ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2020.0308

SARCASM DETECTION ON TWITTER DATA USING SUPPORT VECTOR MACHINE

Ashima Garg and Neelam Duhan
Department of Computer Science and Engineering, J.C. Bose University of Science and Technology, India

Abstract

Sarcasm can change the polarity of a sentence and it becomes the opposite. While sentiment analysis on social media has been widely used, but it is still rare to find sentiments and analyze them, considering the detection of sarcasm in it. Sarcasm detection in sentiment analysis is a challenging task. After successful identification of sarcasm the quality of sentiment analysis improves drastically. Experiments about sentiment analysis by detection of sarcasm are more often found in the language used in context with some special words. Therefore, taking into account research done on English tweets, this study analyzes the sentiment analysis sarcasm in Tweets agreed within context (specific topic) using the interjection and unigram features as features The main task is to detect sarcastic sentences and compare using classification methods namely Support Vector Machine with polynomial kernels. Thereafter incorporating interjection feature words that were expressing one's feelings and intentions and the unigram feature which is a collection of words a single obtained from the corpus automatically. Results of experiments show that the use of interjection features and unigram as detection of sarcasm in tweets using SVM will enhance the accuracy by 91%.

Keywords:
Sarcasm, Sentiment Analysis, Twitter, SVM

1. INTRODUCTION

In this era of globalization, the use of social media is increasing. People use social media as a means of freely expressing their aspirations and opinions on the internet. One of the social media that is often used to express opinions is Twitter. Twitter is considered as a social media that is easy to use and very fast in spreading information. Now the use of Twitter is not only limited to personal interests but is often closely related to several other topics such as political or business interests. Companies or related institutions need opinions or aspirations from Twitter users in general for certain purposes, one of them is by conducting sentiment analysis of these public opinions.

Sentiment analysis is computational research of opinions, sentiments, and emotions that are given textually to determine whether the test is positive or negative. In sarcastic sentence, an event that happened was different where the sentence sarcasm generally actually had a positive meaning but described the opposite situation. Based on the results of research conducted by [2] during the US presidential election, 11% of Twitter users who were active in the presidential election issue used the phrase sarcasm in expressing their opinions. This proves that the use of sarcasm messages on Twitter social media is still often found. According to [3] states that the sentence sarcasm on the topic of food, life, and health is rarely found, however, on the topic of government, brand, or politics the use of sarcastic sentences is often found. Samer [5] stated that sarcasm is used by people because it is considered as a form of conveying emotions with the purpose of entertainment, so the use of sarcasm looks not too serious but the emotion to be conveyed remains implied as focused in [4]. Determining sentiments in the sentence sarcasm is still something that is difficult to do in text processing, even by humans.

Sarcasm is something that is intended to insinuate or offend someone or something. This uses words that are inverted from their intent and have a non-standard writing structure, making it difficult to detect. This is confirmed by the statement of [5] that, characters in sarcastic sentences contrast with each other. If there is sarcasm in the data the results of sentiment analysis will mislead us as there will be ambiguity, and it will affect the results of accuracy and quality as focused by [4] and [6].

Therefore, it is necessary that sarcasm detection is carried out in the analysis process to infer correct sentiments, thus improving accuracy. In this study, the detection of sarcasm is proposed based on text classification with supervised approach learning. The main features used in this study are unigram and interjection features with a method the classifier using Support Vector Machine. While the interjection features it is an important feature in the process of classification of sarcasm sentences by [3].

The limitations of the problem in this study that dataset used are the result of crawling based on the different keywords or contexts to narrow it down the scope of the dataset used so that results analysis will be more leverage. In addition, Polarity results Sentiment analysis is classified into two categories that is sarcasm and positive/negative.

The objectives to be achieved in this research are, to analyze the effect of interjection and unigram features in the classification of sarcasm sentences used in the English language only.

This proposed research completes previous studies by [3], wherein this research interjection features combined with features unigram. These two features are combined so that you can identify sarcastic sentences better, thus potentially giving better results compared to sarcastic sentence classification research.

2. RELATED WORKS

In recent years, sentiment analysis on social media has caught the attention of many researchers. But, along with doing this research, there are many factors that make analyzing the results of sentiment less accurate in its processing. Researchers of [2] focused that one of the factors that challenges sentiment analysis in social media is the use of sarcasm on delivering opinions on social media. Recorded 11% of Twitter users use the language sarcasm in expressing his opinion related to the issue United States presidential election at that time. This is too proven by authors of [3] that, the use of sarcasm on sensitive topics such as politics, brand, and the government is more often found. From his observations of 100 tweets on the topic of food, life, and health there are only 2 tweets indicated sarcasm. Whereas for 100 topic tweets government, brand and politics found 18 tweets sarcasm. Sarcasm itself is a topic of learning from the branch of Psychology, and is declared still difficult to be identified by humans because it has no definite or standard structure. Researchers of [4] categorized the use of sarcasm in 3 ways, namely: (i) Sarcasm with the aim of entertaining, (ii) Sarcasm as a form of delivering frustration or anger, and iii) Sarcasm with a view to avoiding giving clear answers to a question. Along with the growing use of sarcasm on social media, some analytical research sentiment with the detection of sarcasm also began much developed. According to [5] there are 4 characters using sarcasm found namely: i) Sarcasm as a contrasting sentiment, (ii) Sarcasm as the delivery of expression through complex writing, (iii) Sarcasm as a goal delivery of emotions/feelings, and (iv) Sarcasm as written form of expression. Researchers of [10] proposed 4 sets of feature extractions used to detect sarcasm. The first feature is the feature sentiment-relate is used to detect sarcasm.

Bo Pang and Lillian Lee [14] wrote a book that offers a comprehensive idea of the exploration in the concerned area of sentiment analysis. The initial polarity classification of assessments using supervised tactics. The techniques which were explored are backing up Vector Machines (SVMs), Naive Bayes, and Maximum Entropy classifiers; this study used data sets with a different set of functions, for instance, unigrams, bigrams, binary and term frequency feature weights and others. The outcome of their observation was that sentiment classification is not that easy than regular topic-based classification they also concluded that using an SVM classifier with binary unigram based structures generates the best output. A succeeding advancement was the identification and deduction of the neutral portions of documents and the implementation of a polarity classifier on the remaining. This helped to achieve text soundness with contiguous text lengths which were expected to belong to the same subjectivity or objectivity class. Documents were portrayed as charts/graphs with sentences as nodes and attached scores in between as edges. Two supplementary points characterized the subjective and objective nodes. The weights between the two nodes were derived using three various, empirical decomposing tasks. Identifying a partition that reduces the cost function splits the objective from the subjective verdicts. They stated a statistically important enhancement over Naive Bayes standard using the complete text however, with only very minute hike as compared to using a SVM classifier on the overall document.

Researchers of [1] used SVMs and prolonged the feature set for demonstrating the documents with favorable methods from a range of various sources. They hosted structures based on Osgood's Theory of Semantic Differentiation (Osgood, 1967) using Word Net to develop the values of effectiveness, motion, and evaluative of adjectives and Turney's semantic coordination [16]. Their conclusions showcased that using a hybrid SVM classifier, which uses as features the distance of documents from the splitting hyperplane, with all the stated features yields the superlative outcomes.

3. RESEARCH METHODS

Analysis of Sentiments: This study was conducted by classifying tweet sentences in the class of sarcasm or non-sarcasm. The classification process is carried out in several stages, namely data collection and labeling, pre-processing, feature extraction, sentiment classification, performance calculation, and evaluation. Data Collection and Labeling sarcasm; the dataset obtained is then labeled in the sarcasm or non-sarcasm class. From the results of the labeling process, 3892 Tweets were obtained in which 1945 sentences labeled sarcasm and 1947 sentences labeled non-sarcasm. Examples of the tweet dataset obtained can be seen in Table.1.

Table.1. Examples of the tweet dataset obtained

Sl. No	Tweet
1	I think hunger deaths have recorded more than corona deaths in the world #Lockdown
2	Thanks to our incredible Doctors, nurses, police, They are fighting very bravely in this pandemic, Behind the mask they are saving lives in silence
3	Is anyone else feeling like they have been trapped with their mobile phone ever since quarantine started?
4	Lockdown has actually taught me one thing that you can actually be a 'responsible citizen' by being lazy
5	Another master stroke in robbing the middle class is by reducing contribution
6	Corona is on world tour, it will come to your door very soon

Pre-Processing: First of all Tweet dataset is pre-processed. This step is done so that the dataset becomes cleaner and makes the data more efficient for use in subsequent processes, especially in the feature extraction process. The processes to be carried out in this stage include: Conversion of all tweets to lowercase, change two or more spaces to one space only, delete URLs on Tweets, delete symbols on tweets, delete special characters on Twitter namely Hash-tags, Throw words that are unimportant (Stop word removal and change the word affixes to basic words (Lemmatization). The Table.2 is an example of the results of pre-processing that has been done.

Table.2. Raw and Pre-Processes Tweets

Raw Tweets	After Pre-Processing
I think hunger deaths have recorded more than corona deaths in the world #Lockdown	think hunger deaths recorded more corona deaths world
Thanks to our incredible Doctors, nurses, police, They are fighting very bravely in this pandemic, Behind the mask they are saving lives in silence	Thanks incredible Doctors nurses police fighting bravely pandemic mask saving lives silence
Is anyone else feeling like they have been trapped with their mobile phone ever since quarantine started?	feeling trapped mobile phone quarantine started
Lockdown has actually taught me one thing that you can actually be a 'responsible citizen' by being lazy	lockdown actually taught thing responsible citizen lazy
Another master stroke in robbing the middle class is by reducing contribution from 12 % to 10%	Another master stroke robbing middle class reducing PF contribution 12% 10%
Corona is on world tour, it will come to your door very soon	corona world tour door soon

Feature Extraction: Feature extraction was carried out in this study which is the extraction of unigram features and interjection features. Feature Unigram is widely used in research on classification text and can adapt to the corpus used. The unigram feature extraction process is done with cutting Twitter's text token that produces a collection of single and unique words that compiles the dataset with segregation of complexity. The unigram feature extraction process results will generally produce a collection of words on a scale of the big one. An interjection is expressing words a person's feelings and intentions, such as ‘wow’ and ‘Nah!’, or symbolize a mock sound, for example, ‘ouch'. This form usually cannot be affixed or not has syntactic support with other forms. An interjection can also be grouped based on certain characteristics such as irritation interjection, interjection admiration, and others. According to research conducted by [10], the use of the word interjection in a sentence is one of the characteristics of sarcasm. The form of interjection word features used in this study can be seen in Table No. 3. These words are obtained manually based on an analysis of the corpus of data and other information in general.

Table.3. Types of Interjection Features and its Words

Type of Interjection	Word load
detestable interjection	bah, heck, nut, sick
interjection admiration	Ouch, oh, cool, wow, wow
thanksgiving interjection	thank God, fortunately, thank God
interjection of hope	hopefully, hopefully
amazed interjection	ouch, ouch, aye, lol, please, uh, oh, ah
shock interjection	oh dear, really, crazy, are you nut
call interjection	hey, hey, hey, uh, hello, olla
swear interjection	basically, stupid

Sentiment Classification: In this research, the classification process is carried out using Support Vector Machine (SVM). This method uses learning based supervised learning by finding hyper-plane with maximum margin. The aim is to optimize the hyper-plane to be able to classify data accurately. Concept of classification with support vector machine is looking for the best hyper-plane that functions as a separator data class. Support vector machine can work on a high-dimensional dataset using kernel tricks. Support vector machines works on only selected data points that contribute (Support vector) to form the model that will be used in the classification process. The type of kernel used in this study is polynomial. Illustration of the SVM model for class 2 classification is as follows.

Suppose two classes -1 and +1 assumed to be perfectly separated by dimensionless hyper-plane, which is defined in the equation below:

$$\bar{w}\cdot\bar{x}+b=0 \tag{1}$$

A data $\bar{x}$ will be classified as class -1 if it meets inequality and vice versa $\bar{x}$ will be classified as class +1 if it satisfies inequality as:-

$$\bar{w}\cdot\bar{x}+b<0 \tag{2}$$

$$\bar{w}\cdot\bar{x}+b>0 \tag{3}$$

In this case $\bar{w}$ is a weight vector and b is a bias. SVM will look for hyperplane equations that will minimize prediction errors. After getting the value b, the next step is to search hyper-plane to positive/negative classes and use sarcastic class equation as under:

$$\tilde{W}=\sum_i \alpha_i \tilde{S}_1 \tag{4}$$

In this case $\tilde{W}$ represents corresponding feature, $\sum_i \alpha_i$ is the infinite features mapping, and $\tilde{S}_1$ is the support vectors. The following is an illustration of the use of the support vector machine method in separating relevant and irrelevant data.

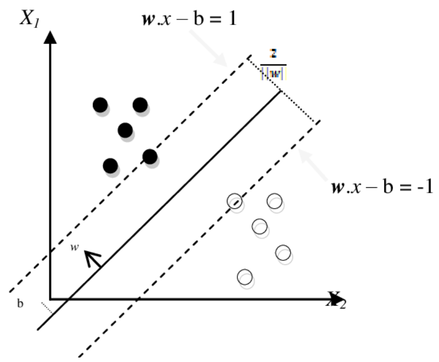

Fig.1. Separating relevant and irrelevant data

Evaluation and Validation: Performance evaluation is done to test the results of classification by measuring the truth value of the system. The statistical measure used to measure the value of system performance is accuracy. Accuracy is the percentage of texts that have been classified correctly by the system. Accuracy is obtained from the calculation results shown in Equation below.

$$Accuracy = (TP+TN)/(TP+FP+TN+FN) \tag{5}$$

where:

TP is the True Positive; is positive data which detected positive.

TN is the True Negative; is the amount of negative data that is detected negatively.

FP is the False Positive; is negative data however detected as positive data.

FN is the False Negative; is positive data however detected as negative data.

Cross-validation (CV): CV is a statistical method can be used to evaluate the performance of a model or algorithm where data is separated into two subsets namely training data and testing data. Model or algorithm trained by a subset of learning and validated by a subset validation. Furthermore the selection of CV types can be based on the size of the dataset. Usually a *K*-fold CV is used because it can reduce computing time steadily maintain the accuracy of the estimate. In this study *K*-fold used is 10-fold, where data is shared into 10 almost identical parts, so we have 10 data subsets to evaluate the performance of the classifier model. For each of the 10 data subsets accordingly, the CV will use 9 fold (90%) for training and 1 fold (10%) for testing like illustrated in Fig.2. The shaded part is dark, is a part of the dataset used as test data, while the part that is not shaded is as data practice it.

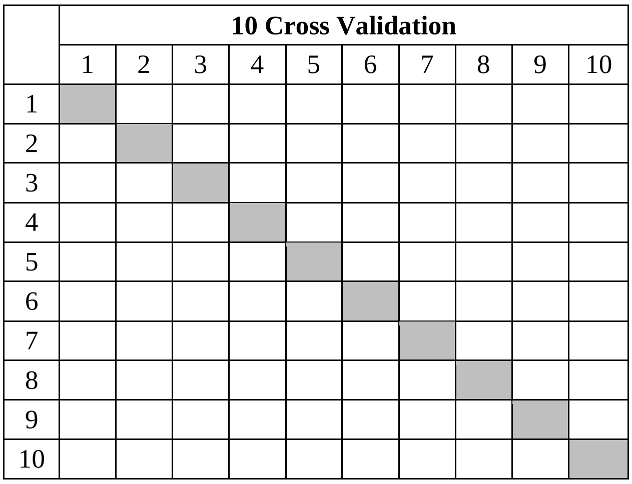

Fig.2. Distribution of training sets and testing sets in 10-cross validation

4. RESULTS AND DISCUSSION

Classification with Support Vector Machine: Support Vector Machine or SVM is a technique which is good in classification by finding hyper plane that can maximize margins between data classes. Hyper plane is useful in separating 2 classes +1 (sarcastic) and class –1 (positive/negative) groups where each class has a pattern. In making decisions with SVM method is used the kernel function $K(x_i,x_d)$. The kernel to use with the research shown in Equation below:

$$K\left(\bar{w}\cdot\bar{x},x_d\right)=\left(X_i^T X_j+C\right),\gamma>0 \tag{6}$$

where, K is the Polynomial Kernel, $\bar{w}$ is the set of weight vectors, $\bar{x}$ is the input vector, C is the margin space and y is the slack variable.

Processing is done on training data Sequential training algorithm is used because it is a simple algorithm without takes a lot of time with calculation stage. The classification process will determine the class of a tweet based on the frequency of occurrence of words from the previous process. As for classification the steps are as follows:

Step 1: Calculate the value of prior probability, at this stage the training data will be calculated with prior probability using the formula:

$$P(c)=\frac{S_C}{P'} \tag{7}$$

Table.4. Example of Conditional Probability Calculation

Class	Prior Probability
Sarcastic	12/101
Happy	12/101
Angry	12/101
Sad	12/101
Fear	12/101

In Table.4 there are five major classes which defined with a prior probability of 0.20 as the weight to classify the initial hyper-plane model for the SVM classifier. Thereinafter, further based on occurrences with other words embedded or marked in tweets the weight based on words will be evaluated which will escalate the position of hyper-plane vide sentiments in tweets.

Step 2: Calculate the value of conditional probability, at this stage the conditional probability of words in each will be calculated using Hyper-plane based on kernel function.

Table.5. Example of Conditional Probability Calculation

Word	Conditional Probability Term in the Class				
(Term)	Sarcastic	Happy	Angry	Sad	Fear
Incredible	10.01667	25.03175	35.51695	45.5137	45.5164
Proof	15.51667	25.51587	35.51695	45.5137	45.5328
Bravely	15.51667	25.51587	35.51695	45.5274	45.5164
The Doctor	15.51667	25.51587	35.51695	45.5274	45.5164
Mask	15.55	25.51587	35.51695	45.5137	45.5164
Awesome	15.51667	25.51587	35.51695	45.5274	45.5164
Stroke	15.51667	25.51587	35.51695	45.5137	45.5164
Support	15.51667	25.53175	35.51695	45.5137	45.5164
Figure	15.51667	25.51587	35.5339	45.5274	5.51639
Reducing	15.51667	25.51587	35.51695	45.5137	5.53279
Lockdown	15.51667	25.51587	35.51695	45.5137	5.53279
Actually	15.51667	25.51587	35.51695	45.5274	5.51639
Results	15.51667	25.51587	35.51695	45.5274	45.5164
Responsible	15.51667	25.51587	35.5339	45.5137	45.5164
Sincere	415.53333	455.51587	35.51695	45.5137	5.51639
Street	15.51639	25.51587	35.51695	45.5137	5.51639
Lazy	15.53333	25.53175	35.5339	45.5274	5.53279
Excellent	15.53333	25.51587	35.51695	45.5137	5.51639
Service	15.51667	25.51587	35.51695	45.5274	5.51639
Robbing	15.53333	25.51587	35.51695	45.5137	5.51639
Contribution	15.51667	25.51587	35.51695	45.5274	5.51639
Behavior	15.51667	25.51587	35.51695	45.5274	445.516
Citizen	15.51667	25.51587	35.5339	45.5137	45.5164
People	15.51667	25.51587	35.51695	45.5274	4445.52
Eat	15.51667	25.54762	35.51695	45.5137	445.516
Angry	15.51667	25.51587	35.55585	45.5137	45.5164
Let	15.55	25.53175	35.5339	45.5137	4445.52
Win	15.51667	25.53175	35.51695	45.5137	445.516
Enemy	15.51667	25.51587	35.51695	45.5274	45.5164
Hell	15.51667	25.51587	35.51695	45.5137	4445.53
Morning	15.51667	25.53175	35.51695	45.5137	445.516
Short	15.51667	25.51587	35.51695	45.5274	45.5164
Full	15.51667	25.51587	35.51695	45.5274	445.516
Period	15.51667	25.51587	35.51695	45.5137	45.5328
Exactly	15.51667	25.51587	35.51695	45.5274	45.5164
Prejudice	15.51667	25.51587	35.51695	45.5274	4445.52
Doubt	15.53333	25.51587	35.51695	45.5137	45.5164
People	15.53333	25.53175	35.51695	45.5137	45.5164

Restless	15.51667	25.51587	35.5339	45.5137	45.5328
Sibling	15.51667	25.53175	35.51695	45.5137	45.5164
Fearless	15.51667	25.51587	35.51695	45.5274	45.5164
Sad	15.51667	25.51587	35.51695	45.5411	45.5164
Happy	15.51667	25.54762	35.51695	45.5137	45.5164
Attack	15.51667	25.51587	35.51695	45.5274	445.516
Noon	15.51667	25.53175	35.51695	45.5137	45.5164
Figure	15.53333	25.51587	35.51695	45.5137	45.5164
Steadfast	15.51667	25.51587	35.5339	45.5137	45.5164
Afraid	15.51667	25.51587	35.51695	45.5137	45.582
Please	15.51667	25.51587	35.51695	45.5274	45.5164
Sincere	15.53333	25.51587	35.51695	45.5137	45.5164

The Table.5 the random weights are assigned to each word under 5 classes respectively. Consequently, while classifying the tweet using SVM the hyper-plane over the kernel method will shift the tweet into the class category based on sentiments using conditional probability and in conjunction with polarity.

Therefore on the basis of the above scenario using SVM classification the below result is drawn in the table 7 for ready reference.

Table.6. Results achieved using SVM

Tweet	Pre-Processing	Classification
I think hunger deaths have recorded more than corona deaths in the world #Lockdown	think hunger deaths recorded more corona deaths world	Negative
Thanks to our incredible Doctors, nurses, police, They are fighting very bravely in this pandemic, Behind the mask they are saving lives in silence	Thanks incredible Doctors nurses police fighting bravely pandemic mask saving lives silence	Positive
Is anyone else feeling like they have been trapped with their mobile phone ever since quarantine started ?	feeling trapped mobile phone quarantine started	Negative
Lockdown has actually taught me one thing that you can actually be a 'responsible citizen' by being lazy	lockdown actually taught thing responsible citizen lazy	Sarcastic
Another master stroke in robbing the middle class is by reducing contribution	master stroke robbing middle class reducing PF contribution 12% 10%	Sarcastic
Corona is on world tour, it will come to your door very soon	corona world tour door soon	Sarcastic

In Table.6, the tweets are pre-processed using the technique namely stop words enabling the removal words which is meaningless thereinafter using conditional probability over kernel method vide SVM the tweets are classified in categories as sarcastic, negative or positive respectively.

Accuracy Test Results: From the Fig.4, we evaluated ~4250 tweets in 5 segments therefore using precision and recall we can calculate the accuracy value derived respectively as follows:

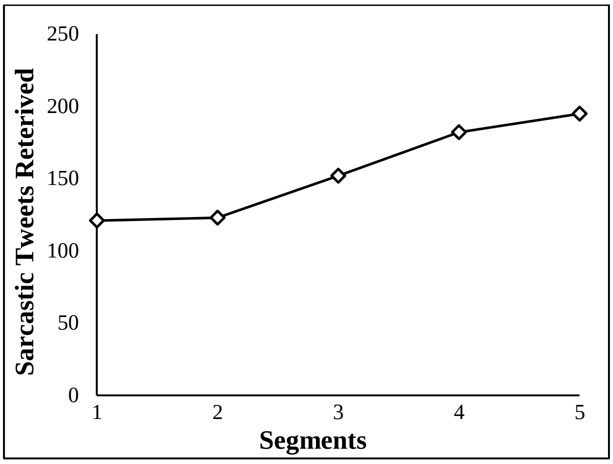

Fig.4. Accuracy Evaluation

5. CONCLUSION AND FUTURE SCOPE

In this research, we have tried to classify the sarcastic sentences in English conversation, supervised learning by using unigram and interjection. One of the things to be discussed is to find out the influence of the interjection and unigram features in the process of classifying sarcastic sentences that are approved by features using classifier methods, namely SVM with polynomial kernels. The results obtained indicate that the interjection and unigram features have a positive increase in classifying the classification of the twittering tone with sarcasm that can be accessed effectively using the unigram feature with interjections. Improvements obtained can be used in the classification method for sarcasm detection. The result obtained by using the interjection and unigram features for the SVM classification method reached a 90.94% accuracy vide K-Cross fold method. For future scope, the same scheme can be integrated with Big-Data over mammoth repositories.

REFERENCES

[1] Peter D. Turney and Patrick Pantel, "From Frequency to Meaning: Vector Space Models of Semantics", *Journal of Artificial Intelligence Research*, Vol. 37, pp. 141-188, 2010.

[2] B. Heredia, J. Prusa and T. Khoshgoftaar, Taghi, "Exploring the Effectiveness of Twitter at Polling the United States 2016 Presidential Election", *Proceedings of IEEE 3rd International Conference on Collaboration and Internet Computing*, pp. 283-290, 2017.

[3] M. Khatoon, W. Mehjabin and S. Chinthamani, "*Sentiment Analysis on Tweets*", IGI Global Publisher, 2019.

[4] R. Gupta, J. Kumar and H. Agrawal, "A Statistical Approach for Sarcasm Detection Using Twitter Data", *Proceedings of*

4th International Conference on Intelligent Computing and Control Systems, pp. 633-638, 2020.

[5] S.M. Sarsam, H.A. Samarraie and A.I. Alzahrani, "Sarcasm Detection using Machine Learning Algorithms in Twitter: A Systematic Review", *International Journal of Market Research*, Vol. 62, No. 2, pp. 1-15, 2020.

[6] P. Dharwal, T. Choudhury and R. Mittal, "Automatic Sarcasm Detection using Feature Selection", *Proceedings of 3rd International Conference on Applied and Theoretical Computing and Communication Technology*, pp. 29-34, 2017.

[7] S. Saha, J. Yadav and P. Ranjan, "Proposed Approach for Sarcasm Detection in Twitter", *Indian Journal of Science and Technology*, Vol. 10, No. 25, pp. 1-8, 2017.

[8] D.K. Bharti and K.S. Babu, "Sarcasm-Detection-in-Twitter-Data -A-Supervised-Approach", *IGI Global Publisher*, 2018.

[9] T. Sariki, G. Kumar and N. Kambhatla, "Sarcasm Detection", *International Journal of Pharmacy and Technology*, Vol. 8, pp. 22163-22181, 2016.

[10] R. Ahuja and S. Sharma, "A Review Paper on Sarcasm Detection", *Proceedings of 3rd International Conference on Advances and Applications in Artificial Intelligence*, pp. 1-5, 2020.

[11] A. Rajadesingan, R. Zafarani and H. Liu, "Sarcasm Detection on Twitter: A Behavioral Modeling Approach", *Proceedings of ACM International Conference on Web Search and Data Mining*, pp. 97-106, 2015.

[12] Bala Durga Dharmavarapu and Jayanag Bayana, "Sarcasm Detection in Twitter using Sentiment Analysis", *International Journal of Recent Technology and Engineering*, Vol. 8, No. 4, pp. 1-13, 2019.

[13] S.K. Bharti, R. Naidu and K.S. Babu, "Hyperbolic Feature-based Sarcasm Detection in Tweets: A Machine Learning Approach", *Proceedings of IEEE International Conference on India Council*, pp. 1-6, 2017.

[14] Pang Bo and Lillian Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1-2, pp. 1-135, 2008.

[15] T. Mullen and N. Collier, "Sentiment Analysis using Support Vector Machines with Diverse Information Sources", *Proceedings of International Conference on Empirical Methods in Natural Language Processing*, pp. 412-418, 2004.