

Deteksi Ujaran Kebencian Bahasa Indonesia Pada Kolom Komentar Instagram Artis Indonesia Menggunakan Metode RoBERTa

Adhe Akram Azhari¹, Dr. Yuliant Sibaroni, S.Si., M.T.², Sri Suryani Prasetyowati M.Si³

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

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹adheakramazhari@students.telkomuniversity.ac.id, ²Yuliant@telkomuniversity.ac.id,

³srisuryani@telkomuniversity.ac.id

Abstrak

Penelitian ini mendeteksi komentar ujaran kebencian dari komentar postingan Instagram dimana metode yang digunakan adalah RoBERTa. Model Roberta dipilih berdasarkan pertimbangan bahwa model ini memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan teks berbahasa Inggris dibandingkan dengan model lainnya, dan kemungkinan memiliki potensi yang baik dalam mendeteksi bahasa Indonesia seperti yang digunakan dalam penelitian ini. Terdapat dua skenario pengujian yaitu full-preprocessing dan non full-preprocessing dimana hasil percobaan menunjukkan bahwa non full-preprocessing memiliki nilai rata-rata akurasi lebih tinggi dibandingkan full-preprocessing, dan nilai rata-rata akurasi non full-preprocessing adalah 85,09%. Full-preprocessing meliputi beberapa tahapan preprocessing yaitu cleansing, case folding, normalization, tokenization, dan stemming. Sedangkan non full-preprocessing mencakup semua proses dalam preprocessing kecuali proses stemming. Ini menunjukkan bahwa RoBERTa memprediksi komentar dengan baik saat tidak menggunakan pemrosesan awal penuh.

Kata kunci : Instagram, RoBERTa, Full-PreProcessing, Non Full-PreProcessing, Ujaran Kebencian
