iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47954

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 10 Issue XII Dec 2022- Available at www.ijraset.com

Sentiment Analysis of Twitter

Prathamesh Raut[1], Rohit Rathod[2], Rohit Tidke[3], Rugved Pande[4], Niraj Rathod[5], Nishant Kulkarni[6]
[1, 2, 3, 4, 5, 6]*Mechanical Department, Vishwakarma Institute of Technology, Pune, India*

Abstract: ***Hate Speech is a widespread problem that degrades a person or people based on their race, religion, gender or disability. This research work proposes a tool to raise awareness on the persistent hate speech in social media platforms. The primary aim of this research work is to highlight the content that promotes violence or hatred against individuals or groups based on religion, gender or ethnicity. Logistic regression is a technique borrowed by machine learning from the field of statistics. It is the go-to method for binary classification problems. Using this algorithm, the model trains itself from the dataset and identifies and displays the sentiment of the tweets. Also, to get the real-time analysis on tweets, Twitter API and libraries such as Tweepy and Textblob are used. The proposed model has the ability to detect the appropriate sentiment with 83.98 percent accuracy. The tool is made free and available for demo use to the public.***
Keywords: APIs, Hate speech, Logistic regression, Sentiments, Twitter

I. INTRODUCTION

The exponential growth of social media such as Twitter and community forums has revolutionized communication and content publishing, but is also increasingly exploited for the propagation of hate speech and the organization of hate-based activities. The anonymity and mobility afforded by such media has made the breeding and spread of hate speech – eventually leading to hate crime – effortless in a virtual land-scape beyond the realms of traditional law enforcement. The term 'hate speech' was formally defined as 'any communication that disparages a person or a group on the basis of some characteristics (to be referred to as types of hate or hate classes) such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics. In the UK, there has been significant increase of hate speech towards the migrant and Muslim communities following recent events including leaving the EU, the Manchester and the London attacks.

In the EU, surveys and reports focusing on young people in the EEA (European Economic Area) region show rising hate speech and related crimes based on religious beliefs, ethnicity, sexual orientation or gender, as 80% of respondents have encountered hate speech online and 40% felt attacked or threatened. Statistics also show that in the US, hate speech and crime is on the rise since the Trump election.

The urgency of this matter has been increasingly recognized, as a range of international initiatives have been launched towards the qualification of the problems and the development of counter-measures. Building effective counter measures for online hate speech requires as the first step, identifying and tracking hate speech online. For years, social media companies such as Twitter, Facebook, and YouTube have been investing hundreds of millions of euros every year on this task, but are still being criticized for not doing enough. This is largely because such efforts are primarily based on manual moderation to identify and delete offensive materials. The process is labor intensive, time consuming, and not sustainable or scalable in reality. A large number of researches has been conducted in recent years to develop automatic methods for hate speech detection in the social media domain. These typically employ semantic content analysis techniques built on Natural Language Processing (NLP) and Machine Learning (ML) methods, both of which are core pillars of the Semantic Web research. The task typically involves classifying textual content into non-hate or hateful, in which case it may also identify the types of the hate speech. Although current methods have reported promising results, we notice that their evaluations are largely biased towards detecting content that is non-hate, as opposed to detecting and classifying real hateful content. A limited number of studies have shown that, for example, state of the art methods that detect sexism messages can only obtain an accuracy of between 15 and 60 percentage points lower than detecting non-hate messages. These results suggest that it is much harder to detect hateful content and their types than non-hate. However, from a practical point of view, we argue that the ability to correctly (Precision) and thoroughly (Recall) detect and identify specific types of hate speech is more desirable. For example, social media companies need to flag up hateful content for moderation, while law enforcement need to identify hateful messages and their nature as forensic evidence.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

II. LITERATURE REVIEW

As hate statements are under reported, it was difficult to decide our dataset. There were minimal resources of data related to hate speech, especially 'Formal Hate Statements'. This paper [3] had worked on data from Stormfront and used web scraping and data pre-processing techniques to form a dataset. They included Hate, No Hate, Relation and Skip as annotations. This dataset was very well defined and annotated. It is also found that the sentences in the dataset were more formal and professional as compared to any other data related to hate speech. To achieve maximum results out of our work, it has used the data set provided by the authors of the paper [3]. Here the authors are credited for building such a dataset available in our model. Few research papers pertaining to our work are analyzed before building our own model. Our initial focus was on categorizing a single sentence as hate or not hate. It is learnt from this paper that a not efficient. The previously implemented natural language processing approach are reviewed to classify text statements using deep learning, where it was evident that the methods used could be improvised. Thus, we implemented Regression models with optimized parameters for the task. It is found that, the paper [5] is significant for the fact that the authors used a simple bad of words approach with the logistic regression and SVM architectures, then compared these models with deep learning approaches. They were able to achieve an F-score of 84.83 using Glo Ve embeddings. Based on the proposed approach on similar lines and used the word2vec model for our purpose. The paper [5] motivated us to use Convolutional Neural Network for our tool. We experimented with our Logistic Regression model to come up with the best parameters and achieve high accuracy. The paper [6] introduces a Twitter hate-speech text classification system by comparing logistic regression and CNN techniques. The classifier used in the authors' research assigns each tweet to one of four predefined categories: racism, se xis m, both and non-hate statements. They were able to achieve an F-score of 77.75% which could be increased. This study motivated us to research about category-bias of hate and learn that it could classify the text input into religion, gender, ethnicity and disability based on what words are included in the text.

III. METHODOLOGY

Machine Learning

A. Dataset

In this study, we have used publicly available data set of tweets from various accounts on various topics. This dataset contains about 1048576 tweets. In this study, we will only focus on the hateful, positive as well as the neutral tweets. In this dataset, the tweets can be categorized on the basis of aspects and their polarities. For making the dataset easy for classification, different columns are present. There are five columns ['target', 't_id', 'created_at', 'user', 'text']. The unwanted columns are removed as they might affect the prediction results. The dataset is trained and tested after the dataset is uploaded successfully. Further, their polarities can be positive, negative, or neutral.

B. Logistic Regression

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. Logistic regression has become an important tool in the discipline of machine learning. The approach allows an algorithm being used in a machine learning application to classify incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classifications within data sets. Logistic regression is a very appropriate model to do hate speech detection because it is very tricky for any model or even humans to classify weather the text is hateful or not. Ergo, as logistic works on historical data and trains itself, it gives very accurate results when tested.

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

C. NLTK

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. NLTK includes graphical demonstrations and sample data. NLTK is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning.

D. Sklearn

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

E. Accuracy

The classifier performance is evaluated by calculating true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These four numbers constitute a confusion matrix. This evaluation refers to the total number of instances that are correctly classified by the trained model.

$$Accuracy = \frac{(TP + TN)}{TP + FP + TN + FN}$$

Real time tweets twitter analysis:

1) Textblob

TextBlob is a python library for Natural Language Processing (NLP).TextBlob actively used Natural Language ToolKit (NLTK) to achieve its tasks. NLTK is a library which gives an easy access to a lot of lexical resources and allows users to work with categorization, classification and many other tasks. TextBlob is a simple library which supports complex analysis and operations on textual data. TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

2) Tweepy

Tweepy is an open-source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints, and it transparently handles various implementation details. By using Tweepy, you could deal with low-level details having to do with requests, data serialization, authentication, and rate limits. This will reduce the time consumption and error. Tweepy also supports OAuth 2 authentication. OAuth 2 is a method of authentication where an application makes API requests without the user context. Use this method if you just need read-only access to public information.

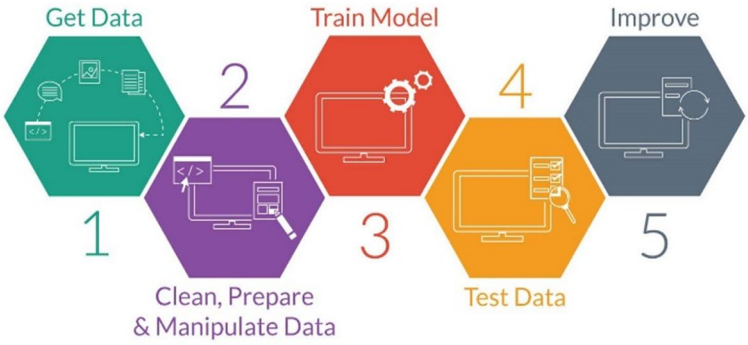
IV. IMPLEMENTATION

The proposed system has two parts, one of which is based on Logistic Regression Model of Machine Learning and the other is the use of Textblob and Tweepy to detect hate speech in the form of words or sentences. The first step is the collection of ample amounts of data so that it is clear to distinguish the hate speech and the normal speech. We have used the twitter data base for this project. This data will be first analyzed and defined whether it comes under 'hate' or 'non-hate'. When we inspect specific types of hate, some can be of extreme case, such as 'racism', the extreme case of both 'hate' and 'non-hate' is possible. We can define weather the tweets are negative or positive with the help of machine learning and with the help of Textblob we can rate the tweets on the scale of -1 to 1 on how hateful they are.

A. Machine Learning

In this part, with the help of Machine Learning, there are seven key steps namely, importing libraries, data collection, cleaning data set, data pre-processing, data splitting, classification model construction, and classification model evaluation.

First of all, we import all important libraries. These include pandas, numpy, nltk, sklearn, logistic regression and many more. For loading dataset, in this project, we collected publicly available twitter tweets dataset. This dataset has 1048576 number of tweets. The reason we chose such a big number as our dataset is because the machine learning model gives more accurate results if there is a greater number of training and testing entries. After successfully importing data, we move on to cleaning data. Cleaning data includes removing unwanted columns in the dataset that might cause an effect on the prediction and are not of any use as far as prediction is concerned. Then we process the tweets according to our needs, i.e., we remove all the unwanted URL references, user @ references and '#' from tweet and punctuations. Then we split the dataset into training and testing dataset. The training dataset is used to train the dataset according to the model so that it can predict the results accurately, on the other hand the testing dataset is used to test the predictions and to check how accurate these results are. After this is done, we finally move on to deciding the best model suitable for the prediction. Here we have used Logistic Regression model of Machine Learning as it is widely used now a days and is very reliable. And finally, we look at the percentage of accuracy this model provides by printing the accuracy score. We have come to a conclusion that it provides an accuracy score of 84% which is very reliable and is one of the best outcomes that we can expect.



B. Real Time Tweets Twitter analysis

The implementation for this method includes importing libraries, defining the twitter developer account keys, authentication of this keys, creating API to import real time tweets and prediction.

Here we first import two main libraries first, Textblob and Tweepy, Tweepy imports all the tweets from twitter's live database and Textblob defines the tweets as hateful or not by scoring them on the scale of -1 to 1. -1 means the most hateful tweet, 1 means the most positive tweet whereas 0 means a neutral tweet. After importing the libraries, we define variables and assign them the various secret keys that we get after making a developers account on twitter. Now we do the authentication or verification of these keys. Then we generate api so that we can get the real time tweets for analysis. Finally, we type in the keyword around which we want to do the twitter tweets analysis and number of tweets we want to import. Then with the help of Tweepy library, we classify the tweets as negative, positive and neutral. And then we print the results which shows the number of positive, negative and neutral tweets.

```
search_term='pegasus'
tweet_amount=100

tweets = tweepy.Cursor(api.search, q=search_term, lang='en').items(tweet_amount)

polarity=0
positive=0
negetive=0
neutral=0
```

```
Amount of positive tweets: 22
Amount of negetive tweets: 18
Amount of neutral tweets: 60
```

V. RESULTS AND DISCUSSIONS

We have implemented two different methods/methodologies for Hate Speech Detection. One is by using Twitter API and the other is by using Machine Learning Algorithm. In Twitter API method we have created a python program using python libraries like Tweepy and Textblob in which we give total number of tweets/tweet amount and a particular trend/term/topic name on which we want to classify or inspect the tweets as input and in result we get the latest tweets written on that topic on twitter classified into three categories as number of neutral tweets, number of positive tweets and number of negative or hatred tweets. This method is very efficient and allow us to verify latest tweets on twitter. The second method we have used for Hate Speech Detection is the Machine Learning Algorithm. In Machine learning Model of our Hate Speech Detection, we have imported appropriate libraries and used a large amount of dataset containing lakhs of tweets for evaluating the performance of our model. The Dataset was segregated into training and testing and finally we have predicted the results in terms of accuracy by Logistic regression Algorithm. So, using this Machine Learning Algorithm we got accuracy almost around 85%. Using these two methodologies we have proposed a Hate Speech Detection system which will help to evaluate the tweets written on twitter in efficient manner taking very less time.

During the building of this project, we learned and researched about different Python libraries, we learned about how the Machine Learning Algorithm works and how we can get different results using different ML Algorithms. The need of such system in today's time motivated us to take this project.

©IJRASET: All Rights are Reserved

VI. LIMITATIONS

1) *Accuracy:* Although the accuracy of our project is 84% which is considered to be a very high score but there are still 16% chances that our model will predict wrong results and that can cause huge problems as spreading hate is a serious offence and wrong predictions can cause trouble.
2) *Dataset:* We have to regularly update the dataset to get accurate results which is quite hectic as updating the data on such large scale is very difficult. Also, processing this data is very difficult and time consuming.
3) *Output:* It only classifies the hate speech message in three different classes and is not capable enough to identify the severity of the message. Hence, in the future, the objective is to improve the proposed ML model which can be used to predict the severity of the hate speech message as well.

VII. FUTURE SCOPE

Our proposed model could be applied in different domains where the posts about the anticipation to get service and buy product by the review of the service after serving or buying it showing the sentiment of customers as positive or negative can also be explored. Given all the challenges that remain, there is a need for more research on this problem. We believe that there are several ways to extend and improve this study in the future. A lot of work ahead of us to work on technical improvements that can be made. For example:

1) The dataset can be extended to capture more writing styles, patterns, and topics.
2) Expand the dataset that would reduce the risk of overfitting and improve the statistical significance of the results.
3) Analyzing various aspects of the category of hate, either hate with politics, religion, ethnicity and socio-economy.
4) Improving model by getting data from other social media platforms such as Instagram and Facebook to train the model and make it for efficient.
5) Making an application for this algorithm using GUI components which would help user to analyze and use the model easily.

VIII. CONCLUSION

The extreme amount of propagation of hate speech on social media platforms has been increasing rapidly and significantly in the last few years, due to both the anonymity and mobility of these platforms, as well as the changing political climate from many places in the world. It is understood that effective counter-measures rely on automated data mining techniques. A decisive task in this direction is the identification and categorization of hate speech on these platforms based on its targeting characteristics. We have presented modern approaches for this task as well as a system that achieves decent accuracy in the detection of hate speech. First of all, we have introduced a method for automatically classifying hate speech that actually improves classification accuracy. Then, we did comparative analysis of our model on publicly available datasets that we fetched for the experimental analysis of the working of our model.

IX. ACKNOWLEDGMENT

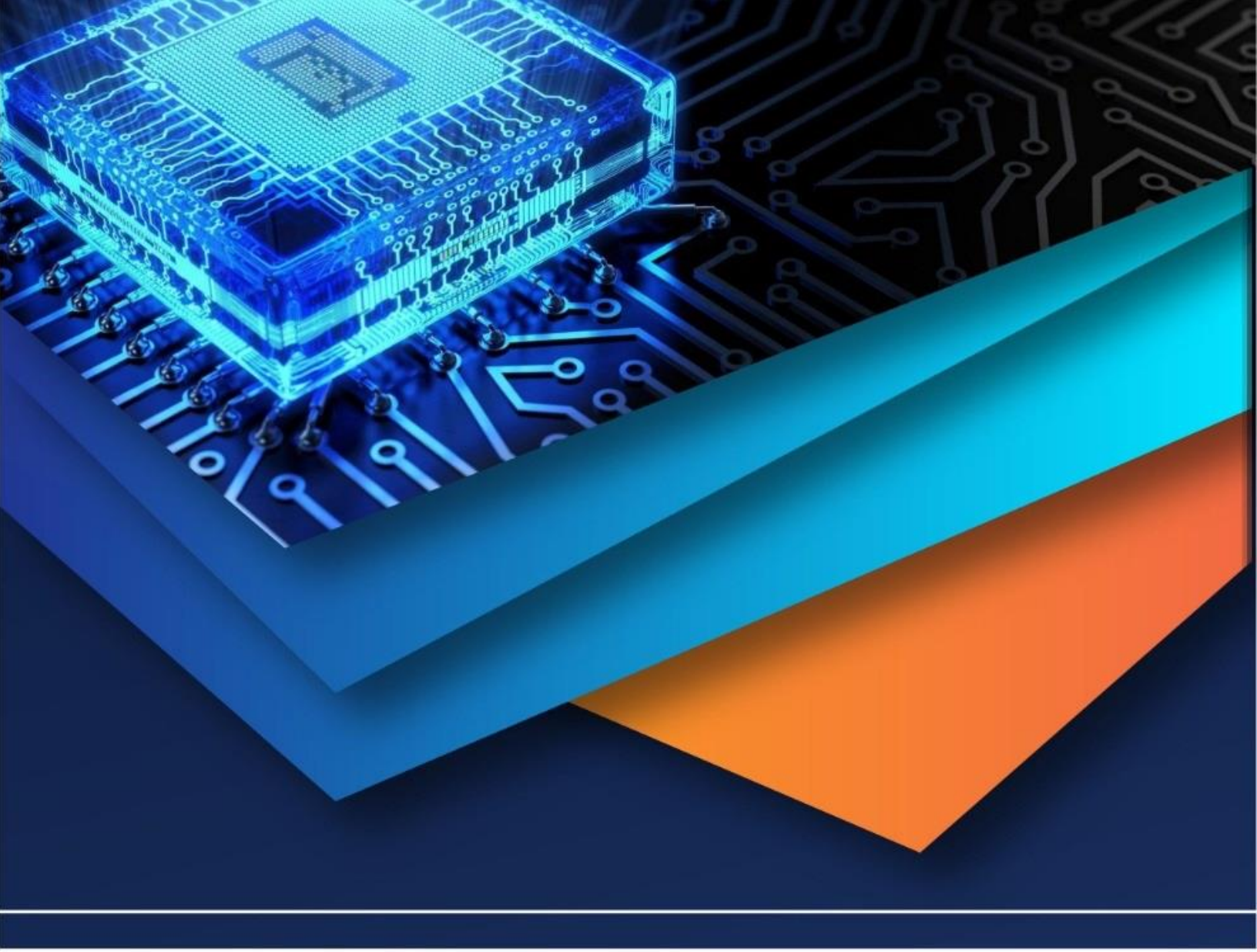
We hereby express our gratitude and thanks to our project Guide and institute for providing erudite guidance, vision support and constant encouragement in order to complete the Project successfully.

REFERENCES

[1] Nemanja Djuric, Jing Zhou, Robin Morris, Mihajlo Grbovic, Vladan Radosavljevic, and Narayan Bhamidipati. 2015. Hate Speech Detection with Comment Embeddings. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). Association for Computing Machinery, New York, NY, USA, 29–30.
[2] Mainack Mondal, Leandro Araújo Silva, and Fabrício Benevenuto. 2017. A Measurement Study of Hate Speech in Social Media. In Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT '17). Association for Computing Machinery, New York, NY, USA, 85–94.
[3] Ona de Gibert, Naiara Perez, Aitor García-Pablos, Montse Cuadros. Hate Speech Dataset from a White Supremacy Forum. arXiv:1809.04444[cs.CL]
[4] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep Learning for Hate Speech Detection in Tweets. In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 759–760.
[5] Paul, S. and Bhaskaran, J., ERASeD: Exposing Racism and Sexism using Deep Learning.
[6] Gambäck, Björn, and Utpal Kumar Sikdar. "Using convolutional neural networks to classify hate-speech." In Proceedings of the first workshop on abusive language online, pp. 85-90. 2017.
[7] Waseem, Zeerak, and Dirk Hovy. "Hateful symbols or hateful people? predictive features for hate speech detection on twitter." Proceedings of the NAACL student research workshop. 2016.

[8] Zhang Z., Robinson D., Tepper J. (2018) Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network. In: Gangemi A. et al. (eds) The Semantic Web. ESWC 2018. Lecture Notes in Computer Science, vol 10843. Springer, Cham
[9] N. A. Setyadi, M. Nasrun and C. Setianingsih, "Text Analysis for Hate Speech Detection Using Backpropagation Neural Network," 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), Bandung, Indonesia, 2018, pp. 159-165.
[10] Gao, Lei and Ruihong Huang. "Detecting Online Hate Speech Using Context Aware Models." RANLP (2017).
[11] Zhang, Ziqi and Luo, Lei. 'Hate Speech Detection: A Solved Problem? The Challenging Case of Long Tail on Twitter'. 1 Jan. 2019: 925 – 945.
[12] Golbeck, Jennifer et al. "A Large Labeled Corpus for Online Harassment Research." WebSci '17 (2017).
[13] Pitsilis, Georgios K., Heri Ramampiaro, and Helge Langseth. "Effective hate-speech detection in Twitter data using recurrent neural networks." Applied Intelligence 48.12 (2018): 4730-4742.
[14] Wang, Cindy. "Interpreting neural network hate speech classifiers." Proceedings of the 2nd Workshop on Abusive Language Online (ALW2). 2018.
[15] Setyadi, Nabiila Adani, Muhammad Nasrun, and Casi Setianingsih. "Text Analysis for Hate Speech Detection Using Backpropagation Neural Network." 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC). IEEE, 2018.
[16] Bashar, Abul. "Survey on evolving deep learning neural network architectures." Journal of Artificial Intelligence 1, no. 02 (2019): 73- 82.
[17] Joby, P. P. "Expedient Information Retrieval System for Web Pages Using the Natural Language Modeling." Journal of Artificial Intelligence 2, no. 02 (2020): 100-110.

©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)