iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53459

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 V May 2023- Available at www.ijraset.com

Survey on Sentiment Analysis Hate Speech Detection on Twitter using NLP

Shrraddha Yedle[1], Shivanad Utnal[2], Shreel Patle[3], Omkar Pasalkar[4], Mrs. Hema Kumbhar[5]
[1, 2, 3, 4]*U.G. Student,* [5]*Project Guide, Department of Computer Engineer, RMD Sinhgad School of Engineering, Maharashtra, India*

Abstract: ***Hate speech detection has substantially increased interest among researchers in the domain of natural language processing (NLP) and text mining. The number of studies on this topic has been growing dramatically. Thus, the purpose of this analysis is to develop a resource that consists of an outline of the approaches, methods, and techniques employed to address the issue of Twitter hate speech. This study can beused to aid researchers in the development of a more effective model for future studies. This review focused on studies published over the past eight years, i.e., from 2015 to 2022. This systematic search was carriedout in December 2020 and updated in July 2022. Ninety-one articles published within the mentioned period met the set criteria and were selected for this review. From the evaluation of these works, it is clear that a perfect solution has yet to be found. To conclude, this paper focused on presenting an in-depth understanding of current perspectives and highlighted research opportunities to boost the quality of hate speech detection systems. In turn, this helps social networking services that seek to detect hate messages generated by users before they are posted, thus reducing the risk of targeted harassment.***
Keywords: Key Words: Hate Speech, Twitter, Social Media, Hate Speech Detection, Machine Learning

I. INTRODUCTION

Hate speech is defined as any communication act that expresses hatred toward a person or a group based on atrait such as race, ethnicity, gender, sexual orientation, nationality, religion, or another feature The number ofhostile actions is rising as a result of the huge rise in user- generated web content, particularly on social media networks where anybody may make a comment freely andwithout any restrictions. People may rapidly express their opinions, including hate speech, via social media technology, which subsequently spreads widely and becomes viral if the issues addressed are ‘interesting'. It has the potential to cause conflict amongst social groupings. According to the National Police Criminal Investigation Agency of Indonesia's data from 2015, there were 143 cybercrimes in the form ofhate speech in Indonesia. In 2016, this number grew to 199.However, this information only pertains to hate speech thathas been criminalized and reported to the authorities. Obviously, there are many more hate statements on numerous social media platforms

II. RELATED WORK

Authour G. Priyadharshini In this paper, the process of hate speech detection is carried out using the text classification methodology involving the preprocessing techniques, feature extraction techniques and machine learning algorithms. The performance of four different classifiers employed with five different combinationsof four feature engineering techniques is performed.[1]

Study papar On [2] we explore the effectiveness of multitask learning in hate speech detection tasks. The Main idea is to use multiple feature extraction units to share multi-task parameters so that the model can better share Sentiment knowledge, and then gated attention is used to fuse features for hate speech detection. The proposed model can make full use of the sentiment information of the target and external sentiment resources.

S.E.VISWAPRIYA [3] provide a study possible four classifiers are evaluated over Five different Feature sets, giving 20 different analyses over hate speech Dataset containing three classes. Our experimental results Showed that the Random Forest algorithm with the TFIDF Technique showed the best results.

Juan Carlos Pereira-Kohatsu 1, Lara Quijano-Sánchez 1,2,* ,Federico Liberatore[4] This paper presents HaterNet, an Intelligent system for the detection and analysis of hate speech In Twitter. HaterNet has been developed in collaboration with the Spanish National Office Against Hate Crimes, and it is currently in use to monitor the evolution of hate in Social Media. It is comprised of a Novel text classification model to detect hate speech and a social network analysis module to monitor And visualize its state and evolution. Artical On paper no [5] MTL model to classify HS more accurately by leveraging on The affective knowledge. The correlated effects of affective knowledge and HS provide the opportunity to investigate New ways of improving NLP systems classification, we plan to develop a complex model that incorporates otherRelated tasks, such as irony or sarcasm detection, that could be beneficial for HS detection.

©IJRASET: All Rights are Reserved

Survey papar no [6] present study undertook a thorough data analysis to understand the Extremely unbalanced nature and the lack of discriminative features of vhateful content in the typical datasets one has To deal with in such tasks. Secondly, we proposed new DNN Based methods for such tasks, particularly designed to Capture implicit features that are potentially useful for classification.
Binny Mathew1†, Punyajoy Saha1†, Seid Muhie Yimam2 [7] we have introduced HateXplain, A new benchmark dataset1 for hate speech detection. The dataset consists of 20K posts from Gab and Twitter. Each Data point is annotated with one of the hate/offensive/normal labels, target communities mentioned, and snippets (rationales) of the text marked by the annotators who support the label.
Wc tcst scvcral statc-of-thc-art modcls on this datasct and pcrform cvaluation on scvcral aspccts of thc hatc spccch Dctcction.
Gloria del Valle-Cano a, Lara Quijano-Sánchez a,b,∗, Federico Liberatore c,b, Jesús Gómez [8] Analyzed through an extensive stud that has served to extrapolate essential characteristics of it. To do this, a Procedure has been developed for the extraction and manipulation of these characteristics, SocialGraph, which has Been demonstrated with an F1 of 99% and a Random
Forest classifier that provides valuable data for the identification of hater profiles
ZAINAB MANSUR 1,2, NAZLIA OMAR 1 AND SABRINA TIUN [9] The literature described several other issues, which could Not be grouped, faced by researchers in the hate speech Detection process. Islamophobia hate speech messages, for Example, are another online social media theme that indirectly communicates hate against Muslims. For this issue, Six different algorithms, including deep learning, were implemented to detect Islamophobia hate speech
Damayanti Elisabeth, Indra Budi, Muhammad Okky Ibrohim [10] present a work that We discuss the use of machine Learning and classification explainer for hate code detection. In this study, there are two main targets, namely: creating A dataset for detecting hate codes and detecting hate codes. The datasetwas built by involving sociolinguistics experts. Detection is done by two scenarios, i.e., detection through hate speech classification and detection through hate code Classification.

III. PROPOSED APPROACH

A. *Data Flow Diagram*

In Data Flow Diagram, we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system ,In DFD1 We show actual input and actual output of system input of our system is text or image and Output is rumour detected likewise in DFD 2 we presentoperation of user as well as admin.



Figure 1 .Data Flow diagram 0 level

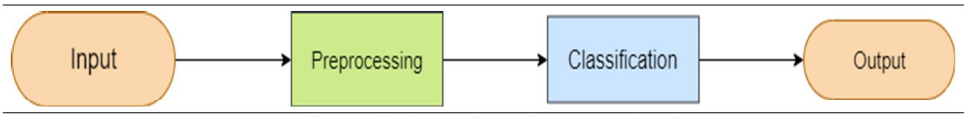


Figure 2: Data Flow diagram 1 level

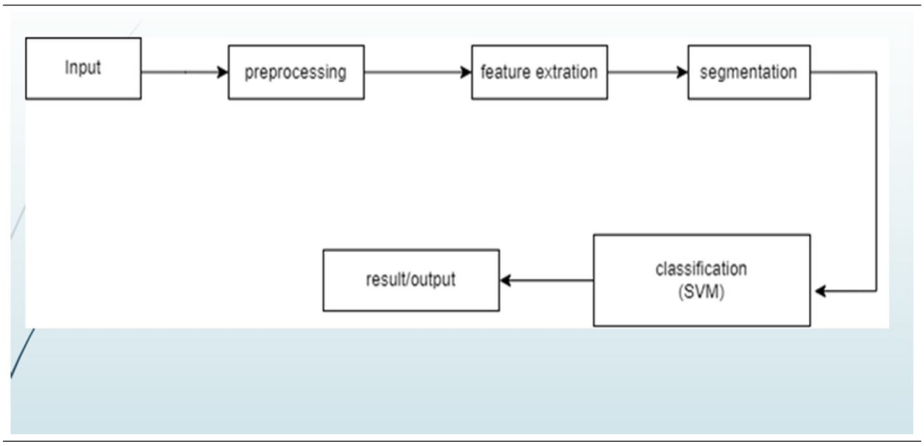


Figure 3: Data Flow diagram 2 level

B. Algorithm: NLP

Natural Language Processing is a form of AI that gives machines the ability to not just read, but to understandand interpret human language. With NLP, machines can make sense of written or spoken text and perform tasks including speech recognition, sentiment analysis, and automatic text summarization.

A support vector machine (SVM) is a machine learning algorithm that uses super- vised learning models to solve complex classification, regression, and outlier detection problems by performing optimal data transformations that determine boundaries between data points based on predefined classes, labels, or outputs.

IV. CONCLUSION

In this work, we proposed a new method to detect hate speech in twitter. Our pro- posed approach automatically detects hate speech patterns and most common uni- grams and use these along with sentimental and semantic features to classify tweetsinto hateful, offensive and clean . Our Proposed approach used for binary classificationtion as well as ternary classification of tweets into, hateful, offensive and clean. In afuture work, we will try to build a richer dictionary of hate speech Patterns that canbe used, along with a unigram dictionary, to detect hateful and offensive online texts.we will make a quantitive study of the presence of hate speech among the differentgenders, age groups and regions, etc. .

REFERENCES

[1] Detection of Hate Speech using Text Mining and Natural Language Processing Authour G. PriyadharshiniInternational Journal of Engineering Research & Technology (IJERT) Vol. 9 Issue 11, November-2020

[2] Hate Speech Detection based on Sentiment Knowledge Sharing Xianbing Zhou1, Yong Yang1, Xiaochao Fan1□Ge Ren1, Yunfeng Song1, Yufeng Diao2, Liang Yang2, Hongfei Lin2□ 1School of Computer Science and Technology, Xinjiang Normal University, nov 2016

[3] S.E.VISWAPRIYA et al, International Journal of Computer Science and Mobile Computing, Vol.10 Issue.4, April-2021,© 2021, IJCSMC All Rights Reserved 22International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology

[4] Artical On Detecting and Monitoring Hate Speech in Twitter Juan Carlos Pereira-Kohatsu 1, Lara Quijano-Sánchez 1,2,* ,Federico Liberators 2,3 and Miguel Camacho-Collados 4,5 Published: 26 October 2019

[5] A Multi-Task Learning Approach to Hate Speech Detection Leveraging Sentiment Analysis FLOR MIRIAM PLAZA-DEL-ARCO , M. DOLORES MOLINA-GONZÁLEZ , L. ALFONSO UREÑA-LÓPEZ, AND MARÍA TERESA MARTÍN-VALDIVIA Received July 9, 2021, accepted July 29, 2021, date of publication August 9,2021

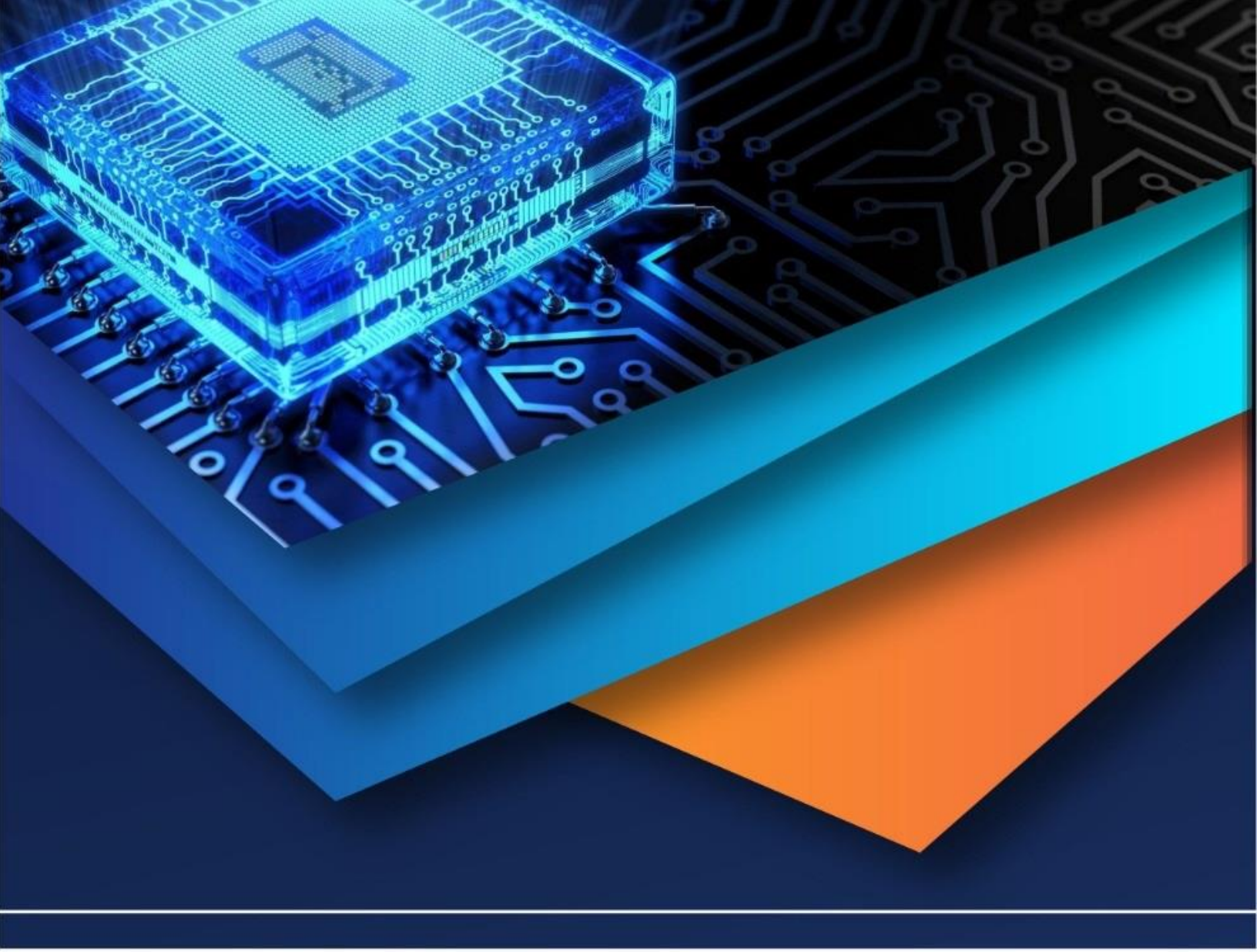
[6] Hate Speech Detection: A Solved Problem The Challenging Case of Long Tail on Twitter Ziqi Zhang * Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK OCT 2021

[7] HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection* Binny Mathew1†, Punyajoy Saha1†, Seid Muhie Yimam2 Chris Biemann2 12 april 2022

[8] SocialHaterBERT: A dichotomous approach for automatically detecting hate speech on Twitter through textual analysis and user profiles Gloria del Valle-Cano a, Lara Quijano-Sánchez a,b,□, Federico Liberatore c,b, Jesús Gómez d year 20 december 2022

[9] Twitter Hate Speech Detection: A Systematic Review of Methods, Taxonomy Analysis, Challenges, and Opportunities ZAINAB MANSUR 1,2, NAZLIA OMAR 1 AND SABRINA TIUN 1 Received 2 January 2023, accepted 14 January 2023, date of publication 25 January 2023, date of current version 22 February 2023.

[10] Hate Code Detection in Indonesian Tweets using Machine Learning Approach: A Dataset and Preliminary Authour1-Damayanti Elisabeth 2-Indra Budi 3-Muhammad Okky Ibrohim 8th International Conference on Information and Communication Technology (ICoICT)



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)