

ABSTRAK

Pemilihan Gubernur Jawa Timur 2024 menjadi topik perbincangan yang luas di Twitter, mencerminkan opini publik terhadap pasangan calon. Penelitian ini menggunakan *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen tweet secara otomatis. Data dikumpulkan melalui *Tweet-Harvest*, kemudian diproses menggunakan teknik *pre-processing* seperti *case folding*, tokenisasi, penghapusan karakter khusus, serta normalisasi kata. Selanjutnya, data direpresentasikan dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk membangun fitur yang digunakan dalam klasifikasi. Model diuji dalam beberapa skenario, termasuk eksperimen dengan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani ketidakseimbangan data, serta pengaruh *stopword removal* terhadap performa model. Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil menunjukkan pendekatan terbaik tanpa *stopword removal* dan tanpa SMOTE, dengan akurasi mencapai 0.67. Keberadaan *stopword* membantu model membedakan sentimen, terutama antara kelas negatif dan netral. Tantangan utama dalam penelitian ini adalah *overlapping* antar kelas yang menyulitkan klasifikasi serta *sparse matrix* dalam representasi TF-IDF. Penelitian selanjutnya dapat mengeksplorasi metode lain serta representasi teks berbasis *Word2Vec* atau *BERT* untuk meningkatkan akurasi klasifikasi.

Kata Kunci: Analisis Sentimen, Pemilu 2024, SMOTE, Support Vector Machine, Stopword Removal, TF-IDF