

Deteksi Ujaran Kebencian Menggunakan CNN dan BiGRU dengan Mekanisme Attention di Twitter

Qomarudin Sifak¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹qomaru@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id

Abstrak

Indonesia menjadi negara yang memimpin dalam penggunaan media sosial terbanyak di Asia. Twitter merupakan salah satu platform media sosial yang paling populer di Indonesia. Twitter adalah media sosial yang biasanya digunakan untuk bertukar pendapat, kritik, dan bercerita. Hal ini sering disalahgunakan oleh pengguna internet, termasuk penyebaran ujaran kebencian yang ditujukan kepada individu atau kelompok. Oleh karena itu, tujuan dari penelitian ini adalah untuk mengatasi permasalahan tersebut dengan membangun sistem pendeteksi ujaran kebencian di Twitter dalam bahasa Indonesia. Penelitian ini menggunakan 69,484 data *tweet* dan mengimplementasikan model *hybrid deep learning* dengan *Convolutional Neural Network* (CNN) dan *Bidirectional Gated Recurrent Unit* (BiGRU). Metode lain yang diterapkan adalah *Bidirectional Encoder Representation from Transformer* (BERT) yang digunakan sebagai *word embedding* untuk membantu sistem lebih memahami konteks dan makna *tweet* dan mekanisme atensi untuk membantu sistem menemukan kata penting dari *tweet*. Penelitian ini menguji delapan pendekatan *hybrid* dalam proses klasifikasi, baik CNN-BiGRU maupun BiGRU-CNN dan menambahkan *attention mechanism* pada model-model tersebut. Hasilnya menunjukkan model *hybrid* BiGRU-CNN dengan *attention mechanism* untuk setiap *layer* mencapai akurasi tertinggi yaitu 88,12%.

Kata kunci : ujaran kebencian, twitter, convolutional neural network, bidirectional gated recurrent unit, attention mechanism