

ABSTRAK

Penelitian ini bertujuan untuk menerapkan algoritma *Random Forest* dalam membangun sistem klasifikasi emosi pada tweet berbahasa Indonesia. *Dataset* yang digunakan terdiri dari 4.403 tweet yang terbagi ke dalam lima kategori emosi, yaitu *anger*, *fear*, *happy*, *love*, dan *sadness*. Tahapan klasifikasi diawali dengan *pre-processing* teks, meliputi normalisasi *slang*, koreksi *typo*, *stopword removal*, dan *stemming*. Selanjutnya, fitur diekstraksi menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Ketidakseimbangan distribusi data antar kelas ditangani menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Model *Random Forest* dikembangkan dengan konfigurasi parameter terbaik berdasarkan hasil uji kombinasi. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 95,74% dengan nilai *f1-score* yang tinggi dan stabil di seluruh kelas emosi. Temuan ini membuktikan bahwa algoritma *Random Forest* mampu memberikan hasil klasifikasi emosi dengan akurasi yang tinggi jika dikombinasikan dengan *Balancing* data dan konfigurasi parameter *Random Forest*. Selanjutnya model ini bisa dibandingkan dengan metode *deep learning* untuk menilai peningkatan kinerja pada data yang lebih kompleks.

Kata Kunci: Klasifikasi Emosi, Tweet Indonesia, *Random Forest*, SMOTE, TF-IDF.

ABSTRACT

This research aims to apply the Random Forest algorithm to build an emotion classification system for Indonesian-language tweets. The dataset consists of 4,403 tweets categorized into five emotion classes: anger, fear, happy, love, and sadness. The classification process begins with text preprocessing, including slang normalization, typo correction, stopwords removal, and stemming. Features are then extracted using the Term Frequency-Inverse Document Frequency (TF-IDF) method. Class imbalance is addressed using the Synthetic Minority Over-sampling Technique (SMOTE). The Random Forest model is developed using the best parameter configuration based on a combination of experimental results. Evaluation results show that the model achieves an accuracy of 95.74% with high and stable F1-scores across all emotion classes. These findings demonstrate that the Random Forest algorithm can deliver high classification accuracy when combined with data balancing and proper parameter configuration. This model can further be compared with deep learning methods to assess potential performance improvements on more complex datasets.

Keywords: *Emotion Classification, Indonesian Tweets, Random Forest, SMOTE, TF-IDF.*