Abstract

Cyberbullying is a form of online bullying that can adversely affect the mental health of the victim. Indonesia, with one of the highest numbers of social media users, often creates a hostile environment due to misuse, including cyberbullying. This research aims to build a system to identify cyberbullying by utilizing a combination of TF-IDF with FastText methods in feature extraction. The method involves TF-IDF, which weighs words based on frequency, and FastText, which captures contextual meaning and semantic relationships. Features are expanded using FastText to capture variations and forms of words, improving data representation. The corpus includes 63,093 IndoNews and 30,000 Tweets, used in FastText feature extraction and expansion. Data is classified using CNN, RNN, and a hybrid CNN-RNN. Results show CNN achieved 85.11% accuracy, an increase of 1.59% from the baseline, RNN achieved 84.60%, an increase of 2.45%, and hybrid CNN-RNN achieved the highest accuracy of 86.32%, an increase of 2%. This demonstrates the hybrid approach, strengthened by combined feature extraction, is more effective in understanding text context and improving accuracy. These findings suggest integrating multiple feature extraction methods enhances machine learning models' performance in identifying cyberbullying, paving the way for robust automated systems to combat online abuse.

Keywords

cyberbullying, feature extraction, hybrid models, FastText, TF-IDF