

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model deteksi sarkasme dalam teks berbahasa Indonesia dengan pendekatan *hybrid* yang menggabungkan metode statistik dan pembelajaran mendalam. Fitur statistik diperoleh melalui perhitungan entropi 3-gram, yang merepresentasikan tingkat ketidakteraturan dalam teks, sementara fitur kontekstual diperoleh melalui representasi *embedding* dari algoritma IndoBERT. Kedua jenis fitur ini kemudian digabungkan dan digunakan sebagai input untuk melatih model klasifikasi menggunakan jaringan saraf tiruan. Dataset yang digunakan berupa kumpulan *tweet* berbahasa Indonesia yang telah dilabeli secara manual ke dalam dua kategori, yaitu sarkasme dan non-sarkasme. Hasil evaluasi menunjukkan bahwa model *hybrid* yang dikembangkan mampu menghasilkan akurasi sebesar 84,7% dan *f1-score* sebesar 84,1%, mengungguli model individu. Sistem ini juga telah diimplementasikan dalam bentuk aplikasi berbasis web yang memungkinkan pengguna melakukan analisis teks secara langsung, dengan fitur tambahan seperti riwayat analisis dan umpan balik pengguna. Penelitian ini menunjukkan bahwa penggabungan fitur statistik dan kontekstual efektif dalam meningkatkan akurasi klasifikasi sarkasme, serta mampu diterapkan dalam sistem nyata yang dapat digunakan oleh masyarakat.

Kata kunci: Sarkasme, Entropi 3-Gram, IndoBERT, Deteksi Teks, Sistem Web, Pemrosesan Bahasa Alami

ABSTRACT

This study aims to develop a sarcasm detection model for Indonesian-language texts using a hybrid approach that combines statistical methods and deep learning. Statistical features are extracted through 3-gram entropy calculations, representing the irregularity level in a text, while contextual features are obtained via embedding representations from the IndoBERT algorithm. These features are then combined and used as input for training a classification model using an artificial neural network. The dataset consists of Indonesian tweets that were manually labeled into two categories: sarcasm and non-sarcasm. Evaluation results show that the hybrid model achieved an accuracy of 84.7% and an f1-score of 84.1%, outperforming individual models. The system was implemented as a functional web-based application that allows users to perform real-time text analysis, supported by additional features such as analysis history and user feedback. This research demonstrates that integrating statistical and contextual features is effective in improving sarcasm classification accuracy and can be successfully applied in real-world applications accessible to the public.

Keywords: Sarcasm, 3-Gram Entropy, IndoBERT, Text Classification, Web System, Natural Language Process