International Journal of
**INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING**

ISSN:2147-6799 www.ijisae.org Original Research Paper

# An Intelligent Data Mining System for Tweet Opinion Analysis using Combined Cluster based Classification Approach

**Abdul Ahad[1], B. Madhuravani[2], Sivaramakrishna Kosuru[3], Syed Mohd Fazal Ul Haque[4], Mohammed Ali Hussain[5]**

**Submitted:** 10/09/2022 **Accepted:** 20/12/2022

***Abstract**:* We're all driven by the need to be heard and to be able to convey our thoughts and ideas clearly. We want to know about and discuss specific issues; in other words, the right to be enlightened and the need to make the proper decisions and choices. It's no secret that microblogging services like Twitter and Facebook are bursting to the seams with information. There is no longer a distinction between "read only" and "read-write" content on the web anymore. As a result of the information contained in micro-blogging posts, comments, and ratings, opinion mining data is extracted from these data points. Sentiment analysis is to identify the ideas, feelings, and attitudes expressed in the source material. This article presents a comparison with similar existing approaches. 85.92 percent accuracy and 82.35% success rate were achieved by the minor architecture. There are F measures of 84.99 percent and the proposed architecture works only with tweets in the English language. In the future, the research will focus on designing an architecture capable of handling many languages. Language to language the accuracy may be different which is uncertain, but when the work is carried out and compared with the state of art of existing approaches our proposed approach seems to be better. When the two methods indicated above are combined, they produce superior results in complicated opinion mining compared to state-of-the-art procedures.

***Keywords:*** *Cluster-based classification, Machine learning, Miner architecture, Supervised learning, Support vector machine, Intelligent Systems.*

## 1. Introduction

Now a day's internet has been a part of every once life, all the websites are keeping in touch with people in the form of social communication networking, it's become the main source for social communication networks websites to interact with all kind of resources such as videos, photos, activities with each other. Social networking is the grouping of individuals based on their own interests into a specific group. Opinion analysis is done using text in form of messages or posts to find whether it is negative or positive or neutral[1,2]. The data is extracted from the twitter API, Twitter data classification and clustering are focus on the difficult problems in the area of machine learning with respect to artificial intelligence. Moreover, there are countless issues related to research in the area of twitter data classification and the role of clustering on text classification in related to the support of accomplishing much better tasks based on the high rate of significance.

Architecture for a Twitter opinion miner has been proposed for classifying the Twitter information. Twitter information is multi-dimensional; one classifier is simply not enough to classify this data accurately. Therefore, the architecture uses ensemble learning to tackle this issue. Considerable preprocessing methods need to be done before opinions are mined from the dataset. Related to tweet classification good research problem is creating the all-around framed portrayal of multiplex relationships by discovering the vector features[3,4]. To analysis certain research problems in data mining; to reduce the dimensionality of the tweet by using the clustering technique. The clustering hallmark is to build a group of similar features dependent on occasion space and going along with them. This defines a proportion of comparability among collision-related features and establish the relationship between recognize features, it is additionally used to characterize similitude among the breakdown of a solitary point occasion that would not, at this point recognized them among their connected features. In a general sense, the features and boundaries identified with the cluster will turn into the normal loads of boundaries dependent on its own integral property.

A Framework for Twitter data mining is introduced in 2014. This paper discusses various challenges offered by sentimental analysis to twitter feeds and presents techniques for investigating the sentiment analysis in actual text mining and the analysis of sentiment done through 3-way classification[1]. It consists of preprocessing techniques such as analysis of idioms; tokenizes correction and termination of ending words. In the end spotting the accuracy of opinions is tested with the proposed algorithm on Twitter datasets. The proposed framework contains real-time access to tweets using Twitter API and a three-level classification to identify the opinion of a tweet. At the first level, an emoticon classifier uses a sample set to identify the emoticons in a tweet and categorize them as positive or negative. A polarity classifier that uses the "bag of words" method which is followed by a

[1] *Anurag University, Hyderabad, Telangana. India*
[2] *MLR Institute of Technology, Dundigal, Hyderabad, Telangana State, India.*
[3] *Andhra Loyola Institute of Engineering and Technology, Andhra Pradesh., India*
[4] *Maulana Azad National Urdu University, Gachibowli Hyderabad Telangana State, India.*
[5] *Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District., Andhra Pradesh, India*

* *Corresponding Author Email: author@email.com*

SentiWordNet classifier in a second level, this classification assigns a score to the tweet which then determines the polarity of the tweet with respect to the interest[2]. At the point when each of the three order strategies are named as unbiased than a tweet having a score of nothing.

A graphical representation of the results is given which compares the results on the basis of the matrices such as Accuracy, Precision, Recall and F-measure. Social media played an important role in developing a personal view of an individual on any topic related to methods of opinion mining of news headings using SentiWordNet[3]. This work focuses on two algorithms to perform the task with each algorithm performing a specified function. Words are proposed using python libraries and each news headline word is assigned a Part of Speech (PoS) tag. Which is further used to determine the sentiment of the headline. PoS tagger is used to identify which word is which part of speech and Lemmatization is done in this algorithm itself. The second algorithm feeds the words to SentiWordNet. This further assigns a positive or negative value to each word and identifies the overall polarity of the headline in order to define its effect or opinion on a certain topic.

The comparison and study on existing techniques for opinion mining including machine learning and lexicon-based approaches were provided along with cross-domain and cross-lingual methods and some valuation metrics[4]. It is been analyzed during research that the machine learning methods like Support Vector Machine and K-Nearest Neighbor obtained the highest accuracy and can be regarded as baseline learning methods, while the lexical-based methods were effective in some of human-labeled documents. The machine learning methods identified the given problem as a classification problem like the SVM and Naïve Bayes methods were used as supervised learning methods for the study of the analysis of tweets, the lexicon-based, dictionary-based and corpus-based methods were used. The complete task of sentimental analysis was decomposed into three tasks such as subjectivity classification, sentiment classification and other complementary tasks[5]. These tasks were done at several levels of granularity like level-1 (words), level-2 (sentence), level-3 (documents), and level-4 (aspect). The dataset used for carrying out the analysis was the publicly available on Twitter.

The study is organized into five sections: First section deals with an introduction about the topic, a brief comment on the problem is described in Second section, a brief description of the suggested method for resolving the problem in addressed in third section, and a discussion of the results achieved using the model proposed in section four. The end of the work is briefly discussed in the fifth portion of the work.The uncertainty which arose due to imposing constraints can be overcome in the proposed approach

## 2. Problem Statement

A Problem is to identifying duplicate tweet detection from a social network. The proposed work used the machine learning model for text-based classification, it is applied not only for testing dataset but also applied on training dataset too. Need to use the classifier and collaborative for the perception of unwanted tweets on the public networks for knowledge-based discovery and best performance[6]. The Support Vector Machine (SVM), Transudative Support Vector Machine (TSVM), Conjunction Support Vector Machine (C-SVM) and Conjunction Transudative Support Vector Machine (C-TSVM) classification approaches demonstrated the efficiency of this model. Many experiments were performed on the cluster classification and upon consulting different integrated clusters, the extensive literature on text classification and machine learning, and the following problems were identified.

The unique algorithm is simply not enough to classify and my opinions from multidimensional tweets. The data from the public networks are prone to sarcasm, spelling mistakes, and internet slang which adversely affect the quality of classification. The efficiency of the machine learning algorithm is dependent on the data set. The scalability was a major issue in opinion mining, increasing the size of the dataset can have an impact on the accuracy of the classifier. Twitter is filled with bots that exploit the machine learning algorithm (such as support vector machine) to their benefits[7]. Such bots adversely affect the accuracy of the machine learning algorithm.

## 3. Proposed System

### 3.1. Opinion Miner Architecture

The tweets were collected from Twitter streaming API according to a given input query shown in figure 1. Tweets were filtered out and stored in JSON format. This required proposing for accurate mining. The tweets were subjected to Twitter Bot or Not for filtered out Bot accounts from the corpus. After filtering out the Bots, the URLs and hashtags were removed; the typographical errors and misspelled words can be corrected by using the spell check feature[8]. An emoticon classifier can be used to extract emoticons from the tweets.

Stemming and lemmatization were done on each word of the tweet for part of speech tagging after which features were extracted for classification. By applying the segmentation on each tweet, we get the top k-ranked segments[9]. This top-k segment was used as features for support vector machine classifier and the tweet was classified. SentiWordNet and emoticon classifiers were applied to each tweet[10]. Scores from the three classifiers were aggregated to generate a Combined Weighted Score (CWS).

### 3.2. Algorithm: Combined Weighted Score Calculation

Input: String of the Query

Output: Predict the Opinions

Query String as input

Retrieve the Tweets form the Twitter API

Filter out the tweets other than English Language

Tweets Filter from bots

Duplicate tweets removed

Tweets save in a JSON file

Preprocess Procedure

- Remove the URL
- Remove the User Name
- Remove the Hash Tags
- Extract the Emoticons
- Replace the slangs and abbreviations
- Apply spell check
- Lemmatization’s
- Stemming
- Extract the Features of classification

End Procedure

Segmentation Procedure (Tweet)

- Generate Segmentation candidates
- Calculate the score of segmentation candidates
- Pick up the top K segments

End Procedure

Classification Procedure

- Classify the Tweet based on Emoticon classifier (EC)
- Classify the tweet using SVM classification
- Classify the Tweet using SentiWordNet(SWN)
- Combine the classification result
- Save the result
- End procedure

- End

To calculate the combined weighted Score by using the following formula

CWS = 0.6 * SVM + 0.4 * SWN + 0.3 * EC

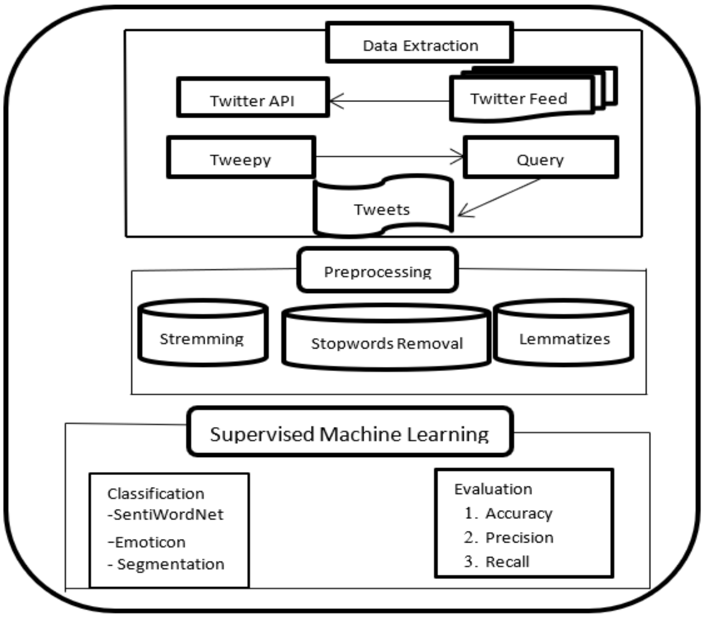


**Figure.1.** Architecture for sentiment Analysis

### 3.3. Algorithm: Text Classification

Considering P categorized a Class. If P is equal to One then label for training sample is

$$Tf = f(x1;y1);:::;f(xn ;yn) \tag{1}$$

Where Tf be the feature vector for $x_i^2$ and $y_i^2$

$$W(fi)=Tf(fi;x) / Df(fi) \tag{2}$$

Where, fi is the feature of regularity and (fi:x) is contingency of document number of times x. The regularity of inverse document is defined as

$$Df(fi) = log2(|D|) / Df(fi) \tag{3}$$

Considering the data set $Df(f_i)$ in which the size of document that contains $f_i$ at best and normalized the unit's length of the feature vector.

***Clustering***: It defines the cluster size and divides the datasets for each cluster. This process performs the mapping with an interesting area of tweets under the classified dataset. The distributed clustering algorithm with bi-selections rehashed in the choice cycle of grouping for both testing and preparing dataset[11]. Then the available datasets are congregated into two individual groups, from which one of the groups is picked out and the other is cut up. This action becomes continuous till it is inferred by number k of which is found in groups. During this interaction in each progression, the bi-chose group of so and its subsequent are planned with k-methods of enhanced arrangement bunch inside the capacity of interior measure[12].

$$Max \sum_{g=1}^{K} \sqrt{sim(u,v)} \tag{4}$$

$$u, v \in \Sigma g$$

Where u, and v addresses the records of two clusters and sim(u,v) discovers the comparability of the archive between them.

***Expansion-Step***: It defines the metadata of each cluster in the feature space for testing and training datasets. That is calculated Tf * Df scheme by its weight of each cluster[13].

***Classification-Step:*** It defines the dataset applied with binary classification using the linear kernel method based on the weights of stack variable g with a default set. This work is implemented by using two primary combined frameworks called Task-1 and Task-2 to pick out the repeated tweets gathered from different datasets. To predict the tweet message whether it is spammed or not in Task-1 and to identify the user tagging with the URL in Task-2. During this process to stores all the posted tweets in a form of a bookmark[14,15]. The main purpose of this work is to broadly predict the tags, user-posted tweets and to identify whether it is spammed or not. The provided datasets are considered in a form, which is collected from active non-spam 3,638 users and spam 46,282 users. The spam is calculated manually by labeling the non-spammers and spammers[16]. A new mathematical model[17]for solving the location-allocation (LA) problem with uncertain parameters. Demand, distance, travel time, and other parameters in classical models may change over time in real-world scenarios. As a result, taking uncertainty into account increases the flexibility of the results and their applications. The goal of this research is to find the best placement for gas stations while keeping an unknown service level constraint in mind. A paradigm for identifying and evaluating causes of inadequacies in previously deployed information system projects is proposed[18]. In a case study, this paradigm was utilized to create fuzzy cognitive maps. The risk assessment stages of the informational system development are represented by the cognitive maps[19], which show the involvement of many stake holders. Uncertain set theory is a branch of uncertainty theory that deals with uncertain sets. The terms membership degree, uncertain set, membership function, uncertainty distribution, independence, anticipated value, Hausdorff distance, critical values, and conditional uncertainty will be discussed in the proposed work[20]. A proposed model of an uncertain inference rule that uses the conditional uncertain set tool to derive consequences from uncertain knowledge or evidence.

The proposed method[21]addresses the issue of clustering three-way data. The application of a direct probabilistic clustering algorithm based on the notion of allotment among fuzzy clusters to a matrix of feeble fuzzy tolerance, which represents a structure of the set of objects under uncertainty, is a method of problem-solving. The approach describes the fundamental concepts of clustering and the design of a direct probabilistic clustering algorithm. The multi-item single-period inventory problem is investigated using the credibility optimization method[22]. The expected value is used in the construction of the profit goal function in a novel risk-neutral inventory issue with uncertain demand. In our inventory problem, we use both discrete and continuous possibility distributions to characterize uncertain demands. The features of a comonotonic fuzzy vector[23] are addressed in general based on the relationship between a fuzzy vector's joint monotone distribution and its marginal monotone distributions. The comonotonic fuzzy vector is a beneficial characteristic of mutually independent fuzzy variables. Several similar characterizations for comonotonic fuzzy vector are established in the case of two-dimensional fuzzy vectors.

While most multiperiodic stock control problems assume that orders are placed at the start of each period or at any time (continuous review) depending on inventory levels, we assume that the periods among both two restocking of several products are identical and independent random variables in this paper. The order quantities in this case are integer types, and each product has two types of space and service level constraints. A model of the situation is first built, in which the shortages are caused by a combination of lead time and emergency orders, with the expenses being shortage, holding, purchase, and transportation. In resilient designs, multi-response surface (MRS) optimization[24,25] is used to discover the best parameters of a system in a satisfactory region and minimize response variation. In this research, an approach for optimizing multi-response surfaces in robust designs is provided, which uses fuzzy set theory to optimize mean and variance at the same time. To begin, a method for constructing a regression model which is based on duplicates of a reaction and aggregate regression models is proposed, with each response being expressed by a fuzzy regression model. The generated regression model incorporates fuzzy coefficients that account for data uncertainty.

## 4. Experimental Results

The proposed system was tested many times for various subsets of the corpus. The results presented in the following sections are an average of five independent runs on the dataset of 7000 samples with 90/10 splits. That is 90% of samples were used for training and 10% for testing. The process took an average time of 1010 seconds. A set of bigrams and trigrams was created using green corpus which contains 4,583 and 4,585 entered respectively. For each tweet, if a segment exists then it was selected otherwise it was rejected. The Selected segments were those that had priority in sentiment analysis. Top K segments were extracted from the green corpus of processed Tweets and these were used as features for SVM Classification. The sample positive and negative tweets after processing are depicted in Figure 2.

Positive Tweet:
Tweet: I hope that one day I will get best gift in my god.
Native Tweet:
Tweet: Feeling very sad after getting my result.

**Figure 2.** Positive and Negative Tweets

Table 1 represents the list of top k segments extracted from the tweets according to their probability. The first column represents the features and the second column represents the corresponding opinion score. The 0:1 value is a (-ve) opinion whereas 1:0 being (+ve) opinion. The third column describes the corresponding SVM probability of the segment. These segments are the features used by the classifier.

**Table 1.** List of top K extracted segments

| | | |
|---|---|---|
| Bad= True | 0.1 | 10.5 |
| Excited= True | 1.0 | 13.2 |
| Loved= True | 1.0 | 12.2 |
| Enjoy=True | 1.0 | 10.5 |
| Sad= True | 0.1 | 20.5 |
| Especially = True | 1.0 | 10.2 |
| Thanks =True | 1.0 | 15.2 |
| Remember=True | 1.0 | 13.2 |
| Ready=True | 1.0 | 12.5 |
| Wonderful=True | 1.0 | 10.2 |

In identifying the positive and negative emoticons according to the regular expressions the emoticon classifier works. The score of sentiment is assigned accordingly. Figure 3 shows the result of the emoticon classifier.

Tweet: Good news today for Babu Special mention to @abc123for an amazing win @ quiz on data mining☺#career
Sentiment: 1
Prediction:1
Tweet: I miss how happy I used to be........:(
Sentiment: -1
Prediction: -1

**Figure 3.** Emoticon Classifier Results

SentiWordNet Classifier is used for assigning the sentiment score to each tweet. Positive sentiment has a positive score and negative sentiment has a negative. Figure 4 depicts the score of each sample tweet using the SentiWordNet classifier.

Tweet: The way to start an Exam. Start reading and all students should follow this
Sentiment:1
Prediction:1
Tweet: Another day another virus attack and people want to import these savages to Italy.
Sentiment:-1
Prediction:-1
Negative Tweets:
Tweet: feeling very #bad and kind?#disappointed
Processed Tweet: Feeling very bad and disappointed

**Figure 4**. SentiWordNet Classifier result

In all these experimental results the opinion dataset was preprocessed and the frame cases related to the experimental study is as follows:

*Case 1*: Tweets without URLs and tags by removal stop words and stemming.
*Case 2*: Tweets with removal URLs and tags where stemming and stop words applied.
*Case 3*: Tweets with URLs and tags where stemming and stop words not applicable.

*Case 4*: Tweets with URLs and tags where elimination of stemming and stop words not applied.

*Case 5:* Tweets with URLs and without labels were stemming and stop words ends are applied.

*Case 6*: Tweets with URLs and without labels were stemming and stop words are taken out what's more, are been not applied.

*Case 7*: Tweets having labels and evacuation of URLs where expulsion of stop words and stemming is applied.

*Case 8*: Tweets by evacuation of URLs and comprehensive of labels where stop words and stemming words are not taken out.

Table 2 shows the trial results and furthermore it gives a specific pattern to contrasting the standard consequences of Support Vector Machine (SVM) and Transudative Support Vector Machine (TSVM) Classifier, the C-TVSM and C-SVM compares to be close to the genuine execution of TSVM or SVM classifier, when it has been utilized related by the proposed group composing calculation. Figure 5 shows the test examination of different classifiers.

**Table 2.** Comparison of performance measures using various techniques

| | SVM | TSVM | C-SVM | C-TSVM |
|---|---|---|---|---|
| Case 1 | 80.04 | 87.14 | 80.74 | 89.27 |
| Case 2 | 80.32 | 83.78 | 80.93 | 91.10 |
| Case 3 | 82.78 | 90.24 | 82.87 | 91.66 |
| Case 4 | 83.16 | 81.14 | 83.46 | 92.52 |
| Case 5 | 82.78 | 91.05 | 83.63 | 90.75 |
| Case 6 | 82.40 | 90.92 | 82.28 | 89.44 |
| Case 7 | 81.59 | 90.21 | 82.44 | 90.34 |
| Case 8 | 81.67 | 91.14 | 80.68 | 91.35 |

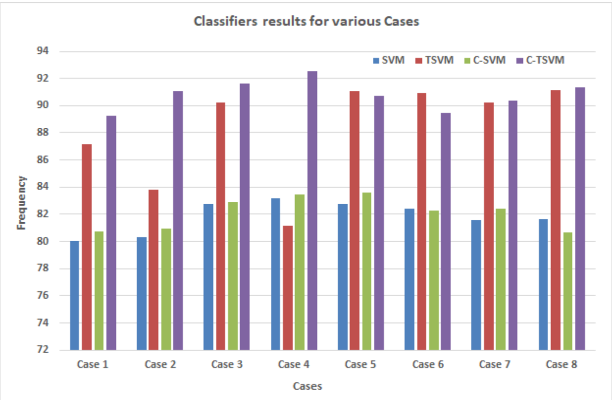


**Figure 5.** Bar graphs representation of various cases of classifiers

At first, the estimations of C-SVM and C-TSVM characterization demonstrate that the classifier will perform far superior contrasted with the SVM and TSVM standard classifier. C-TVSM classifier will perform best than different classifiers. Figure 6 shows the exactness investigation aftereffects of different classifiers.

**Table 3. Accuracy comparative Results**

| | SVM | TSVM | C-SVM | C-TSVM |
|---|---|---|---|---|
| Case 1 | 80.12 | 81.40 | 87.24 | 86.10 |
| Case 2 | 81.12 | 80.10 | 84.25 | 88.20 |
| Case 3 | 81.27 | 82.45 | 90.30 | 87.42 |
| Case 4 | 82.15 | 84.10 | 86.10 | 90.32 |
| Case 5 | 81.68 | 83.50 | 88.14 | 87.64 |
| Case 6 | 81.06 | 84.17 | 89.72 | 89.12 |
| Case 7 | 82.18 | 84.23 | 88.12 | 87.10 |
| Case 8 | 82.66 | 81.65 | 89.10 | 87.12 |

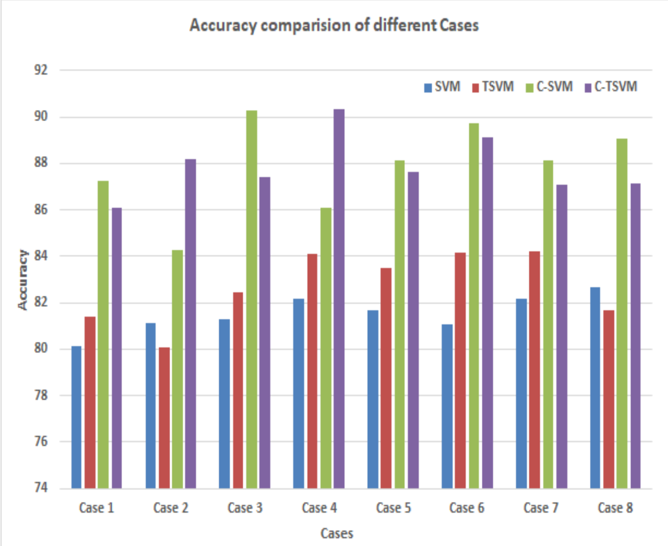


**Figure 6.** Accuracy comparative Results graph

The aftereffects of best and acquired 95.35% tweets with labels in URLs are remembered for tweet datasets and some stop words and stemming is taken out. The incorporation and exclusion of labels in URLs will influence the presentation of datasets. The incorporation of labels in tweet dataset as referenced in Case 3 and Case 4. The between bunch arrangement execution is higher in Case 1 and Case 2 when the labels in URLs are taken out. In the event that 5, Case 6, Case 7, and Case 8 a superior execution is noted, when URLs do exist at that point labels are prohibited from dataset.

Training Time: 1080.45
Testing Time:362.85
Accuracy:0.858

**Figure 7.** SentiWordNet Classifier Result

Figure 7 shows a better positive result compared to Support Vector Machine and Transductive Support Vector Machine classification. The Conjunction Support Vector Machine produces an intermediate output better than both the classifiers.

## 5. Conclusion

Tweet sentiment analysis uses a binary classification system to determine if a message's sentiment is favorable or negative. To sort them, scientists have relied on well-established methods such as Naive Bayes, Maximum Entropy, and SVM. To sort them, scientists have relied on well-established methods such as Naive Bayes, Maximum Entropy, and SVM A cluster-based classifier paired with a single server queuing paradigm offers better classification accuracies than a stand-alone. In this paper, we'll talk about how to extract and analyze feelings and opinions using opinion mining and opinion-based classification.

The results of the proposed work have been used to tweet datasets for a specific system while discovering spam tweets on friendly networks like Twitter. Our suggested Miner Architecture was capable of achieving 85.92 percent accuracy and 85.32 percent re-call. Recurrence is 84.99 percent. At the basic level, an emoticon classifier employs a sample set to determine if a tweet contains positive or negative emotions. A polarity classifier that uses the "bag of words" technique, which follows a SentiWordNet classifier in a subsequent level, this classification assigns a score to the tweet, which then determines the polarity of the tweet with respect to interest. A better-shared system rather than TSVN or SVM technique to revealing spam tweets has been shown susceptible to the trial aftereffects of a cooperative order approach with SVM/TSVM classifiers. In the future, the research will aim to create a framework that can manage multiple languages. The proposed method is only suitable for English-language opinion mining, but we can extract it for other languages.

## Conflicts of interest

The authors declare no conflicts of interest.

## References


[1] S. Bashir, U. Qamar, M. Y. Javed, *An ensemble based decision support framework for intelligent heart disease diagnosis*, in: International conference on information society (i-Society 2014), IEEE, 2014, pp. 259-264.

[2] B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, in: Mining Text Data, 2012.

[3] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: *an enhanced lexical resource for sentiment analysis and opinion mining*, in: Lrec, Vol. 10, 2010, pp. 2200-2204.

[4] Abdul Abdul, Yalavarthi Suresh Babu, and Ali Hussain, A New Approach for Integrating Social Data into Groups of Interest Springer Series, 978-81-322-2755-7, 2016.

[5] Bulent Tutmez et al., Assessment of Uncertainty in Geological Sites Based on Data Clustering and Conditional Probabilities, *Journal of Uncertain Systems,*1(3), pp.207-221, 2007

[6] S. A. A. Hridoy, M. T. Ekram, M. S. Islam, F. Ahmed, R. M. Rahman, *Localized twitter opinion mining using sentiment analysis*, *Decision Analytics* 2 (1) (2015) 1-19.

[7] G. I. Webb, J. R. Boughton, Z. Wang, Not so naive bayes: aggregating one-dependence estimators, Machine learning 58 (1) (2005) 5-24.

[8] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, Botornot: *A system to evaluate social bots*, in: Proceedings of the 25th international conference companion on world wide web, 2016, pp. 273-274.

[9] N Rao, Kantipudi MVV Prasad, An Evaluation of Data Security for Telemedicine Application Development, *International Journal of Computer Applications*, 79(1)(2013).

[10] Abdul Abdul, Yalavarthi Suresh Babu, and Ali Hussain, Multi-Level Tweets Classification and Mining using Machine Learning Approach, *Journal of Engineering and Science*, 1818-7803, 2018.

[11] M. Bayomi, K. Levacher, M. R. Ghorab, S. Lawless, Ontoseg: A novel approach to text segmentation using ontological similarity,*2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, IEEE, 2015, pp. 1274-1283.

[12] F. H. Khan, S. Bashir, U. Qamar, Tom, Twitter opinion mining framework using hybrid classication scheme, *Decision support systems* 57 (2014) 245-295 257.

[13] E. Golpar-Rabooki, S. Zarghamifar, J. Rezaeenour, *Feature extraction in opinion mining through persian reviews*, Journal of AI and Data Mining 3 (2) (2015) 169-179.

[14] D Subbarao, Kantipudi MVV Prasad, M Arun Kumar, The Influence of Electronic Communication on Machine Learning, *International Journal of Advanced Research in Computer Science*, 2(3) (2011)

[15] A. Esuli, F. Sebastiani, Sentiwordnet: A publicly available lexical resource 300 for opinion mining., *in: LREC*, 6 (2006) 417-422.

[16] G. Hesamian ,J. Chachi "On Similarity Measures for Fuzzy Sets with Applications to Pattern Recognition, Decision Making, Clustering, and Approximate Reasoning" Journal of Uncertain Systems Vol.11, No.1, pp.35-48, 2017.

[17] V.R. Ghezavati et al., "Designing Location-Allocation Model in a Service Network considering Chance-Constrained Programming: A Queuing Based Analysis" Journal of Uncertain Systems Vol.4, No.2, pp.116-122, 2010.

[18] Yee Ming Chen, Meng-Jong Goan, Pei-Ru Cheng "Uncertainty and Risk Analysis in Information System Projects Development "Journal of Uncertain Systems Vol.4, No.1, pp.34-46, 2010.

[19] Hans Schjær-Jacobsen, Numerical Calculation of Economic Uncertainty by Intervals and Fuzzy Numbers" Journal of Uncertain Systems Vol.4, No.1, pp.47-58, 2010.

[20] Baoding Liu "Uncertain Set Theory and Uncertain Inference Rule with Application to Uncertain Control "Journal of Uncertain Systems Vol.4, No.2, pp.83-98, 2010.

[21] Dmitri A. Viattchenin., "An Outline for a Heuristic Approach to Possibilistic Clustering of the Three-Way Data" Journal of Uncertain Systems Vol.3, No.1, pp.64–80, 2009.

[22] Ya-Nan Li1, Ying L., "Optimizing Fuzzy Multi item Single-period Inventory Problem under Risk-neutral Criterion" Journal of Uncertain Systems Vol.10, No.2, pp.130-141, 2016.

[23] Xueqin Feng, Yankui Liu., "Characterizing Credibilistic Comonotonicity of Fuzzy Vector in Fuzzy Decision Systems" Journal of Uncertain Systems Vol.10, No.4, pp.312-320, 2016.

[24] Ata Allah Taleizadeh, Seyed Taghi Akhavan Niaki, Gholamreza Jalali Naini., "Optimizing Multiproduct Multiconstraint Inventory Control Systems with Stochastic Period Length and Emergency Order" Journal of Uncertain Systems Vol.7, No.1, pp.58-71, 2013.

[25] Mahdi Bashiri, Seyed Javad Hosseininezhad., "A Fuzzy Programming for Optimizing Multi Response Surface in Robust Designs" Journal of Uncertain Systems Vol.3, No.3, pp.163-173, 2009.