ABSTRACT

Cyberbullying is one of the growing social issues alongside the rapid expansion of social media platforms. The impacts of cyberbullying, such as hate comments, can have significant psychological and emotional effects on victims. This research highlights the importance of developing accurate hate comment detection systems to mitigate these effects. The goal of this study is to analyze the performance of the IndoBERT and Cendol models in detecting hate comments within cyberbullying cases. The methodology includes collecting datasets containing keywords related to hate comments and bullying, as well as testing the IndoBERT and Cendol models. Evaluation metrics such as accuracy, precision, recall, and F1-Score are used.

The findings reveal that both IndoBERT and Cendol models demonstrate competitive performance in detecting hate comments. IndoBERT excels in handling formal text, while Cendol is more effective in understanding informal language and local contexts. A survey conducted with 328 respondents identified 64 keywords associated with cyberbullying. The IndoBERT model achieved the highest accuracy of 90.7% at the 5th epoch with a learning rate of 10^{-5} and a batch size of 8. Meanwhile, the Cendol model reached an accuracy of 90.6% at the 5th epoch with a learning rate of 10^{-4} and a batch size of 2. The social media platform used in this research is X (formerly Twitter), which is widely used for interaction and opinion sharing.

Although both models exhibit strong performance, there is a potential risk of overfitting, which is a crucial concern for future research. Overfitting can reduce a model's ability to generalize to unseen data. This study makes an important contribution to the development of Indonesian language hate speech detection systems, which can be applied not only on X but also on other social media platforms such as Facebook, Instagram, and TikTok to mitigate the negative impacts of hate comments across various platforms.

Keyword: *Cyberbullying*, *Hate Comments*, IndoBERT, Cendol, NLP.