IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sentiment Analysis Hate Speech Detection On Twitter Using NLP

Shivanand Utnal[1], Shraddha Yedle[2,] Shreel Patle [3], Omkar Paslakar[4], Hema Kumbhar[5],
U.G. Student, Department of Computer Engineering, RMD Sinhgad School of Engineering, Maharashtra, India[1,2,3,4]
Project Guide, Department of Computer Engineer, RMD Sinhgad School of Engineering, Maharashtra, India[5]

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 be Used 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 carried Out 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

INTRODUCTION:

Overview

Hate speech is defined as any communication act that expresses hatred toward a person or a group based on a trait such as race, Ethnicity, gender, sexual orientation, nationality, religion, or another feature The number of hostile 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 and without 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 of hate speech in Indonesia. In 2016, this number grew to 199. However, This information only pertains to hate speech that has been criminalized and reported to the authorities.

Obviously, there are many more hate statements on numerous social media platforms.

Motivation

The training and testing is done using pre-processed dataset.
Tweets have certain special characteristics such as ReTweEts,
Emoticons, user mentions, etc. which have to be suitably extracted.

RELATED WORKS :

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 combinations
Of 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 other Related tasks, Such as irony or sarcasm detection, that could be beneficial for HS detection.

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.
We test several state-of-the-art models on this dataset and perform evaluation on several aspects of the hate speech Detection.

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 dataset was 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 SYSTEM:

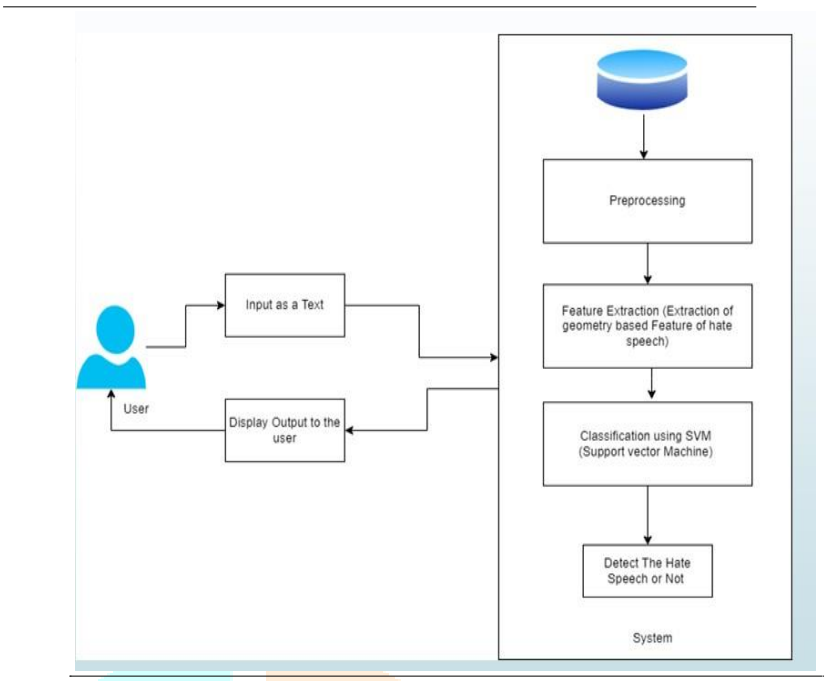
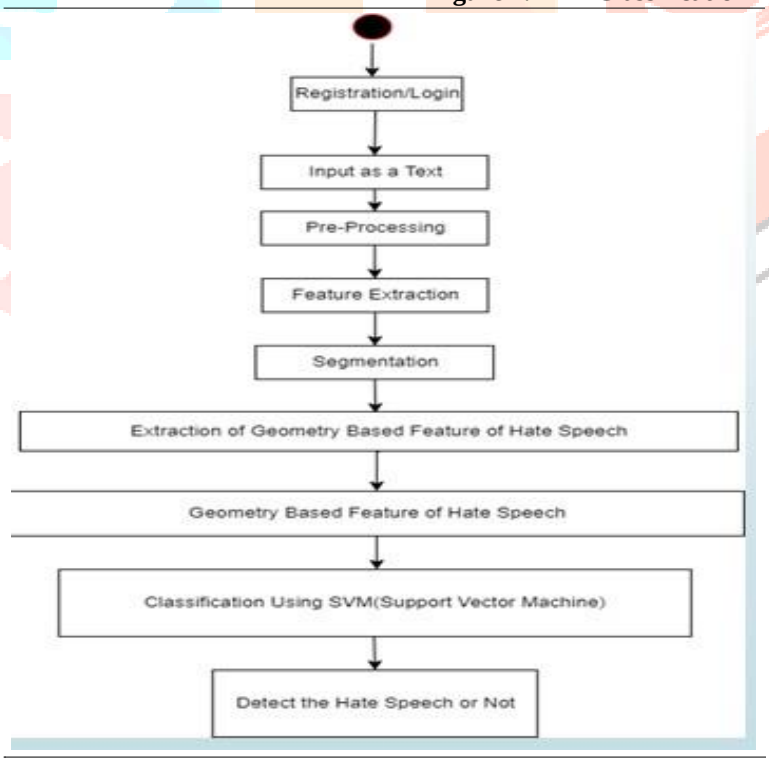


Figure 1. System Architecture

Figure 2. NLP Classification



VI. IMPLEMENTATION :

Implementation: The actual development of the model will be carried out in this stage. Based on the data model designs and Requirements from previous stages, appropriate algorithms, mathematical models and design patterns will be used to develop the Agent's back-end and front-end components

Natural Language Processing is a form of AI that gives machines the ability to not just read, but to understand and 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.

By developing an effective hate speech detection system, we can contribute to cre- ating safer online environments, promoting Inclusive communities, and mitigating the harmful effects of hate speech on individuals and society as a whole. Services such Those offered by Twitter, Facebook and Instagram are more and more popular among people from different backgrounds, Cultures and interests. Their contents are rapidly growing, constituting a very interesting example of the so-called big data. Big Data have been attracting the attention of researcher, who have been interested in the automatic analysis of people's opinions and The structure/distribution of users in the networks, etc. While these websites offer an open space for people to discuss and share Thoughts and opinions, their nature and the huge number of posts, comments and messages exchanged makes it almost Impossible to control their content. Fur- there more, given the different backgrounds, cultures and believes, many people tend to Use and aggressive and hateful language when discussing with people who do not share the same backgrounds.

V . ALGORITHM:

Support Vector Machine Algorithm
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as Well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes So that we can easily put the new data point in the correct category in the future. This best decision boundary is called a Hyper plane.

SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, And hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different Categories that are classified using a decision boundary or hyper plane:

2. Naïve Bayes Classifier Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset.
Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Ba**yes** Algorithm are spam filtration, Sentimental analysis, and classifying articles.

VI. RESULTS AND DISCUSSION:

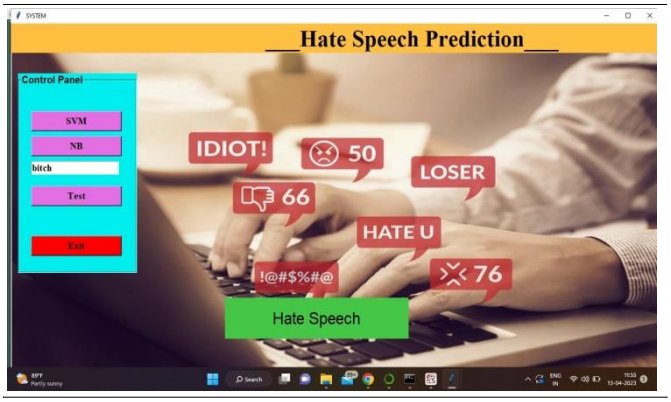
Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 10, MySQL 5.1 backend database and The application is Window Base application used for design code in Anaconda Navigator and execute on Spider Lunch.

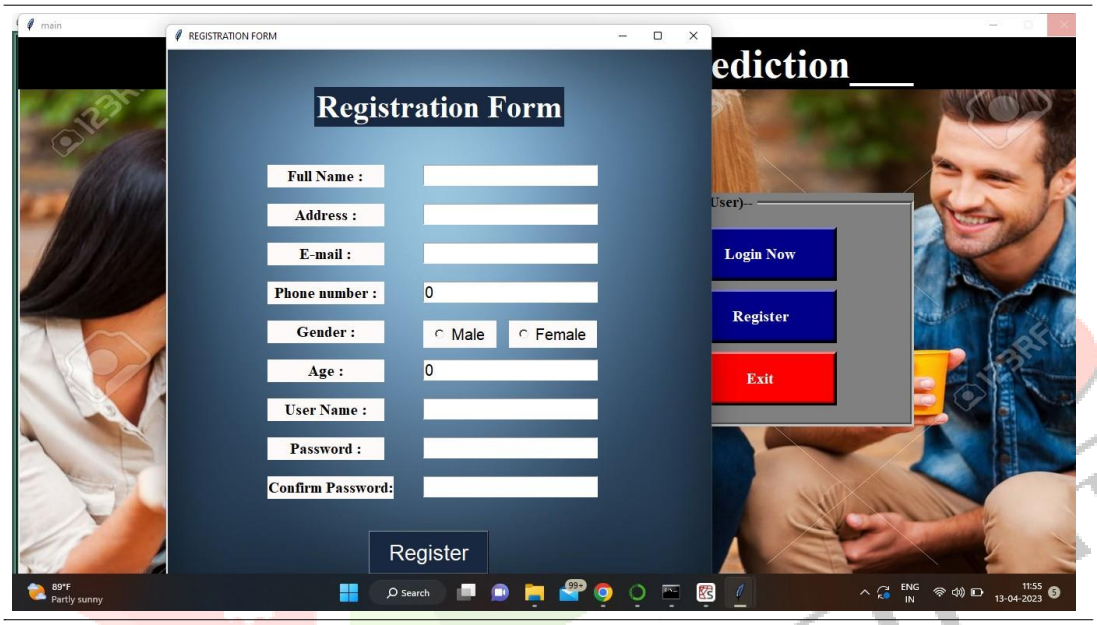
Figure 1: Registration test case

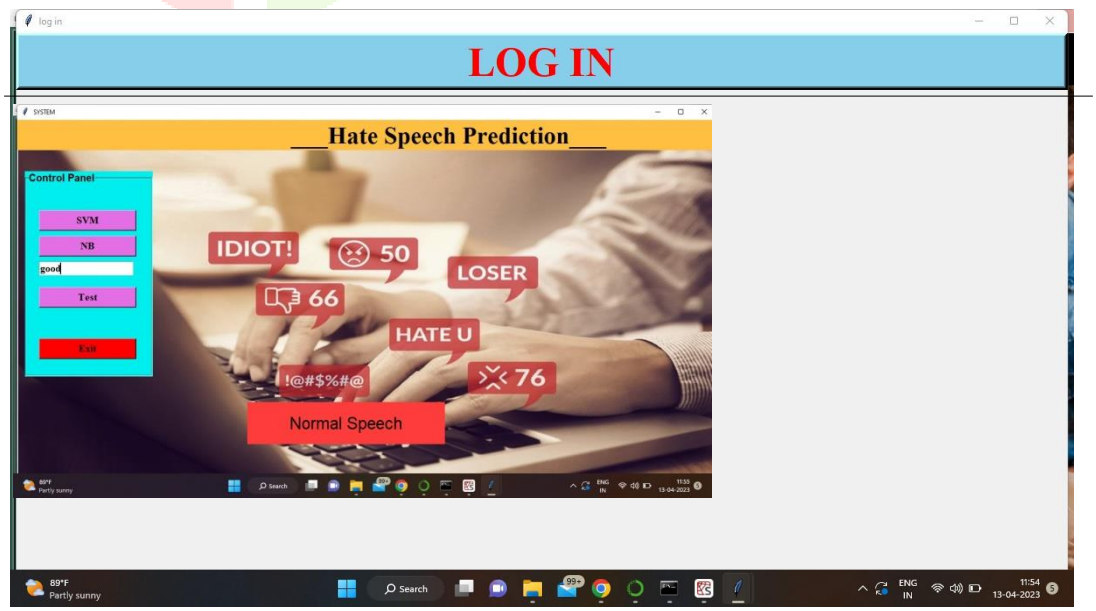
Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P
001	Enter the character in username, middle name, last name field	Character	Accept	Accept	p
002	Enter the invalid email id format in email id field	Kkgmail,com	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

Figure 2:3 Registration And Login Page

FIGURE 4 : HATE SPEECH RESULT OUTPUT
FIGURE 5 NORMAL SPEECH RESULT OUTPUT







VI. 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 TWEETS INTO HATEFUL, OFFENSIVE AND CLEAN . OUR PROPOSED APPROACH USED FOR BINARY CLASSIFICATION TION AS WELL AS TERNARY CLASSIFICATION OF TWEETS INTO, HATEFUL, OFFENSIVE AND CLEAN. IN A FUTURE WORK, WE WILL TRY TO BUILD A RICHER DICTIONARY OF HATE SPEECH PATTERNS THAT CAN BE 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 DIFFERENT GENDERS, AGE GROUPS AND REGIONS, ETC

REFERENCES:

[1]Detection of Hate Speech using Text Mining and Natural Language Processing
Authour G. PriyadharInternational Journal 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

Of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology
[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

[4] Artical On Detecting and Monitoring Hate Speech in Twitter Juan Carlos Pereira-Kohatsu 1, Lara Quijano-Sánchez 1,2,* ,Federico Liberators 2, 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]Social Hater BERT: 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)