

ABSTRAK

Penelitian ini dilatarbelakangi oleh meningkatnya jumlah ulasan pelanggan Toko Batik Madura pada platform e-commerce yang menyebabkan analisis sentimen secara manual menjadi tidak efisien. Permasalahan yang dihadapi meliputi tingginya dimensi data teks, ketidakseimbangan kelas sentimen, serta kebutuhan akan algoritma klasifikasi yang stabil dan optimal. Oleh karena itu, Support Vector Machine (SVM) dipilih karena memiliki kemampuan generalisasi yang baik pada data berdimensi tinggi. Untuk memperoleh konfigurasi terbaik, dilakukan pengujian beberapa kernel, yaitu Linear, Radial Basis Function (RBF), Polynomial, dan Sigmoid. Selain itu, pendekatan ensemble learning menggunakan Bootstrap Aggregating (Bagging) diterapkan untuk meningkatkan stabilitas model dan sensitivitas terhadap kelas minoritas. Data ulasan pelanggan diproses melalui tahapan preprocessing teks yang meliputi normalisasi, tokenisasi, stopword removal, dan stemming, kemudian direpresentasikan menggunakan Term Frequency–Inverse Document Frequency (TF-IDF). Ketidakseimbangan kelas ditangani menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa SVM kernel Linear dengan parameter $C = 1$ menghasilkan performa terbaik dengan akurasi 0.91, precision 0.81, recall 0.87, dan F1-score 0.83. Penerapan Bagging-SVM tidak meningkatkan akurasi secara signifikan, namun mampu meningkatkan nilai recall hingga 0.88 sehingga lebih sensitif terhadap kelas sentimen minoritas.

Kata Kunci: *Analisis Sentimen, Support Vector Machine, Bagging-SVM, TF-IDF, SMOTE.*

ABSTRACT

This research is motivated by the increasing number of customer reviews of Madura Batik stores on e-commerce platforms, which makes manual sentiment analysis inefficient. The main challenges include high-dimensional textual data, sentiment class imbalance, and the need for a stable and optimal classification algorithm. Therefore, Support Vector Machine (SVM) was selected due to its strong generalization capability in high-dimensional data. Several kernel functions—Linear, Radial Basis Function (RBF), Polynomial, and Sigmoid—were evaluated to obtain the best configuration. In addition, an ensemble learning approach using Bootstrap Aggregating (Bagging) was applied to improve model stability and sensitivity to minority sentiment classes. Customer review data were processed through text preprocessing stages, including normalization, tokenization, stopword removal, and stemming, and were represented using Term Frequency–Inverse Document Frequency (TF-IDF). Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). The experimental results show that the Linear kernel SVM with parameter $C = 1$ achieved the best performance, with an accuracy of 0.91, precision of 0.81, recall of 0.87, and an F1-score of 0.83. Although the Bagging-SVM approach did not significantly improve accuracy, it increased recall to 0.88, indicating better sensitivity to minority sentiment classes.

Keywords: Sentiment Analysis, Support Vector Machine, Bagging-SVM, TF-IDF, SMOTE.