

ABSTRAK

Cyberbullying di platform media sosial telah menyebar luas di masyarakat. Cyberbullying dapat terjadi dalam berbagai bentuk, termasuk ujaran kebencian, trolling, konten dewasa, rasisme, pelecehan, atau kata-kata kasar. Salah satu platform media sosial yang memiliki banyak pelaku cyberbullying adalah twitter yang telah berganti nama menjadi 'X'. Sifat anonim dari platform 'X' ini memungkinkan pengguna dari seluruh dunia untuk melakukan cyberbullying karena mereka dapat dengan bebas membagikan pemikiran dan ekspresi mereka tanpa harus mempertanggungjawabkan identitas mereka. Tujuan dari penelitian ini adalah untuk melihat bagaimana pengaruh semantik dari IndoBERT terhadap model hybrid deep learning yang digunakan untuk mendeteksi cyberbullying pada platform 'X'. Penelitian ini menggunakan 30.084 tweet dengan pendekatan hybrid deep learning yang menggabungkan CNN dan LSTM. Dalam skenario IndoBERT, fitur IndoBERT pertama-tama digabungkan dengan TF-IDF, kemudian diperluas menggunakan FastText sebelum diterapkan pada model hybrid deep learning. Hasil pengujian menghasilkan tingkat akurasi oleh: CNN (79,41%), LSTM (78,77%), CNN-LSTM (81,18%), CNN-LSTM-IndoBERT (82,05%). Hasil dari penelitian ini menunjukkan bahwa model hybrid-IndoBERT meningkatkan akurasi sebesar 2,64% dibandingkan dengan baseline awal CNN dan 3,28% dibandingkan dengan baseline awal LSTM. Penelitian ini telah berhasil mengidentifikasi cyberbullying di platform 'X', yang merupakan langkah penting dalam upaya membangun lingkungan yang lebih aman di platform 'X'. Kontribusi utama dari penelitian ini adalah untuk mengeksplorasi pengaruh fitur semantik IndoBERT pada model hybrid deep learning untuk deteksi cyberbullying dan mengintegrasikan ekstraksi fitur TF-IDF serta ekspansi fitur FastText untuk meningkatkan performa klasifikasi teks.

Kata Kunci: cyberbullying; hybrid deep learning ; indoBERT; TF-IDF; fasttext