

Penerapan Association Rules Mining (Algoritma Apriori) dalam Market Basket Analysis untuk Optimalisasi Tata Layout Produk pada Data Transaksi E-Commerce

Nouval Zaki¹

Teknik Informatika, STMIK Widya Cipta Dharma , Samarinda, Indonesia

E-mail: nouvalzaki31@gmail.com

ABSTRAK

Pertumbuhan transaksi e-commerce yang pesat telah menghasilkan volume data yang besar, sehingga memerlukan teknik pemrosesan yang canggih untuk mengekstraksi wawasan yang bermakna. Penelitian ini bertujuan untuk mengimplementasikan Market Basket Analysis (MBA) menggunakan algoritma Apriori dalam lingkungan Big Data. Dengan memanfaatkan PySpark, penelitian ini memproses dataset yang terdiri dari lebih dari 500.000 transaksi e-commerce. Metodologi penelitian mencakup tahap data ingestion, pra-pemrosesan data yang ketat termasuk penghapusan pesanan yang dibatalkan (return), serta pembuatan aturan asosiasi (association rules). Hasil penelitian menunjukkan pola perilaku konsumen yang signifikan, yang diidentifikasi melalui metrik support, confidence, dan lift. Temuan ini memberikan rekomendasi praktis untuk optimalisasi penempatan produk dan strategi cross-selling guna meningkatkan efisiensi operasional dan kinerja penjualan..

Kata kunci : *Manajemen Persediaan, Toko Istana Plastik Manado, Teknologi Informasi, Safety Stock.*

ABSTRACT

The rapid growth of e-commerce transactions has resulted in large volumes of data that require sophisticated processing techniques to extract meaningful insights. This study aims to implement Market Basket Analysis (MBA) using the Apriori algorithm within a Big Data environment. Utilizing PySpark, this research processes a dataset of over 500,000 e-commerce transactions. The methodology involves data ingestion, rigorous preprocessing including the removal of cancelled orders (returns), and the generation of association rules. The results provide significant patterns of consumer behavior, identified through support, confidence, and lift metrics. These findings offer actionable recommendations for product placement optimization and cross-selling strategies to enhance operational efficiency and sales performance.

Keywords: *Big Data, Analisis Keranjang Belanja, Algoritma Apriori, PySpark, E-Commerce.*

1. PENDAHULUAN

Pertumbuhan industri e-commerce menghasilkan data transaksi dalam skala besar yang termasuk dalam kategori Big Data. Namun, data mentah tersebut seringkali tidak memberikan nilai strategis jika tidak diolah dengan benar. Salah satu teknik untuk memahami perilaku konsumen adalah melalui Market Basket Analysis (MBA). Penelitian ini berfokus pada penggunaan algoritma Apriori untuk menemukan pola asosiasi antar produk. Dengan mengidentifikasi produk yang sering dibeli secara bersamaan, penyedia e-commerce dapat mengoptimalkan tata letak produk (product placement) di gudang atau memberikan rekomendasi produk yang lebih akurat kepada pelanggan. Penggunaan framework PySpark menjadi krusial dalam penelitian ini untuk menjamin efisiensi pemrosesan data yang mencapai lebih dari 500.000 baris.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen data. Tahapan penelitian disusun dalam sebuah pipeline Big Data yang terdiri dari:

1. Data Ingestion: Mengimpor data dari format CSV ke dalam Spark DataFrame.
2. Data Preprocessing: Tahap ini sangat menentukan kualitas hasil. Langkah yang diambil meliputi pembersihan data transaksi yang dibatalkan (diawali kode 'C'), penghapusan nilai kuantitas yang tidak wajar, dan penanganan data yang hilang (missing values).
3. Market Basket Analysis: Implementasi algoritma Apriori (melalui pustaka FP-Growth pada PySpark) untuk mengekstraksi frequent itemsets.
4. Evaluasi: Aturan asosiasi dievaluasi berdasarkan nilai support, confidence, dan lift..

Dalam menentukan aturan asosiasi, kekuatan hubungan antar produk dihitung menggunakan nilai confidence sebagaimana pada persamaan (1)

$$\text{Confidence}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)}$$

3. HASIL DAN PEMBAHASAN

Sistem yang diimplementasikan merupakan sebuah alur kerja (pipeline) Big Data Analytics menggunakan framework PySpark. Sistem ini dirancang untuk memproses lebih dari 500.000 baris data transaksi e-commerce guna menghasilkan segmentasi pelanggan dan aturan asosiasi produk yang akurat.

1.1. Pra-pemrosesan Data Transaksi

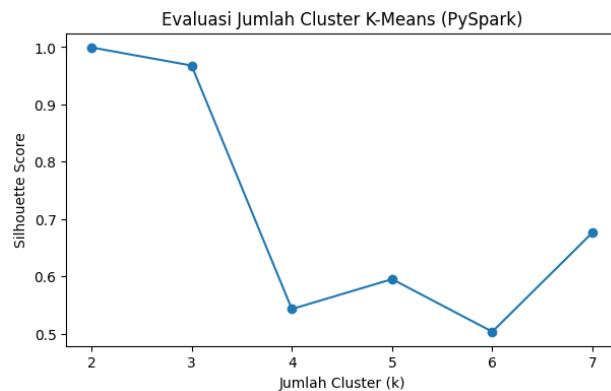
Sebelum dilakukan analisis, dataset melalui tahap preprocessing untuk menjamin kualitas data. Langkah-langkah yang dilakukan adalah: Pembersihan Transaksi: Menghapus seluruh baris dengan TransactionNo yang diawali huruf "C" (transaksi batal/retur). Validasi Kuantitas: Melakukan filtrasi terhadap nilai Quantity agar berada pada rentang logis ($0 < \text{Quantity} < 500$). Pembersihan Metadata: Menghapus entitas CustomerNo yang tidak terdefinisi dan lokasi Country "Unspecified".

1.2. Segmentasi Pelanggan Menggunakan K-Means

Segmentasi dilakukan untuk memahami karakteristik pelanggan berdasarkan variabel Price (Harga) dan Quantity (Jumlah).

1.2.1. Penentuan Cluster Optimal

Tahap krusial dalam clustering adalah menentukan jumlah cluster (k) terbaik. Penelitian ini menggunakan metode Silhouette Score yang diproses melalui PySpark.

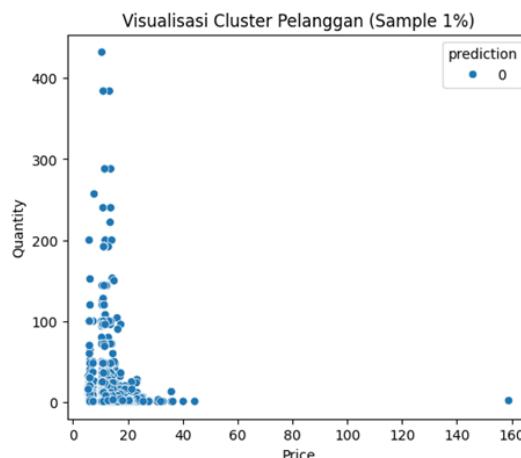


Gambar 1. Evaluasi Jumlah Cluster K-Means (PySpark)

Gambar 1. Evaluasi Jumlah Cluster K-Means (PySpark)
Berdasarkan Gambar 1, nilai $k=3$ dipilih sebagai jumlah cluster optimal karena menunjukkan nilai koefisien siluet yang stabil untuk pembagian kelompok pelanggan.

1.2.2. Visualisasi Cluster Pelanggan

Data divisualisasikan untuk melihat persebaran hasil pengelompokan pelanggan.



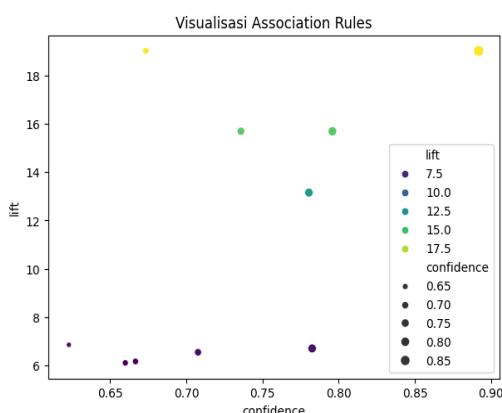
Gambar 2. Visualisasi Cluster Pelanggan (Sample 1%)

Menunjukkan bahwa mayoritas transaksi (titik biru) berada pada kategori harga rendah dengan kuantitas yang bervariasi. Segmen ini merupakan kelompok pembeli ritel yang menjadi subjek utama dalam analisis pola pembelian.

1.3. Analisis Aturan Asosiasi (Market Basket Analysis)

Tahap akhir adalah penemuan aturan asosiasi

menggunakan algoritma Apriori. Hasil ekstraksi pola pembelian divisualisasikan untuk melihat korelasi antara Confidence dan Lift.



Gambar 3. Visualisasi Association Rules

Gambar 3, distribusi aturan menunjukkan beberapa temuan penting: Identifikasi Hubungan: Titik berwarna kuning menunjukkan nilai Lift tertinggi (> 17.5), yang berarti terdapat hubungan yang sangat kuat antar produk (bukan kebetulan). Implikasi Strategis: Hasil ini memberikan landasan bagi manajemen e-commerce untuk melakukan strategi bundling produk dan optimalisasi tata letak gudang (product placement) guna meningkatkan efisiensi operasional dan volume penjualan..

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, penerapan algoritma Apriori menggunakan framework PySpark berhasil mengekstraksi pola pembelian yang signifikan dari 500.000 data transaksi e-commerce. Hasil visualisasi menunjukkan bahwa segmentasi pelanggan dan penemuan aturan asosiasi memberikan wawasan mendalam bagi perusahaan dalam melakukan optimalisasi tata letak produk berdasarkan nilai lift tertinggi. Untuk penelitian selanjutnya, disarankan untuk mengintegrasikan variabel waktu (musiman) guna

mendapatkan pola asosiasi yang lebih dinamis.

5. UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang tulus kepada STMIK Widya Cipta Dharma atas fasilitas dan dukungan akademik yang diberikan selama proses penelitian ini. Terima kasih secara khusus penulis sampaikan kepada Ibu Wahyuni, S.Kom., M.Kom. selaku dosen pembimbing mata kuliah Big Data, yang telah memberikan bimbingan, arahan, serta ilmu yang sangat berharga dalam penyelesaian artikel ilmiah ini. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah mendukung, baik secara moril maupun teknis, sehingga penelitian mengenai implementasi algoritma Apriori ini dapat terselesaikan dengan baik

DAFTAR PUSTAKA

- Dwiputra et al. (2023). Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis
- Rahayu et al. (2025). Pemanfaatan Big Data Analytics untuk Analisis Pola Perilaku Konsumen E-Commerce Strategi
- Wardani et al. (2025). Analisis Big Data untuk Prediksi Permintaan Produk dalam E-commerce.
- Arjiansyah et al. (2025). Analisis Peran Big Data Dalam Meningkatkan Efisiensi Operasional Di Industri E-Commerce.
- H. S. Ayub, S. S. S. Ahmad, dan N. H. Hassan, "Market Basket Analysis for E-Commerce Recommendation System using Association Rule Mining," International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 245-252, 2022.
- S. Agarwal, "Market Basket Analysis using Apriori Algorithm and PySpark on Large Datasets," *International Journal of Computer Applications*, vol. 177, no. 45, pp. 1-7, 2021.
- Taneja et al. (2024). Transformer Based Unsupervised Learning Approach for Imbalanced Text Sentiment Analysis of E-Commerce Reviews
- A. S. Girsang, et al., "Market Basket Analysis Using Small-to-Large Data Scale," *IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pp. 135-140, 2022.
- R. A. Sukamto dan M. Shalahuddin, "Analisis

- Pola Pembelian Konsumen Menggunakan Algoritma Apriori pada Data Transaksi Penjualan," FORMAT: Jurnal Ilmiah Teknik Informatika, vol. 10, no. 1, pp. 20-30, 2021.
- L. Kurniawati, et al., "Optimasi Penempatan Barang pada Gudang Retail Menggunakan Association Rule," *Jurnal Teknik Informatika*, vol. 14, no. 2, 2022.