

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325843750>

# A COMPREHENSIVE STUDY ON SARCASM DETECTION TECHNIQUES IN SENTIMENT ANALYSIS

Research · June 2018

CITATION

1

READS

1,261

2 authors:



Dr. Vadivu G  
SRM University

38 PUBLICATIONS 49 CITATIONS

[SEE PROFILE](#)



Sindhu Chandra Sekharan  
SRM University

13 PUBLICATIONS 11 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Cloud Computing [View project](#)



cyber threat intelligence [View project](#)

## A COMPREHENSIVE STUDY ON SARCASM DETECTION TECHNIQUES IN SENTIMENT ANALYSIS

<sup>1</sup>Sindhu. C, <sup>2</sup>G.Vadivu, <sup>3</sup>Mandala Vishal Rao<sup>1,3</sup>Department of Computer Science and Engineering, <sup>2</sup>Department of Information Technology  
SRM Institute of Science and Technology

Chennai, India

<sup>1</sup>sindhucmaa@gmail.com

**Abstract:** During past few years, there has been a lot of increase in opinionated textual data in social media over the Internet. Sentiment Analysis is used to analyze the opinionated text. It helps us to understand (text) the emotion behind the writer. It is facing many challenges and Sarcasm detection is one of the major challenges in it. Sarcasm is the unconventional way of conveying a message which conflicts the context. It can lead to a state of ambiguity. Data pre-processing is one of the primal works implemented by many researchers. Many data preprocessing techniques such as tokenization, stemming and lemmatization, removal of stop words is done by many researchers. Several research works have been done on sarcasm detection. Many feature extraction techniques were implemented. Several classifiers are used in various researchers such as Support Vector Machine (SVM), Naïve Bayes, AdaBoost, Random Forest etc. Results are included in papers such as accuracy, precision, recall, F-score reflects how better the model is. This paper will brief various methodology and techniques used in sarcastic text detection for Sentiment Analysis.

**Index Terms:** Irony, Satire, Sentiment Analysis, Sarcasm detection, Gradient Boost.

### 1. Introduction

Natural Language Processing (NLP) is one of the important domains in artificial intelligence. It acts as a platform between the computer and human languages. It helps in making the machine understand, analyze and interpret the data. It helps in querying the datasets and provide an answer. It helps in not only understand the text or speech but also the context behind it. It works for structured and unstructured data. The linguistic structure depends on various factors like social context, regional dialects, slang etc. NLP is facing few challenges in this field. Sentiment Analysis is one of the important fields in NLP which deals with analyzing the context.

Sentiment Analysis is the process of analyzing the opinions expressed by the writer and determine the attitude towards the topic. It is used to classify the polarity of a document or an opinionated text. The

intensity of the text can also be further classified by Sentiment Analysis. Several analyses can be performed using sentiment analysis. These analytics can be used to determine and retrieve various levels of sentiment. The analysis process is examined on various individual entities i.e., by words or phrases in the document. It provides a quick understanding of the writer's attitudes. It is sometimes known as opinion mining where it speaks about a particular entity and discusses the feedbacks of it. Several data pre-processing and classification techniques used in Sentiment Analysis. Sentiment words convey a positive or a negative meaning. There are few key challenges faced by Sentiment Analysis such as Entity named recognition, Anaphora recognition, parsing, sarcasm detection and many others.

Many people these days express their opinions on various social websites. People have started to express their emotion in sarcasm. Sarcasm is one of the leading challenges faced in Sentiment Analysis. Sarcasm is an indirect manner of conveying a message. It is basically a bitter expression which is conveyed. Sarcasm can also reflect a state of ambivalence. It contradicts the meaning, in the context which is said. Sarcasm can be expressed in many ways. It can be expressed in speech and text. Sarcasm can be conveyed through various ways like a direct conversation, speech, text etc. In direct conversation, facial expression and body gestures provide the hint of sarcasm. In the speech, sarcasm can be inferred if there is any change in tone. In the text, it is difficult to identify sarcasm compared to other methods, but, it can be conveyed using a capital letter, excessive usage of exclamatory marks, exaggeration, usage of emoticons etc. It can be reflected using rating of stars by using a hyperbole and providing less number of stars. There are various applications of sarcastic text detection. It is used for letting the reviewer know the intent of the writer and the context in which it is said. Sarcasm is more predominant in the places where there are capital letters, emoticons, exclamation marks etc. Sarcasm detection is one of the important tasks in sentiment analysis. In Twitter, it helps to understand the intent behind a tweet. Twitter also acts as a tool for the prediction of Election results [23]. In Amazon and shopping websites, it helps to understand the review of

the product. In various social websites, like Facebook, Instagram etc. it helps to understand the opinionated comments. The consumer's preferences and opinions can be analyzed in order to understand the market behavior for a better consumer experience [12].

Though Irony, Sarcasm, and Satire look the same, there is a subtle difference between them. The Irony is referred as a gap between reality and expectations, Sarcasm is a state of mockery and satire is an exaggeration.

**Example: This question is making me go crazy.**

**In case of Sarcasm:** Consider a person, when a question is posed to him if he is not able to answer it, and don't even want the questioner to know that this person doesn't know the answer. Later, he replies that this question makes me go crazy. Which means, please don't ask me these kinds of questions.

**In case of Irony:** Consider a person who is very confident in answering any question posed to him. After the question is posed, he is unable to answer the question. Then he gives a reply, this question makes me go crazy which means that the question which is asked makes the person unanswerable. So, this question created a gap between the person's expectation and reality, which serves as Irony.

**In case of Satire:** Consider a person who is able to answer the questions without taking any stress. If a question is posed to him and it makes him think very deeply, then the person says This question is making me go crazy. Which indirectly means that he is taking a lot of stress in answering the question.

## 2. Related Work

A comment, tweet, review, feedback made by many people have got different characteristics. They depend on various aspects such as geographical area, current news, trending information, age, gender etc. [9]. Researchers agree that they are two types of irony namely verbal or sarcasm and situational irony [4]. A common syntactic feature is the Part of Speech (POS) tag which may be associated with words in a document. All the categories of the feature are likely to play a role with respect to the identification of features indicative of sarcasm [7]. Behavioral scientists and linguist's sarcasm is studied well [2]. Several Antonyms pairs are derived using reasoning rules in Serbian WordNet ontology. Antonyms pairs, Positive polarity, ordered sequence of sentiment tags, Irony markers and Parts of speech tagging are used [8]. [10] discusses the researcher's non-context related features and classifying them into a pre-defined set of classes including opinions, deals, events and private messages.

The ratio of emotional words is computed in [1]. Many researchers have classified words as positive, negative and neutral. The intensity of words and provided them a rating of 1-5 where 1 represents less

positive or less negative and 5 represents more positive or more negative [11], [25]. Emoticons can also reflect the nature of status. The basic intuition is that the orientation of sentiment words and emoticons are same when they occur in common. For orientation identification, there are two approaches namely corpora and dictionary-based approaches. Conjunctions are used to join the sentences when positive conjunctions like and is used when they give rise to a positive orientation if negative conjunctions like but are used then it gives rise to opposite orientation. A Network is constructed with synonyms in wordnet [11].

Tweets in twitter relating to movies are collected from different cities across different countries over a certain period of time. They are classified into 3 types namely negative sentiment, positive sentiment, and cognitive statement. Using UH-filter the meaningless tweets are removed. [12]. Emoticons are also one of the main sources for the formation of a sarcastic sentence. Emoticons are common and strong signals of sentiment expression. Sentiment polarity has to be considered [13]. It also discusses researchers saying that emoticons play an important role in training machine learning classifiers and constructing sentiment lexicon. Initially, word2vec algorithm is used which converts the text into a vector space, later k-means algorithm is used for clustering. A brief about emoticons and what do they mean is discussed in table-2 [13]. Due to the implementation of map-reduce, the model has attained time complexity and performance. The execution duration is reduced from 15 to 9 minutes [16]. Both Twitter and Amazon datasets are trained on punctuations, patterns enrich punctuation, enrich patterns, semi-supervised Identification algorithm. Star rating is also considered. [18].

Slang words are also used to analyze the sentiment. Classification of text into subjective and objective is done. Sentiment slang is identified using subjective sentences. The Polarity of the score is determined using weighted inverse document frequency [20]. There are 8 types of basic emotions classified namely joy, disgust, trust, surprise, anger, anticipation, sadness and fear, which are extracted from Sentiment Analysis and Social Cognition Engine Emotional Lexicon (SEANCE EmoLex) [6]. A hashtag tokenizer has been developed for GATE (A software which helps in solving various text processing). A new algorithm is developed. A token is formed and matched against Linux dictionary which is converted into GATE dictionary. The Viterbi-like algorithm is used for the best possible match. If there are the combination of matches without a gap, then they are converted into tokens and hashtag is removed [24].

Sarcasm can also be represented using numerical. For example, it is very comfortable to wake up at 4 in the morning. The above example can indirectly mean that it is not very comfortable waking up 4 in the morning [27].

Important

Some researchers have basically classified reviews into 8 classes and then mapped them to the sarcastic sentence to extract the patterns as discussed in table-1 [3]. By using emoticons, there is a good relationship between the perceived sentiment of the tweet and the implicit sentiment of an emoticon. It is showed the emoticon has an impact on the sentiment of the complete tweet. Two methods of lexicon generation approaches are used. The first method uses semantic

association with seed lexicon to derive word sentiment scores. The second method uses word frequencies in both positive and negative frequencies to iteratively refine seed lexicon [19]. Emoticons are used in various sarcastic statements are discussed in table-2 [13]. Table 3 discusses about the various paper's tasks, datasets, features extraction, Classifications used, and their results

**Table 1.** Classes of sarcastic text

Class No	Feature	Example
Class 1	Co-existence of positive and negative	It is not sweet at all, it is called delicate sweetness.
Class 2	Positive sentence followed by a negative sentence and vice versa	Sweet and delicious oranges. They are rotten in a delivery box.
Class 3	A dilemma in the sentence.	I was thinking whether to buy the product because price is (good/expensive).
Class 4	Positive phrase followed by a negative phrase and vice versa.	Delivery was good but the product was not.
Class 5	Comparison between bad and worse meaning in the situation.	I regret that I bought this product, but it's better than losing money again.
Class 6	Comparison with a better product.	I know more delicious oranges that are sold nearby.
Class 7	It implies a negative meaning to the target product in the review	If this were a disposable product, I would get satisfaction about it.
Class 8	No specific positive and negative point.	I can use it so so.

**Table 2.** Emoticons in sarcastic text

Clusters	Emoticons	Sample words from cluster
A	: ) :D =)	Happy birthday good thanks fantastic wonderful lovely amazing awesome
B	: ) :- ) :- ) :-D =D ; P =]	Smile friends favorite music heart kind positive coffee
C	: ( : / XD : ' ) : ' ( : - ( D : ; ( : - / : ! : \	Sorry sad miss hate shit ugly broke late sick
D	: P ; D : - P : ] : p	Can't lol never feel look what
E	( :	Please love iloveyou goodnight follow
F	XP	Shoot stuck fatally
G	DX	#music smartphone camera
H	8)	Party happiest weekend playing top happiness friday spring

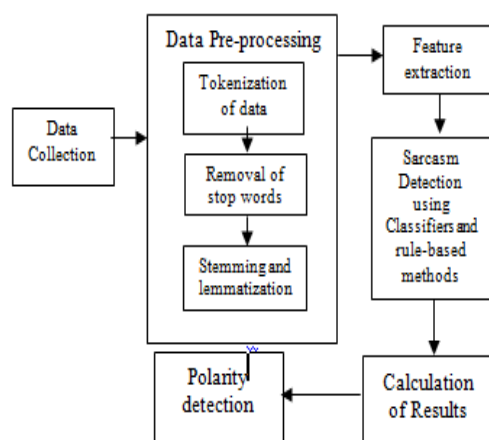
### 3. Identifying Sarcasm in Sentiment analysis

The general work flow in sarcastic text detection is as shown in figure 1.

#### A. Data Collection:

Data is retrieved by many researchers using #sarcasm in Twitter API [1], [5], [7]. Data is crawled from

amazon product reviews in [4]. A study was made on the usage of datasets. Dataset retrieved through Twitter API has the most number of implementations of their model. The Twitter dataset is used by most of them followed by Amazon. Amazon dataset is mainly classified on product attributes such as Product description, image, brand, price, information. Product ratings actually help us to identify the sarcasm for the text review provided.



**Figure 1.** General Workflow of sarcasm detection in sentiment analysis

### B. Data Pre-processing:

Pre-processing of data is majorly done in three steps. Step 1 – Tokenization of data. In this process, we tokenize the sentences into words. Step 2 – Stemming and lemmatization. In this process, the words are brought into present tense. Step 3- Removal of stop words. In this process, we remove the stop words which are not using the detection process example – Articles.

#### i. Tokenization of data:

In this process, after the data is retrieved from the dataset. The data which is taken is in the form of sentences and phrases. Now, these sentences and phrases are tokenized into words, so that it is easily understandable.

#### ii. Removal of stop words:

Stop words are the words that the search engine has programmed to ignore when both indexing and retrieving of entries. Articles are the best examples of stop words.

#### iii. Stemming and Lemmatization:

Stemming and lemmatization is the process of converting the words into their root words so that they can be analyzed as a single item.

#### iv. P-O-S tagging:

P-O-S tagging basically stands for parts- of -speech tagging. This is one of the most used in sarcasm detection. It helps us to classify the words into a verb, adverb, adjective, noun etc. So that when there is too much of sarcasm then that leads to a hint of the presence of sarcasm i.e., by providing too much of praise in the sentence [5], [25].

### C. Feature Extraction:

There are several techniques to extract features from text. In various papers sarcasm is hinted in different ways like when a positive sentence is followed by a

negative sentence, the positive phrase is followed by a negative phrase and vice-versa. Other cases like when there is a dilemma in a sentence, when there is a comparison between bad and worse situations, when there is a comparison with a better product, when it implies a negative meaning to the target in the product review, when there is neither a specific positive nor specific negative point, in all these cases, it hints to be a sarcastic text, as discussed in table-1. Furthermore, when a part of the text in the review is mentioned in capital letters, when there is an excessive use of exclamation marks, usage of emoticons, when there is too much of exaggeration which turns out to be a hyperbole, providing a less star rating and a good review.

Several features are retrieved from the opinionated text. **Sarcasm as a wit**, it discusses the exclamation marks used in the text. **Sarcasm as a whimper**, it discusses the exaggeration of a sentence. **Sarcasm as avoidance**, in this the text is of complicated sentences, uncommon words, and unusual expression [1].

#### 1. Term – Frequency:

Term-frequency determines the number of times a particular word appears in the document [15].

#### 2. Feature Presence:

Feature presence specifies whether a particular feature appears in the document or not. 0 indicates the absence of the feature and 1 indicated the presence of the feature in the document [15].

#### 3. Term Frequency–Inverse Document Frequency (TF-IDF):

TF-IDF is the frequency of occurrence of a term in the document [15]. It is used as frequency measure of feature [2]. The required features are extracted and vectorized using TF-IDF [7].

#### 4. Weighted Term Frequency-Inverse Document Frequency:

It is used in the calculation of sentiment score using slang words with the help of weighted TF-IDF is done. If the score is greater than zero then a positive opinion can be expected, if it is less than zero then the negative opinion is expected [20]. TF-IDF scores are obtained by weighting scheme on each EmoNgram [6].

#### 5. Delta Term Frequency-Inverse Document Frequency(TF-IDF):

It makes an easy computation, understanding, and implementation. SVM's are used to show delta IF-IDF increases the accuracy of sentiment analysis. It works on both sentiment polarity and subjectivity detection. It can also work on various sizes of the document [26]. The weight of the slang score is efficiently based on delta TF-IDF.

#### 6. N-gram:

N-gram is a contiguous sequence of tokens in computational linguistics and probability. N-gram is basically a sequence, the sequence is as a unigram, bigram, trigram etc. It deals with having common words before a particular word. If there is one common word

important

then it is called a unigram, if there are two common words then it is called bigram and so on. Labels are assigned to unigram [19]. The possibility of n-gram combination is considered using n-gram.

7. *Word2vec:*

Word2vec produces a vector space using the large input of text. To define the representation of words including emoticons Word2vec is used [13].

8. *Pattern-Related:*

The Supervised learning method is which leans towards sarcastic patterns based on parts-of-speech. If the pattern of high frequency words. Length of the pattern is also taken into consideration [1]. Generalized patterns are based on three situations namely exact match, no match and partial overlap [21].

**D. Sarcasm detection using classifiers and rule-based methods:**

Several machine learning classifiers such as SVM, Naïve Bayes, AdaBoost, gradient boosting, Random forest etc. are used. Several Rule-based methods are also used. It is discussed in detail in Section IV.

**E. Polarity Detection:**

During this process, the polarity of the review or statement is identified and labeled it as sarcastic or not.

**F. Calculation of Results:**

In the process calculation of results i.e., the accuracy, precision, recall, and F-score are done.

**4. Classifiers, Clustering and Rule based methods**

**A. Classifiers Used:**

Several classification techniques are used by many researchers during their experiments.

*i. Naïve Bayes:*

Naïve Bayes computes the posterior probability of class based on the distribution of word regardless of their position in document [2], [6], [7]. Twitter streaming API and a

**Table 3.** Survey on papers

Paper	Task	Dataset	Feature Extraction	Classifications used	Result
[1]	To find the effectiveness of the model	Retrieved from Twitter API	Sarcasm as a wit, Sarcasm as a whimper, Sarcasm as an avoidance	Sentiment related features, punctuation related features, lexical and syntactic features, pattern related features.	Accuracy – 83.1%, precision – 91.1%, Recall – 73.4%, F-Score – 81.3 %
[2]	To find the accuracy of proposed model.	online review sites, media sites, and microblogging sites	Term presence, Term Frequency, Term Frequency–Inverse Document Frequency (TF-IDF).	Naive Bayes Classifier, Maximum Entropy, SVM, 10 cross validation	Accuracy - 50%
[3]	Identify sarcasm	sarcasm-labeled corpus and features based on punctuations such as “!” and “?”	bag-of-words.	N gram, Boosting rules and rejection rules	F-score Baseline - 0.012, F-score Their method - 0.053
[4]	Relate comments with ratings	Twitter and Amazon	Emoticons, heavy punctuation marks, quotation marks, positive interjections	semi-automatic procedure	Presence of irony does not affect readers.

[5]	To find the better classifier among the used.	Twitter Streaming API	Word tokenization, POS tagging, stemming, lemmatization	Random forest, decision tree, logistic regression, Naïve Bayes, gradient boost.	Gradient boost provides highest accuracy percent
[6]	To find the better classifier.	News article dataset	Emotion, sentiment, bag of sorted emotion, SentiNet, TF-IDF.	Regression, SVM, Naïve Bayes, Bagging, AdaBoost, Random Forest, ExtraTree, Gradient Boosting.	Random forest - 0.724
[7]	Find feature that gives a better result	original bilingual corpus	Lexical, Pragmatic, Prosodic, Syntactic, Idiosyncratic	N - grams, SVM	syntactic category F-measure = 0.847

predefined training set to classify tweets which evaluate on Naïve Bayes [5], [22].

#### ii. Random Forest:

Using Random Forest researchers have checked the performances of classification of each set of features apart [1], [6]. Random Forest Classifier works by building multiple decision trees and obtaining class labels in [5].

#### iii. Decision Tree:

Functional Trees maximized the univariate and multivariate form of decision trees with a linear function as used in constructive induction learning in [2], [7].

#### iv. AdaBoost:

AdaBoost is one of the meta-algorithm in machine learning in machine learning. It is also known as adaptive boosting. It is used by the researchers in [5], [6].

#### v. Gradient Boost:

Gradient boosting is a machine learning technique which is being used for classification and regression problems. It is used to ensemble weak prediction models. It is used by researchers in [6].

#### vi. Support Vector Machines (SVM):

SVM is used to recognize satire articles [1]. Non-linear SVM was used as the classification model because it had been shown to perform well in the context of similar domains found in the literature [6], [7], [22].

#### vii. Maximum Entropy:

Maximum entropy is one of the classification techniques which aims to attain uniformity. Maximum entropy classifier is used [12], [22].

#### viii. Cross-Validation:

The concept of cross-validation is to make the sample of data into k equal subsamples of data for training and testing it. If the sample is made into k equal parts, then it is called k-fold cross-validation. 5-fold cross validation for both training and testing classifiers [13]. 5-fold cross validation scores are calculated on SentiSense dataset [14]. Experiments are carried out using 10-fold cross-validation [7].

### B. Clustering:

#### i. K-means Clustering:

To understand the exact meaning of emoticon, k-means clustering is used [13]. [14]. K-means clustering is done with k=5 [25].

### C. Rule-Based Methods

Rule-based methods are also used as a part of sarcasm detection. Rule-based methods include Semantic, Syntactic, Pragmatic, Prosodic etc.

#### i. Lexical Method:

In lexical method object, action and characteristics are reduced to lexical features such as noun, verb, adjective respectively. English dictionaries are used to correct the misspelled words and remove meaningless words. Lexical features are retrieved in the form of n-grams [7].

#### ii. Semantic Method:

Semantics is basically the meaning of a language. Natural language processing helps to know how people



think and communicate their views. Semantic matching is compared with graph-based matching to give rise a score which is used to detect the level of sarcasm [10]. Semantic processing is used for generating a meaningful review [25].

#### iii. Syntactic Method:

The syntactic method basically follows the set of rules which are used to govern the structure of the sentence. Four groups of parts of speech are considered i.e., noun, verb, adjective, and adverb. Bilingual texts are considered, from translated corpus each token is tagged and built on 36 different tags and each tag to the corresponding group. Example – noun singular, noun plural, proper noun singular, proper noun plural. Word-tag pair is used to represent syntactic feature for better performance [7].

#### iv. Pragmatic Method:

Punctuation marks are examined as pragmatic features. Excessive usage of punctuation marks suggests the presence of sarcasm. To avoid dispersion maximum length of punctuation marks are reduced to three characters [7].

#### v. Prosodic Method:

Prosodic provides the rhythm and tune of the speech. [10] discusses the researchers who proposed an approach for automatic sarcasm detection using spectral, contextual and prosodic cues. Interjections differs according to the language used and are found in the language used [7]

#### vi. Idiosyncratic:

Idiosyncratic is focussing on the distinctive element. A syntax rule of noun-adposition-noun to formed to identify the idiosyncratic phraseology. Noun, adposition, and noun are tagged by parts of speech tagger. Example – head of cabbage.

Every word in the document is very important. They can be three terms which help us to understand the importance of a particular term.

### 5. Discussion

Experiments, using the Support Vector Machine (SVM) model, recorded the best accuracy of 54.1% for sarcasm detection using a negation and interjection feature combination [7]. The best one is Random Forest (0.724) out of the all classifications performed. The most reliable one in base classifiers is SVM which offers 0.682 F1 scores [6]. The best accuracy provided by the model 86.1% and precision is 68.6% in [8]. In proposed system accuracy, precision, recall, f-score is 83.1%, 91.1%, 73.4, 81.3% respectively in [1]. People are able to judge whether the review is sarcastic or not depending on the stars attached to the review in [4]. An average of 70% of articles in English have used

unigram and has achieved an average accuracy of 68% [2]. Few sarcastic sentences did not fit in the defined syntactic patterns due to the removal of particles and expression as a result, accuracy is decreased [3]. 177 idiosyncratic features are identified [7].

Accuracy is calculated as, for unsupervised learning using point mutual information and words is 68.5%, for unsupervised learning using point mutual information and emoticons is 78.7%, for supervised it is 81.5% [11]. Naïve Bayes unigram provides an accuracy of 79%, and that of bigram provides 64% accuracy, Maximum accuracy is provided by maximum entropy unigram of 84% [12]. First approach Parsing-based lexicon generation algorithm (PBLGA) provides the precision, recall, and F-score as 0.89, 0.81, 0.84 respectively. The second approach is based on the occurrence of interjection words which provided a precision, recall, and F-score as 0.85, 0.96 and 0.90 respectively [17].

Emoticons, when used in sarcasm detection, are to be used with caution. There are few emoticons which provide us reliable information about sentiment analysis and a large group are complicated. Accuracy with emoticon is calculated as 0.78 and without emoticon as 0.61 [13]. KNN provided the highest F1-score of 0.796 [14]. The proposed model attained precision, recall and F-score of 0.714, 0.51, 0.586 respectively [16]. Semi-supervised identification algorithm provided the highest precision, accuracy, and F-score as 0.912, 0.947 and 0.827 respectively.

Bosnian, Croatian, Serbian, Montenegrin (represented as BCMS). All these languages are closely related to each other. Initially, each tweet is classified as BCMS or not\_BCMS manually. 8 different thresholds are classified, negatively classified are marked as ncl and positively classified are marked with cl. Several rules were even introduced. Annotators are used which places the tweets into different classes namely BCMS, not\_BCMS, ironic, not\_ironic [8]. Precision in hashtag tokenizer and sarcasm detection is 98% and 91% respectively. Sarcasm detection polarity is 80%.

### 6. Conclusion

Several data pre-processing techniques were used. Researchers worked on various classifiers and results are provided. A comparative research on these classifiers was done to know which classifiers provide the better results. The random forest has the best performance in [1], Gradient Boosting in [5], Ensemble classifiers [6]. Combination of syntactic, prosodic and pragmatic provides the better performance [7]. A developed classifier with 5 features namely positive sentiment polarity, positive sentiment words, parts-of-speech tagging, irony markers and ordered sequence of sentiment tags has achieved highest accuracy [8].



Lexicons are generated from both emoticons and emotion seed words [19].

### References

- [1] Mondher Bouazizi and Tomoaki Otsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," *IEEE Access* Volume 4, 2016. pp. 5477-5488.
- [2] Anandkumar D. Dave and Prof. Nikita P. Desai, "A Comprehensive Study of Classification Techniques for Sarcasm Detection on Textual Data," in *Proc. International Conference on Electrical, Electronics, and Optimization Techniques*, 2016, pp. 1985-1991.
- [3] Satoshi Hiai and Kazutaka Shimada, "A Sarcasm Extraction Method Based on Patterns of Evaluation Expressions," in *Proc International Congress on Advanced Applied Informatics*, 2016, pp. 31-36.
- [4] Filatova, Elena. "Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing." *LREC*. 2012, pp. 392-398.
- [5] Anurkarsh G Prasad; Sanjana S, Skanda M Bhat, B S Harish. "Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data" in *Proc. International Conference on Knowledge Engineering and Applications*, 2017, pp. 1-5.
- [6] Pyae Phy Thu and Than Nwe Aung. "Effective Analysis of Emotion-Based Satire Detection Model on Various Machine Learning Algorithms," in *Proc. IEEE 6<sup>th</sup> Global Conference on Consumer Electronics*, 2017.
- [7] Mohd Suhairi Md Suhaimin, Mohd Hanafi Ahmad Hijazi, Rayner Alfred and Frans Coenen. "Natural Language Processing Based Features for Sarcasm Detection: An Investigation Using Bilingual Social Media Texts," in *Proc. International Conference on Information Technology*, 2017, pp. 703-709.
- [8] Miljana Mladenović, Cvetana Krstev, Jelena Mitrović, Ranka Stanković, "Using Lexical Resources for Irony and Sarcasm Classification," *Proceedings of the 8<sup>th</sup> Balkan Conference in Informatics*.
- [9] Setra Genyang Wicana, Taha Yasin İbisoglu and Uraz Yavanoglu, "A Review on Sarcasm Detection from Machine-Learning Perspective," in *Proc. International conference on Semantic Computing*, 2017, pp. 469-476.
- [10] Manoj Y. Manohar and Prof. Pallavi Kulkarni, "Improvement Sarcasm Analysis using NLP and Corpus based Approach," in *Proc. International Conference on Intelligent Computing and Control Systems*, 2017, pp. 618-622.
- [11] Shuigui Huang, Wenwen Han, Xirong Que and Wendong Wang, "Polarity Identification of Sentiment Words based on Emoticons," *International Conference on Computational Intelligence and Security*, 2017. pp 134-138.
- [12] U. R. Hodeghatta, "Sentiment analysis of Hollywood movies on Twitter," in *Proc. IEEE/ACM ASONAM*, Aug. 2013, pp. 1401-1404.
- [13] Hao Wang, Jorge A. Castanon, "Sentiment Expression via Emoticons on Social Media," in *Proc. International Conference on Big Data*, 2015, pp. 2404-2408.
- [14] Michael Sejr Schlichtkrull, "Learning Affective Projections for Emoticons on Twitter," in *Proc. International Conference on Cognitive Infocommunications*, 2015, pp. 539-543.
- [15] Ms. Payal Yadav and Prof. Dhatri Pandya, "SentiReview: Sentiment Analysis based on Text and Emoticons," *International Conference on Innovative Mechanisms for Industry Applications*, 2017, pp 467-472.
- [16] Archana. R and S. Chitrakala, "Explicit Sarcasm Handling in Emotion Level Computation of Tweets – A Big Data Approach," in *Proc. International Conference on Computing and Communications Technologies* 2017, pp. 106-110.
- [17] Santosh Kumar Bharti, Korra Sathya Babu and Sanjay Kumar Jena, "Parsing-based Sarcasm Sentiment Recognition in Twitter Data," in *Proc. International Conference on Advances in Social Networks Analysis and Mining*, 2015, pp. 1373-1380.
- [18] Dmitry Davidov, Oren Tsur and Ari Rappoport, "Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon," in *Proc. Fourteenth Conference on Computational Natural Language Learning*, pp. 107-116.
- [19] M. Boia, B. Faltings, C.-C. Musat, and P. Pu, "A :) Is worth a thousand words: How people attach sentiment to emoticons and words in tweets," in *Proc. Int. Conf. Soc. Comput.*, Sep. 2013, pp. 345-350.
- [20] K. Manuel, K. V. Indukuri, and P. R. Krishna, "Analyzing internet slang for sentiment mining," in *Proc. 2nd Vaagdevi Int. Conf. Inform. Technol. Real World Problems*, Dec. 2010, pp. 9-11.
- [21] A. Joshi, P. Bhattacharyya, and M. J. Carman. (Feb. 2016). "Automatic sarcasm detection: A survey." [Online]. Available: <https://arxiv.org/abs/1602.03426>
- [22] B. Pang, L. Lillian, and V. Shivakumar, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. ACL Conf. Empirical Methods Natural Lang. Process.*, vol. 10. Jul. 2002, pp. 79-86.

- [23] J. M. Soler, F. Cuartero, and M. Roblizo, "Twitter as a tool for pre- dicting elections results," in Proc. IEEE/ACM ASONAM, Aug. 2012, pp. 1194\_1200.
- [24] D. Maynard and M. A. Greenwood, "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis," in Proc. 9<sup>th</sup> Int. Conf. Lang. Resour. Eval., May 2014, pp. 4238\_4243.
- [25] S. Homocanu, M. Loster, C. Lo\_, and W.-T. Balke, "Will I like it? Providing product overviews based on opinion excerpts," in Proc. IEEE CEC, Sep. 2011, pp. 26\_33.
- [26] Justin Martineau, and Tim Finin, "Delta TFIDF: An Improved Feature Space for Sentiment Analysis," in Proc.AAAI International Conference on Weblogs and Social Media, May. 2009.
- [27] Lakshya Kumar, Arpan Somani, Pushpak Bhattacharyya "Having 2 hours to write a paper is fun!": Detecting Sarcasm in Numerical Portions of Text, arXiv:1709.01950v1 [cs.CL] 6 Sep 2017.
- [28] S.V.Manikanthan and D.Sugandhi " Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel " International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 ,Issue 1 –MARCH 2014.
- [29] T.Padmapriya, Ms. N. Dhivya, Ms U. Udhayamathi, "Minimizing Communication Cost In Wireless Sensor Networks To Avoid Packet Retransmission", International Innovative Research Journal of Engineering and Technology, Vol. 2, Special Issue, pp. 38-42.

