

INTISARI.

Media sosial X menjadi sarana utama masyarakat dalam menyampaikan opini, termasuk dalam Bahasa Jawa. Namun, analisis sentimen terhadap teks berbahasa Jawa masih terbatas karena karakteristik linguistiknya yang berbeda serta minimnya model yang secara khusus menangani bahasa daerah. Permasalahan utama penelitian ini adalah bagaimana mengklasifikasikan sentimen Bahasa Jawa secara efektif serta membandingkan kinerja model IndoBERT dan mBERT. Penelitian ini menerapkan pendekatan *deep learning* berbasis *transformer* untuk analisis sentimen tiga kelas, yaitu positif, negatif, dan netral. Metode penelitian meliputi pengumpulan data dari media sosial X, pelabelan sentimen secara manual, *preprocessing* data berupa *cleaning data* dan *case folding*, serta pembagian data latih dan data uji dengan rasio 80:20. Proses pemodelan dilakukan menggunakan IndoBERT dan mBERT, kemudian dievaluasi menggunakan *confusion matrix* dengan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kedua model mampu melakukan klasifikasi sentimen dengan baik, namun IndoBERT menghasilkan performa yang lebih unggul dengan akurasi 0.7959 dibandingkan mBERT yang hanya 0.7643. Hal ini disebabkan IndoBERT lebih mampu menangkap karakteristik linguistik bahasa lokal. Kesimpulannya, IndoBERT lebih efektif digunakan untuk analisis sentimen Bahasa Jawa pada media sosial X dibandingkan mBERT.

Kata kunci: Analisis sentimen, Bahasa Jawa, IndoBERT, mBERT, media sosial X

ABSTRACT

X social media serves as a primary platform for the public to express opinions, including content written in the Javanese language. However, sentiment analysis of Javanese text remains limited due to its distinct linguistic characteristics and the lack of models specifically designed to handle regional languages. The main problem addressed in this study is how to effectively classify Javanese sentiment and compare the performance of IndoBERT and Multilingual BERT (mBERT). This study applies a transformer-based deep learning approach for three-class sentiment analysis, namely positive, negative, and neutral. The research methodology includes data collection from X social media, manual sentiment labeling, data preprocessing consisting of text cleaning and case folding, and splitting the dataset into training and testing sets with an 80:20 ratio. The modeling process is conducted using IndoBERT and mBERT, and the models are evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results show that both models perform well in sentiment classification; however, IndoBERT achieves superior performance with an accuracy of 0.7959 compared to mBERT, which achieves an accuracy of 0.7643. This improvement is attributed to IndoBERT's ability to better capture local linguistic characteristics. In conclusion, IndoBERT is more effective than mBERT for Javanese sentiment analysis on X social media.

Keywords: sentiment analysis, Javanese language, IndoBERT, mBERT, X social media