IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

DETECTION OF CYBERBULLYING ON SOCIAL MEDIA USING MACHINE LEARNING AND DEEP LEARNING

[1]Mrs. P. SWATHI Assistant Professor, [2]V. ALEKHYA, [3]R. SHAMILI, [4]SK. SAFANA
[5]B. NIKHILA GAYATHRI
EMAIL: swathipuvvula126@gmail.com
Vijaya Institute of Technology for Women
(Affiliated to J.N.T.U Kakinada, Approved by A.I.C.T.E, New Delhi)
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ABSTRACT

The rapid expansion of social media platforms has revolutionized communication, allowing individuals to connect and share opinions globally. However, this growth has also led to an alarming rise in cyberbullying, where individuals are subjected to harassment, hate speech, and online abuse. Traditional methods of detecting cyberbullying are often inadequate due to the vast amount of user-generated content and the evolving nature of online threats. To address this challenge, machine learning techniques have been explored to develop automated systems for detecting and mitigating cyberbullying in real time. This study focuses on utilizing machine learning algorithms to analyze social media posts and identify potential instances of cyberbullying. Various natural language processing (NLP) techniques, including sentiment analysis and keyword extraction, are employed to detect harmful or offensive language. Additionally, supervised learning models such as Support Vector Machines (SVM), Random Forest, and deep learning approaches like Recurrent Neural Networks (RNN) and transformers are leveraged to classify textual content effectively. The integration of these techniques enhances the system's ability to recognize patterns associated with cyberbullying. A crucial aspect of this research is the development of a comprehensive dataset that encompasses diverse examples of cyberbullying across multiple social media platforms.

1. INTRODUCTION

The proliferation of social media platforms has significantly transformed the way people communicate, interact, and share information. Platforms like Facebook, Twitter, Instagram, and TikTok have enabled individuals from different parts of the world to connect instantly, breaking geographical barriers and fostering digital communities. The accessibility and convenience of social media have made it an integral part of daily life, with billions of users actively engaging in content creation, discussion, and entertainment.

However, the increasing use of social media has also given rise to various challenges, one of the most concerning being cyberbullying. Cyberbullying refers to the use of digital platforms to harass, threaten, or embarrass individuals. Unlike traditional forms of bullying, cyberbullying can occur anonymously, making it more difficult to identify and address. With the

ease of sharing messages, images, and videos, harmful content can spread rapidly, causing significant emotional and psychological distress to victims.

1.1 The Impact of Cyberbullying

Cyberbullying has far-reaching consequences, particularly for young individuals who are more vulnerable to online harassment. Studies have shown that victims of cyberbullying often experience anxiety, depression, low self-esteem, and in severe cases, suicidal thoughts. Unlike physical bullying, which is limited to specific locations such as schools or workplaces, cyberbullying can follow individuals wherever they go, as long as they have access to digital devices. This persistent exposure to harmful content exacerbates the psychological effects, making it challenging for victims to escape the cycle of abuse.

2. LITERATURE REVIEW

The rise of social media and online communication platforms has introduced new opportunities for interaction and information sharing. However, it has also given rise to the problem of cyberbullying, which involves the use of digital platforms to harass, intimidate, or embarrass individuals. Unlike traditional bullying, cyberbullying can occur anonymously and has a wider reach, making it more difficult to prevent and control.

Researchers and technology experts have explored various methods to detect and mitigate cyberbullying, including manual moderation, keyword-based filtering, and machine learning based approaches. The evolution of artificial intelligence (AI) and natural language processing (NLP) has opened new possibilities for automated cyberbullying detection, enhancing the ability of systems to identify and respond to harmful online behaviour. This literature survey provides an overview of existing research in the field of cyberbullying detection, examining different approaches, challenges, and advancements in machine learning for content moderation.

2.1 Traditional Approaches to Cyberbullying Detection

Before the advent of AI-driven solutions, traditional approaches to cyberbullying detection primarily relied on manual reporting and rule-based filtering techniques. Social media platforms implemented community guidelines that allowed users to report cyberbullying incidents, which were then reviewed by human moderators. However, this method was time-consuming, requiring human intervention to assess every reported post.

2.2 Machine Learning-Based Approaches

Machine learning has emerged as a promising solution for cyberbullying detection, enabling automated classification of text-based interactions. Various supervised and unsupervised learning techniques have been applied to improve the accuracy of cyberbullying detection models.

2.3 Natural Language Processing (NLP) in Cyberbullying Detection

Natural language processing (NLP) techniques have played a crucial role in improving the accuracy of cyberbullying detection models. NLP enables computers to understand and interpret human language, making it possible to analyze

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

online conversations and identify harmful intent.

3. PROPOSED SYSTEM:

The increasing prevalence of cyberbullying on social media has led to significant psychological and social consequences for individuals, particularly teenagers and young adults. Existing cyberbullying detection systems rely on traditional methods such as manual moderation, keyword filtering, and sentiment analysis, which have proven to be ineffective in accurately identifying harmful content. These limitations highlight the need for a more advanced, automated, and context-aware cyberbullying detection system.

The proposed system leverages state-of-the-art machine learning and natural language processing (NLP) techniques to enhance the accuracy and efficiency of cyberbullying detection. It integrates deep learning models, real-time content analysis, sentiment detection, and multi-modal analysis to improve the identification of harmful content. The system also incorporates ethical AI principles, ensuring fairness, privacy protection, and transparency in decision-making.

4. RESULT

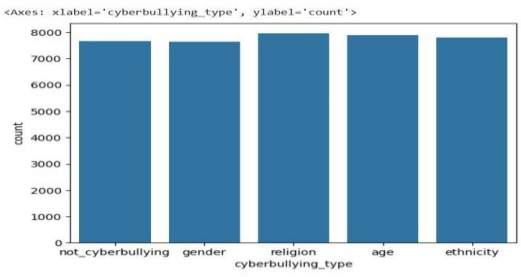

Fig: Cyberbullying Type

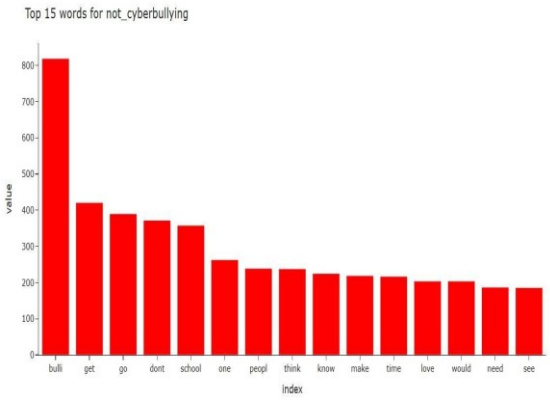

Fig: Words of Cyberbullying types

CONCLUSION

In conclusion, Cyberbullying has become a widespread issue on social media platforms, affecting millions of users worldwide. Unlike traditional bullying, cyberbullying occurs in digital spaces where anonymity and rapid content sharing make it difficult to control. The negative consequences of cyberbullying, including emotional distress, mental health issues, and social isolation, highlight the need for an effective detection and prevention system.

REFERENCE:

1. Dinakar, K., Reichart, R., & Lieberman, H. (2011). Modeling the detection of textual cyberbullying. *Fifth International AAAI Conference on Weblogs and Social Media.*

2. Hinduja, S., & Patchin, J. W. (2010). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior, 29(2), 129-156.*

ISSN 2347–3657
Volume 13, Issue 2, 2025

3. Dadvar, M., Trieschnigg, D., Ordelman, R., & de Jong, F. (2013). Improving cyberbullying detection with user context. *European Conference on Information Retrieval, 693-696.*
4. Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media.*

Kshirsagar, M., Veeramachaneni, S., & Yang, Y. (2018). Leveraging linguistic structures for cyberbullying detection. *International Conference on Computational Linguistics (COLING), 3745-3756.*

5. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26, 3111-3119.*
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30.*
7. Rajadesingan, A., Zafarani, R., & Liu, H. (2015). Sarcasm detection on Twitter: A behavioral modeling approach. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 97-106.*
8. Rieger, A., Kumar, I., Martinez, G., & Reisert, D. (2021). Ethical considerations in AI-based cyberbullying detection. *Journal of Artificial Intelligence Research, 70, 1-15.*
9. Zhang, Z., Robinson, D., & Tepper, J. (2020). Detecting hate speech on Twitter using BERT and multi-view learning. *Proceedings of the 24th Conference on Computational Natural Language Learning, 438-449.*