iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50265

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: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Hate Speech Detection Using Machine Learning

Suraj Futane[1], Twinkal Bandwal[2], Dnyaneshwari Dhonde[3], Sakshi Gudmewar[4], Aishwarya kadam[5]
[1, 2, 3, 4]*Students,* [5]*Asst.Professor, Department of Information of Technology Engineering Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India*

Abstract: ***Twitter's central goal is to enable everybody to make and share thoughts and data, and to communicate their suppositions and convictions without boundaries. Twitter's job is to serve the public discussion, which requires portrayal of a different scope of points of view. Yet, it does not advance viciousness against or straightforwardly assault or undermine others based on race, nationality, public cause, rank, sexual direction, age, inability, or genuine illness. Hate Speech can hurt a person or a community. So, it is not appropriate to use hate speech. Now, due to increase in social media usage, hate speech is very commonly used on these platforms. So, it is not possible to identify hate speeches manually. So, it is essnetial to develop an automated hate speech detection model and this resaech work shows different approaches of Natural Language Processing for classification of Hate Speech through Machine Learning Algorithms.***
Keywords: Logistic Regression, SVM, Tf-Idf, Random Forest, Hate Speech.

I. INTRODUCTION

Due to increasing scale of social media, people are using social media platforms to post their views. Giving opinions which are harsh or rude to someone directly on face is a difficult task. So, people feel it is safe over internet to abuse or post something offensive to others. So, they feel secured posting such content on the internet. Due to this the use of hate speech over the social media is increasing daily. So, as to handle such a large data of users over social media, automatic detection of hate speech methods are required. In this paper we use machine learning methods to classify whether hate speech or not. There are a number of machine learning applications, One of them is for text based classification. Each instance or here we can say each tweet is represented using the same set of features used by machine learning algorithms. There are two types of problems solved machine learning algorithms, supervised and unsupervised. Supervised learning is the task of training model based on given dataset containing both set of features and labels. Though unsupervised learning is the training system function in which data set is neither categorized nor named. Supervised learning is further divided into two types regression and classification based on labels of dataset.Here we concerned only about classification. Classification machine algorithms used categorical dataset and are used to classify the class/category of the unknown instance. Various machine learning application includes task that can be set up as supervised. We aim to do this task by applying supervised classification methods like Support vector machine, logistic regression and random forest on labeled hate speech dataset. Each instance is represented in form of vector the length of vector is dependent on the method used For representation of tweets. In this paper we used two types approaches for vector representation of a tweet, vector term frequency-inverse document frequency(tf-idf) and bag of words .-classes data. The XGBoost algorithm uses the exact greedy algorithm to find the best split.

II. RELEATED WORK

[1] The paper is based on AI:Finding a text's viewpoint can be done by conducting a sentiment analysis. It also goes by the names emotion Al and opinion mining (AI). People describe their experiences with events in social media comments, and they are curious to discover whether most other people felt the same way about the same event. Sentiment Analysis can be used to classify the data.Sentiment analysis sorts unstructured text comments about events, products, etc. that have been uploaded by various users into groups according to whether they are good, negative, or neutral.

[2]The paper is based on sentimental analysis:Identifying the sentiment polarity of a particular feature of the text is the goal of aspect-based sentiment analysis. The significance of the information interplay between an aspect word and context has been recognised in earlier research. However, the majority of current information interaction techniques are coarse-grained, which causes some information to be lost. Additionally, the majority of techniques disregard the significance of position data in determining the sentiment polarity of the aspect.

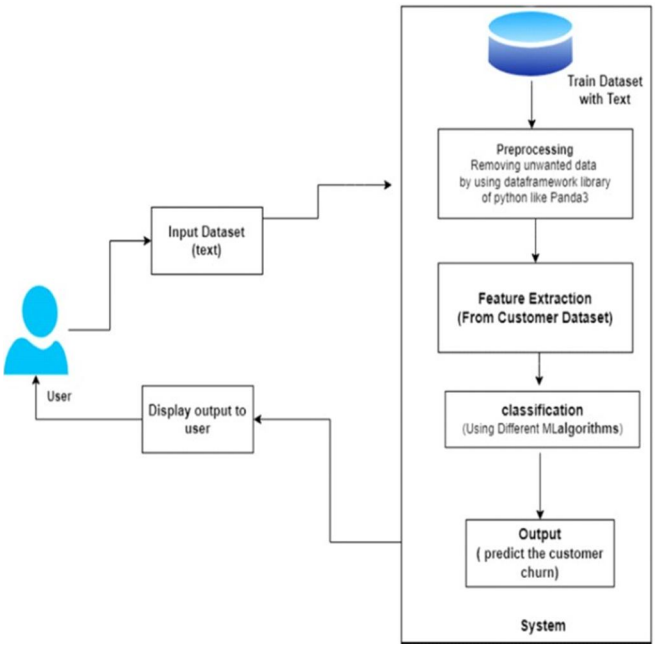
[3]The paper is based on social network:The paIn recent years, due to the vigorous development of social network media, a large amount of social data can be obtained through social media. This trend has made it easier for researchers in natural language

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

processing to obtain textual research materials, and has further stimulated the development of natural language processing. Emotion recognition is an important task in the field of natural language processing, and emotion recognition helps improve interaction in social. [4]The paper is based on NLP:In natural language processing, aspect-level sentiment classification is a popular research area (NLP). How to create efficient algorithms to model the relationships between aspects and opinion words that appear in a sentence is one of the major challenges. The graph convolutional networks (GCNs) achieve the promising results among the various methods suggested in the literature because of their strong ability to capture the large distance between the aspects and the opinion words.
[5]The paper is based on social media:At present, with the growing number of Web 2.0 platforms such as Instagram, Facebook, and Twitter, users honestly communicate their opinions and ideas about events, services, and products. Owing to this rise in the number of social platforms and their extensive use by people, enormous amounts of data are produced hourly.

III. METHODOLOGY

In this section, the steps taken to predict customer churn are first described in general, then the actions taken in each step are reviewed. In the first step, the dataset is pre-processed. In this step, actions such as deleting rows with empty values, converting data format to processable format for algorithms, data normalization is performed and in the feature extraction step, some features of the table are extracted. Using balancing data algorithms, the data is balanced and the models begin to be trained and then perform the test phase. Finally, the performance of the models is compared based on the three balancing methods. Fig.1 shows the steps that have been taken.



1) About the Dataset The data used in this study was published on the Kaggle website, which is owned by a company in the telecom industry. In this dataset, there are 3333 rows and 21 columns, and the target column is called Churn. In this dataset, the customer specifications are presented, and the amount of customers' use of each of the company's services, which includes Voice Mail, SMS, and calls, is mentioned during the day and night.
2) Data preprocessing A dataset needs to be preprocessed before entering the models. Data preprocessing was performed in the following steps
3) By exploring the dataset, the null and missing values identified and Each row containing these values were deleted.
4) Feature extraction Feature selection is a process of searching for meaningful features for analysis. Using this method, it is possible to remove features that have duplicate information, and by recognizing the effective features, reduce the complexity of the data and increase the computational speed of the machine learning model. At this stage, some features have been removed.
5) Data balancing In classification models, it is assumed that the number of samples is evenly distributed in different classes. If this is not the case and the samples are imbalanced divided between the classes.
6) Output we'll predict percentage of customer churn.On basis on analysis find flaws to minimize the churn prediction.

IV. ALGORITHMS

A. *Support Vector Machine(SVM)*

Support Vector Machine It is one of the most popularized Supervised Learning algorithm, which is used for Classification as well as Regression problems. However, basically, it is used for Classification problems in Machine Learning scenario. The intent of the SVM algorithm is to create the best decision boundary that can segregate n-dimensional space into classes so that it can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane of SVM.

B. *Random Forest Algorithm*

Random Forest classifier is a learning method that operates by constructing multiple decision trees and the final decision is made based on the majority of the trees and is chosen by the random forest. It is a tree-shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, instance, or reaction. Using of Random Forest Algorithm is one of the main advantages is that it reduces the risk of over fitting and the required training time. Additionally, it also offers a high level of accuracy. It runs efficiently in large databases and produces almost accurate predictions by approximating missing data.

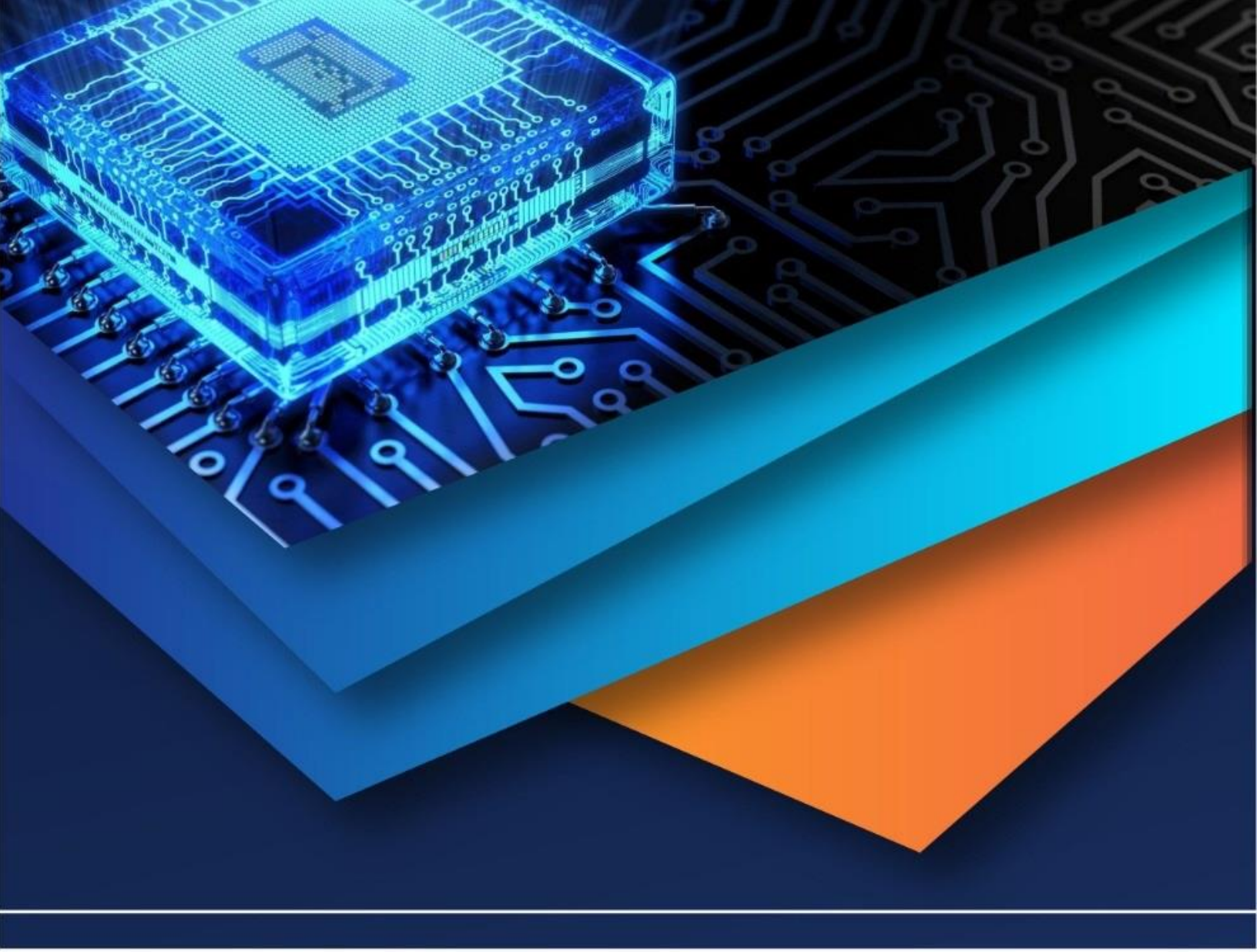
V. CONCLUSION

In routine life, as the usage of social media is increased everyone seems to think like they can speak or write anything they want. Due to this thinking hate speech has been increased so it becomes necessary to automate the process of classifying the hate speech data. To simplify the process of classifying of hate speech we have used machine learning approach to detect hate speech from the twitter data. For this we have used tf-idf and bag of words methods to extract feature from the tweets. To classify hate speech from the tweets we have implemented machine learning algorithms like SVM, Logistic Regression and Random Forest. We can conclude from the results obtained that by using Data without preprocessing and machine learning models with default parameters, Random Forest with bag of words gives best performance with 0.6580 F1 Score and 0.9629 Accuracy Score. But as explained earlier only obtaining highest accuracy is not enough when we are dealing with imbalance class dataset. For that we have used here F1 score which is quite low for data without preprocessing. To improve this we have used some preprocessing steps and gridsearch to obtained best parameter for machine learning model. After preprocessing and using gridsearch SVM with Tf-IDF gives best performance with 0.7488 F1 Score and 0.9668 Accuracy Score.Tfidf feature extraction model derives superlative accuracy in comparison to bag of words model because bag of words just count the frequency of words and used it as a vector but tf-idf model uses ratio of term frequency to the document frequency. Limitations of this approach is that it can be only applied to the twitter dataset so to detect hate speech from big data can be a challenge. In future f1 score and accuracy can be improved. More machine learning techniques needs to be explored. Also different method needs to be applied to handle the imbalance class dataset.

REFRENCES

[1] Abdullah Alsaeedi, Mohammad Zubair Khan, “A Study on Sentiment Analysis Techniques of Twitter Data”, International Journal of Advanced Computer Science and Applications, Vol.10, No.2, 2019.
[2] Suchita V Wawre, Sachin N Deshmukh, “Sentiment Classification using Machine Learning Techniques”, Internaional Journal of Science and Research (IJSR), Vol.6, 2015.
[3] Ali Hasan, Sana Moin, Ahamad Karim and Shahaboddin Shamshirband, “Machine Learning-Based Sentiment Analysis for Twitter Accounts”, Journal mca, 16 January 2018, Accepted 24 February 2018, Published: 27 Febryary 2018.
[4] Vishal A. Kharde, S. S. Sonawane, “Sentiment Analysis of Twitter Data: A survey of Techniques”, International Journal of Computer Applications (0975-8887) Volume 139, No.11, April 2016 .
[5] Suchita V Wawre, Sachin N Deshmukh, “Sentimental Analysis of Movie Review using Machine Learning Algorithm with Tuned Hyperparameter”, International Journal of Innovative Research in Computer and Communication Engineering, Vol.4, Issue 6, June 2016
[6] Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainydin, “Sentiment Analysis Techniques in Recent Works”, Science and Information Conference 2015 July page no. 28-30, 2015.
[7] Bac LeHuy, Nguyen, “Twitter Sentiment Analysis Using Machine Learning Techniques”, conference paper, Advanced computaional methods for knowledge engineering,volume 358, pp 279-289, 2015.
[8] Alec Go, Richa Bhayani, Lei Huang, “Twitter Sentiment Classication using Distant Supervision”, Semantic Scholar, page no. 57-61, published 2009.
[9] Richa Sharma, Shweta Nigam, Rekha Jain, “Polarity Detection at Sentence Level”, International Journal of Computer Applications (0975 – 8887), Volume 86 – No 11, January 2014,29.
[10] Farhan Hassan Khan, Saba Bashir, Usman Qamar,“TOM: Twitter opinion mining framework using Macine Learning.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)