IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Target-Oriented Investigation of Online Abusive Attacks

Prof. C. S. Jaybhaye, Manish More ,Shreyash Mandalik ,Yogita Tarade, Prachi Thakare
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *The exponential growth of online platforms has led to an alarming increase in toxic behaviors such as hate speech, cyberbullying, and harassment, which traditional moderation tools often fail to mitigate effectively. This paper presents a real-time, multi-layered system for online abusive content detection, integrating advanced Natural Language Processing (NLP), Machine Learning (ML), and social network analysis (SNA). By leveraging deep contextual models like BERT and incorporating behavioral profiling, the system detects nuanced, context-dependent abuse with high accuracy. A MERN-based full-stack application interfaces with a Python microservice to process user-generated content in real-time, flag abusive messages, and notify moderators. Experimental results on the Jigsaw Toxic Comment dataset demonstrate over 90% accuracy, with sub-300ms inference latency. The system supports multi-label classification of abuse types, automated moderation workflows, and scalable backend infrastructure. Limitations include language constraints and challenges in detecting sarcasm or low-frequency abuse types. The proposed framework offers a practical, deployable solution toward safer and more inclusive digital communication environments*

Keywords: Online Abuse Detection, Hate Speech, Natural Language Processing (NLP), Machine Learning, BERT, Real-Time Monitoring, Social Network Analysis, MERN Stack, Cyberbullying, Content Moderation, Behavioral Profiling, Deep Learning, Toxic Comment Classification, Automated Moderation System

DOI: 10.48175/IJARSCT-27590