

Abstrak

Saat ini banyak contoh yang biasa digunakan dalam pengembangan masalah deep learning, seperti Transformers. Penelitian ini menggunakan salah satu arsitektur dari Transformers yaitu IndoBERT, IndoBERT itu sendiri terdiri dari BERT yaitu Bidirectional Encoder Representations from Transformers yang biasa digunakan untuk masalah deep learning. Metode IndoBERT diimplementasikan untuk mendeteksi penggunaan kalimat abusive pada teks bahasa Indonesia. Dataset yang digunakan dalam penelitian ini memiliki ketidakseimbangan jumlah data pada setiap kelas sehingga akan dilakukan penambahan data untuk mengetahui pengaruh penambahan jumlah data terhadap kinerja hasil arsitektur tersebut. Tahapan pengerjaan dalam penelitian ini dimulai dari dataset, pra-pemrosesan data, pembuatan model dengan metode IndoBERT untuk mendeteksi kalimat abusive, pelatihan dan pengujian. Pengujian dilakukan terhadap arsitektur KNN, SVM, Naive Bayes, BERT Multilingual Base dan BERT Base lalu dibandingkan dengan IndoBERT. Hasil pengujian menunjukkan bahwa selain BERT Multilingual Base, BERT Base dan IndoBERT hanya dapat memprediksi terhadap kelas mayoritas sehingga dilakukan penambahan penggunaan dataset. Hasil pengujian menunjukkan IndoBERT dapat lebih baik dalam mengklasifikasikan kalimat abusive pada teks bahasa Indonesia. Di model BERT Base, berhasil menghasilkan nilai F1 Score untuk semua kelas sebesar 0.6842 IndoBERT sudah dapat menghasilkan nilai F1 Score untuk semua kelas lebih baik.

Kata kunci : Kalimat Abusive, IndoBERT, F1 Score, SVM, Naive Bayes, BERT
