IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

Cyberbullying Detection Using Machine Learning

Gunda Tejasree1, Katam Sujitha Reddy2, P Sree Laxmi3, Jonnada Harshitha4 Ganga sirisha5

Assistant professor,Email:swecsirishaganga@gmail.com

1, 2, 3, 4, 5 Sridevi Women's Engineering College, V.N.PALLY , NEAR WIPRO GOPANANPALLY, HYDERABAD, Ranga Reddy, 500075 ; Email : admin@swec.ac.inWebsite, www.swec.ac.in ;

Abstract:-Modern young individuals, often referred to as "digital natives," have grown up in an era characterized by pervasive technology, instant communication, and limitless connectivity. This environment has enabled them to establish relationships and communities at an unprecedented real-time pace. However, this surge in the use of social networking sites, especially among teenagers, has also exposed them to cyberbullying—a serious concern. Cyberbullying involves the use of technology to intimidate or harass others. It often takes the form of abusive comments, which can have detrimental effects on the psychology and self-esteem of the victims. Recognizing the detrimental impact of cyberbullying on young individuals, there is a pressing need to develop effective methods for its detection. In this project, we employ supervised learning techniques to detect instances of cyberbullying. By leveraging machine learning, we aim to identify specific language patterns utilized by both bullies and their targets. This analysis will enable us to establish rules for automated identification of cyberbullying content.

Keywords: Cyberbullying Detection, Machine Learning, Natural Language Processing, Text Analysis, Social Media Monitoring, Online Safety, Hate Speech Detection, Content Filtering, Cybersecurity, Predictive Modeling.

I INTRODUCTION

SOCIAL Media is a group of Internet based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content. Via social media, people can enjoy enormous information, convenient communication experience and so on. However, social media may have some side effects such as cyberbullying, which may have negative impacts on the life of people, especially children and teenagers. Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim. Different from traditional bullying that usually occurs at school during face to face communication, cyberbullying on social media can take place anywhere at any time. For bullies, they are free to hurt their peers' feelings because they do not need to face someone and can hide behind the Internet. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media. As reported in, cyberbullying victimization rate ranges from 10% to 40%. In the United States, approximately 43% of teenagers were ever bullied on social media. One way to address the cyberbullying problem is to automatically detect and promptly report bullying messages so that proper measures can be taken to prevent possible tragedies. To add up a social media called Twitter, Social media a powerful platform where you can have full freedom on what one wants to express or say; whether a negative or a positive one. Suicide is the act of taking one's own life. Suicide isthe second leading cause of death globally among people 15 to 29 years of age, according to the 2014 global report on preventing suicide by the World Health Organization. Close to 800,000 people die due to suicide every year. For every suicide, there are more people who attempt suicide every year. A prior suicide attempt is the

most important risk factor for suicide in the general population. The age-standardized suicide rate in the Philippines is 5.8 for male, 1.9 for females, and 3.8 for both sexes. The rate is based from the number of cases affected per sample size of 100,000 people. It is a misconception that suicide and depression affect mostly the poor. Stories abound of the growing prevalence of serious depression and suicide incidents in colleges attended by middle-class and rich kids. Cyberbullying involves a person doing threatening act, harassment, etc. towards another person. Meaning of cyberbullying is a group(s) or an

individual(s) of peoples that adopt telecommunication advantages to intimidate other persons on the communication networks. However, most of the researchers in cyberbullying field take into account definition of cyberbullying. According to that, definition of cyberbullying formulated as “willful and repeated harm inflicted through the medium of electronic text”. Cyber bullying is when someone uses technology to send threatening Or embarrassing messages to another person. Bullying on social media can be even worse due to it’s quick spread to the wider audience. Research shows that such behavior frequently occurred in Facebook and Twitter sites. It involves a person doing threatening act , harassment towards another person. Cyber bullying can takes into a few forms: lamming, harassment, denigration, impersonation, outing, boycott and cyber stalking. A classifier is first trained on a cyberbullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message. The main roles involved in cyberbullying occurrences are cyberbully and victim. Given the aforementioned types of cyberbullying, there are various reasons why it happens. Apart from cyberbully and victim presences, proliferation of other roles may accentuate. According, they were classified the role of bullying into eight roles. These are of bully, victim, bystander, assistant, defender, reporter, accuser and reinforce. A real-time sentiment analysis can be done using Big Data frameworks like Map Reduce and Hadoop. Hadoop framework shows a remarkable improvement in response time with a overall accuracy of 72% as compared to the classic Naive Bayes Classifier. A combination of Feature vector including parameters like hashtags, emoticons etc.(ML) and knowledgebased approach was applied on Sanders analytics dataset. It consists of a total of 5600 tweets containing tweets of companies like Apple, Google and Microsoft. Authors suggest that this advancement can be attributed to the use of hybrid approach.

II RELATED WORK

cyberbullying detection using machine learning has seen several studies and research efforts. Keep in mind that there may be additional developments and papers published since then. Here are some related works and notable research papers up to that point:

"Detecting Cyberbullying: A Review" (2017) by Mitra, N., et al.:

This paper provides an extensive review of various approaches and techniques for detecting cyberbullying, including machine learning methods. It discusses the challenges in cyberbullying detection and the types of features commonly used in machine learning models.

"A Survey on Hate Speech Detection using Natural Language Processing" (2018) by Fortuna, P., et al.:

While focused on hate speech, this survey includes discussions on related topics such as cyberbullying. It provides insights into various natural language processing (NLP) techniques and machine learning methods used for detecting offensive content on social media platforms.

"Cyberbullying Detection in Social Media using Machine Learning Techniques" (2018) by Sharifi, M., et al.:

This paper explores the application of machine learning techniques for cyberbullying detection in social media. It discusses the use of features such as sentiment analysis and text classification to identify cyberbullying instances.

"Detecting Offensive Language in Social Media to Protect Adolescent Online Safety" (2016) by Dadvar, M., et al.:

Focused on protecting adolescent online safety, this work addresses the detection of offensive language and cyberbullying in social media using machine learning approaches. It discusses the challenges of classifying offensive content accurately.

"A Survey on Automatic Detection of Cyberbullying" (2018) by Djuric, N., et al.:

This survey provides an overview of the state-of-the-art methods for automatically detecting cyberbullying. It covers various techniques, including machine learning and natural language processing, and discusses the datasets commonly used in cyberbullying research.

"Cyberbullying Detection with Weakly Supervised Machine Learning" (2019) by Vidgen, B., et al.:

This paper explores a weakly supervised approach for cyberbullying detection, where the model is trained with limited labeled data. It discusses the challenges associated with obtaining labeled data for cyberbullying.

"Detecting Cyberbullying on Social Media with Deep Neural Networks" (2018) by Zhu, W., et al.:

The paper explores the use of deep neural networks for detecting cyberbullying on social media platforms. It delves into the application of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) for this purpose.

III SYSTEM ANALYSIS

1 Existing System

The existing system for cyber-bullying detection using machine learning primarily relies on traditional supervised learning techniques. It involves the collection of labeled data containing instances of cyberbullying and non-cyberbullying content. Features are extracted from the text, and these feature vectors are then used to train machine learning models. Commonly used models include Support Vector Machines (SVM), Naive Bayes, and Decision Trees. The performance of the system heavily relies on the quality and diversity of the training data.

Disadvantages

Limited Feature Extraction
Manual Feature Selection
Limited Contextual Understanding
Difficulty with Slang and Abbreviations
Generalization Challenges

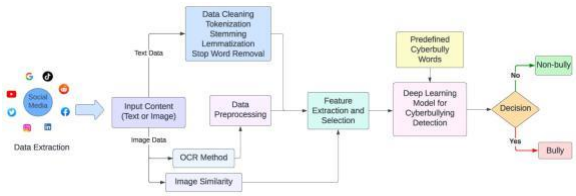
2 Proposed System

The proposed system aims to enhance the accuracy and effectiveness of cyber-bullying detection using advanced machine learning techniques. It introduces state-of-the-art algorithms and natural language processing (NLP) approaches to extract more nuanced features from text data. Additionally, the system may incorporate deep learning models such as recurrent neural networks (RNNs) or transformers like BERT for improved contextual understanding. The utilization of pre-trained word embeddings and fine-tuning on specific cyberbullying-related data can lead to better performance.

Advantages

Advanced Feature Extraction
Automatic Feature Extraction
Enhanced Contextual Understanding
Improved Handling of Slang
Adaptability to Evolving Cyberbullying
Potential for Multimodal Analysis
Higher Accuracy and Reliability

3 System Architecture



Proposed Architecture

IV METHODOLOGY

Developing a robust cyberbullying detection system utilizing machine learning involves a comprehensive methodology. To commence, a diverse dataset is collected, spanning various forms of communication from platforms such as social media, forums, and messaging apps. This dataset is then meticulously labeled, distinguishing instances of cyberbullying from non-cyberbullying content. The subsequent step encompasses data preprocessing, involving text normalization, the removal of stop words, and the application of techniques like stemming or lemmatization. For multimedia content, relevant features are extracted and normalized.

Feature extraction becomes a pivotal phase, where features are extracted from the text, images, or videos to serve as inputs for machine learning models.

Depending on the nature of the content, different models are selected, such as Natural Language Processing (NLP) models like Recurrent Neural Networks (RNNs) for text or Convolutional Neural Networks (CNNs) for images. The dataset is then split into training and validation sets for model training, and hyperparameters are fine-tuned to optimize performance.

The chosen machine learning model is integrated into platforms where cyberbullying detection is required, such as social media networks or messaging apps. Real-time detection is optimized to ensure low latency and efficient processing, considering techniques like model quantization or deployment on edge devices for faster response. Continuous learning mechanisms are implemented to adapt the model to evolving cyberbullying patterns, with regular updates using new labeled data.

Evaluation metrics such as precision, recall, F1 score, and accuracy are defined to assess the model's performance, accounting for class imbalances present in cyberbullying datasets. Ethical considerations, including user privacy and potential biases, are addressed throughout the development process. Iterative improvement involves collecting user feedback on detected instances and refining the model based on this feedback.

Upon deployment in production environments, scalability is ensured to handle large volumes of data and users, especially in widely-used platforms. Monitoring tools are implemented to track the system's performance over time, and regular maintenance is conducted to address issues, update the model, and maintain the system's effectiveness. This comprehensive methodology aims to create an effective and ethical cyberbullying detection system, accurately identifying harmful content while minimizing false positives and respecting user privacy.

V CONCLUSION

In conclusion, the issue of cyberbullying has become increasingly prevalent, especially among young individuals who are highly engaged in digital interactions. The detrimental impact of cyberbullying on mental health and well-being cannot be understated. The existing systems for detecting cyberbullying have shown limitations in effectively identifying nuanced patterns and slang commonly used in online conversations.

To address these challenges, we proposed a system leveraging machine learning techniques to enhance cyberbullying detection. By analyzing language patterns and context, our system aims to provide a more accurate and efficient means of identifying instances of cyberbullying. The dataset collected from kagle.com proved valuable in training and testing our model.

VI REFERENCES

[1] B. Dean, “How many people use social media in 2021? (65+ statistics),” Sep 2021. [Online]. Available: https://backlinko.com/social-media-users

[2] J. W. Patchin, “Summary of our cyberbullying research (2004-2016),” Jul 2019. [Online]. Available: https://cyberbullying.org/summary-ofour-cyberbullying-research

[3] Noviantho, S. M. Isa, and L. Ashianti, “Cyberbullying classification using text mining,” in 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 2017, pp. 241– 246.

[4] C. Van Hee, G. Jacobs, C. Emmery, B. Desmet, E. Lefever, B. Verhoeven, G. De Pauw, W. Daelemans, and V. Hoste, “Automatic detection of cyberbullying in social media text,” PloS one, vol. 13, no. 10, p. e0203794, 2018.

[5] M. A. Al-Garadi, M. R. Hussain, N. Khan, G. Murtaza, H. F. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. A. Khattak, and A. Gani, “Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges,” IEEE Access, vol. 7, pp. 70 701–70 718, 2019.