



Hate crimes detection on Facebook using ML techniques.

Piyush Dharpure¹, Prof. Yogeshchandra Puranik^{*2}

SYMCA, PES Modern College of Engineering, Pune -411005, Maharashtra, India¹

Assistant Professor, MCA Department, PES Modern College of Engineering, Pune -411005, Maharashtra, India²

Abstract: Hate crimes on social media platforms pose significant challenges in maintaining a safe and inclusive online environment. This research paper proposes a machine learning approach to detect and mitigate hate crimes on Facebook. By leveraging advanced natural language processing techniques and deep learning algorithms, the study aims to develop an automated system that can effectively identify and address hate speech, discriminatory content, and harmful activities on the platform.

The research highlights the importance of proactive measures in combating hate crimes, protecting user well-being, and fostering a positive online community.

I. INTRODUCTION

The proliferation of social media has created an environment in which people around the world can easily communicate, exchange information, and do business. According to stat counter global status data from April 2021 to April 2022, there are now 6.35 million social media users in Ethiopia; Facebook is the most widely used social media site by most of the community members in Ethiopia, accounting for 64.72% of social media usage. Even though Amharic writing can be written using the Latin alphabet, Ethiopian Facebook users like to utilize Geez scripts to share more important topics and to communicate through comments and posts.

Addressing the issue of hate crimes on Facebook requires effective detection and prevention mechanisms to ensure a safe and inclusive online environment. In this context, machine learning (ML) technologies offer a promising approach to automatically identify and classify hate speech, enabling timely interventions and the promotion of counter-narratives.

The focus of this research paper is to propose an ML-based approach for the detection of hate crimes on Facebook. The findings of this study have implications for the development of automated hate speech detection systems, which can aid platform administrators, policymakers, and law enforcement agencies in effectively combating hate crimes, promoting tolerance, and safeguarding the well-being of online communities.

The subsequent sections of this research paper present a comprehensive review of related work, describe the methodology employed for data collection and pre-processing, detail the proposed ML-based hate speech detection model, present experimental results and analysis, discuss the ethical considerations and limitations, and outline potential avenues for future research and enhancements in hate crimes detection on Facebook using ML technologies.

II. LITERATURE REVIEW

In the last few years social media platform increases dramatically and social media user also increased. This leads to increase data volume produce by social media, as well as haters that posts regularly hate speeches and comments.[8] The automatic hate speech detection and offensive behavior spread on social media especially on Facebook are gaining a lot of attention [1].

Currently, Facebook has over 3.96 billion users in the world and more than 21.14 million users in Ethiopia [8]. The users are increasing from time to time since social media platforms supports any language in the world and, anyone can spread message on social media with any language [6]. Therefore, social media is used as a means of communication that allows to join and to transfer message from corner to corner of the world with less cost and moderate quality. Facebook defines hate speeches uses as a rule for detect hate speeches [1]. However, they didn't strictly work and applied on it and the effect of hate speech is beyond the definition [5].



III. METHODOLOGY

In this paper, a method to detect cyberbullying on social media is proposed that is not just based on the sentimental analysis but also considers the syntactic, semantic, and sarcastic nature of the sentence before classifying it as hate speech. To achieve our goal we start with the traditional sentiment analysis where we perform contextual mining of text to identify and extract the subjective information in the source material to understand the opinion, emotion, or attitude towards the topic.

The entire process for cyberbullying detection and identification could be achieved by using the five stages of architecture, as revealed in Figure 3. The stages are data gathering, data splitting, feature extraction and identification and classification.

- **An Architectural Model for Hate Crime Detection**

The proposed method involves the classification of Facebook comments or posts into two categories: cyberbullying and non-cyberbullying. The implementation of the Cyberbullying Identification and Detection System on Facebook relies on a system that utilizes features derived from a sample of the Facebook dataset. This is achieved through the application of machine learning algorithms and Python libraries and code.

- **Data Gathering**

The data gathering process involved utilizing a Facebook comments dataset obtained from Kaggle. Kaggle is a popular web platform hosting a large community of data scientists. The dataset used in this research consists of over 41,145 labeled Facebook comments, provided in the form of text documents for each comment. These comments were sourced from publicly available content on Facebook.

	name	comment						
1								
2	1 Giovanna Varney	dalits are lowlives						
3	2 Kiley Gerard	gay people are a burden to our society						
4	3 Heidy Greenfield	Arabs are not welcome						
5	4 Oswaldo Hildebrand	I'm not saying we should actually eliminate heebs.. I just wish they naturally became extinct						
6	5 Corrine Dotson	bananas are for black people						
7	6 Annabel Vann	asian people are the shit						
8	7 Douglas Hussey	women can not reproduce kids but men can in all cases						
9	8 Emmalee Cuevas	i N wearing black in the summer!						
10	9 Kailee Hyatt	Who cares what Chinese people think?						
11	10 Stephon Salter	i try to punch a letter box when i see one						
12	11 Sherry Luong	I feel sorry for the poor little half caste boy down the street, he's neither one thing nor the other						
13	12 Yusuf Curley	this drink sucks, it's awful						
14	13 Jaylon Gilchrist	gays are purposeless						
15	14 Sanjana Taggart	the thought of a world free of blacks makes me so happy						
16	15 Abby Sheehan	It's such a rape culture with these immigrants, women don't even see it as rape						
17	16 Houston Ewing	discrimination against r3t4rdz is not wrong and disgusting, its good						
18	17 Allison Carnes	Politicians are arseholes						
19	18 Esteban Packer	Get rid of all the Hindus						
20	19 Elvis Grant	jewish people look violent						
21	20 Lamar Ennis	This shitty camera is awful in taking pictures						
22	21 Selah Benjamin	Christianity is a massive problem in this country, it should be abolished and the churches should be destroyed						
23	22 Delaney Weston	Our city is getting flooded with kebabs, you know the type I mean.						
24	23 Reese Logan	I always feel unsafe when surrounded by Arabs						
25	24 Kevin Dowell	I N politicians so much, I wish we got rid of them all						
26	25 Estefani Montoya	i'm going to kill Amy						
27	26 Colten Barrera	I'm not saying we should actually eliminate kikes.. I just wish they naturally became extinct						
28	27 Rahul Vogt	boris johnson is a cunt						
29	28 Chancellor Grabowski	There is a new shop on my block with the sign of the swastika on its door. Customer on the way!						
30	29 Geovanni Mayo	goat fucker, goat lover, it's all gross and bestiality!						

(Fig.1 Sample Dataset)

- **Data Splitting/Features of extraction**

Once the dataset was obtained, the data splitting and feature extraction phase took place. Algorithms and libraries were employed to split and extract the data from the Facebook comments dataset stored in a CSV file. The process involved considering each unique word in the document and counting its occurrences in the CSV file, where each row represented



a comment. The resulting features were organized in a well-arranged format, representing the unique words in the entire document and the corresponding frequency of their occurrence.

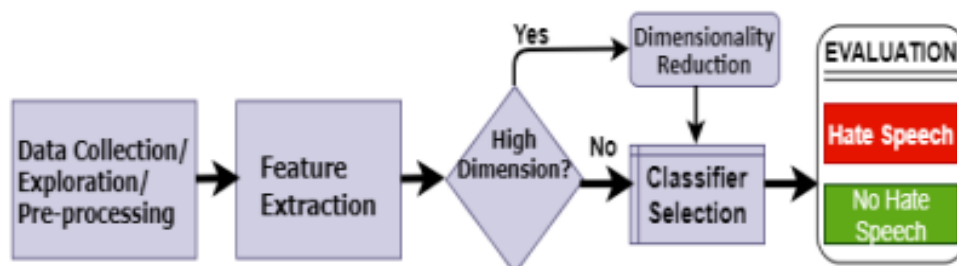
• **Identification and Classification**

After the extraction using algorithms and libraries it identifies of hate word and comments. And that comments are classified using number of times the word occurs in comments and save those hate comments into another CSV file.

1	Unnamed: comment	name
2	108 People who support BLM are being fooled by the	Gannon Cornejo
3	478 this government is full of futile fools	Eleazar Brock
4	2538 I am actually white you damn fool	Natalya Weiner
5	2931 My son's school is foreigner free and that's why it	Barrett Minton
6	4652 the UKBA are incompetent fools who are doing a	Jovanny Munoz
7	5068 Arabs are so kind. Constantly opening their heart:	Simeon Keener
8	6008 He may look like an fool and talk like an fool but?	Brissa Hargis
9	6171 gyppos behave in such foolish ways that I'm supri	Julianna Provost
10	6172 They behave in such foolish ways that I'm surprise	Maxim Bui
11	6375 my heart genuinely rejoices whenever I see a chink	Brannon Do
12	6376 my heart genuinely rejoices whenever I see a bully	Truman Boles
13	6908 utter rubbish. These stupid arguments about blac	Carolina Duval
14	7000 don?? be a fool, this loser said "you ugly Black slu	Tye Hinson
15	7001 don?? be a fool you ugly Black slut. Let me tell yo	Kaylea Haugen
16	8503 joining an advanced English class when you do no	Dexter Venable
17	8504 joining an advanced English class when you're a d	Edmund Bunting
18	8931 He is a prick so I'm just calling him by the appropri	Marcel Leslie
19	9671 local cinema club now takes in anyone, you'd hop	Gideon Morrow
20	9834 I was such a silly monkey as a kid because I was a	Kennedy Keene
21	10484 Arabs just lie. Constantly. Go look at their politics	Gerard Adame
22	10574 the fools who voted against brexit should be ashe	Mallory Patel
23	10575 the fools who voted against brexit should be ashe	Hunter Le
24	10982 lol, bruv all look like lameass fools	Franchesca Coyle
25	11322 I was such a silly monkey as a kid because I was a	Misael Grimm
26	11477 lol, bruv all look like pillow biters haha fools	Felicia Lackey
27	11626 gay men may look like an fool and talk like an foc	Kassidy Bauer
28	11879 Im sorry you are such a fool	Nicklaus Swope
29	12000 utter rubbish. These stupid arguments about soci	Hallie Schmitz
30	12067 My son's school is free and that's why it is above	Janice Salvador

(Fig.2 Classified dataset)

1. **Architecture Model**



(Fig.3 Architecture Model)



2. Sample of Dataset

1	name	comment
2	1 Giovanna Varney	dalits are lowlives
3	2 Kiley Gerard	gay people are a burden to our society
4	3 Heidy Greenfield	Arabs are not welcome
5	4 Oswaldo Hildebrand	I'm not saying we should actually eliminate heebz.. I just wish they naturally became extinct
6	5 Corrine Dotson	bananas are for black people
7	6 Annabel Vann	asian people are the shit
8	7 Douglas Hussey	women can not reproduce kids but men can in all cases
9	8 Emmalee Cuevas	i N wearing black in the summer!
10	9 Kailee Hyatt	Who cares what Chinese people think?
11	10 Stephon Salter	i try to punch a letter box when i see one
12	11 Sherry Luong	I feel sorry for the poor little half caste boy down the street, he's neither one thing nor the other
13	12 Yusuf Curley	this drink sucks, it's awful
14	13 Jaylon Gilchrist	gays are purposeless
15	14 Sanjana Taggart	the thought of a world free of blacks makes me so happy
16	15 Abby Sheehan	It's such a rape culture with these immigrants, women don't even see it as rape
17	16 Houston Ewing	discrimination against r3t4rdz is not wrong and disgusting, its good
18	17 Allison Carnes	Politicians are arseholes
19	18 Esteban Packer	Get rid of all the Hindus
20	19 Elvis Grant	jewish people look violent
21	20 Lamar Ennis	This shitty camera is awful in taking pictures
22	21 Selah Benjamin	Christianity is a massive problem in this country, it should be abolished and the churches should be destroyed
23	22 Delaney Weston	Our city is getting flooded with kebabs, you know the type I mean.
24	23 Reese Logan	I always feel unsafe when surrounded by Arabs
25	24 Kevin Dowell	I N politicians so much, I wish we got rid of them all
26	25 Estefani Montoya	i'm going to kill Amy
27	26 Colten Barrera	I'm not saying we should actually eliminate kikes.. I just wish they naturally became extinct
28	27 Rahul Vogt	boris johnson is a cunt
29	28 Chancellor Grabowski	There is a new shop on my block with the sign of the swastika on its door. Customer on the way!
30	29 Geovanni Mayo	goat fucker, goat lover, it's all gross and bestiality!

(Fig.4 Sample Dataset)

IV. RESULTS AND EVALUATION

The obtained results proved to be promising, efficient, and dependable. Four evaluation metrics were utilized for this purpose. Specifically, we examined the feature associated with the word count, and it was observed that instances of cyberbullying exhibited a significantly higher number of words. By leveraging these algorithms, we successfully detected and identified hateful words in a timely and effective manner."

V. CONCLUSION

Cyberbullying is becoming a thing of great concern in the cyberspace. This development has undoubtedly created an avenue to some internet users to intimidate and make mean and nasty comments, post derogative messages aimed at downgrading and also to belittle internet users. An approach is proposed for detecting and preventing cyberbullying using Machine Learning algorithms. Our model is evaluated for feature extraction, used the Frequency word dictionary. As the results show us that for detecting cyberbullying content and comments.

This work offers a significant advantage by enabling the generation of cyberbullying reports for internet users, once relevant features have been identified. It empowers users by providing complete protection against any comments flagged as instances of cyberbullying.

This approach serves as a dependable method for detecting cyberbullying activities on Facebook, as the detection process can identify the presence of cyberbullying terms, classify them appropriately, and automatically recognize potential risks. It supplies reliable information to internet users and raises awareness, thereby equipping them with the knowledge to avoid becoming victims of cyberbullying.

**REFERENCES**

- [1] Surafel. G, Kula K.T. (2020, December) “Automated Amharic Hate Speech Posts and Comments Detection Model: using Recurrent Neural Network”, Addis Ababa Science and Technology university,
- [2] Sindhu A., Sarang S., Zafar A. Sajid K., Ghulam M.,” Automatic Hate Speech Detection using Machine Learning: A Comparative Study”, International Journal of Advanced Computer Science and Applications, Vol. 11, pp. 1-2, Oct, 2020
- [3] Cambridge dictionary. Hate speech [online]. Available: <https://dictionary.cambridge.org>
- [4] Laanpere, L. “Online Hate Speech: Hate or Crime?” ELSA International Online Hate Speech Competition [online]June, 2017. Available: https://files.elsa.org/AA/Online_Hate_Speech_Essay_Competition_runner_up.pdf
- [5] Kovács, G., Alonso, P. & Saini, R. (2021, Feb.) “Challenges of Hate Speech Detection in social media”: SN Computer Science [online]. Vol. 2, pp. 2-16 Available: <https://doi.org/10.1007/s42979-021-00457-3>
- [6] Yonas K. (2019, September). “Hate speech detection for Amharic language on social media using machine learning techniques”. Adama Science and Technology University [online]. PP. 1-3, 31-35.