

ABSTRAK

Media sosial, terutama Twitter, telah menjadi wadah penting untuk berkomunikasi, berbagi opini, dan menyampaikan pandangan mengenai berbagai isu sosial dan politik. Namun, platform ini juga seringkali mencakup konten negatif seperti kekerasan verbal, yang dapat memicu emosi dan ketidaknyamanan penggunanya. Oleh karena itu, analisis sentimen terhadap konten-konten di Twitter terkait kekerasan verbal menjadi penting. Penelitian ini bertujuan untuk menganalisis sentimen kekerasan verbal Bahasa Indonesia pada Twitter dengan menggunakan model INDOBERT dan mengevaluasi hasilnya menggunakan confusion matrix, selain itu penelitian ini sebagai upaya penulis dalam berkontribusi dalam mengatasi masalah kekerasan verbal di media sosial dan kontribusi terhadap ilmu NLP(Natural Language Processing).

Penelitian ini dimulai dengan proses crawling, labeling dataset secara manual, dan labeling dataset menggunakan metode supervised learning, yang kemudian dataset akan di analisa model INDOBERT. Temuan utama penelitian ini adalah labeling secara supervised learning hanya akan efektif pada data yang berimbang dan baik. Temuan selanjutnya menunjukkan bahwa penggunaan model INDOBERT untuk analisis sentimen kekerasan verbal Bahasa Indonesia mencapai akurasi sebesar 72%, lebih baik dibandingkan dengan menggunakan model BERT dengan akurasi 69% pada data yang sama dan *hyperparameter* yang sama. Hasil ini menunjukkan bahwa INDOBERT lebih efektif dalam memproses bahasa Indonesia.