Abstract

The rapid growth of social media has brought significant benefits but has also raised critical concerns, such as cyberbullying, which poses a serious threat to mental well-being. Although several previous studies have addressed this issue, the majority have focused on approaches based on Convolutional Neural Networks (CNN). This research explores the effectiveness of cyberbullying detection using advanced graph-based machine learning techniques, specifically Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). Text normalization, TF-IDF, and other preprocessing methods for feature extraction are applied to prepare the social media datasets for classification. Experimental results indicate that GCN outperforms GAT across all critical metrics, achieving an accuracy of 94.87%, due to its effectiveness in capturing complex relationships within graph data. In contrast, GAT reached a peak accuracy of 87.79% and an F1-score of 79.27%. These results not only highlight the superior performance of GCN in metric evaluation but also open opportunities for practical implementation in automated cyberbullying detection. Furthermore, they provide a foundation for future studies to explore parameter optimization, application to diverse datasets, and the adoption of advanced GNN architectures to further improve detection accuracy.

Keywords— cyberbullying detection, graph neural networks, social network analysis, text classification.