International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1162-1164 www.ijaem.net ISSN: 2395-5252

Fight Back Cyberbullying with NLP on Sns Using Code Text Approach

Prof.Dr.Vivek N. Waghmare, Kunal Gupta, Kanchan Ghule, Ankita Dhondge, Shruti Dive

Information Technology Department, Sandip Institute of Technology and Research Centre Nashik, India

Date of Submission: 09-03-2023 Date of Acceptance: 18-03-2023

ABSTRACT:This research paper is about the important problem of rising hate and offensive Contents made against people or communities on social media is addressed by this proposed system. Cyberbullying behaviour has drawn more and more attention as social media use has grown. Teen suicide has been linked to cyberbullying, among other serious and harmful effects on a person's life. Automatically detecting bullying content based on suitable machine learning and natural language processing algorithms is one efficient way to lessen and end cyberbullying. The negative effects of cyberbullying on social media are horrific and might even result in the unfortunate victims' deaths.The behaviour of the victims also changes due to this, which affects their Emotions, self-confidence and a sense of fear is also seen in such people. Thus, a complete solution is required for this problem. Cyber bullying needs to stop.The issue can be resolved by employing a machine learning approach to detect and prevent it, but this needs to be done from a different angle.

Keywords: Hate speech; social networking site; natural language processing; text classification; machine learning

I. INTRODUCTION

Cyberbullying behaviour has drawn more and more attention as social media use has grown. Cyberbullying can have major, detrimental effects on a person's life and even encourage young suicide. Automatically detecting bullying content based on suitable machine learning and natural language processing algorithms is one efficient way to lessen and end cyberbullying. Nevertheless, many of the methods currently used in the research are only standard text categorization models that do not take bullying features into account.

II. RELATED WORK

[1] Glenn Sterner, Diane Felmlee "The Social Networks of Cyberbullying on Twitter "

In this research they applies a social network perspective to the issue of cyber aggression or cyberbully, on the social media platform Twitter.Because of the potential for anonymity and the simplicity with which so many people can join in the harassment of victims, cyber violence is particularly dangerous. Utilizing a comparative case study methodology, the authors examined thousands of Tweets to explore the use of denigrating slurs and insults contained in public tweets that target an individual‟s gender, race, or sexual orientation. Findings indicate cyber aggression on Twitter to be extensive and often extremely offensive, with the potential for serious, deleterious consequences for its victims.

[2] Aditya Bohra, Deepanshu Vijay, Vinay Singh. "A Dataset of Hindi -English Code-Mixed Social Media Text for Hate Speech Detection."

The tweets are annotated with the language at word level and the class they belong to (Hate Speech or Normal Speech). They also propose a supervised classification system for detecting hate speech in the text using various character levels, word levels, and lexicon-based features. With the recent surge in the amount of user-generated social media data, there has been a tremendous scope in automated text analysis in the domain of computational linguistics. The popularity of opinion-rich online resources like review forums and microblogging sites has encouraged users to express and convey their thoughts all across the world in real time.

[3] Ravindra Nayak and Raviraj Joshi. "L3Cube-HingCorpus and HingBERT: A Code Mixed Hindi-English Dataset and BERT Language Models."

In this literature Code-mixed NLP has been extensively studied. As pre-trained transformerbased architectures are gaining popularity, they observe that real code-mixing data

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1162-1164 www.ijaem.net ISSN: 2395-5252

are scarce to pre-train large language models. They present L3Cube-HingCorpus, the first large-scale real Hindi-English code mixed data in a Roman script. It consists of 52.93M sentences and 1.04B SITRC, Department of Information Technology Engineering 2022-23 Page 12 tokens, scraped from Twitter. They further present HingBERT, HingMBERT, HingRoBERTa, and HingGPT. The BERT models have been pre-trained on code mixed HingCorpus using masked language modelling objectives. They show the effectiveness of these BERT models on the subsequent downstream tasks like code-mixed sentiment analysis, POS tagging, NER, and LID from the GLUECoS benchmark. The HingGPT is a GPT2 based generative transformer model capable of generating full tweets.

[4] Singh Kushagra, Indira Sen, Ponnurangam Kumaraguru "A Twitter Corpus for Hindi-English Code Mixed POS Tagging."

Code-mixing is a linguistic phenomenon that occurs when many languages are employed simultaneously and is becoming more widespread in multilingual society. The prevalence of code-mixed information on social media has also increased the demand for tools that can recognise it automatically. Any Natural Language Processing (NLP) pipeline must include automatic Parts-of-Speech (POS) tagging, yet there is a dearth of annotated data on which to train such models.In this work, they present a unique language tagged and POS-tagged dataset of code-mixed EnglishHindi tweets related to five incidents in India that led to a lot of Twitter activity.

[5] Tommy K.H. Chan, Christy M.K. Cheung, Zach W.Y. Lee. "Cyberbullying on social networking sites: A literature review and future."

Social networking site cyberbullying is a new social problem that has received a lot of scholarly attention. By a literature review and analyses, this study aims to consolidate the existing body of knowledge. They first go over the nature, trends in research, and theoretical underpinnings.They then develop an Integrative framework based on social cognitive theory to synthesize what known and identify what remains to be learned, with a focus on the triadic reciprocal relationships between perpetration, victims, and bystander. They discuss the key findings and highlight opportunities for future research they conclude this paper by noting.

III. PROBLEM STATEMENT

As the name suggests, cyberbullying involves using the internet to bully people who are either known or unknown to the bully. Extreme outbursts of wrath and even suicide attempts have resulted from cyberbullying for those affected. Hate speech detection in social media texts is an important Natural language processing task, which has several crucial applications like sentiment analysis, investigating cyberbullying, and examining socio-political controversies. While relevant research has been done independently on code-mixed social media texts and hates speech detection, the work on detecting hate speech in Hindi English code-mixed social media text is not feasible for the current time & age.

IV. OBJECTIVES

The objectives of the system are:

- To make the internet a safe place for society.
- To establish an early system to auto-report cyberbully incidents.
- To create an Ai to identify hate speech in code-texted Indian language.
- To have an active way of dealing with cybercrime at some level.

V. SURVEY RESULTS DISCUSSION

As per our knowledge, there exists no sentiment classifier for code-mixed social media text. To achieve our goal and complete objectives we will be creating initial prototypes. As one of the first steps to achieve this, we will be creating API services for our Code-Mixed semantic analyzer. A semantic analyzer will judge the given text and score it in the range from 1 to 10 for hate speech contains. For the next steps, we will be creating the SNS (Social Network Sites) scrapper to auto-feed the comments from posts and provide feedback on the given comments. As the last step in prototyping, we will be making some custom actions using the cloud functions such as auto deleting and auto reporting. The actions can have abilities on reporting the collected crime based on the hate speech score to the public as well as government forums and authorities. After having a proof of concept we collaboratewith social media influencers and the appropriate people to test the system. After the initial test, we will be creating a data scraper for popular SNS like Instagram Comments. As well as auto-delete and auto-report actions of above mentioned SNS. We will mark it as the end of phase one and try to collaborate with companies, governments, and SNS sites to improve our product.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1162-1164 www.ijaem.net ISSN: 2395-5252

VI. CONCLUSION

Cyberbullying is a serious issue, and likely any form of bullying it can have long term effects on its victims. Our project will grow and help individuals to be aware of Cyber bullies. Parents, teachers and children must work together to prevent Cyberbullying and to make Internet a safer place for all. To have social harmony and reduced depression and other mental illness caused by cyberbullying. Thus Creating counter measures needs an active funnel to take proper actions against the cyber bullies and to create an unsupervised classification system for detecting hate speech in the text using various character levels, word level attributes as well as emoji in Hindi-English code text language to create an open source action and plugin system using cloud technology to take suitable actions on the results.

VII. FUTURE SCOPE

The aim of the study are to (I) describe the traits of those who engage in cyberbullying and (II) specify the kinds of tools used to gauge cyberbullying on social media. Those who bully, those who are targeted, and those who witness bullying, are all susceptible to long-term, social and emotional problems. Researchers frequently discover a correlation between being bullied and depression and suicide.

REFERENCES

[1]. Glenn Sterner, Diane Felmlee. "The Social Networks of Cyberbullying on Twitter."
[2]. Aditya Bohra, Deepanshu Vijay, Vinay Singh. "A Dataset of Hindi -English Code-Mixed Social
[3]. Media Text for Hate Speech Detection." [3]Ravindra Nayak and Raviraj Joshi. "L3Cube-HingCorpus and HingBERT: A Code Mixed Hindi-English Dataset and BERT Language Models."
[4]. Kushagra Singh, Indira Sen, Ponnurangam Kumaraguru. "A Twitter Corpus for Hindi-English Code Mixed POS Tagging."
[5]. Shivang Chopra, Ramit Sawhney. "Hindi-English Hate Speech Detection: Author Profiling, Debiasing, and Practical Perspectives.
[6]. Dr.A.K.Jaithunbi, Gollapudi Lavanya, Dondapati Vindhya Smitha , Bandi Yoshna. "Detecting Twitter Cyberbullying Using Machine Learning."
[7]. Rui Zhao, Anna Zhou, Kezhi Mao."Automatic Detection of Cyberbullying on Social Networks based on Bullying Features."
[8]. Tommy K.H. Chan, Christy M.K. Cheung, Zach W.Y. Lee. "Cyberbullying on social networking sites: A literature review and future."
[9]. Pavitar Parkash Singh, Vijay Kumar, Majid Sadeeq. "Cyber Bullying as an Outcome of Social Media Usage: A Literature Review"
[10]. Renee Garett, Lynwood R. Lord, and Sean D. Young. "Associations between social media and cyberbullying: a review of the literature."