ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis perbandingan metode *Support Vector Machine* (SVM) dan *Naive Bayes* dalam mengklasifikasikan sentimen ulasan produk *e-commerce* pada platform Tokopedia menggunakan data *web scraping* sebanyak 571 ulasan periode tahun 2024. Data mencakup variabel teks ulasan, tanggal penulisan, dan *username* yang diproses melalui *preprocessing* teks (*text cleaning*, *stopword removal*, *stemming* dengan Sastrawi), *auto-labeling* menggunakan pendekatan *lexicon-based*, dan *feature extraction* TF-IDF yang menghasilkan fitur diskriminatif untuk klasifikasi sentimen. Pembagian data dilakukan secara *stratified* dengan proporsi *training* (80%) dan *testing* (20%) pada *dataset* hasil filtering klasifikasi *binary* (positif vs negatif) yang terdiri dari 114 ulasan dengan komposisi 18 ulasan positif dan 96 ulasan negatif. Hasil penelitian menunjukkan perbedaan performa yang signifikan, dimana *Support Vector Machine* dengan *linear kernel* menghasilkan *accuracy* 85,96%, *precision* 83,97%, *recall* 85,96%, dan *F1-score* 84,24%, sedangkan *Multinomial Naive Bayes* menghasilkan *accuracy* 84,21%, *precision* 70,91%, *recall* 84,21%, dan *F1-score* 76,99% pada *testing set*. Berdasarkan nilai *F1-score* dari kedua metode menunjukkan bahwa *F1-score* SVM 84,24% lebih tinggi dari *F1-score Naive Bayes* 76,99%, dengan selisih 7,25 poin persentase, sehingga algoritma SVM lebih baik diterapkan untuk mengklasifikasikan sentimen ulasan produk *e-commerce*. Karakteristik *imbalanced dataset* (84,2% negatif vs 15,8% positif) menunjukkan bahwa SVM mampu mengidentifikasi kedua kelas sentimen dengan performa yang lebih seimbang, sementara *Naive Bayes* mengalami bias ekstrem dengan memprediksi seluruh *testing samples* sebagai sentimen negatif dan gagal mengidentifikasi kelas positif. Analisis *confusion matrix* mengkonfirmasi superioritas SVM dalam menangani ketidakseimbangan kelas dan memberikan kemampuan diskriminasi yang lebih *robust* untuk implementasi sistem klasifikasi sentimen pada platform Tokopedia.

Kata Kunci: Analisis Sentimen, *E-commerce*, *Naive Bayes*, *Support Vector Machine*, *Text Mining*

ABSTRACT

This research aims to conduct a comparative analysis of Support Vector Machine (SVM) and Naive Bayes methods in classifying sentiment of e-commerce product reviews on the Tokopedia platform using web scraping data of 571 reviews from 2024. The data includes text review variables, writing dates, and usernames that were processed through text preprocessing (text cleaning, stopword removal, stemming with Sastrawi), auto-labeling using a lexicon-based approach, and TF-IDF feature extraction that generates discriminative features for sentiment classification. Data splitting was performed using stratified method with training (80%) and testing (20%) proportions on the filtered binary classification dataset (positive vs negative) consisting of 114 reviews with a composition of 18 positive reviews and 96 negative reviews. The research results show significant performance differences, where Support Vector Machine with linear kernel achieved accuracy 85.96%, precision 83.97%, recall 85.96%, and F1-score 84.24%, while Multinomial Naive Bayes achieved accuracy 84.21%, precision 70.91%, recall 84.21%, and F1-score 76.99% on the testing set. Based on the F1-score values from both methods, SVM's F1-score of 84.24% is higher than Naive Bayes' F1-score of 76.99%, with a difference of 7.25 percentage points, indicating that the SVM algorithm is better applied for classifying sentiment of e-commerce product reviews. The characteristics of the imbalanced dataset (84.2% negative vs 15.8% positive) demonstrate that SVM can identify both sentiment classes with more balanced performance, while Naive Bayes experiences extreme bias by predicting all testing samples as negative sentiment and failing to identify the positive class. Confusion matrix analysis confirms SVM's superiority in handling class imbalance and providing more robust discriminative capabilities for implementing sentiment classification systems on the Tokopedia platform.

Keywords: *E-commerce, Naive Bayes, Sentiment Analysis, Support Vector Machine, Text Mining*