

Deteksi Ujaran Kebencian pada Komentar Twitter Indonesia Menggunakan *Convolutional Neural Network*(CNN) dan *FastText Word Embedding*

Fadhilah Nadia Puteri¹, Yuliant Sibaroni², Fitriyani³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹fadhilahnadiap@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id,

³fitriyani@telkomuniversity.ac.id

Abstrak

Ujaran kebencian merupakan masalah yang sering muncul di Indonesia, termasuk di platform media sosial seperti Twitter. Ujaran kebencian mengacu pada segala bentuk komunikasi, baik lisan, tertulis, maupun simbolik, yang mungkin menyinggung, mengancam, atau menghina individu atau kelompok berdasarkan atribut seperti agama, ras, etnisitas, orientasi seksual, atau karakteristik lainnya. Adanya kebebasan berekspresi dan berkomunikasi di media sosial memicu penyebaran ujaran kebencian dengan cepat dan luas. Untuk menghindari hal ini, diperlukan sistem yang dapat mendeteksi ujaran kebencian di media sosial. Deep learning memiliki potensi lebih baik dalam mengenali dan menganalisis pola bahasa yang mencerminkan ujaran kebencian dalam teks. Dalam penelitian sebelumnya, akurasi yang diperoleh adalah 73,2% dengan menggunakan metode Convolutional Neural Network. Penelitian ini mengusulkan sebuah sistem deteksi ujaran kebencian menggunakan model Convolutional Neural Network dan word embedding FastText. Performa model klasifikasi Convolutional Neural Network dan FastText sebagai word embedding memberikan hasil yang sangat baik dalam mendeteksi ujaran kebencian, dengan melibatkan proses K-Fold Cross Validation untuk menentukan nilai dropout yang tepat, dapat mencapai nilai akurasi 80%. Nilai akurasi yang dihasilkan dapat menjadi tolak ukur bahwa model yang telah dibangun mampu menghindari penyebaran ujaran kebencian di media sosial.

Kata kunci : ujaran kebencian, twitter, deep learning, convolutional neural network, fasttext

