iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32227

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Sarcasm Detection using Support Vector Machine

Yash R Mahadik
Academic Research Student, Department of Information Technology, B.K. Birla College of Arts, Science and Commerce (Autonomous) Kalyan, Thane, India

Abstract: ***Sarcasm is the use of language where people used positive words in order to convey a negative message. Sarcasm is an example of what some researchers call uncertain speaking, where what is said by the people is actually differ from what is meant (especially so as to insult or wit somebody, to indicate irritation, or to be funny). sentiment analysis on social media has been extensively used, however it's still rare to look out sentiments and analyse them, considering the detection of satire in it. Satire detection in sentiment analysis is a difficult task to do. The most used way for sarcasm detection is Machine learning approach. This paper proposed the used of linear and radial basis function kernels of support vector machine algorithm for sarcasm detection and compared them to find out which kernel provide best accuracy. This paper will help many online businesses to improved present services of business for enhance more profit and customer satisfaction.***
Keywords: Sarcasm detection, machine learning, support vector machine, linear kernel, rbf kernel

I. INTRODUCTION

Natural Language process (NLP) is one amongst the necessary domains in computing. It acts as a platform between the pc and human languages. It helps in creating the machine perceive, analyse and interpret the info. It helps in querying the datasets and supply a solution. It helps in not solely perceive the text or speech however conjointly the context behind it. It works for structured and unstructured information. The linguistic structure depends on varied factors like social context, regional dialects, slang etc. NLP is facing few challenges during this field. Sentiment Analysis is one amongst the necessary fields in NLP that deals with analysing the context.

The present world categorized as a datadriven world. Thanks to advancements in networking, social platform and micro-blogging, sites like Facebook and Twitter introduced. These playing vital roles. This social media created an open platform to all or any for communication. In this, present era users are generating vast data within the social media. Sarcasm occurs frequently in user-generated content on the social platform.

Humans have a social nature. Social nature means we have a tendency to act with one another in positive, friendly ways and it additionally implies that we all know approach to manipulate others terribly. Sarcasm, this is every absolutely humorous and negatively nasty, performs a essential 1/2 of in human social interaction. Sarcasm detection may be a terribly slim analysis field in NLP, a selected case of sentiment analysis wherever rather than detective work a sentiment within the whole spectrum, the main focus is on sarcasm. So the task of this field is to notice if a given text is sarcastic or not.

II. LITERATURE REVIEW

In [1], Wandra, K. et al. discussed sarcasm, types of sarcasm in detail and the various challenge in sarcasm detection. The most used way for sarcasm detection is Machine learning based approach. Author used many machine learning techniques like Naive Bayes, Maximum entropy and support Vector Machine who achieved great success in text categorization. In [2], C. S. et al. gives idea about various techniques and methods used for sarcasm detection.

Author explained Many data pre-processing techniques, several classifiers and many feature extraction techniques. In [4], Nagaraj Devi, D. et al. used four machine learning classification methods for detection of sarcasm in plain text. First classification is Support vector classification, second is Logistic regression, third is Naive bayes classification and fourth is Decision tree classification. In [6], Sharma, S. et al. proposed cross-domain sarcasm detection of tweets for detecting sarcastic content in online reviews. Author conducted their experiment on Amazon product review dataset specifically the humor corpus version1 having regarding 2000 reviews. Author used Support vector machine and neural network using lexical, pragmatic linguistic incongruity and context incongruity features. In [7], Parveen, S. et al. proposed a method to detect sarcasm in tweet that makes use of the different components of twitter.

Author used four classes of options that covers differing kinds of wittiness which was accustomed classify whether or not the tweets area unit barbed or non-sarcastic. For classifications author used Naive bayes, Maximum entropy and Support vector machine. Author trained Classifiers into four forms of feature that were sentiment-related options, punctuation-related options, grammar options and pattern options.

In [8], Arora, H. et al. proposed new optimized business model to analyse online reviews of all the users to improved present services of business for enhance more profit and customer satisfaction. The proposed system filters reviews and classifies them using sentimental technique with the help of deep learning. The main object of this research lies in helping tourism industries to understand the social sentiment of their brand, product or service while monitoring online conversations. In [9], Kaushik, A. et al. discussed about an overview of the sentimental analysis approach of opinion mining with techniques and tools. Author explained assessment of sentiment with two ways direct opinions and comparison, types of feature used for opinion mining, opinion retrieval, opinion classification, opinion summarization, techniques of opinion mining which includes supervised machine learning, unsupervised learning and case based reasoning, semantic orientation which includes corpus based approach and dictionary based approach, tools used in opinion mining which includes review seer tool, internet fountain, red opal, opinion observer. In [13], Sentamilselvan, K. et al. proposed the naive bayes Classifier and logistic regression to perform sentiment analysis and classify them on the basis of better accuracy of categorized technique. Firstly, study about the twitter data investment was discussed. Secondly, methods to read dataset, pre-processing of data and feature extraction, and finding accuracy have been described. Thirdly, the result of this research is presented. Fourthly, conclusion and future work is provided. In [19], Sivaprakasam, L. et al. presents the concept of sarcasm detection on textual data. It put forward the various challenges pertaining to sarcasm which makes it difficult to detect. Author only explained research work that was present in various papers that have proposed techniques to identify sarcasm.

III. METHODOLOGY

SVM or Support Vector Machine may be a linear model for classification and regression issues. It will solve linear and non-linear issues and work well for several sensible issues. The algorithmic program creates a line or a hyperplane that separates the info into categories.

SVM algorithm uses a collection of mathematical function which is known as kernel. The kernel takes data as input and transforms it into a particular form. Different svm algorithm use unique forms of kernel. The kernels used in this paper are:

1) Linear Kernel: For dividing the data using single line the linear kernel is used. The linear kernel is faster than any other kernel while training the svm model.

2) Radial Basis Function (RBF) Kernel: RBF (Radial Basis Function) kernel is another common kernel technique employed in SVM models for additional. RBF kernel may be a perform whose price depends on the space from the origin or from some purpose.

Dataset: The dataset used for detection namely Sarcasm On Reddit were downloaded from kaggle website. The dataset were divided into training and testing data with ratio 80:20 for training set and testing set accordingly.

IV. EXPERIMENT

The dataset used for detection is taken from kaggel website. Machine learning svm algorithm is used for detection. The dataset is divided in training and testing data. In figure 1 and 3 training data is applied and in figure 2 and 4 testing data is applied.

After using the linear kernel on figure 1, in the first column the model is able to predict 16630 comments as sarcastic and wrongly predict 987 sarcastic comments as non-sarcastic. In the second column the model is able to predict that 8802 comments are non-sarcastic and wrongly predict 3493 comments as sarcastic.

After using the linear kernel on figure 2, in the first column the model is able to predict 3610 comments as sarcastic and wrongly predict 779 sarcastic comments as non-sarcastic. In the second column the model is able to predict that 1468 comments as non-sarcastic and wrongly predict 1621 comments as sarcastic.

After using the rbf kernel on figure 3, in the first column the model is able to predict 17062 comments as sarcastic and wrongly predict 555 sarcastic comments as non-sarcastic. In the second column the model is able to predict that 3803 comments as non-sarcastic and wrongly predict 8492 comments as sarcastic.

After using the rbf kernel on figure 4, in the first column the model is able to predict 4137 comments as sarcastic and wrongly predict 252 sarcastic comments as non-sarcastic. In the second column the model is able to predict that 777 comments as non-sarcastic and wrongly predict 2312 comments as sarcastic.

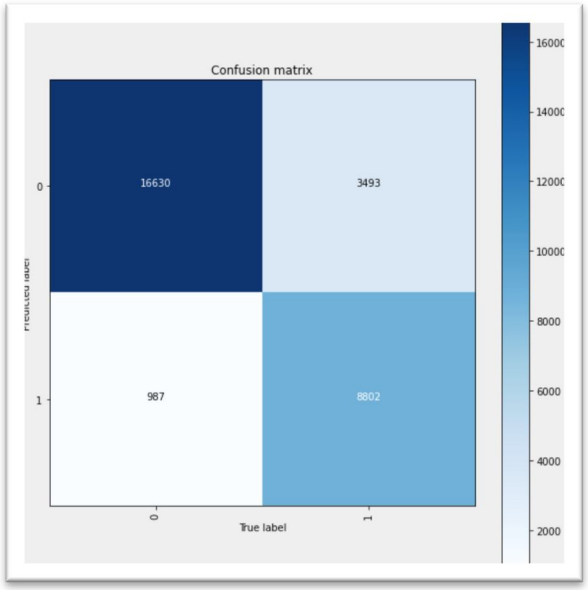


Fig. 1

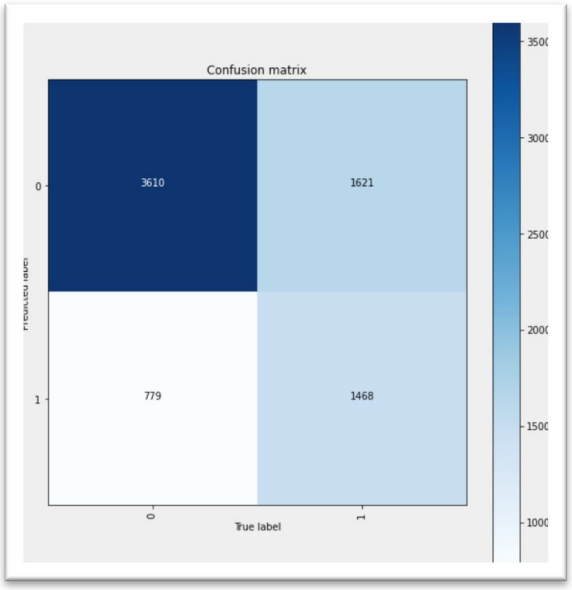


Fig. 2

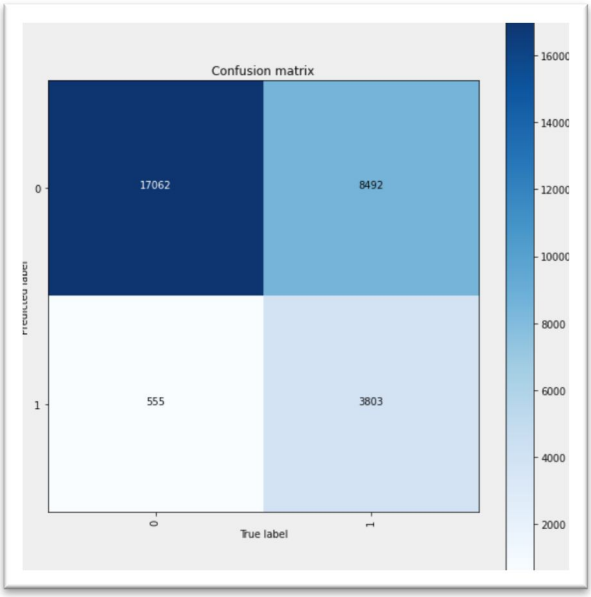


Fig. 3

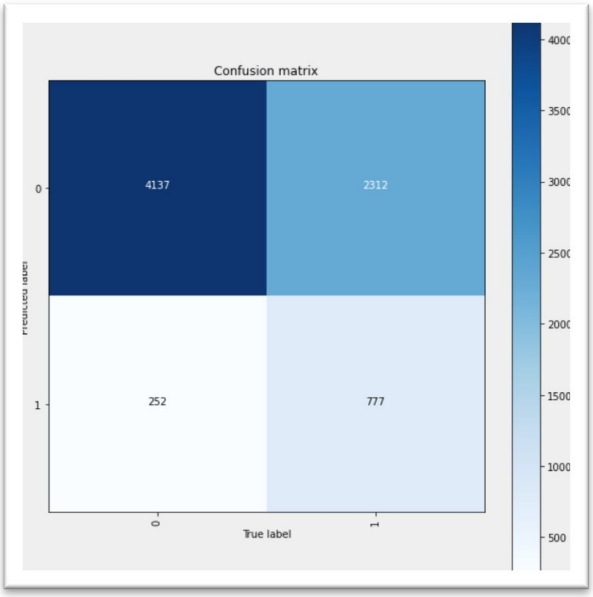


Fig. 4

V. CONCLUSION

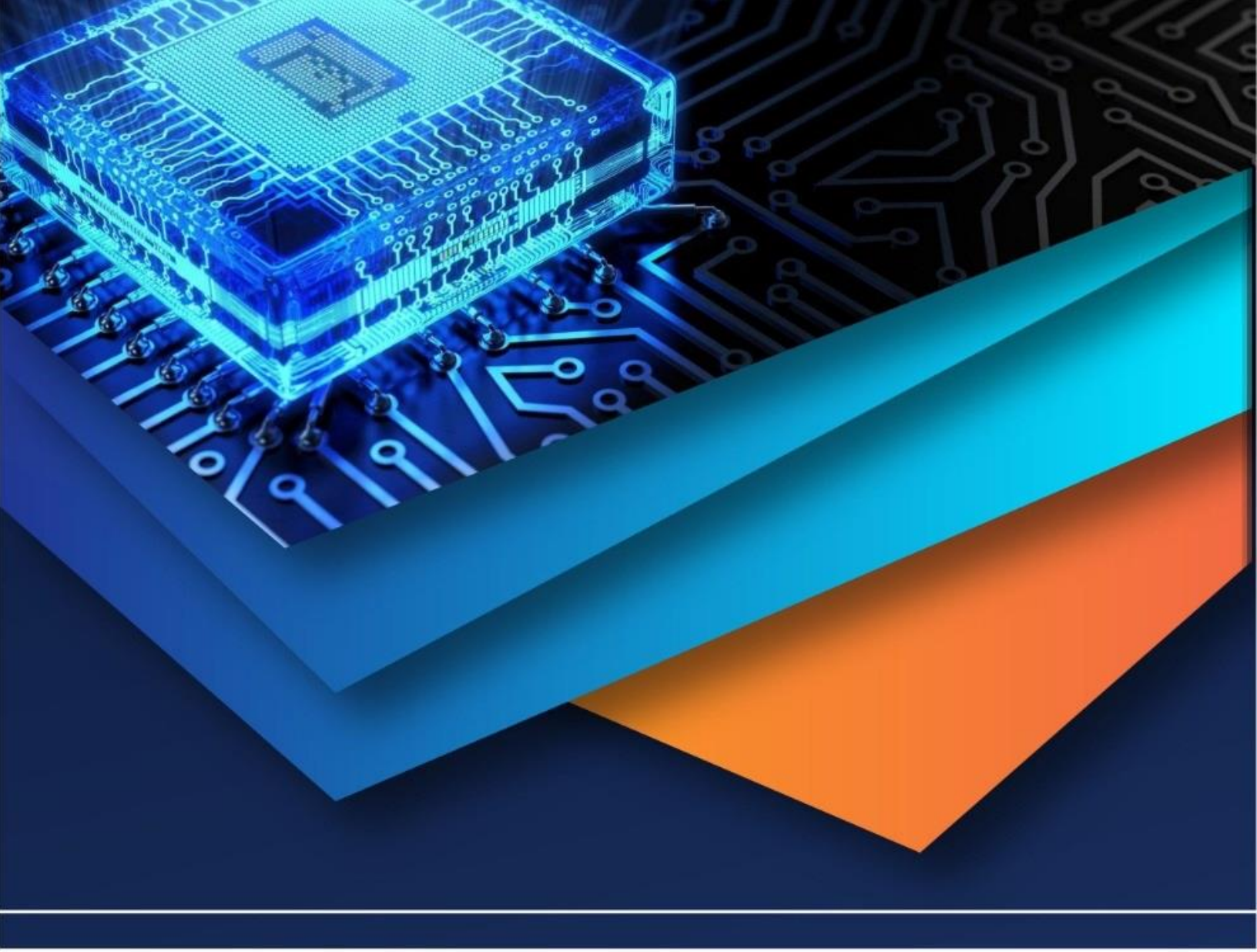
From the above experiment, it is observed that in the training model linear kernel provide 85% accuracy and rbf kernel provide 69% accuracy and in the testing model linear kernel provide 67% accuracy and rbf kernel provide 65% accuracy. Finally, it is concluded that the linear kernel provide the better accuracy than the rbf kernel.

VI. ACKNOWLEDGEMENT

I would like to express my special thanks to my teacher Prof. Swapna Augustine Nikale, Department of Information Technology of B.K. Birla College of Arts, Science and Commerce (Autonomous) Kalyan, Thane.

REFERENCES

[1] Wandra, K., & Barot, M. (2017). SARCASM DETECTION IN SENTIMENTAL ANALYSIS. International Journal of Current Engineering and Scientific Research, 4(9), 24–32. http://troindia.in/journal/ijcesr/vol4iss9part4/24-32.pdf
[2] C, S., Vadieu, G., & Rao, M. (2018). A COMPREHENSIVE STUDY ON SARCASM DETECTION TECHNIQUES IN SENTIMENTAL ANALYSIS. International Journal of Pure and Applied Mathematics, 118, 433–442. https://acadpubl.eu/hub/2018-118-22/articles/22a/63.pdf
[3] Dharmavarapu, B., & Bayana, J. (2019). SARCASM DETECTION IN TWITTER USING SENTIMENTAL ANALYSIS. International Journal of Recent Technology and Engineering, 8(1), 642–644. https://www.ijrte.org/wp-content/uploads/papers/v8i1s4/A11180681S419.pdf
[4] Nagarjana Devi, D., Rajanikanth, T., & Balaram, V. (2018). SARCASM DETECTION IN PLAIN TEXT USING MACHINE LEARNING. International Journal of Engineering Research in Computer Science and Engineering, 5(4), 225–228. http://ijercse.com/specissue/aprilissue/40.pdf
[5] Vijayalaksmi, N., & Senthilrajan, A. (2017). A HYBRID APPROACH FOR SARCASM DETECTION OF SOCIAL MEDIA DATA. International Journal of Scientific and Research Publications, 7(5), 327–336. http://www.ijsrp.org/research-paper-0517/ijsrp-p6537.pdf
[6] Sharma, S., & Chakraverty, S. (2018). SARCASM DETECTION IN ONLINE REVIEW TEXT. ICTACT Journal on Soft Computing, 8(3), 1674–1679. http://ictactjournals.in/paper/IJSC_Vol_8_Iss_3_Paper_5_1674_1679.pdf
[7] Parveen, S., & Deshmukh, S. (2017). OPINION MINING IN TWITTER - SARCASM DETECTION. International Research Journal of Engineering and Technology, 4(10), 201–204. https://www.irjet.net/archives/V4/i10/IRJET-V4I1037.pdf
[8] Arora, H., & Bansal, M. (2020). DEVELOPING A MODEL FOR SENTIMENT ANLYSIS TECHNIQUE IN THE FIELD OF TOURISM USING DEEP LEARNING. International Journal of Recent Technology and Engineering, 8(6), 456–462. https://www.ijrte.org/wp-content/uploads/papers/v8i6/F7390038620.pdf
[9] Kaushik, A., Kaushik, A., & Naithani, S. (2015). A STUDY ON SENTIMENT ANALYSIS : METHODS AND TOOLS. International Journal of Science and Research, 4(12), 287–292. https://pdfs.semanticscholar.org/c151/dfad8c1bf88b0afc716758c77d533ded7dd0.pdf
[10] Alsaeedi, A., & Khan, Z. (2019). A STUDY ON SENTIMENT ANALYSIS TECHNIQUES OF TWITTER DATA. International Journal of Advanced Computer Science and Applications, 10, 361–374. https://thesai.org/Downloads/Volume10No2/Paper_48-A_Study_on_Sentiment_Analysis_Techniques.pdf
[11] Nagarjana, D., VijayBhaskar, K., & Pavan, Y. (2019). SURVEY ON DETECTION OF SARCASM IN SENTIMENT ANALYSIS. International Journal of Engineering Research and Application, 9(10), 39–44. http://www.ijera.com/papers/vol9no10/Series-2/H0910023944.pdf
[12] Bindra, K., & Gupta, A. (2016). TWEET SARCASM : MECHANISM OF SARCASM DETECTION IN TWITTER. International Journal of Computer Science and Information Technologies, 7(1), 215–217. http://ijcsit.com/docs/Volume%207/vol7issue1/ijcsit2016070148.pdf
[13] Sentamilselvan, K., Aneri, D., Athithiya, C., & Kumar, P. (2020). TWITTER SENTIMENT ANALYSIS USING MACHINE LEARNING TECHNIQUES. International Journal of Engineering and Advanced Technologies, 9(3), 4205–4209. https://www.ijeat.org/wp-content/uploads/papers/v9i3/C6281029320.pdf
[14] Kumar, A., Kumar, R., & Katiyar, V. (2019). A COMPARATIVE ANALYSIS OF SARCASM DETECTION. International Journal of Recent Engineering Research and Development, 4(8), 104–108. http://ijrerd.com/papers/v4-i8/16-IJRERD-D157.pdf
[15] J, A., & Ravikumar, P. (2018). AN EFFECTIVE APPROACH FOR SARCASM DETECTION IN TEXT DATA FOR SENTIMENTAL ANALYSIS. International Journal of Engineering and Technology, 7(4), 136–138. https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.sciencepubco.com/index.php/ijet/article/download/23825/11946&ved=2ahUKEwioguGb4fLrAhWjoekKHRomDXQQFjABegQIAhAB&usg=AOvVaw3v49_l9y1MplZ-BCEEelvS
[16] Bholane, S., & Gore, D. (2015). A SURVEY ON SENTIMENT ANALYSIS ON TWITTER DATA USING DIFFERENT TECHNIQUES. International Journal of Computer Science and Information Technologies, 6(6), 5358–5362. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.735.7484&rep=rep1&type=pdf
[17] Parveen, S., Surnar, A., & Sonawane, S. (2017). OPINION MINING IN TWITTER : HOW TO MAKE USE OF SARCASM TO ENHANCE SENTIMENT ANALYSIS : A REVIEW. International Journal of Computer Engineering & Technology, 6(6), 863–865. http://ijarcet.org/wp-content/uploads/IJARCET-VOL-6-ISSUE-6-863-865.pdf
[18] Jansi, K., Sajja, P., & Goyal, P. (2018). AN EXTENSIVE SURVEY ON SARCASM DETECTION USING VARIOUS CLASSIFIERS. International Journal of Pure and Applied Mathematics, 119(12), 13183–13186. https://acadpubl.eu/hub/2018-119-12/articles/5/1200.pdf
[19] Sivaprakasam, L., & Jayaprakash, A. (2017). A STUDY ON SARCASM DETECTION ALGORITHMS. International Journal of Engineering Technology Science and Research, 4(9), 1362–1366. http://www.ijetsr.com/images/short_pdf/1506172758_1362-1366-ieteb306_ijetsr.pdf
[20] Garg, A., & Duhan N. (2020). SARCASM DETECTION ON TWITTER DATA USING SUPPORT VECTOR MACHINE. ICTACT Journal of soft computing, 10(4), 2165-2170. http://ictactjournals.in/paper/IJSC_Vol_10_Iss_4_Paper_7_2165_2170.pdf
[21] https://www.kaggle.com/danofer/sarcasm



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY