the above method, the weight wp of a punctuation feature p will be stated as in which the N-gram will be denoted by *Mw*, the word "s maximum value will be denoted by *Mng*, and the pattern feature groups will be denoted by *Mpa*thus maximizing the *wp*with average weights of other features.

e) Key-based feature - In the Key-based feature, a list with 18000 words are used as shown in [14] along with the strength of the sentiment within the range of 1 to -1. Depending upon the weight of key based feature, the strength of the word will be calculated [15].

4. Evaluating the performance with the results of the experiment

In this part, in view of the aftereffect of Sentiment Classification Algorithm (SCA) and Support Vector Machine, performances of different algorithms are evaluated. Our critical boundaries to assess performances are Recall, Accuracy, and Precision along with a couple of open-source AI library [16]-[18] during the evaluation of algorithms performances assessment. Here the exactness of all calculations on 1000 tweets are given and four attributes whose are precision, review, F-Score, and Accuracy are observed as shown in Table 1.

Precision likewise called positive predictive value is the small portion of recovered occasions that are pertinent while recalling otherwise called sensitivity is the fraction of relevant instances that are understanding and measure of relevance. From precision and recall, the F- score is determined. As indicated by the result, an outline of performance evaluation for various sizes of the dataset is considered but for better examination, the result must be compared with different parameters which are effortlessly addressed through graph diagrams below.

a) Performance Evaluation Based on Accuracy

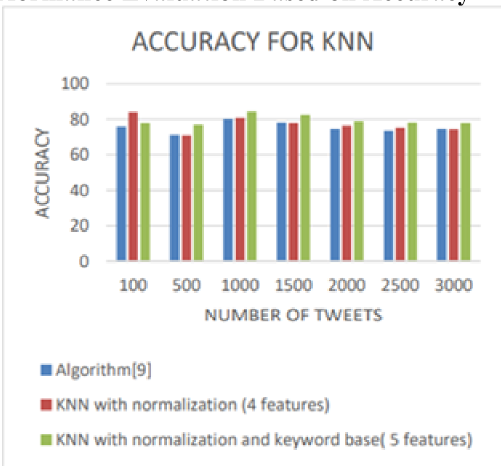

Figure 1: Shows KNN accuracy

The different KNN versions for different dataset sizes with the accuracy is shown with a graph in above mentioned Figure 1, and it can be seen that when the original KNN version is used with 4 features, the accuracy will be higher for all the datasets with dataset normalized. If a keyword-based feature was added, the accuracy will be much increased.

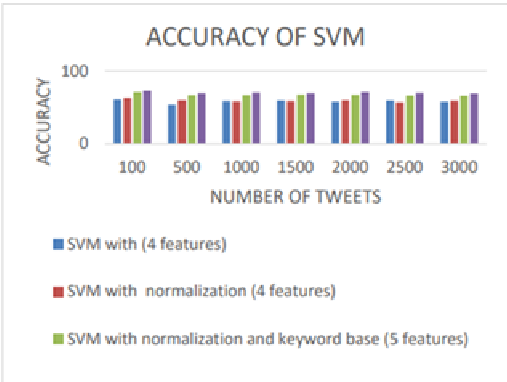

Figure 2: Shows SVM's accuracy

SVM"s accuracy is shown in Figure 2 and it can be observed that SVM accuracy will be 60% with 4 features for different dataset sizes. Once the dataset is normalized, accuracy will be increasing and more and more when added with another feature. Final accuracy of 70% was achieved with grid search methodology for every dataset.

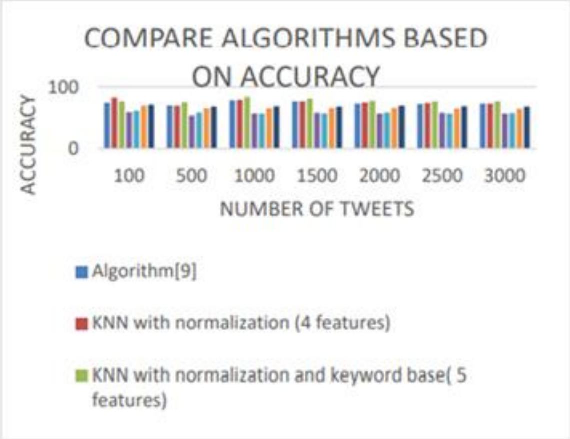

Figure 3: Shows the comparing algorithms based on accuracy

The KNN accuracy for its different versions will be higher than the SVM accuracy for different versions and it is shown in Figure 2 and it shows that compared to SVM, KNN is performing better on all datasets. Sometimes, a model having lower level of accuracy would be considered for predicting the power on the problem. If there is a large class imbalance problem, the model will not be useful and hence, added measures such as recall and precision are considered for classifier evaluation.

b)Evaluating the performance depending upon the precision

During the process of assessment of the classifiers, it is must that the precision should be considered. Lower precision indicates bulk quantity of false positives and gives higher accuracy with lower value of exactness. Precision must be analyzed when the priority is higher than the exactness. It can be done by using the number of positives operated from false and true positives.

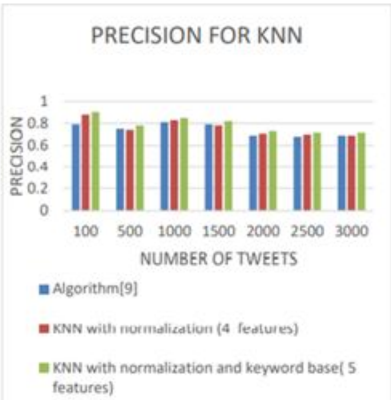

Figure 4: Shows KNN's precision

The original KNN"s precision is lower when compared with updated KNN"s precision as seen in Figure 4 with 4 features with normalized dataset. By adding more features, the precision will be increasing.

The precision details for multiple SVM versions are shown below in Figure 5 and when measured with 4 features, the precision will be quite increasing thereby normalizing it. When more features are added, the value will be increasing more with the grid search.

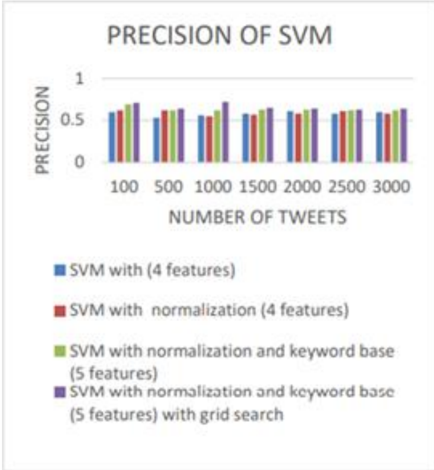

Figure 5: Shows SVM precision

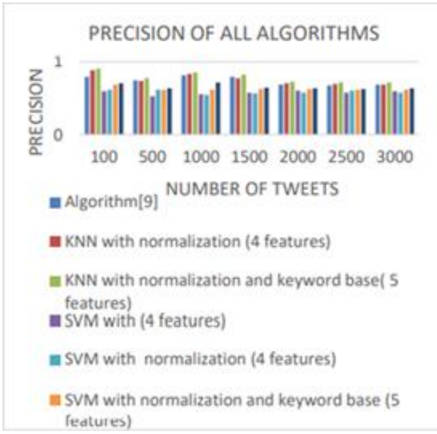

Figure 6: Shows comparing algorithms based on precision

It can be concluded with the Figure 6 that the precision will be higher for multiple types of KNN than SVM thus exhibiting the better performance than the SVM.

c) Performance Evaluation Based on Recall

In performance evaluations of classifier, recall is playing a vital role as a measure of classifiers completeness. On the other side, the lower recall will be representing the false negatives.

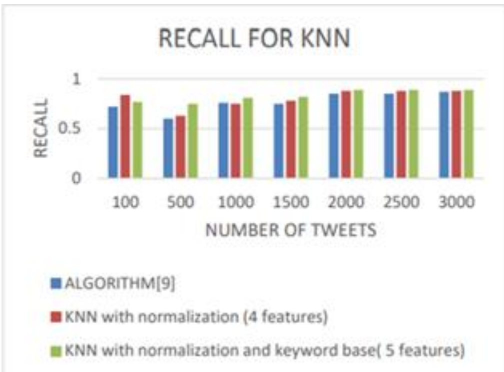

Figure 7: Shows the KNN recall

The recall can be evaluated using the true positives by dividing them with false negatives and the true positives and from Figure 7, it can be observed that a good value of recall will be provided by different KNN versions with different sized dataset and when this is modified with modifying the KNN version, the recall will be increasing simultaneously to 90%.

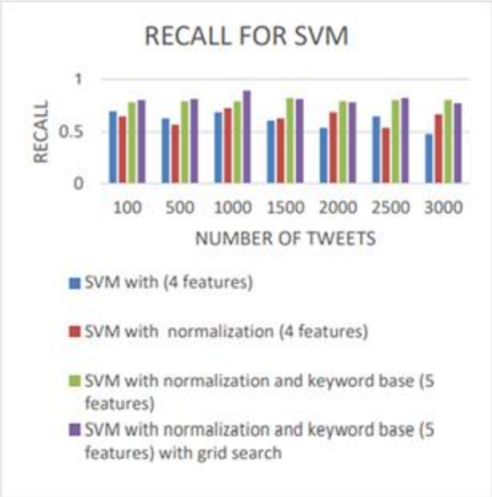

Figure 8: Shows the recall for SVM

From the Figure 8, it can be observed that the recall is good for different SVM versions and it will be increasing with upgraded version. Recall is found to be 60% to 70% for SVM with 4 features without normalization and with normalization versions. When more features are added, it will be increasing up to 80%.

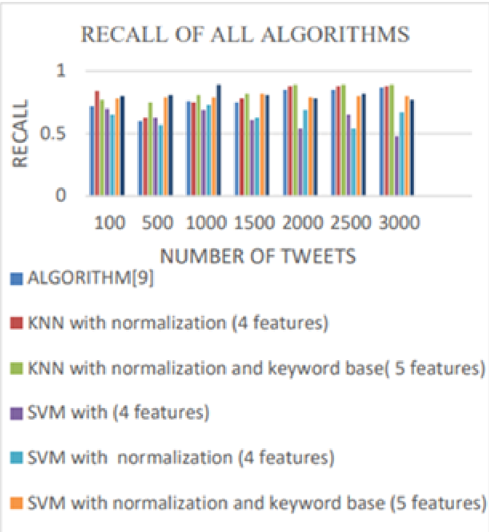

Figure 9: Shows the comparing algorithms based on recall

From the comparison between multiple types of KNN and SVM in Figure 9, it can be said that the percentages of the last two types of SVM and KNN are same closely with each other.

d)ROC Graph for Performance Evaluation

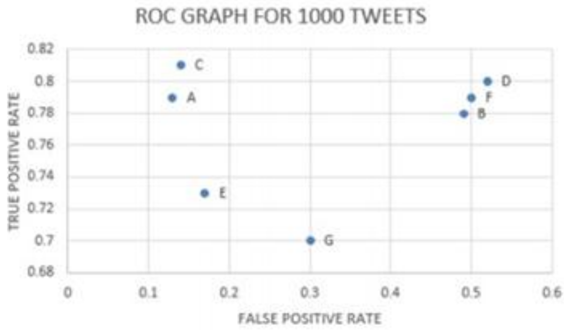

Figure 10: Shows the ROC graph for 1000 tweets

Depending upon their performance receiver operating characteristics, choosing of classifiers will be done using the ROC graph method [19] in which the genuine positive rate (TPR) will be plotted against the false-positive rate (FPR). The affectability or likelihood of identification genuine positive rate will be demonstrated whereas the dropout or likelihood of false alert will be indicated with the bogus positive rate. The algorithm functioning with 1000 Tweets is as shown in Figure 10. From the ROC graph, it can be observed that the algorithm performance was better if the true positive rate is higher and the false positive rate if lower.

A 2D algorithm plane will be plotted with the false positive rate on X-axis and the true positive rate on Y-axis and the performance will be better than the other algorithms. With the data point in the northwest corner, the performance will be better when compared with other algorithms with o data point fulfilling the requirements. From all the graphs, it can be said that KNN is performing better than SVM due to the reasons like SVM will be better with higher dimensions but, it cannot perform better in this experiment as five features are used. When the k-fold-cross-validation algorithm is used, highly imbalanced datasets will be produced with the difference between higher positive tweets and negative tweets. It is ever difficult to find the best c and gamma pair for any particular dataset using the grid search algorithm. A hyper plane between the classes will be assumed with SVM and sometimes it will be difficult to determine.

5.Conclusion and Future Work

Dependent upon the micro-blogging, analysis of sentiment is still in development, for example, positive sentiment is “It is a Nice Day!” and negative sentiment is “it is a horrible day!” This paper was proposed to find out the negative sentiment on Twitter data and the positive sentiment on Twitter data. In this proposed work, a simple model has been proposed with features like pattern feature, keyword-based feature, N-gram feature, word feature, and punctuation feature. Also, a machine-learning algorithm SVM (Support Vector Machine and a KNN classifier are being utilized for calculating the accuracy of all algorithms and it was found that the performance of SCA is better than the SVM.

In the future, many related issues will be studied relatively thereby adding the additional features to improve the performance by considering the English tweets and no other emoticons tweets. Hence, several language tweets will be added in the next work along with the emoticon's tweets. Aside from this, an attempt will be made to distinguish another assessment mark of person and simultaneously, with a major measure of tweets. Hence, it can be said that our future work will be accompanying activities like:

1. Additional features – Detection of sentiment will be done by adding additional features for providing excellent results.
2. Functioning with more tweets – currently, text tweets are used but, in the future, more emoticons tweets will be used. Additionally, a dataset containing the large tweets will be considered.
3. Working with numerous dialects – in the future, different languages will be used along with the current Java languages and Java JAR files.
4. Evaluating the performance & computing the exactness – For calculating accuracy and evaluating the performance, a confusion matrix will be used currently, and in the future, machine learning algorithms will be implemented.
5. Working with world issues – Given a productive conclusion name, an attempt will be made to perceive how it very well may be applied to tackling real-world problems. Concerning model, foreseeing official political decision, assessing item notoriety, and so on
6. Focus on recognizing another opinion mark – both the positive and negative sentiment labels will be used in future work.

Hence in this proposed paper, the main focus was laid over the sentiment analysis in a general category such as positive sentiment and negative sentiment in the field of analyzing the sentiments which can be compared with the other domains.

References

[1] Amole A, Nain S. Performance Evaluation of Machine Learning Classifiers in Sentiment Mining.
[2] Al-Radaideh QA, Al-Qudah GY. Application of rough set-based feature selection for Arabic sentiment analysis. Cognitive Computation. 2017 Aug; 9 (4): 436-45.
[3] Bouazizi M, Ohtsuki T. A pattern-based approach for multi-class sentiment analysis in Twitter. IEEE Access. 2017 Aug18; 5: 20617-39.
[4] Bouazizi M, Ohtsuki T. Multi-class sentiment analysis in Twitter: What if classification is not the answer. IEEE Access. 2018 Oct18; 6: 64486-502.
[5] Frost SL. *How simulation can illuminate pedagogical and system design issues in dynamic open ended learning environments (*Doctoral dissertation, University of Saskatchewan).
[6] Jianqiang Z, Xiaolin G. Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access. 2017 Feb 22; 5: 2870-9.
[7] Koç SŞ, Özer M, Toroslu İH, Davulcu H, Jordan J. Triadic co-clustering of users, issues and sentiments in political tweets. Expert Systems with Applications. 2018 Jun 15; 100: 79-94.
[8] Katta P, Hegde NP. A Hybrid Adaptive neuro-fuzzy interface and support vector machine-based sentiment analysis on political twitter data. International Journal of Intelligent Engineering and Systems. 2019; 12 (1): 165-73.
[9] Kisan HS, Kisan HA, Suresh AP. Collective intelligence & sentimental analysis of twitter data by using Standford NLP libraries with software as a service (SaaS). In2016 IEEE international conference on computational intelligence and computing research (ICCIC) 2016 Dec 15 (pp. 1-4). IEEE.
[10] Liu D, Lei L. The appeal to political sentiment: An analysis of Donald Trump‟s and Hillary Clinton‟s speech themes and discourse strategies in the 2016 US presidential election. Discourse, context & media. 2018 Oct 1; 25: 143-52.
[11] Liu H, Cocea M. Fuzzy rule-based systems for interpretable sentiment analysis. In2017 Ninth International Conference on Advanced Computational Intelligence (ICACI) 2017 Feb 4 (pp. 129-136). IEEE.
[12] Padmaja K, Hegde NP. Twitter sentiment analysis using adaptive neuro-fuzzy inference system with genetic algorithm. In2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019 Mar 27 (pp. 498-503). IEEE.
[13] Shahana PH, Omman B. Evaluation of features on sentimental analysis. Procedia Computer Science. 2015 Jan 1; 46: 1585-92.
[14] Tripathy A, Agrawal A, Rath SK. Classification of Sentimental Reviews Using Machine Learning Techniques. Procedia Computer Science. 2015 Jan 1; 57: 821-9.
[15] Wang Z, Cui X, Gao L, Yin Q, Ke L, Zhang S. A hybrid model of sentimental entity recognition on mobile social media. EURASIP Journal on Wireless Communications and Networking. 2016 Dec; 2016 (1): 1-2.
[16] Watanabe H, Bouazizi M, Ohtsuki T. Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. IEEE access. 2018 Feb 15; 6: 13825-35.
[17] Yang M, Qu Q, Chen X, Guo C, Shen Y, Lei K. Feature-enhanced attention network for target-dependent sentiment classification. Neurocomputing. 2018 Sep 13; 307: 91-7.