SARCASTIC APPROACHES BASED ON MACHINE LEARNING CLASSIFIER WITH BIG DATA

[1]Roobini S, [2]Prasanna Venkatesh R, [3]Ravi Kumar P, [4]Vinoth R

[1,2,3]Students, [4]Assistant Professor
Department of Information Technology
Gojan School of Business and Technology, Redhills, Chennai-600 052.

***Abstract*: Sarcasm is a refined type of incongruity generally utilized in interpersonal organizations and microblogging sites. It is now and then wont to pass on understood information inside the message an individual communicates. Mockery can be utilized for different capacities like analysis or joke. Be that as it may, it's grave in any event, for people to recognize. Along these lines, perceiving harsh proclamations is appallingly useful to improve programmed estimation investigation of data gathered from miniature writing for a blog sites or interpersonal organizations. Feeling investigation alludes to the Identification and total of mentalities and assessments communicated by Internet clients towards a Specific point. In this paper, we will in general propose an example put together way to deal with notice Sarcasm with respect to Twitter. We propose four arrangements of alternatives that cowl the different types of Sarcasm we tend to characterized. We utilize those to group tweets as harsh and non-wry.**

II.INTRODUCTION:

Sarcasm is part of human nature and perhaps an evolutionarily noble entity. It is the routine of remarks that undoubtedly refer to the opposite of what the individuals say and made in order to miffed someone's feelings or to disparage something in a hysterical way. The understanding the delicacy of this practice needs second order elucidation of the narrator's or author's objectives; different parts of the brain must slog together to understand sarcasm. Sarcasm appears to work out the brain more than genuine testimonials do. Sarcasm has a two-faced quality: it's both comical and means. So, the researchers show curiosity in sarcasm detection of social media text, especially in tweets. The rapid growth of tweets leads to critical in the analysis of data. It is also known as opinion mining that derives the opinion of a person or attitude of a speaker. Many researchers focus their interest towards sentimental analysis particularly in the field of the social network from the past few years. Machine learning methods and algorithms pave a new way for sentiment analysis particularly sarcasm detection by providing a set of algorithms and procedures. The modern-day world can be labeled as a data-driven world. With the invention of mobile devices and advancements in networking technology, there has been an exponential increase in data being generated by a single device in the network.

III.EXISTING SYSTEM:

The extraction method of sarcastic sentences in product reviews. Sarcasm, which expresses a negative meaning with positive words, often lead to mistakes in sentiment analysis. Therefore, sarcasm detection is an important task in sentiment analysis. For our method, we collected sarcastic sentences to analyze them in advance. We manually labeled 70 sentences as sarcastic sentences from 10,000 reviews. We generated extraction rules on the basis of the analysis of the sentences. The rate of sarcastic sentences contained in reviews was low (70/10,000). However, 21 sarcastic sentences appeared in 233 reviews with 1-point, which is the worst point in this review dataset. In other words, approximately 10 contained sarcastic sentences. This fact denotes that the detection of sarcastic sentences lead to the improvement of sentiment analysis, namely positive-negative identification, because conventional PN identification methods without sarcastic detection can not recognize the polarity of the reviews correctly. This result shows a significance of sarcasm extraction even if the number of sarcastic sentences in reviews is small. In the experiment, we compared our method with a baseline based on a simple rule. As a result, Our method outperformed the baseline However, some approaches to extract sarcastic sentences have, such as Riloff's method. Comparison with state-of-the-art methods is important future work to evaluate our method. In Addition, the accuracy of our method was insufficient, especially the precision rate. The result is due to the lack of analysis. Although we analyzed sarcastic sentences in our data, the data contains only 70 sarcastic sentences. Collecting new sarcastic sentences and analyzing the sentences manually are important.

IV.PROPOSED SYSTEM:

Manual analysis of numerous sentences is costly. Therefore, generating rules automatically becomes necessary. In this we are going to use large dataset and Naive Bayes classifiers to get the maximum accuracy. The classifiers used here is Naive Bayes for predicting the accuracy rate higher. This classifier comes under supervised learning. In this project we use 21 special features along with usual unigrams and bigrams for classification. These 21 features were divided in to 4 categories:

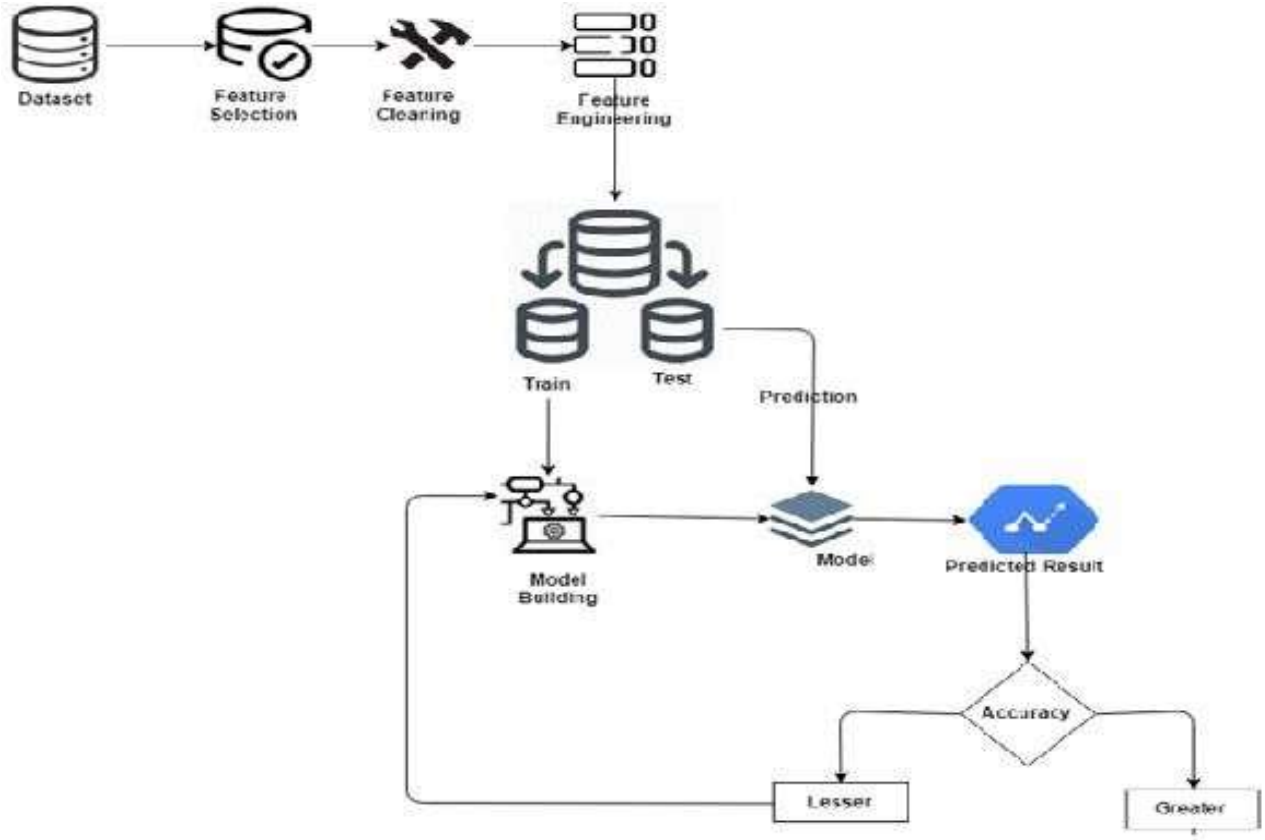
1. Text expression-based features
2. Emotion-based feature
3. Familiarity-based feature
4. Contrast-based feature

So that the accuracy of our sarcasm detection will be improved.

V.DISADVANTAGES:

- The sarcastic sentence in the dataset is very low
- Without sarcastic sentence we cannot recognize the polarity of reviews correctly
- Here the accuracy of our method is insufficient
- Riloff's method is used here for predicting and accuracy purpose

VI. ARCHITECTURE DIAGRAM:



VII. MODULE EXPLANATION:

The benefit of using these words based on their entropy score in the feature-set is that we were able to reduce uncertainty in the prediction outcome as these words have a different impact of frequency count in sarcasm and non-sarcasm tweets. Extracting Lightweight Features -After collecting 400,000 labelled tweets, we extracted around 350,000 English tweets Feature selection is based on which features will make an impact in our project and which feature we don't need to use. The Features we need to use is extracted from the dataset and other features are left as it is. The Feature can be multiple class as well as Single feature so we need to decide how our feature should come. Analyzing of the data helps in screaming of the data carefully which can rectify misleading results. After the feature extraction we'll search for the any null parameters in my data's. If there is any null parameters there means we need to fill the parameters with the related content. In this process the steaming of the words are done. The similar words are been considered as one and the model is being build. After the Preprocessing we need to build the model according to the features if our feature has labels means model can be built as Supervised algorithm if my feature doesn't have the labels means we need to use Unsupervised learning algorithms if there is combination means Semi supervised Ensemble Model should be used. Supervised learning is the task of inferring a function from labeled training data. By fitting to the labeled training set, we want to find the most optimal model parameters to predict unknown labels on other objects (test set). If the label is a real number, we call the task *regression*. If the label is from the limited number of values, where these values are unordered, then it's *classification*. In unsupervised learning we have less information about objects, in particular, the train set is unlabeled. What is our goal now? It's possible to observe some similarities between groups of objects and include them in appropriate clusters. Some objects can differ hugely from all clusters, in this way we assume these objects to be anomalies. Semisupervised learning tasks include both problems we described earlier: they use labeled and unlabeled data. That is a great opportunity for those who can't afford labeling their data. The method allows us to significantly improve accuracy, because we can use unlabeled data in the train set with a small amount of labeled data. Reinforcement learning is not like any of our previous tasks because we don't have labeled or unlabeled datasets here. RL is an area of machine learning concerned with how software agents ought to take actions in some environment to maximize some notion of cumulative reward.

VIII.CONCLUSION AND FUTURE ENHANCEMENT:

The way of up the existent caustic remark detection algorithms by as well as higher preprocessing and text mining techniques like emoji and slang detection area unit given. For classifying tweets as sarcasm and no sarcasm there are various techniques used, however, the paper takes up a classification algorithm and suggests various improvements that directly contribute to the advance of accuracy. The project derived analytical views from a social media dataset and also filtered out or reverses analyzed sarcastic tweets to achieve a comprehensive accuracy in the classification of the info that's given. The model has been tested in time period and may capture live streaming tweets by filtering through hash tags so perform immediate classification.

REFERENCES

[1] Wicana, SetraGenyang, TahaYasinİbisoglu, and UrazYavanoglu. "A Review on sarcasm detection from machine-learning perspective." 2017 IEEE 11th International Conference on Semantic Computing (ICSC). IEEE, 2017.

[2] Selvan, LokmanyathilakGovindanSankar, and Teng-Sheng Moh. "A framework for fast-feedback opinion mining on Twitter data streams." 2015 International Conference on Collaboration Technologies and Systems (CTS). IEEE, 2015.

[3] Suhaimin, MohdSuhairiMd, et al. "Natural language processing-based features for sarcasm detection: An investigation using bilingual social media texts." 2017 8th International Conference on Information Technology (ICIT). IEEE, 2017.

[4] Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P., & Carman, M. (2016). Are word embedding-based features useful for sarcasm detection arXiv preprint arXiv:1610.00883.

[5] Dave, Anandkumar D., and Nikita P. Desai. "A comprehensive study of classification techniques for sarcasm detection on textual data." 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, 2016.

[6] Hiai, Satoshi, and Kazutaka Shimada. "A sarcasm extraction method based on patterns of evaluation expressions." 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). IEEE, 2016.

[7] Dave, A. D., & Desai, N. P. (2016, March). A comprehensive study of classification techniques for sarcasm detection on textual data. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 1985-1991). IEEE.

[8] Zhang, Meishan, Yue Zhang, and Guohong Fu. "Tweet sarcasm detection using deep neural network." Proceedings of COLING 2016, The 26th International Conference on Computational Linguistics: Technical Papers. 2016.

[9] Bharti, Santosh Kumar, et al. "Sarcasm analysis on twitter data using machine learning approaches." Trends in Social Network Analysis. Springer, Cham, 2017. 51-76.

[10] Ahmad, Tanvir, et al. "Satire detection from web documents using machine learning methods." 2014 International Conference on Soft Computing and Machine Intelligence. IEEE, 2014.

[11] Dmitry Davidov, Oren Tsur and Ari Rappoport, Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon

[12] Ellen Riloff, AshequlQadir, PrafullaSurve, Lalindra De Silva, Nathan Gilbert and Ruihong Huang, Sarcasm as Contrast between a Positive Sentiment and Negative Situation

[13] Roberto Gonzalez-Ibanez, SmarandaMuresan and Nina Wacholder, Identifying Sarcasm in Twitter: A Closer Look

[14] R. Vinoth, L. J. Deborah, P. Vijayakumar and N. Kumar, "Secure Multifactor Authenticated Key Agreement Scheme for Industrial IoT," in IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3801-3811, 1 March1, 2021, doi: 10.1109/JIOT.2020.3024703.