

ABSTRACT

The development of information technology encourages the government to utilize digital media to improve public services, such as through the Jabar Quick Response (JQR) program launched to handle humanitarian issues and disasters quickly via social media. This study aims to compare the performance of three BERT models, namely IndoBERT, MultilingualBERT (mBERT), and IndoROBERTa in analyzing sentiment on tweets related to the JQR program. The evaluation was conducted using accuracy, precision, recall, and F1-Score metrics. The results showed that IndoBERT and IndoROBERTa both obtained the best results with 93% accuracy. Although all three recorded identical values on the key metrics, there were differences in their performance characteristics. IndoBERT had a higher precision (macro) of 91%, indicating the accuracy of the model in predicting the recognized classes, while IndoROBERTa recorded a higher recall (macro) value of 89%, indicating its ability to recognize the entire class distribution more evenly, including minority classes. The mBERT model showed the lowest performance with an accuracy of 92%, F1-score macro of 87%, and F1-score weighted of 92%. Based on these results, IndoROBERTa can be considered as a more representative model if the evaluation priority is directed towards the balance of recognizing all classes proportionally in the sentiment analysis task of social media data.

Keywords: Accuracy, IndoBERT, IndoROBERTa, mBERT, Sentiment Analysis

ABSTRAK

Perkembangan teknologi informasi mendorong pemerintah untuk memanfaatkan media digital guna meningkatkan kualitas layanan publik, salah satunya melalui program Jabar *Quick Response* (JQR) yang bertujuan menangani isu kemanusiaan dan kebencanaan secara cepat melalui media sosial. Penelitian ini bertujuan untuk membandingkan performa tiga model yaitu IndoBERT, MultilingualBERT (mBERT), dan IndoROBERTa dalam melakukan analisis sentimen terhadap tweet yang berkaitan dengan program JQR. Evaluasi dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score (macro dan weighted)*. Hasil penelitian menunjukkan bahwa IndoBERT dan IndoROBERTa sama-sama memperoleh hasil terbaik dengan akurasi sebesar 93%. Meskipun ketiganya mencatatkan nilai yang identik pada metrik utama tersebut, terdapat perbedaan dalam karakteristik kinerjanya. IndoBERT memiliki *precision (macro)* yang lebih tinggi sebesar 91%, mengindikasikan ketepatan model dalam memprediksi kelas yang dikenali, sedangkan IndoROBERTa mencatatkan nilai *recall (macro)* yang lebih tinggi sebesar 89%, menunjukkan kemampuannya dalam mengenali seluruh distribusi kelas secara lebih merata, termasuk kelas minoritas. Model mBERT menunjukkan performa paling rendah dengan akurasi sebesar 92%, *F1-score macro* sebesar 87%, dan *F1-score weighted* sebesar 92%. Berdasarkan hasil tersebut, IndoROBERTa dapat dianggap sebagai model yang lebih representatif apabila prioritas evaluasi diarahkan pada keseimbangan dalam mengenali seluruh kelas secara proporsional dalam tugas analisis sentimen terhadap data media sosial.

Kata kunci: Analisis Sentimen, Akurasi, IndoBERT, IndoROBERTa, mBERT.