

PENERAPAN WEB SCRAPING DAN KLASIFIKASI SUPPORT VECTOR MACHINE UNTUK MENENTUKAN PRODUK TERLARIS SERTA TREN PASAR BERDASARKAN KATEGORI PADA MARKETPLACE

Mohammad Nafiis Septiano

ABSTRAK

Perkembangan e-commerce di Indonesia yang pesat mendorong kebutuhan untuk menganalisis data penjualan guna mengidentifikasi tren pasar dan produk terlaris. Penelitian ini menggunakan metode *web scraping* pada platform Tokopedia untuk mengumpulkan data produk yang relevan. Data yang dikumpulkan mencakup kategori, nama produk, harga, jumlah terjual, rating, dan lokasi toko. Proses pengambilan data ini memanfaatkan Selenium dan BeautifulSoup, diikuti dengan tahapan *data preprocessing* seperti pembersihan duplikasi, transformasi tipe data, penentuan label klasifikasi biner, serta pembagian data latih dan uji. Metode klasifikasi yang diterapkan adalah Support Vector Machine (SVM) dengan pendekatan *supervised learning*, di mana produk dikategorikan menjadi terlaris atau tidak terlaris berdasarkan jumlah penjualannya. Analisis tambahan dilakukan melalui visualisasi distribusi penjualan, tren pasar harian dan mingguan, serta identifikasi lima produk teratas pada setiap kategori. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi 87,59% dalam mengklasifikasikan produk. Visualisasi tren penjualan juga memberikan wawasan tentang perbedaan preferensi konsumen antar kategori. Temuan ini diharapkan dapat dimanfaatkan oleh para pelaku bisnis sebagai dasar untuk pengambilan keputusan strategis, seperti penentuan stok, strategi promosi, dan pengembangan produk yang sesuai dengan kebutuhan pasar.

Kata Kunci: *Web Scraping*, Selenium, BeautifulSoup, Data Preprocessing Tokopedia, Klasifikasi, Support Vector Machine, Analisis Tren Pasar.

**APPLICATION OF WEB SCRAPING AND SUPPORT VECTOR MACHINE
CLASSIFICATION TO DETERMINE BEST-SELLING PRODUCTS AND
MARKET TRENDS BASED ON CATEGORIES IN A MARKETPLACE**

Mohammad Nafiis Septiano

ABSTRACT

The rapid growth of e-commerce in Indonesia has increased the need for sales data analysis to identify market trends and best-selling products. This study applies a web scraping method on the Tokopedia platform to collect relevant product data, including category, product name, price, number of units sold, rating, and store location. Data collection was carried out using Selenium and BeautifulSoup, followed by data preprocessing steps such as duplicate removal, data type transformation, binary label classification, and dataset splitting into training and testing sets. The classification method employed is Support Vector Machine (SVM) with a supervised learning approach, where products are categorized as best-selling or not based on their sales volume. Additional analyses include visualization of sales distribution, daily and weekly market trends, and identification of the top five products in each category. The results show that the SVM model achieved an accuracy of 87.59% in classifying products. Furthermore, sales trend visualizations provide insights into consumer preferences across categories. These findings are expected to serve as a strategic reference for businesses in making data-driven decisions regarding inventory management, promotional strategies, and product development aligned with market needs.

Keywords: *Web Scraping, Selenium, BeautifulSoup, Data Preprocessing, Tokopedia, Classification, Support Vector Machine, Market Trend Analysis.*