Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Effective Twitter Sentiment Analysis System with Ensemble Classifiers and Feature Selection

Supriya Sameer Nalawade[1], Akshay Gajanan Bhosale[2]

[1]Research Scholar, Electronics Department, Sanjay Ghodawat University, Kolhapur, Maharashtra, India, supriya32119@gmail.com
[2]Electronics and Electrical Department, Sanjay Ghodawat University, Kolhapur, Maharashtra, India, akshay.bhosale@sanjayghodawatuniversity.ac.in

One of the most intriguing areas of research these days is sentiment analysis from Twitter. In order to create such systems, it blends data mining methodologies with natural language processing techniques. We presented an effective Twitter sentiment analysis method in this study. A machine learning model was developed using the suggested approach to identify both good and negative tweets. During the training phase, our model employed various methods to represent the input labelled tweets using various feature sets. For more accurate results, the classifier ensemble is shown various basis classifiers throughout the classification phase. The suggested technique may be used to gauge users' opinions based on their tweets, which is highly beneficial for a variety of uses, including product reviews, political polarity identification, and marketing.

Keywords: Opinion mining, Sentiment analysis, Classifier ensemble, Feature selection, Information gain.

1. Introduction

These days, the most popular online activities on the Internet are blogging, microblogging, social media communications, and chat. One of the most widely used microblogging platforms, Twitter contains a vast volume of both organized and unstructured data, making it one of the greatest sources of user-generated material. In accordance with their interests, people can express their views and ideas on many issues in their posted tweets.

Sentiment analysis, also known as opinion mining, is the process of identifying, detecting, and extracting views, feelings, and attitudes about certain topics by combining data mining, text mining, and web mining approaches. It might be used with a variety of data sources, including news, blogs, and reviews [1]. In a variety of application fields, including customer satisfaction [4], movie reviews [3], marketing campaign evaluation [2], and many more, identifying user opinions is very valuable information.

Twitter sentiment analysis differs from other social media sites in a number of ways: (i) users often use brief tweets to convey their status and mood; (ii) users may utilize emoticons and abbreviations to reduce character space; and (iii) feature engineering problems lead to several language representation concerns [5]. In this work, we investigated several feature set combinations that may be effectively applied to twitter representation. The retrieved features may result in a complicated calculation challenge because of the large dimension of the produced features vector, as is the case with the majority of text mining systems. In order to address this issue, features transformation techniques like feature hashing [7] and feature selection strategies like information gain and mutual information [6], etc.

Based on the content of the tweet, it may be categorized as either good or negative. If the text body of the tweet expresses a feeling, it is classified as neutral. Our consideration of the sentiment analysis system as a classification issue resulted from this. We should examine the collections of input tweets in this challenge and categorize them based on the attitudes that are present in each one. Additionally, an ensemble—a grouping of several classifiers—is utilized to create a single classifier that combines the advantages of each individual classifier. We used a majority voting ensemble in this research, which aggregates the results from three basic classifiers.

This paper's primary goal is to provide an effective system for Twitter sentiment analysis that makes use of the majority voting ensemble classifier and information gain as a feature selection method. The following research issues are addressed by implementing the suggested system and assessing its accuracy: (i) Which well-known feature set records the best accuracy? (ii) Does the use of information gain result in improved performance? (iii) Does the ensemble model outperform the individual classifiers in terms of accuracy? Furthermore, what elements influence its performance?

This paper's remaining sections are arranged as follows: The most pertinent work on the sentiment analysis problem is found in Sect. 2. The suggested Twitter sentiment analysis method is briefly described in Section 3. Section 4 presents the experimental data and commentary. Finally, conclusions are made in Sect. 5.

2 Related Work

It is suggested that sentiment analysis be used to determine users' polarity toward a certain topic based on their reviews, comments, or opinions. News stories, blogs, product reviews, microblogs, and forums have all used this subject. Owing to the wealth of research in this field, Ravi and Ravi [9] provided a thorough overview of the tasks, methodologies, and uses of opinion mining, which included a distinct section for sentiment analysis in general. Another review that addressed sentiment analysis techniques on Twitter data with a comparative examination of the current methodologies was given by Kharde and Sonawane [10].

Ghiassi et al. [11] used a supervised feature selection method using n-grams and statistical analysis to create a Twitter-specific vocabulary for sentiment analysis. 3440 manually gathered and annotated tweets from Justin Bieber's Twitter account were used to evaluate their suggested model. According on their experimental findings, their suggested model obtained 95.1% accuracy and marginally surpassed the conventional SVM classifier.

One of the difficulties with the current sentiment analysis techniques is choosing the salient characteristics set. The tweet text has a variety of characteristics that might be retrieved, but which combination yields the best accuracy rate? Agrawal and Mittal [12] have investigated a number of feature extraction and selection methods in order to identify the salient characteristics of a sentiment analysis based on machine learning. To determine the semantic orientation of each extracted feature and gauge the overall polarity of the input text, they integrated corpus-based and lexicon-based methodologies.

The feature-engineering challenge on Twitter sentiment categorization was examined by Agrawal et al. [13]. A variety of characteristics, including unigram, POSfeatures, senti-features, and tree kernel model, were merged in their feature sets. SVM was applied to the various feature set combinations for the classification challenge. They used a set of 11,875 manually annotated tweets to test their suggested approach. Their findings showed that, at around 75.39%, the feature set that included the unigram and sentifeatures had the best accuracy rate.

A single classifier was employed to complete the classification job in the majority of the machine learning-based sentiment analysis techniques that were described. For instance, the Naïve Bayes (NB) approach was adopted by Saif et al. [16] due to its strong performance in text classification issues, whereas Zhang et al. [14] and Mohammad et al. [15] employed the Support Vector Machines (SVM) technique. The classifier ensemble technique, on the other hand, was developed to train several classifiers and integrate their findings to address a single classification issue. By merging many classifiers, this method attempted to address some of the issues with the individual classifiers and produced a generalized decision boundary for the classification input [17]. It is not guaranteed that the performance of the classifier ensemble is always better than the individual classifiers combined in it, but in some cases, it reduces the risk of selecting inefficient classifier with the unseen data [18].

The underlying classifiers utilized in each of the suggested classifier ensembles and how their choices are combined vary. For instance, the majority voting ensemble was employed in the work of Lin and Koltz [19] and Rodríguez-Penago et al. [20]. The weighted voting ensemble with trained Naïve Bayes classifiers was employed by Clark et al. [21]. Furthermore, Hassan et al. [5] suggested a bootstrap approach that used around six basic classifiers and mixed various dataset, feature, and classifier parameters. Another combination rule was recently introduced by Da Silva et al. [22]. For each class, they determined the ensemble classifier's final judgment by averaging the probabilities generated by four classifiers.

3 The Proposed System

The elements of the suggested Twitter sentiment analysis system will be briefly explained in this section. The suggested approach operates in two stages, training and classification, as seen in Fig. 1. The training phase's objective is to develop the classification model based on the input labelled tweet collections to differentiate between good and bad tweets. During the classification stage, the newly unlabelled tweets will be given a positive or negative label by the trained classification model. Pre-processing, feature extraction, feature selection, and the classification model for sentiment analysis are the four phases that make up the system.

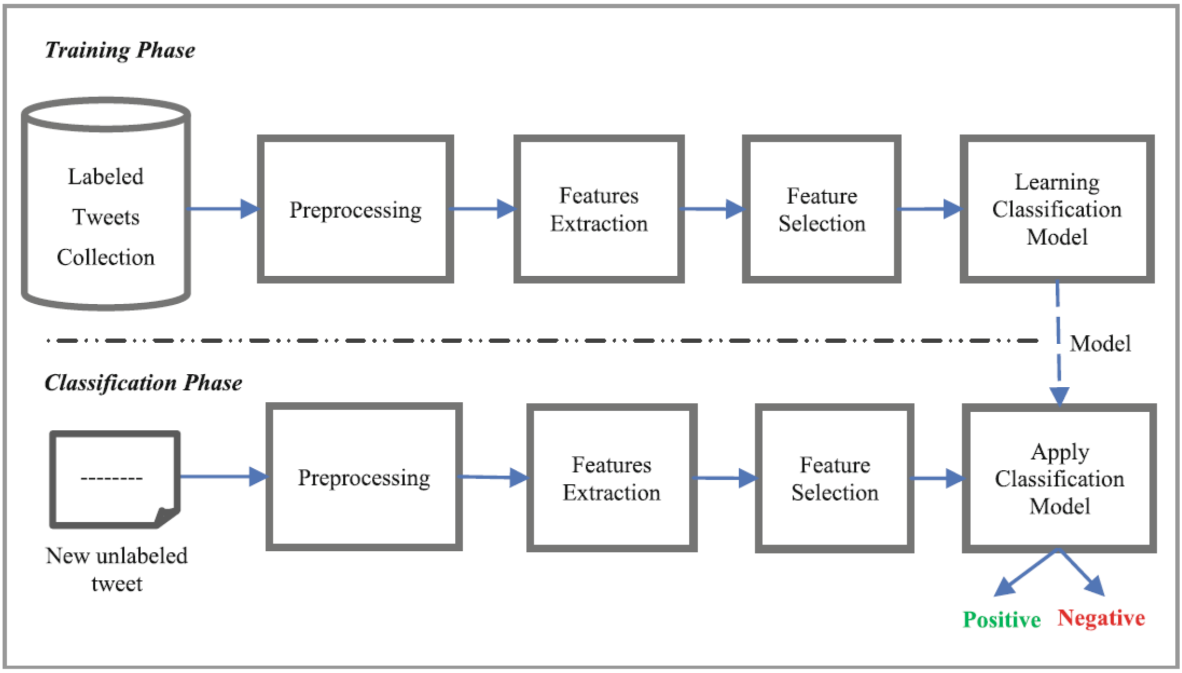


Fig. 1. Overview of the proposed system

3.1 Preprocessing

This stage's primary goal is to handle the input tweet text using natural language processing techniques so that it is appropriate for the subsequent step, which involves accurately extracting the features. Fig. 2 displays the comprehensive block diagram for the preprocessing stage together with an example tweet.

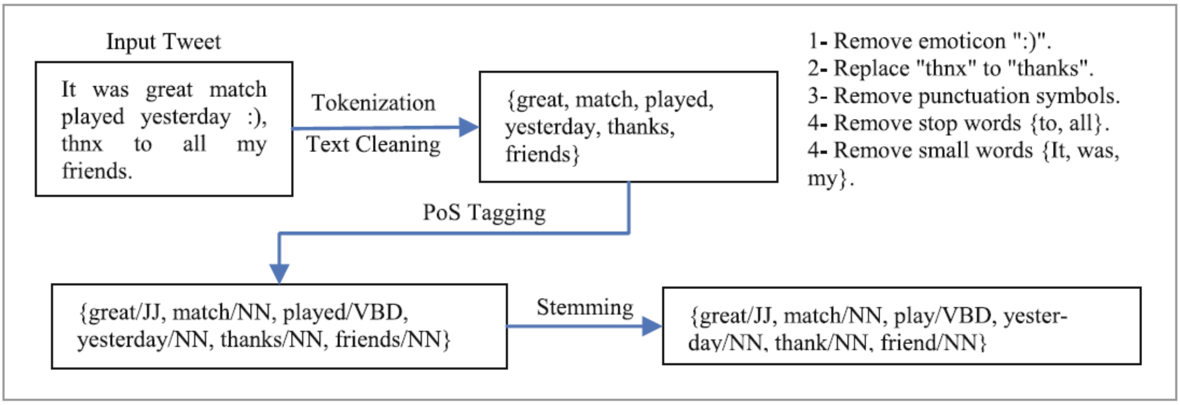

Input Tweet
It was great match played yesterday :), thnx to all my friends.
Tokenization
Text Cleaning
{great, match, played, yesterday, thanks, friends}
1- Remove emoticon ":)".
2- Replace "thnx" to "thanks".
3- Remove punctuation symbols.
4- Remove stop words {to, all}.
4- Remove small words {It, was, my}.
PoS Tagging
{great/JJ, match/NN, played/VBD, yesterday/NN, thanks/NN, friends/NN}
Stemming
{great/JJ, match/NN, play/VBD, yester-day/NN, thank/NN, friend/NN}

Fig. 2. Example for tweet text preprocessing

Tokenization, Text Cleaning, PoS Tagging, and Stemming are the four sub-steps that make up the preprocessing stage. Tokenizing, or dividing the input text into distinct phrases (referred to as tokens), was the first step in the preprocessing process. A word, acronym, hyperlink, emoji, or other punctuation mark that is often used in tweets can be represented by each token.

The second phase, called "Text Cleaning," is in charge of eliminating any extraneous language from the tweet's actual content. As seen in Fig. 2, following the tokenization and text cleaning procedures, the example input tweet "It was great match played yesterday :), thnx to all my friends" is converted into a list of terms that is {great, match, played, yesterday, thanks, friends}.

The third step is PoS Tagging in which we extract the part of speech tags for the input text. For example, word such as “great” is tagged with “JJ” because it is an adjective word. The final step is to stem the words to their original root in order to reduce the initial set of words representing the input tweet text. For example, the word “played” is transformed into the stem word “play”. The final words of the preprocessing step for the example tweet is {great/JJ, match/NN, play/VBD, yesterday/NN, thank/NN, friend/NN}.

3.2 Feature Extraction

To represent the input tweet text, a number of characteristics must be retrieved. We shall outline the many feature kinds that the suggested system would employ in this section. Extracting every potential stemmed word—also known as a term or token—from the input text is the simplest and most conventional method of text representation. This process is known as the Bag-of-words (BoW). The BoW in this work includes all of the unique bigrams (two consecutive word phrases) and unigrams (single word terms). For instance, the Unigram features "great," "match," "play," "yesterday," "thank," and "friend" are used in the tweet "great match play yesterday thank friend." "Great_match," "match_play," "play_yesterday," "yesterday_thank," and "thank_friend" are the bigram characteristics.

Certain words can convey the writer's viewpoint. While terms like terrible, bad, and harmful are examples of negative opinions, words like great, terrific, amazing, and exceptional may convey positive opinions. The opinion lexicon compiled by Liu et al. [23], which includes a list of 2006 positive and 4783 negative terms, was utilized in this work. Both positive and negative words are counted as Lexicon-based qualities for every tweet. As lexicon-based traits, the sample tweet has two positive words, "great" and "thank," and zero negative terms.

The part-of-speech tags for the words that were extracted are saved throughout the preprocessing stage. As PoS features, we count the amount of nouns, verbs, adjectives, and adverbs. Four nouns ("match," "yesterday," "thank," and "friend"), one verb ("play"), one adjective ("great"), and zero adverbs make up the retrieved PoS characteristics for the sample tweet.

Emoticons are symbols that, in the writer's perspective, symbolize certain states. In this phase, we compiled a list of emoticons that are often used on social media, particularly in tweets. There are 16 neutral, 77 negative, and 112 positive emoticon symbols in the list. The Emoticons features for each tweet in the collection are the number of emoticons detected in each state. The above tweet has zero negative and neutral emoticons and just one positive one, :).

3.3 Feature Selection

characteristics that were extracted. The majority of these characteristics are seen in the BoW bigrams and unigrams. The quantity of unique phrases in the input tweets collection significantly enhanced the vector's dimension. The majority of text processing systems, including sentiment analysis systems, suffer from the curse of excessive dimensionality. In this instance, we reduced the dimension of the output feature vector by using Information Gain (IG) as a feature selection strategy. In the suggested approach, each feature's information gain weight is determined using Equation (1), and features with weights greater than 0.01 are chosen.

Consider the input tweets collection with class attribute C that has two classes

$\{C_1$ = positive and C_2 = negative$\}$. For any given feature x, the information gain (IG) is

calculated by:

$$IG(x) = -\sum_{j=1}^{2} P(C_j) \log\left(P(C_j)\right) + P(x) \sum_{j=1}^{2} P(C_j|x) \log\left(P(C_j|x)\right) + P(\bar{x}) \sum_{j=1}^{2} P(C_j|\bar{x}) \log\left(P(C_j|\bar{x})\right)$$

Where, P(Cj) is the fraction of tweets labeled with class Cj, P(x) is the fraction of tweets in which feature x occurs and P(Cj|x)is the fraction of tweets with class Cj that has feature x.

3.4 Classification Model for Sentiment Analysis

Building a classification model that can effectively distinguish between tweets with positive and negative labels during the training phase is the primary stage in the suggested approach. Such a model might be constructed using a variety of machine learning approaches. We used SVM, NB, and LR as basis learners in a majority voting ensemble classifier that we built in the suggested system. These methods are widely used and perform exceptionally well in text categorization tasks. Fig. 3 provides a summary of the majority voting classifier ensemble that is employed in the suggested system.

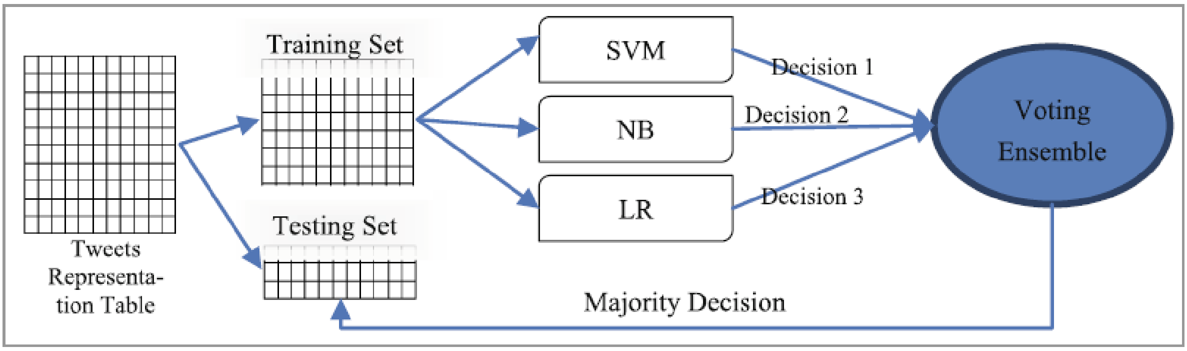


Fig. 3. Majority voting classifier ensemble

The extracted features from the collection of input tweets are divided into two groups, the training set and the testing set, as seen in Fig. 3. Each classifier receives the training set in order to record its judgment. The voting ensemble's ultimate decision output is thus the three classifiers reached a majority decision. Since this is the majority choice, the voting ensemble model will view this tweet as a good one. Lastly, the accuracy of the constructed classifier ensemble model is verified using the testing set.

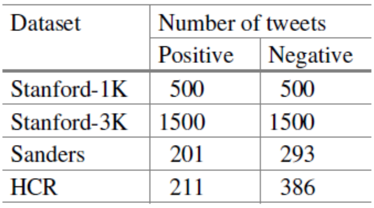
4 Experimental Results and Discussion

After the suggested approach is put into practice, its accuracy is evaluated using a variety of well-known datasets related to sentiment analysis on Twitter. The Stanford CoreNLP package is used to facilitate the Java implementation of the preprocessing and feature extraction procedures. The RapidMiner® program is used to implement the classification ensemble model and feature selection. Every experiment was carried out on a computer running 64-bit Windows 7 Enterprise Edition® with an Intel® Core TM i7-3770 CPU operating at 3.40 GHz and 8.00 GB of RAM.

4.1 Datasets

Four datasets are used in evaluating the designed experiments in order to evaluate the performance of the proposed system. The distribution of the positive and negative polarities in the used datasets is shown in Table 1.

Table 1. Distribution of positive and negative polarities in the datasets



Dataset	Number of tweets	
	Positive	Negative
Stanford-1K	500	500
Stanford-3K	1500	1500
Sanders	201	293
HCR	211	386

About 1.6 million tweets (800,000 positive and 800,000 negative) make up the Stanford Twitter Sentiment Corpus, which was gathered using a scraper that makes certain requests to the Twitter API [24]. Because of computing constraints, we did not use the entire training dataset in our trials. Two sample datasets, Stanford-1K and Stanford-3K, each containing 1000 and 3000 tweets, are obtained by unified sampling.

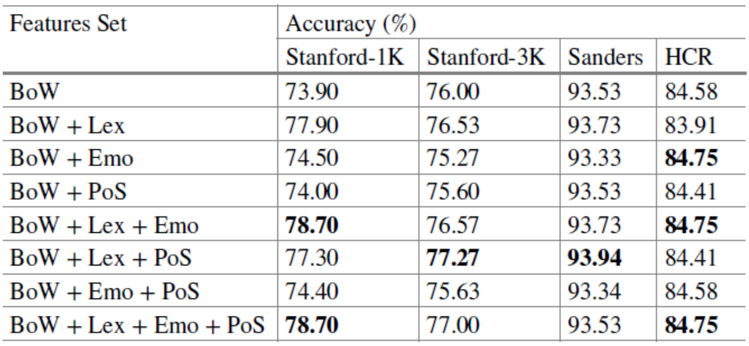
The final dataset, known as the Sanders Dataset [25], has around 5513 manually categorized tweets with four labels: irrelevant, neutral, negative, and positive. Four search terms—@apple, #google, #microsoft, and #twitter—are utilized with the Twitter API. Since the majority of these tweets are presently invalid or removed, we were unable to get them all. Only positive and negative tagged tweets—roughly 201 positive and 293 negative tweets—are of relevance to us in our research.

The Health Care Reform (HCR) Dataset is the fourth dataset. Crawling tweets using the hashtag "#hcr" in March 2010 is how this dataset was gathered [26]. A portion of these tweets have been manually classified as neutral, negative, and favorable. Only positive and negative tweets—roughly 597 total—are of relevance to us in this experiment (211 positive and 386 negative).

4.2 Sentiment Analysis Results

Set of Important Features. Answering the first research question and identifying the key feature set that yields the maximum accuracy are the goals of this investigation. The accuracy of the suggested approach utilizing the majority voting ensemble model is shown in Table 2. The various combinations of Bagof-Words (BoW), Lexicon-based features (Lex), Emoticon-based features (Emo), and PoS features (PoS) are used to represent each dataset.

Table 2. Accuracy for different combinations of features sets



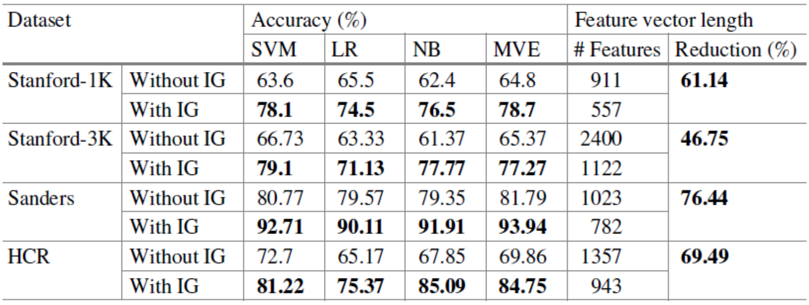
Features Set	Accuracy (%)			
	Stanford-1K	Stanford-3K	Sanders	HCR
BoW	73.90	76.00	93.53	84.58
BoW + Lex	77.90	76.53	93.73	83.91
BoW + Emo	74.50	75.27	93.33	**84.75**
BoW + PoS	74.00	75.60	93.53	84.41
BoW + Lex + Emo	**78.70**	76.57	93.73	**84.75**
BoW + Lex + PoS	77.30	**77.27**	**93.94**	84.41
BoW + Emo + PoS	74.40	75.63	93.34	84.58
BoW + Lex + Emo + PoS	**78.70**	77.00	93.53	**84.75**

Table 2 illustrates that there is not much of a difference in accuracy between BoW and other feature set combinations. Additionally, we see that the accuracy is higher when utilizing the feature set that contains all of the characteristics (BoW + Lex + Emo + PoS), as demonstrated by the Stanford-1K and HCR datasets. When the full features set is used, the accuracy of the suggested method is still extremely good in the Stanford-3K and Sanders datasets, with a very tiny margin from the best accuracy. This enables us to conclude that while Emo features may not significantly improve overall accuracy, Lex and PoS features are useful supplements to conventional BoW characteristics.

Information Gain (IG) is used. The experiment's objectives are to address the second research question and investigate how the information acquisition approach affects the suggested system in terms of both dimension reduction and accuracy improvement. The experiment's first section reports the accuracy of the Majority Voting Ensemble (MVE) model and the standalone classifiers (SVM, LR, and NB) in two scenarios: one without information gain and one with it.

Since the information gain approach is primarily used to choose the features that better fit the specified classes, we are interested in evaluating the reduction ratio that is produced after using it in the second half of the experiment. The reported accuracy for each dataset and the feature vector length before and after applying the IG approach are compared in Table 3.

Table 3. Accuracy comparison after using IG with different classifiers

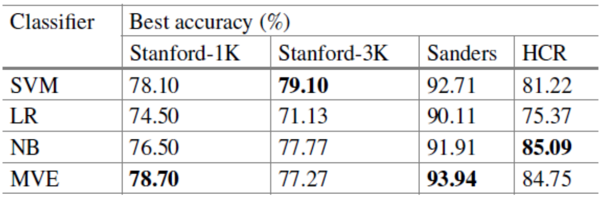


Dataset		Accuracy (%)				Feature vector length	
		SVM	LR	NB	MVE	# Features	Reduction (%)
Stanford-1K	Without IG	63.6	65.5	62.4	64.8	911	**61.14**
	With IG	**78.1**	**74.5**	**76.5**	**78.7**	557	
Stanford-3K	Without IG	66.73	63.33	61.37	65.37	2400	**46.75**
	With IG	**79.1**	**71.13**	**77.77**	**77.27**	1122	
Sanders	Without IG	80.77	79.57	79.35	81.79	1023	**76.44**
	With IG	**92.71**	**90.11**	**91.91**	**93.94**	782	
HCR	Without IG	72.7	65.17	67.85	69.86	1357	**69.49**
	With IG	**81.22**	**75.37**	**85.09**	**84.75**	943	

Table 3 makes it evident that applying the information gain strategy improved the suggested system's accuracy by an average of 15% across all classifiers. Additionally, each dataset's feature vector length is decreased by about utilizing the IG approach. The average is 63.45%. As a result, the classifiers use less computing power to differentiate between positive and negative classifications. Based on these findings, we can say that applying the information gain (IG) approach significantly lowers the feature vector's dimension while simultaneously improving the model classifier's performance.

Assessment of the Majority Voting Ensemble (MVE). Our goal in this experiment is to assess the MVE model's performance in order to respond to the third research question. Every classifier's performance is evaluated across all datasets and parameters, that obtained the greatest level of accuracy are noted. The accuracy of the MVE model is also recorded for comparison with other methods, as indicated in Table 4, and is assessed by merging the choices of the standalone classifiers with the ideal parameters.

Table 4. Accuracy comparison for different classifiers



Classifier	Best accuracy (%)			
	Stanford-1K	Stanford-3K	Sanders	HCR
SVM	78.10	**79.10**	92.71	81.22
LR	74.50	71.13	90.11	75.37
NB	76.50	77.77	91.91	**85.09**
MVE	**78.70**	77.27	**93.94**	84.75

Table 4 demonstrates that while LR produces the poorest accuracy results, SVM and NB classifiers perform well across a range of datasets. When it comes to the Majority Voting Ensemble (MVE) approach, its effectiveness is influenced by the distinct classifiers. For instance, MVE outperforms the basis classifiers and achieves higher results with accuracy of 78.70% and 93.94%, respectively, in the Stanford-1K and Sanders datasets when the base classifiers perform well. In contrast to SVM and NB findings, LR obtains a relatively low accuracy of around 75.37% in the HCR dataset. We can observe that MVE attempts to recover from such a decline and attains an accuracy of around 84.75%, which is little less than the NB classifier's maximum accuracy of roughly 85.09%.

These findings show that when the outputs of these classifiers are close together, MVE performs best, surpassing even the individual classifiers in accuracy. When one classifier performs poorly, MVE attempts to recover from such performance and attains favourable outcomes that are really close to the greatest ones.

5 Conclusions and Future Work

We presented an effective Twitter sentiment analysis method in this study. The suggested system represented input labeled tweets with various feature sets using various approaches. Early pruning is done on the unimportant and unnecessary characteristics utilizing. The feature selection method known as Information Gain (IG). To carry out the classification job, which is in charge of identifying the output sentiment polarity, the classifier ensemble is constructed from a diverse group of basic classifiers. The most popular tweet datasets were utilized in several trials to examine the suggested system's performance in various areas.

The three primary research topics in this paper were addressed by the experimental findings. First, the accuracy of the ensemble model and individual classifiers was increased by an average of 15% with the use of the IG feature selection approach. Secondly, the ensemble model attempted to integrate the basic classifiers' performance, however if one classifier wasn't appropriate for the dataset being utilized, the results may be impacted. Third, we can see that when combined with the BoW features, the reported outcomes of the lexicon-based and PoS features improved the classifiers' accuracy. However, there were not as many additions to emoticon-based functionality.

We could modify the feature extraction and classification processes in the future to effectively identify "neutral" tweets and incorporate them into the suggested system. Additionally, the suggested system might be modified to be a multilingual system in order to enable tweets in other languages, for Arabic [27].

References

1. Gaur, M., Pruthi, J.: A survey on sentiment analysis and opinion mining. Int. J. Curr. Eng. Technol. 7(2), 444–446 (2017)
2. Li, Y.-M., Li, T.-Y.: Deriving market intelligence from microblogs. Decis. Support Syst. 55(1), 206–217 (2013)
3. Rui, H., Liu, Y., Whinston, A.: Whose and what chatter matters? The effect of tweets on movie sales. Decis. Support Syst. 55(4), 863–870 (2013)
4. Kang, D., Park, Y.: Review-based measurement of customer satisfaction in mobile service: sentiment analysis and VIKOR approach. Expert Syst. Appl. 41(4), 1041–1050 (2014) Efficient Twitter Sentiment Analysis System M. M. Fouad et al.
5. Hassan, A., Abbasi, A., Zeng, D.: Twitter sentiment analysis: a bootstrap ensemble framework. In: The International Conference on Social Computing (SocialCom), Alexandria, VA (2013)
6. Manning, C., Raghvan, P., Schutze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
7. Lin, J., Kolcz, A.: Large-scale machine learning at twitter. In: The International Conference on Management of Data (SIGMOD 2012), New York, NY, USA (2012)
8. Vinodhini, G., Chandrasekaran, R.: Sentiment classification using principal component analysis based neural network model. In: The International Conference on Information Communication and Embedded Systems (ICICES 2014), Chennai, India (2014)
9. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches and

applications. Knowl. Based Syst. 89, 14–46 (2015)

10. Kharde, V., Sonawane, S.: Sentiment analysis of twitter data: a survey of techniques. Int. J.Comput. Appl. 139(11), 5–15 (2016)
11. Ghiassi, M., Skinner, J., Zimbra, D.: Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert Syst. Appl. 40, 6266–6282 (2013)
12. Agarwal, B., Mittal, N.: Prominent Feature Extraction for Sentiment Analysis. Socio-Affective Computing Series. Springer International Publishing (2016)
13. Agrawal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, P.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media (LSM 2011),Stroudsburg, PA, USA (2011)
14. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining lexicon-based and learningbased methods for twitter sentiment analysis. HP Laboratories (2011)
15. Mohammad, S., Kiritchenko, S., Zhu, X.: NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. In: Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013), Atlanta, GA, USA (2013)
16. Saif, H., He, Y., Alani, H.: Semantic sentiment analysis of twitter. In: Proceedings of the 11thInternational Conference on the Semantic Web (ISWC 2012), Berlin, Heidelberg (2012)
17. Kuncheva, L.: Combining Pattern Classifiers: Methods and Algorithms, 2nd edn. Wiley, New York (2014)
18. Fumera, G., Roli, F.: A theoretical and experimental analysis of linear combiners for multiple classifier systems. IEEE Trans. Pattern Anal. Mach. Intell. 27(6), 942–956 (2005)
19. Lin, J., Kolsz, A.: Large-scale machine learning at twitter. In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD 2012), New York,NY, USA (2012)
20. Rodríguez-Penagos, C., Atserias, J., Codina-Filba, J., Garcıa-Narbona, D., Grivolla, J., Lambert, P., Saur, R.: FBM: combining lexicon-based ML and heuristics for social media polarities. In: Proceedings of the Seventh International Workshop on Semantic Evaluation, Atlanta, GA, USA (2013)
21. Clark, S., Wicentwoski, R.: SwatCS: combining simple classifiers with estimated accuracy. In: Proceedings of the Seventh International Workshop on Semantic Evaluation, Atlanta, GA, USA (2013)
22. Da Silva, N., Hruschka, E., Hruschka Jr., E.: Tweet sentiment analysis with classifier ensembles. Decis. Support Syst. 66, 170–179 (2014)
23. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: 14th International World Wide Web Conference, Chiba, Japan (2005)
24. Stanford Twitter Sentiment Corpus. http://help.sentiment140.com/for-students. Accessed May 2017
25. Sanders Dataset. http://www.sananalytics.com/lab/. Accessed May 2017
26. Speriosu, M., Sudan, N., Upadhyay, S., Baldridge, J.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the First Workshop on Unsupervised Learning in NLP (EMNLP 2011), Stroudsburg, PA, USA (2011)
27. Mostafa, A.M.: An evaluation of sentiment analysis and classification algorithms for Arabic textual data. Int. J. Comput. Appl. 158(2), 29–36 (2017)