

ABSTRAK

Tren belanja online membuat berbagai brand *marketplace* di Indonesia menerapkan strategi pemasaran terbaiknya untuk menarik minat pelanggan, salah satunya program promosi *event* belanja. Shopee dan Tokopedia merupakan dua *brand marketplace* teratas di Indonesia dengan pengunjung terbanyak berdasarkan data Similarweb tahun 2021. Pengguna *marketplace* juga saling bertukar informasi serta pengalaman seputar promosi *event* belanja *marketplace* melalui media sosial (Twitter). Tujuan dari penelitian ini adalah membangun model analisis sentimen yang mampu mengklasifikasikan *tweets* masyarakat terkait dengan program promosi *event* belanja yang dilakukan oleh Shopee dan Tokopedia. Proses pengambilan data *tweets* dilakukan menggunakan teknik *scraping*. Penelitian ini menggunakan data *tweets* pada periode yang telah ditentukan. Data set dibagi menjadi tiga, data set Shopee, data set Tokopedia, dan data set Gabungan Tokopedia dan Shopee. Rangkaian *text preprocessing* yang dilakukan yaitu *case folding*, *tokenizing*, *filtering*, normalisasi kata, dan *stemming*. Pembobotan kata dilakukan dengan TF-IDF, *Support Vector Machine* sebagai algoritma pengklasifikasian, *Grid Search* untuk mencari parameter optimal, dan *K-Fold Cross Validation* serta *Confusion Matrix* untuk validasi dan pengujian model. Berdasarkan hasil analisis dan observasi, penelitian ini mengidentifikasi *event* belanja pada Shopee yaitu tanggal 25, *flash sale*, gratis ongkir, COD, tanggal kembar, dan Shopee 12.12. Sedangkan untuk Tokopedia yaitu tanggal 25, kejar diskon, bebas ongkir, COD, WIB, dan Tokopedia 12.12. Berdasarkan hasil pelabelan data, distribusi sentimen masyarakat untuk program promosi *event* belanja Tokopedia cenderung positif, Shopee cenderung negatif, serta sentimen masyarakat terhadap program promosi *event* belanja kedua *marketplace* didominasi oleh sentimen positif. Dari hasil pengujian yang dilakukan, model yang menggunakan data set Shopee yaitu Skenario 3 dan Skenario 4 mendapat nilai akurasi tertinggi sebesar 72.12% dan 71.52%. Adapun dari hasil pencarian parameter terbaik menggunakan *Grid Search*, pengujian model menghasilkan nilai selisih tertinggi dari sebelumnya menggunakan parameter *default* untuk data set Tokopedia 1.44% dan data set Shopee sebesar 0.54%.

Kata kunci: *Support Vector Machine*, *Grid Search*, *K-Fold Cross Validation*, Analisis Sentimen, *Twitter*, *Marketplace*, Program Promosi *Event* Belanja

ABSTRACT

Online shopping trends make various brand marketplaces in Indonesia apply their best marketing strategies to attract customers, one of which is a shopping event promotion program. Shopee and Tokopedia are the two top marketplace brands in Indonesia with the most visitors based on Similarweb data in 2021. Marketplace users also exchange informations and experiences about the promotion of marketplace shopping events through social media (Twitter). The purpose of this research is to build a sentiment analysis model that is able to classify tweets related to shopping event promotion programs conducted by Shopee and Tokopedia. The process of retrieving tweets data using scraping techniques. This research used tweets data at a predetermined period. The data sets are divided into three, Shopee data set, Tokopedia data set, Shopee and Tokopedia combined data set. The process of text preprocessing are case folding, tokenizing, filtering, word normalization, and stemming. Word weighting is done with TF-IDF, Support Vector Machine as a classification algorithm, Grid Search to find optimal parameters, K-Fold Cross Validation for validation and Confusion Matrix for model testing. Based on the results of analysis and observation, this research identifies shopping events on Shopee, namely tanggal 25, flash sale, gratis ongkir, COD, tanggal kembar, and Shopee 12.12. As for Tokopedia, tanggal 25, kejar diskon, bebas ongkir, COD, WIB, and Tokopedia 12.12. Based on the results of data labeling, the distribution of public sentiment for the Tokopedia shopping event promotion program tends to be positive, Shopee tends to be negative, and public sentiment for the shopping event promotion program of both marketplaces is dominated by positive sentiment. From the test results, the model using the data set Shopee namely Scenario 3 and Scenario 4 got the highest accuracy value of 72.12% and 71.52%. As for the best parameter search results using Grid Search, model testing resulted in the highest difference value from the previous using the default parameters for Tokopedia data set of 1.44% and Shopee data set of 0.54%.

Keywords: Support Vector Machine, Grid Search, K-Fold Cross Validation, Sentiment Analysis, Twitter, Marketplace, Shopping Event Promotion Program