

ABSTRAK

Pertumbuhan bisnis *e-commerce* yang pesat seperti Xavera Store menghasilkan jumlah ulasan konsumen yang besar tetapi sering kali tidak konsisten dalam hal kualitas, di mana banyak data ulasan yang kosong meskipun konsumen memberikan penilaian. Penelitian ini bertujuan untuk melaksanakan analisis sentimen terhadap ulasan produk hijab dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Kernel Linear*. Mengingat terbatasnya jumlah *dataset* dan adanya ulasan yang tidak memiliki teks, penelitian ini menerapkan strategi substitusi nilai penilaian sebagai pengganti ulasan yang kosong guna menjaga integritas data. Proses ekstraksi fitur dilakukan dengan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Hasil dari pengujian menunjukkan bahwa model SVM berhasil mencapai tingkat Akurasi Total sebesar 98,47% dengan pembagian data 80% untuk data latih dan 20% untuk data uji (262 data uji). Model menunjukkan kinerja yang sangat baik pada kelas positif dengan F1-score 0,99, tetapi mengalami tantangan pada kelas negatif dengan presisi 0,70 sebagai akibat dari ketidakseimbangan data (*class imbalance*). Penelitian ini menyimpulkan bahwa pemakaian rating sebagai pengganti ulasan kosong adalah langkah yang efektif dalam memperluas jangkauan data pada *dataset* yang ukuran kecil, serta algoritma SVM terbukti sangat andal dalam mengklasifikasikan sentimen konsumen secara otomatis.

Kata kunci: *Sentiment Analysis*, *Rating*, Support Vector Machine, TF-IDF

ABSTRACT

The rapid growth of e-commerce businesses such as Xavera Store generates a significant volume of customer reviews, yet they are often inconsistent in quality, with many instances of empty review text despite customers providing ratings. This study aims to implement sentiment analysis on hijab product reviews using the Support Vector Machine (SVM) algorithm with a Linear Kernel. Given the limited dataset size and the presence of textless reviews, this research applies a rating-based substitution strategy to replace empty reviews to maintain data integrity. The feature extraction process was conducted using Term Frequency-Inverse Document Frequency (TF-IDF). The experimental results demonstrate that the SVM model achieved an Overall Accuracy of 98.47% using an 80:20 data split ratio (262 test data). The model showed exceptional performance in the positive class with an F1-score of 0.99, although it faced challenges in the negative class with a precision of 0.70 due to extreme class imbalance. This study concludes that utilizing ratings as a substitute for empty reviews is effective in expanding data coverage within small-scale datasets, and the SVM algorithm is proven reliable for automated customer sentiment classification.

Keywords: Sentiment Analysis, Support Vector Machine, Rating, TF-IDF