

ABSTRAK

Twitter adalah layanan media sosial yang penggunaannya dapat membuat, memposting, memperbarui, dan membaca pesan singkat yang disebut *tweet*. Namun, Twitter sering disalahgunakan oleh para penggunaannya dengan memposting pesan negatif (*tweet*) yang mengandung *cyberbullying*. *Cyberbullying* berdampak buruk pada kesehatan mental para korban, dengan beberapa kasus yang secara tragis mengarah pada bunuh diri karena stres yang hebat. Oleh karena itu, perlu dilakukan tindakan *preventif*, antara lain dengan membuat sistem pendeteksi *cyberbullying* di Twitter. Penelitian ini mengusulkan pendekatan *hybrid deep learning*, ekspansi fitur menggunakan Word2Vec, dan ekstraksi fitur menggunakan TF-IDF dalam membangun sistem deteksi *cyberbullying* di Twitter berbahasa Indonesia. Serangkaian skenario pengujian dilakukan pada sistem yang dibangun dengan menggunakan 29,085 data *tweet* berbahasa Indonesia. Hasil akhir dari penelitian ini menunjukkan akurasi terbaik diperoleh untuk CNN-LSTM *hybrid* dengan akurasi sebesar 79,26% dan untuk *hybrid* LSTM-CNN dengan akurasi sebesar 79,48%. Hasil ini membuktikan bahwa menggabungkan model *hybrid*, Word2Vec sebagai ekstensi fitur, dan TF-IDF sebagai ekstraksi fitur memberikan skor akurasi terbaik dibandingkan dengan model *deep learning* lainnya. Dengan demikian, penelitian ini berhasil mendeteksi *cyberbullying* di Twitter sehingga dapat dimanfaatkan untuk menciptakan lingkungan media sosial yang lebih sehat bagi penggunaannya.

Kata kunci: deteksi *cyberbullying*, *hybrid deep learning*, ekspansi fitur, Word2Vec