

ABSTRAK

Media sosial, terutama Twitter, menjadi platform utama bagi masyarakat Indonesia untuk mengekspresikan emosi dan berbagi pendapat. Dengan lebih dari 24 juta pengguna aktif di Indonesia, volume data yang dihasilkan sangat besar setiap harinya. Analisis emosi pada *tweet* menjadi penting untuk memahami dinamika sentimen publik terhadap berbagai peristiwa dan kebijakan. Namun, pemrosesan teks bahasa Indonesia pada *tweet* menghadapi tantangan besar, terutama karena penggunaan bahasa informal (slang) yang meluas. Penggunaan slang dalam teks sosial media dapat menurunkan akurasi model *Natural Language Processing* (NLP) jika tidak dilakukan normalisasi yang tepat. Selain itu, distribusi kelas dalam klasifikasi emosi pada *tweet* sering kali tidak seimbang, dengan beberapa emosi lebih jarang muncul daripada yang lain, menyebabkan bias dalam prediksi model.

Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan mengimplementasikan normalisasi *slang word* dan teknik *balancing* kelas menggunakan *Random Oversampling* (ROS) untuk meningkatkan akurasi klasifikasi emosi pada *tweet* berbahasa Indonesia menggunakan model *IndoBERTweet*. Normalisasi *slang word* bertujuan untuk menggantikan slang word dengan bentuk formal yang lebih mudah dipahami oleh model, sedangkan teknik *balancing* kelas digunakan untuk mengurangi bias terhadap kelas mayoritas. Model *IndoBERTweet* yang telah dilatih khusus untuk memahami teks informal berbahasa Indonesia diharapkan dapat menunjukkan peningkatan performa melalui penerapan kedua teknik tersebut.

Hasil penelitian menunjukkan bahwa penerapan normalisasi *slang word* meningkatkan akurasi model dari 92,57% menjadi 93,02%. Teknik *balancing* kelas lebih lanjut meningkatkan akurasi menjadi 93,67%, dan kombinasi kedua teknik tersebut menghasilkan akurasi tertinggi sebesar 94,75%. Penelitian ini membuktikan bahwa normalisasi *slang word* dan *balancing* kelas secara efektif meningkatkan kinerja model *IndoBERTweet* dalam klasifikasi emosi pada *tweet* berbahasa Indonesia, serta memberikan kontribusi dalam pengembangan sistem NLP yang lebih akurat untuk aplikasi analisis sentimen dan pemantauan opini publik.

Kata Kunci: Klasifikasi Emosi, Normalisasi Slang, *Random Oversampling*, *IndoBERTweet*

ABSTRACT

Social media, particularly Twitter, has become a primary platform for Indonesians to express emotions and share opinions. With more than 24 million active users in Indonesia, a massive volume of data is generated daily. Emotion analysis on tweets is essential to understand public sentiment dynamics toward various events and policies. However, processing Indonesian-language text on tweets faces significant challenges, especially due to the widespread use of informal language (slang). The use of slang in social media text can reduce the accuracy of Natural Language Processing (NLP) models if proper text normalization is not applied. In addition, class distribution in emotion classification on tweets is often imbalanced, where some emotions appear much less frequently than others, leading to biased model predictions.

This study aims to address these issues by implementing slang word normalization and class balancing techniques using Random Oversampling (ROS) to improve the accuracy of emotion classification on Indonesian-language tweets using the IndoBERTweet model. Slang word normalization aims to replace informal words with formal forms that are more understandable by the model, while class balancing techniques are applied to reduce bias toward majority classes. IndoBERTweet, a pre-trained model specifically trained on Indonesian tweets, is expected to show improved performance through the implementation of these two techniques.

The results show that applying slang word normalization increased the model's accuracy from 92,57% to 93,02%. The class balancing technique further improved accuracy to 93,67%, and the combination of both techniques achieved the highest accuracy of 94,75%. This study demonstrates that the combination of slang word normalization and class balancing effectively enhances the performance of the IndoBERTweet model in emotion classification on Indonesian-language tweets and contributes to the development of more accurate NLP systems for sentiment analysis and public opinion monitoring applications.

Keywords: *Emotion Classification, Normalization, Random Oversampling, IndoBERTweet.*