Chapter 8
Enhancing Online Security:
A Cyberbullying Detection and Prevention System for Monitoring Abusive Language

J. Jesy Janet Kumari
https://orcid.org/0000-0002-3318-1632
The Oxford College of Engineering, Bangalore, India

S. Thangam
https://orcid.org/0000-0003-2251-3651
Amrita School of Computing, Amrita Vishwa Vidyapeetham, Bengaluru, India

ABSTRACT

Cyberbullying is the use of technology to harass, threaten, embarrass, or target another person. Cyberbullying is a growing problem in our digital age, and it can have severe consequences for victims. Online bullying can be particularly damaging and upsetting because it's usually anonymous or hard to trace. It's also hard to control, and the person being victimized has no idea how many people have seen the messages or posts. They can be easier to commit than other acts of bullying because the bully doesn't have to confront their target in person. Detecting cyberbullying and preventing it is crucial in creating a safe online environment for all. The work involves the use of natural language processing and sentiment analysis to find abusive language and negative sentiment in online conversations. Before the comment is posted the detected instances of cyberbullying can be warned to the sender and prevented from being sent in a conversation.

DOI: 10.4018/979-8-3373-2282-7.ch008

1. INTRODUCTION

Cyberbullying is the use of technology as a medium to bully someone. Social networking sites provide a fertile medium for bullies, and teens and young adults who use these sites are vulnerable to attack. Through machine learning, we can detect language and patterns used by bullies and their victims and develop rules to automatically detect cyberbullying content (Islam et al., 2020). Cyberbullying is harassing, threatening, embarrassing someone, or targeting sharing about that person through technology. Cyberbullying actions, which are more common among young children and young people, can also be seen by adults. In such cases, severe legal actions are imposed on adults, such as prison sentences.

Cyberbullying affects victims both emotionally and psychologically (Desai et al., 2021). Social media also allows bullies to exploit anonymity to satisfy their evil deeds. Things became even more serious when bullying incidents occurred continuously over time. So, preventing this from happening helps the victim. The prevalence of information cannot be predicted because of the wide range of possibilities and the speed of new media. The perpetrators, so called "cyber harassment", can act anonymously and often consider themselves safe because of this anonymity.

In most cases, the identity of the perpetrator (Nizam et al., 2024) is presented very differently from the reality. Age or external image are not important criteria for cyberbullying. This can happen both between people of the same age and between people of different ages. There is a possibility of unintentional cyberbullying, as thoughtless actions without awareness of the consequences can hurt the people involved. The abuser often does not see these reactions and is unaware of the scale of the actions. Given the consequences of cyberbullying on its victims (Wulandari et al., 2024), it is imperative to find the proper actions to detect and prevent it. Machine learning is one of the successful approaches that learn from data and creates a model that automatically classifies right actions. Machine learning can be useful to detect language patterns of bullies and thus can generate a pattern to detect acts of cyberbullying (Islam et al., 2020).

The main purpose of this work is to create a safe and healthy environment for individuals to engage in digital communication. Cyberbullying has profound consequences on mental health and can lead to anxiety, depression, and suicide. It can also affect academic performance, social relationships, and self-esteem. Therefore, it is crucial to show and prevent cyberbullying behavior to mitigate these negative effects. Machine learning (Mahesh et al., 2021) techniques have the potential to improve the detection and prevention of cyberbullying by analyzing enormous amounts of data and identifying patterns that may not be apparent to human analysts.

The scope of this work is to explore the use of machine learning techniques in the detection and prevention of cyberbullying. Specifically, the work will focus on