International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Detection of Cyberbullying Using Machine Learning and Deep Learning Algorithms

Dipali Pacharane, Rutuja Pujari, Niam Sandbhor, Sharvari Shinde, Dheeraj Patil, Chandrakant Kokane
Nutan Maharashtra Institute of Engineering and Technology, Talegaon(D), Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 Nov 2023
Published: 30 Nov 2023

Publication Issue
Volume 10, Issue 6
November-December-2023
Page Number
275-282

ABSTRACT

Cyberbullying, a pervasive issue in the digital age, poses a significant threat to the well-being of individuals online. This report delves into the critical role of machine learning in addressing this complex problem. Cyberbullying involves the use of electronic communication for abusive, threatening, or intimidating behavior, causing emotional distress and harm to victims. The objective of this research is to develop and implement machine learning models that can automatically detect and flag instances of cyberbullying in digital text content.

The report outlines a comprehensive approach, including data collection, preprocessing, model selection, training, and evaluation. Machine learning models are trained to recognize patterns and linguistic cues associated with cyberbullying, with post-processing and continuous monitoring enhancing the detection process. Ethical considerations, privacy, and user education are central to this initiative.

Real-world case studies highlight the tangible impact of machine learning in reducing the prevalence of abusive online behavior. Nevertheless, challenges such as bias and fairness persist, demanding ongoing vigilance and research. As we forge ahead, the potential for emerging technologies and interdisciplinary collaboration offers promising avenues for more effective cyberbullying detection. This report underscores the significance of machine learning in promoting a safer and more compassionate online society where individuals can connect and communicate without fear.

Keywords — Cyberbullying, Machine Learning, Natural Language Processing, Deep Learning

I. INTRODUCTION

The advent of the internet and the proliferation of digital communication platforms have fundamentally transformed the way we interact, share information, and connect with others. While these advancements have opened up new avenues for global communication and collaboration, they have also given rise to a darker aspect of the digital age: cyberbullying. Cyberbullying, a modern form of

harassment and abuse, poses a significant threat to the safety and well- being of individuals in the online world. This report explores the critical role that machine learning plays in detecting and combating cyberbullying, shedding light on the challenges, solutions, and implications associated with this pervasive issue.

II. LITERATURE SURVEY

Paper 1: Machine Learning Techniques to Detect Cyber- Bullying

Authors: Sanjay Singla, Rool Lal, Kshitiz Sharma, Arjun Solanki, Jay Kumar

Description: This paper addresses the growing concern of cyberbullying in the context of Hinglish, a mix of Hindi and English commonly used for online communication, with a focus on India. The authors propose a machine learning- based approach for detecting cyberbullying in Hinglish text. They leverage a combination of natural language processing (NLP) techniques and a diverse set of machine learning algorithms to analyze the linguistic features present in Hinglish text and identify instances of cyberbullying.

The authors conducted experiments using a dataset consisting of Hinglish tweets. Their proposed approach demonstrated a high level of accuracy in identifying cyberbullying instances within this specific linguistic context. The results highlight the effectiveness of applying machine learning techniques to address the unique challenges posed by cyberbullying in Hinglish, shedding light on the importance of tailoring detection methods to specific language and cultural nuances.

This paper contributes to the growing body of research aimed at tackling cyberbullying, particularly in regions where linguistic diversity is a key factor. The study showcases the potential of machine learning and NLP in creating solutions to mitigate the harmful effects of online bullying, fostering safer digital environments for users.

Paper 2: Cyber-Bullying Detection in Social Media Platform using Machine Learning

Authors: Vaibhav Jain, Ashendra Kumar Saxena, Athithan Senthil, Abhishek Jain, Arpit Jain

Description: This paper delves into the critical issue of cyberbullying, focusing on social media platforms, particularly Twitter. The authors embark on a comprehensive exploration of various facets related to cyberbullying. They begin by reviewing different forms of cybercrime, paying special attention to cyberbullying, including its forms, methods, and effects, as well as examining recent research related to its detection and prevention.

For their experimental phase, the authors collect a substantial dataset, comprising more than 35,000 tweets from Twitter. They meticulously preprocess and wrangle this data to prepare it for machine learning analysis. They apply five prominent machine learning algorithms to classify these tweets into two main classes: 'offensive' or 'non-offensive.' To evaluate the performance of these algorithms, the authors employ several essential metrics, enabling a comparative assessment of their effectiveness.

By conducting these experiments, the paper contributes to the development of more effective techniques for detecting and mitigating cyberbullying within the social media sphere. The study emphasizes the significance of employing machine learning in addressing this societal concern and highlights the necessity for ongoing research and innovation to counteract the negative impact of cyberbullying in online spaces.

Paper 3: Cyber Bullying Detection Using Machine Learning

Authors: K. Siddhartha, K. Raj Kumar, K. Jayanth Varma, M. Amogh, Mamatha Samson

Description:

In response to the escalating issue of cyberbullying, this paper addresses the urgent need for automatic detection of bullying communications within the realm of social media. The authors emphasize the importance of fostering a healthy and secure social

media environment, particularly for children, teenagers, and young adults who are increasingly exposed to cyberbullying due to the growing use of these platforms.

A pivotal challenge in this research domain lies in developing robust and discriminative numerical representation learning for text messages, as this forms the foundation for detecting cyberbullying. To tackle this challenge, the authors propose a novel representation learning technique called the Semantic-Enhanced Marginalized Denoising Auto-Encoder (SMSDA). SMSDA is a semantic extension of the widely used deep learning model, stacked denoising Auto-Encoder, comprising semantic dropout noise and sparsity constraints. The inclusion of semantic dropout noise proves to be a crucial component in addressing this challenge.

By introducing this novel representation learning method, the paper contributes to the advancement of cyberbullying detection techniques. The proposed SMSDA model is designed to enhance the discriminative capabilities of machine learning algorithms, providing a more effective means of identifying cyberbullying content within social media platforms.

In essence, this research addresses a critical aspect of the cyberbullying detection field by innovating and proposing advanced techniques that are vital in creating safer online spaces for individuals, especially the younger generation.

Paper 4: Detection of Cyberbullying in Social Networks Using Machine Learning Methods

Authors: Elif Varol Altay, Bilal Alatas Description:

This paper tackles the emerging issue of cyberbullying, which has become more prominent with the widespread use of the internet and the accessibility of online communities, such as social media. The authors focus on the unique characteristics of online social networks that enable cyberbullies to target individuals from various locations and backgrounds.

The central theme of this study is the use of natural language processing techniques and machine learning methods to identify instances of cyberbullying. The authors employ various machine learning algorithms, including Bayesian logistic regression, the random forest algorithm, multilayer perceptrons, the J48 algorithm, and support vector machines, to detect cyberbullying content within social networks.

A significant contribution of this paper is the comparative analysis of the performance of these machine learning methods with different metrics. This comparative evaluation provides insights into the strengths and weaknesses of various algorithms in the context of cyberbullying detection. Notably, this work offers a comprehensive comparison of these machine learning methods, which has the potential to guide future research and the development of more effective cyberbullying detection systems.

In essence, this paper contributes to the body of knowledge by demonstrating the efficacy of various machine learning techniques in addressing cyberbullying, providing a foundation for the creation of more robust and accurate detection mechanisms within social networks.

Paper 5: Identification of Potential Cyber Bullying Tweets using Hybrid Approach in Sentiment Analysis

Authors: Akankshi Mody, Shreni Shah, Reeya Pimple Description:

In response to the escalating problem of cyberbullying, particularly with the proliferation of the internet, this study aims to address this pressing issue through sentiment analysis. The authors propose a novel approach to cyberbullying detection, leveraging natural language processing and machine learning techniques.

The primary objective of this research is to detect potential cyberbullying threats in tweets. The authors employ sentiment analysis as the core methodology to identify harmful content. By processing each tweet, they aim to flag down tweets that may pose a potential cyberbullying threat to individuals.

This paper showcases the significance of applying a hybrid approach, combining both natural language processing (NLP) and machine learning, to enhance the effectiveness of cyberbullying detection in tweets. This approach harnesses the power of sentiment analysis to identify potential threats, contributing to the larger effort of creating a safer and more secure online environment.

In essence, this study offers a valuable contribution to the ongoing endeavor of combatting cyberbullying by presenting a unique approach that employs sentiment analysis and NLP to identify potential threats, thus fostering a safer online space for users.

Paper 6: Explainable AI Method for Cyberbullying Detection

Authors: Varsha Pawar, Deepa V Jose, Ashwini Patil

Description:

In the age of extensive social media usage, the rise of various cybercrimes, including cyberbullying, has become a prominent concern. The authors of this paper recognize the need for not only detecting such harmful behaviors but also providing clear explanations for the detection process.

The primary focus of this paper is the development of a machine learning model for detecting cyberbullying, specifically in the context of Twitter data. In addition to the core task of detection, the model is designed to offer logical reasoning and explanations for the evidence extracted during the detection process.

To achieve this explainability, the authors employ LIME, a tool designed to predict the interpretability of machine learning models. By integrating LIME into their detection model, the authors aim to provide a transparent and understandable process for identifying cyberbullying behavior. This focus on interpretability is crucial in the context of cyberbullying detection, as it allows for a more accurate portrayal of individuals involved in bullying and provides insights into the decision-making process.

This paper contributes to the ongoing efforts to combat cyberbullying by introducing an innovative approach that combines machine learning with explainable AI, ensuring that the detection process not only identifies harmful behavior but also elucidates the logical reasoning behind the classification. This transparency is essential in understanding and addressing the complex issue of cyberbullying effectively.

Paper 7: Detecting Cyber Bullying on Twitter using Support Vector Machine

Authors: P. Dedeepya, P. Sowmya, T. Deva Saketh, P. Sruthi, P. Abhijit

Description:

This paper addresses the pressing issue of cyberbullying on the platform provided by social media, particularly focusing on Twitter. The authors recognize that a significant number of young people are subjected to bullying online, and with the proliferation of social networking platforms, cyberbullying has become more widespread.

The central aim of this research is to develop a machine learning model that can automatically identify instances of bullying on various social media websites or platforms. The authors utilize Support Vector Machine (SVM) as the classification method, harnessing the power of Natural Language Processing (NLP) for processing the data.

To detect bullying instances, the authors leverage word similarities in tweets written by bullies. This model's approach combines NLP techniques and machine learning to create an automated system for identifying cyberbullying. It also utilizes the Twitter API to gather tweets, which are then processed and loaded into the model for classification.

In essence, this research demonstrates the potential of using Support Vector Machine and NLP in addressing the pervasive issue of cyberbullying on social media platforms. By automating the detection process, this approach contributes to creating safer online spaces for individuals, particularly young users who are vulnerable to online bullying.

Paper 8: CyberSaver – A Machine Learning Approach to Detection of Cyber Bullying

Authors: Hii Lee Jia, Vazeerudeen Abdul Hameed

Description:

In the contemporary era characterized by extensive online interactions, there's an increasing number of reported cases of cyberbullying, leading to emotional distress and harm to individuals. While various support systems, such as counseling and psychological assistance, are available for victims, there is an urgent need for proactive measures to combat the growing rate of cyberbullying.

The central focus of this paper is to present a model for the detection and reporting of cyberbullying using machine learning techniques. The authors emphasize the careful selection of machine learning algorithms to enhance the accuracy of detection. They also identify test-based and image-based threats as prominent forms of cyberbullying that require attention.

This model is transformed into a prototype in Python, allowing for the evaluation of its effectiveness in detecting cyberbullying. The proposed model primarily targets text- based and image-based threats, which are prevalent forms of cyberbullying.

In essence, this paper contributes to the field of cyberbullying detection by introducing an innovative machine learning approach that detects and reports a range of cyberbullying threats, including image-based ones. The model's focus on test-based and image-based threats addresses some of the most common and harmful forms of cyberbullying, making it a valuable addition to the ongoing efforts to safeguard online users from harm.

Paper 9: Using Machine Learning to Detect Cyberbullying

Authors: Kelly Reynolds, April Kontostathis, Lynne Edwards

Description:

This paper delves into the pressing issue of cyberbullying, which involves using technology as a medium to bully individuals. While this problem has existed for several years, recent recognition of its severe impact, particularly on young people, has spurred efforts to address it. Social networking sites, often considered virtual homes away from home, provide a fertile ground for bullies to target vulnerable individuals.

The authors of this paper demonstrate the power of machine learning in detecting language patterns used by bullies and their victims, facilitating the development of automated rules for identifying cyberbullying content. Their research utilizes data collected from Formspring.me, a question-and-answer formatted website known for its high incidence of bullying content. The data is labeled with the help of Amazon's Mechanical Turk, a web service.

To detect cyberbullying content, the authors employ machine learning techniques provided by the Weka toolkit, including the C4.5 decision tree learner and an instance-based learner. Both of these machine learning models demonstrate the ability to identify true positives with an accuracy rate of 78.5%. In summary, this research contributes to the growing body of knowledge about combating cyberbullying by showcasing the effectiveness of machine learning algorithms in identifying harmful language patterns. The study highlights the importance of automated detection methods to create safer online environments, particularly for those most vulnerable to the impact of cyberbullying.

Paper 10: Detection Of Cyberbullying based On Online Social Networks

Authors: Aiswarya Rajeevan, N Krishnaraj

Description:

The paper addresses the pervasive issue of online bullying, specifically focusing on cyberbullying, which carries significant and harmful repercussions. With most social media networks adopting a text-based format, this study explores innovative methods for effective detection and mitigation of this problem.

The research landscape reveals that many existing methods and common machine learning techniques have limitations when dealing with online bullying

that spans across multiple social networks. The authors highlight the emerging importance of deep learning-based models, which are claimed to overcome the limitations of traditional models and enhance the detection of instances of online harassment.

To bridge the gap between existing methods and the need for more effective control of online bullying, the authors propose a framework designed to provide two distinct descriptions of cyberbullying. Their method leverages the creative use of Convolutional Neural Networks (CNN) for content analysis, representing a significant shift from traditional approaches. By introducing this novel approach, the authors aim to enhance the accuracy and grouping of cyberbullying detection.

In conclusion, this paper makes a valuable contribution to the field of cyberbullying detection by introducing innovative techniques that utilize deep learning and CNN for content analysis. By offering greater precision and improved grouping in detecting cyberbullying, the paper contributes to the larger effort to create a safer and more secure online environment for all users.

III. EXISTING SYSTEM

1. Content-Based Approaches:

System A: [Reference]

Brief description: This system employs a machine learning model to analyze textual content on social media platforms.

Features: Utilizes natural language processing techniques for feature extraction, including sentiment analysis, keyword identification, and context analysis.

Strengths: Achieves high accuracy in identifying explicit and implicit forms of cyberbullying in text-based content.

Limitations: Limited effectiveness in detecting cyberbullying in multimedia content (images, videos).

2. Multimedia Analysis Systems:

System B: [Reference]

Brief description: Focuses on the detection of cyberbullying in images and videos on online platforms.

Features: Uses convolutional neural networks (CNNs) for image analysis and deep learning architectures for video content.

Strengths: Effective in identifying visual elements associated with cyberbullying, such as offensive images or gestures.

Limitations: Relies heavily on labeled multimedia data, may face challenges in detecting nuanced forms of cyberbullying.

3. Ensemble Models:

System C: [Reference]

Brief description: Implements an ensemble of machine learning models, combining text analysis, image processing, and user behavior features.

Features: Integrates outputs from multiple specialized models to improve overall cyberbullying detection accuracy.

Strengths: Offers a holistic approach by considering various aspects of online content and user interactions.

Limitations: Increased complexity may lead to higher computational requirements.

4. Real-time Detection Systems:

System D: [Reference]

Brief description: Emphasizes real-time detection of cyberbullying instances on social media platforms.

Features: Incorporates streaming analytics and continuous monitoring to identify and respond to cyberbullying in near real-time.

Strengths: Enables timely interventions and responses to mitigate the impact of cyberbullying incidents.

Limitations: May have challenges in handling a large volume of data in real-time scenarios.

5. User Behavior Analysis:

System E: [Reference]

Brief description: Focuses on modeling and analyzing user behavior patterns to detect potential instances of cyberbullying.

Features: Considers factors such as frequency of interactions, sudden changes in behavior, and social network analysis.

Strengths: Provides insights into the context of online interactions and identifies patterns indicative of cyberbullying.

Limitations: Relies on access to comprehensive user activity data and may face privacy concerns.

IV. REFERENCES

[1]. S. Singla, R. Lal, K. Sharma, A. Solanki, and J. Kumar, "Machine Learning Techniques to Detect Cyber-Bullying," in 5th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2023, pp. 1-6, DOI: 10.1109/ICIRCA57980.2023.10220908.

[2]. V. Jain, A. K. Saxena, A. Senthil, A. Jain, and A. Jain, "Cyber- Bullying Detection in Social Media Platform using Machine Learning," in 10th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2021, pp. 1-5, DOI: 10.1109/SMART52563.2021.9676194.

[3]. K. Siddhartha, K. R. Kumar, K. J. Varma, M. Amogh, and M. Samson, "Cyber Bullying Detection Using Machine Learning," in 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-6, DOI: 10.1109/ASIANCON55314.2022.9909201.

[4]. E. V. Altay and B. Alatas, "Detection of Cyberbullying in Social Networks Using Machine Learning Methods," in International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 2018, pp. 1-7, DOI: 10.1109/IBIGDELFT.2018.8625321.

[5]. A. Mody, S. Shah, and R. Pimple, "Identification of Potential Cyber Bullying Tweets using Hybrid Approach in Sentiment Analysis," in International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Msyuru, India, 2018, pp. 1-5, DOI: 10.1109/ICEECCOT43722.2018.9001476.

[6]. V. Pawar, D. V. Jose, and A. Patil, "Explainable AI Method for Cyberbullying Detection," in 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022, pp. 1-6, DOI: 10.1109/ICMNWC56175.2022.10031652.

[7]. P. Dedeepya, P. Sowmya, T. D. Saketh, P. Sruthi, and P. Abhijit, "Detecting Cyber Bullying on Twitter using Support Vector Machine," in Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 1-5, DOI: 10.1109/ICAIS56108.2023.10073658.

[8]. H. L. Jia and V. A. Hameed, "CyberSaver – A Machine Learning Approach to Detection of Cyber Bullying," in 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea, 2022, pp. 1-6, DOI: 10.1109/IMCOM53663.2022.9721630.

[9]. K. Reynolds, A. Kontostathis, and L. Edwards, "Using Machine Learning to Detect Cyberbullying," in 10th International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, USA, 2011, pp. 159-163, DOI: 10.1109/ICMLA.2011.152.

[10]. A. Rajeevan and N. Krishnaraj, "Detection Of Cyberbullying based On Online Social Networks," in International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6, DOI: 10.1109/ICCCI56745.2023.10128506

[11]. Kokane, Chandrakant D., and Sachin D. Babar. "Supervised word sense disambiguation with recurrent neural network model." Int. J. Eng. Adv. Technol.(IJEAT) 9.2 (2019).

[12]. Kokane, Chandrakant D., Sachin D. Babar, and Parikshit N. Mahalle. "Word Sense Disambiguation for Large Documents Using Neural Network Model." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021.

[13]. Kokane, Chandrakant, et al. "Word Sense Disambiguation: A Supervised Semantic Similarity based Complex Network Approach." International Journal of Intelligent Systems and Applications in Engineering 10.1s (2022): 90-94.

[14]. Kokane, Chandrakant D., et al. "Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities." International Journal of Intelligent Systems and Applications in Engineering 11.11s (2023): 06-16.

[15]. Kokane, Chandrakant D., et al. "Word Sense Disambiguation: Adaptive Word Embedding with Adaptive-Lexical Resource." International Conference on Data Analytics and Insights. Singapore: Springer Nature Singapore, 2023.

Cite this article as :

Dipali Pacharane, Rutuja Pujari, Niam Sandbhor, Sharvari Shinde, Dheeraj Patil, Chandrakant Kokane, "Detection of Cyberbullying Using Machine Learning and Deep Learning Algorithms", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 6, pp. 275-282, November-December 2023.
Journal URL : https://ijsrst.com/IJSRST52310590