

ABSTRAK

Pratiwi, Rystiana L. D. 2026. Analisis Perbandingan Representasi Fitur TF-IDF, *FastText*, dan BERT Pada Model *Support Vector Machine* Untuk Klasifikasi Sentimen Instagram Folkative. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Moh. Dasuki, M.Kom; Dr. Reni Umilasari, S.Pd, M.Si.

Analisis sentimen pada media sosial memiliki tantangan tersendiri karena penggunaan bahasa informal, bahasa gaul, sarkasme, dan kesalahan dalam penulisan yang umum ditemukan pada platform seperti Instagram. Penelitian ini bertujuan untuk menerapkan dan membandingkan tiga metode representasi fitur teks, yaitu TF-IDF sebagai *baseline* serta *word embedding FastText* dan BERT (*Bidirectional Encoder Representations from Transformers*), pada model *Support Vector Machine* (SVM) untuk klasifikasi sentimen komentar akun Instagram Folkative. Data yang digunakan berjumlah 11.914 komentar yang telah melalui proses *pre-processing*. Untuk pelabelan data, digunakan metode semi-otomatis berbasis *lexicon-based* yang divalidasi oleh pakar Bahasa Indonesia. Optimasi model dilakukan menggunakan *GridSearchCV* dan validasi model melalui *Stratified 5-Fold Cross Validation*. Hasil penelitian menunjukkan bahwa TF-IDF dengan SVM memperoleh performa terbaik dengan akurasi 87,61% menggunakan kernel RBF, diikuti *FastText* dengan SVM sebesar 87,54% menggunakan kernel RBF, dan BERT dengan SVM sebesar 86,38% menggunakan kernel RBF. Hasil ini menunjukkan bahwa metode yang lebih sederhana tetap dapat memberikan hasil terbaik apabila sesuai dengan karakteristik data. TF-IDF lebih efektif dalam mengklasifikasikan komentar karena komentar pada akun Instagram Folkative cenderung singkat dan mengandung kata sentimen yang eksplisit, sedangkan *FastText* dan BERT belum mampu memberikan peningkatan performa yang lebih tinggi pada dataset ini.

Kata kunci: Analisis Sentimen, Instagram, TF-IDF, *FastText*, BERT, *Support Vector Machine*

ABSTRACT

Pratiwi, Rystiana L. D. 2026. *Comparative Analysis of TF-IDF, FastText, and BERT Feature Representations in Support Vector Machine Models for Folkative Instagram Sentiment Classification*. Undergraduate Thesis. Bachelor Program. Informatics Engineering Study Program. Universitas Muhammadiyah Jember.

Advisors: Moh. Dasuki, M.Kom; Dr. Reni Umilasari, S.Pd, M.Si.

Sentiment analysis on social media presents its own challenges due to the use of informal language, slang, sarcasm, and typos commonly found on platforms like Instagram. This study aims to apply and compare three text feature representation methods, namely TF-IDF as a baseline and FastText word embedding and BERT (Bidirectional Encoder Representations from Transformers), on a Support Vector Machine (SVM) model for sentiment classification of comments on the Folkative Instagram account. The data used amounted to 11,914 comments that had gone through a pre-processing process. For data labeling, a semi-automatic lexicon-based method was used that was validated by Indonesian language experts. Model optimization was performed using GridSearchCV and model validation through Stratified 5-Fold Cross Validation. The results showed that TF-IDF with SVM achieved the best performance with an accuracy of 87.61% using the RBF kernel, followed by FastText with SVM at 87.54% using the RBF kernel, and BERT with SVM at 86.38% using the RBF kernel. These results demonstrate that simpler methods can still provide superior results if they match the characteristics of the data. TF-IDF was more effective in classifying comments because comments on the Folkative Instagram account tend to be short and contain explicit sentiment, while FastText and BERT were unable to provide significant performance improvements on this dataset.

Keywords: Sentiment analysis, Instagram, TF-IDF, FastText, BERT, Support Vector Machine