

Prediksi Preferensi Tren Belanja dalam E-Commerce Menggunakan Metode K-Means Clustering dan Algoritma Apriori

¹Zia Ulhaq Azzahro, ²Eka Fauziah, ³Putri Widya Ayu Septi Wulandari, ⁴Rindi Febri Wulandari, ⁵Aidina Ristyawan

¹⁻⁵Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: ¹azzahrazia182@gmail.com, ²ekafau603@gmail.com, ³putriwidyacantikk@gmail.com,
⁴rindyfebriwulandari@gmail.com, ⁵aidinaristi@unpkediri.ac.id

Penulis Korespondens : Aidina Ristyawan

Abstrak— Pesatnya perkembangan industri e-commerce telah meningkatkan kebutuhan untuk memahami lebih dalam perilaku dan preferensi konsumen. Penelitian ini bertujuan untuk memprediksi tren belanja pengguna e-commerce dengan mengintegrasikan metode K-Means Clustering dan algoritma Apriori. K-Means digunakan untuk mengelompokkan data pengguna berdasarkan karakteristik belanja yang serupa, seperti frekuensi transaksi, jenis produk, dan waktu pembelian. Setelah pengelompokan, algoritma Apriori diterapkan untuk mengidentifikasi pola asosiasi dalam kebiasaan belanja setiap klaster yang terbentuk. Hasil penelitian menunjukkan bahwa kombinasi kedua metode ini efektif dalam mengidentifikasi segmentasi pasar yang lebih rinci dan menghasilkan aturan asosiasi yang relevan untuk mendukung pengambilan keputusan strategis, seperti rekomendasi produk dan perencanaan promosi. Dengan demikian, pendekatan ini dapat berfungsi sebagai alat prediktif untuk meningkatkan layanan personalisasi dalam e-commerce.

Kata Kunci— e-commerce, K-Means Clustering, algoritma Apriori, tren belanja

Abstract— The rapid growth of the e-commerce industry has increased the need for a deeper understanding of consumer behavior and preferences. This study aims to predict shopping trends among e-commerce users by integrating the K-Means Clustering method and the Apriori algorithm. K-Means is used to cluster user data based on similar shopping characteristics, such as transaction frequency, product types, and purchase timing. After clustering, the Apriori algorithm is applied to identify association patterns within the shopping behavior of each formed cluster. The results of the study show that the combination of these two methods is effective in identifying more detailed market segmentation and generating relevant association rules to support strategic decision-making, such as product recommendations and promotional planning. Thus, this approach can serve as a predictive tool to enhance personalized services in e-commerce.

Keywords— e-commerce, K-Means Clustering, Apriori algorithm, shopping trends

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mengubah perilaku konsumen dalam berbelanja, dengan e-commerce menjadi pilihan utama karena kemudahan akses dan variasi produk[1]. Melimpahnya data transaksi mendorong pemanfaatan teknik data mining untuk

pengambilan keputusan bisnis yang lebih tepat[2], [3]. Namun, banyak pelaku e-commerce masih kesulitan mengelola data secara optimal[4]. Algoritma K-Means terbukti efektif untuk mengelompokkan konsumen berdasarkan preferensi dan memprediksi tren produk, baik untuk sistem rekomendasi maupun pengolahan data visual[5], [6]. Selain itu, desain UI/UX yang baik juga penting dalam menarik pelanggan. Pranata et al. [7] menyatakan bahwa penggunaan aplikasi desain seperti Figma membantu UMKM menciptakan antarmuka toko online yang lebih intuitif dan mudah dikelola.

Teknik data mining seperti clustering dan association rule mining efektif dalam mengolah big data untuk memahami segmen konsumen. Apriyanto dan Sitio[1] menyebut metode K-Means membantu merancang strategi pemasaran yang personal dan efisien. Adani et al.[8] menambahkan bahwa metode ini dapat menemukan pola tersembunyi dalam data yang tidak terlihat secara manual. Apriori digunakan untuk mengidentifikasi produk yang sering dibeli bersamaan, mendukung strategi seperti bundling dan cross-selling. Selain itu, keberhasilan penerapan data mining dalam berbagai domain, termasuk bidang medis, menunjukkan bahwa pemilihan metode preprocessing dan teknik penyeimbangan data seperti SMOTE-Tomek dapat meningkatkan akurasi analisis pola dan segmentasi[9]. Menurut Styawati et al.[10], algoritme ini membantu meningkatkan efektivitas pemasaran melalui analisis pola transaksi. Integrasi antara K-Means dan Apriori juga dilakukan oleh Putri dan Purnomo[11], dengan K-Means menyaring data yang lebih relevan untuk dianalisis lebih lanjut menggunakan Apriori.

Samsudin et al.[12] menyebut K-Means efektif mengelompokkan barang agar sesuai permintaan pelanggan. Bagdja et al.[13] menunjukkan bahwa penerapannya dalam toko online meningkatkan akurasi promosi berdasarkan tren tiap klaster. Apriana dan Yuliansyah [14] menyimpulkan bahwa clustering dan Apriori mendukung keputusan bisnis berbasis data di pasar digital. Selvia dan Ramadhanu[15] membuktikan K-Means juga efektif dalam pengolahan citra untuk klasifikasi produk seperti sepatu. Berdasarkan hal tersebut, penelitian ini bertujuan mengembangkan pendekatan prediktif untuk mengidentifikasi preferensi belanja konsumen melalui kombinasi K-Means dan Apriori, guna menghasilkan segmentasi pelanggan dan pola hubungan antarproduk sebagai dasar strategi pemasaran adaptif.

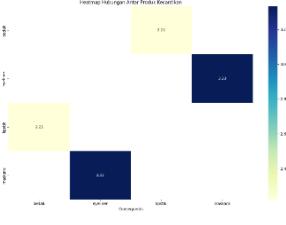
II. METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan teknik *data mining* untuk menganalisis tren preferensi belanja konsumen dalam platform *e-commerce*. Dua algoritma utama yang digunakan adalah K-Means untuk proses pengelompokan (klustering) dan Apriori untuk penemuan aturan asosiasi. Sumber data yang digunakan berasal dari **Online Shoppers Purchasing Intention Dataset** yang tersedia di **UCI Machine Learning Repository**, yang terdiri dari 12.330 entri dan 18 atribut, mencakup baik atribut numerik maupun kategoris[2]. Pada tahap pra-pemrosesan data, dilakukan penghapusan terhadap atribut-atribut yang dianggap tidak relevan, seperti *Operating System* dan *Browser*. Selain itu, atribut numerik seperti *BounceRateScore* didiskritisasi[2], dan proses normalisasi diterapkan sebelum tahap klustering[4]. Segmentasi konsumen dilakukan menggunakan algoritma K-Means, dengan jumlah klaster yang ditentukan melalui metode *Elbow*, serta menggunakan ukuran jarak Euclidean sebagai dasar pengelompokan[4]. Setelah klaster terbentuk, algoritma Apriori digunakan untuk mengekstraksi aturan asosiasi berdasarkan nilai *support* dan *confidence*, yang menghasilkan *frequent itemsets* serta pola hubungan seperti “jika X maka Y”. Apriori sendiri merupakan salah satu algoritma

paling populer dalam pencarian pola frekuensi tinggi, yaitu pola item dalam basis data yang memiliki frekuensi di atas ambang batas minimum *support*[16].

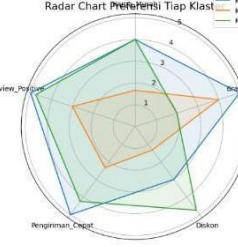
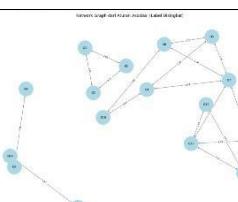
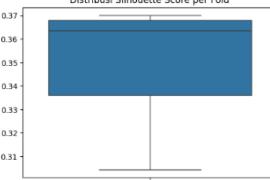
Implementasi seluruh proses dilakukan menggunakan bahasa pemrograman Python, dengan fokus evaluasi diarahkan pada interpretabilitas hasil klaster serta kekuatan aturan asosiasi yang dihasilkan. Selain itu, teknik SMOTE-Tomek juga diterapkan untuk menyeimbangkan data. SMOTE-Tomek bekerja dengan meningkatkan jumlah data pada kelas minoritas dan menghapus sampel dari kedua kelas, sehingga menghasilkan kumpulan data yang seimbang serta memiliki struktur klaster yang lebih terdefinisi dengan baik[10].

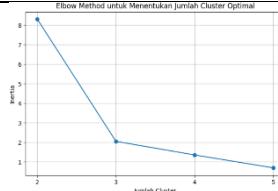
III. HASIL DAN PEMBAHASAN

No.	Nama Experimen	Tampilan	Keterangan
1.	Kombinasi K-Means dan Apriori		<p>Hasil pada tabel menunjukkan asosiasi antar pernyataan responden dalam Cluster 0 berdasarkan analisis <i>association rule mining</i>. Aturan dengan nilai <i>confidence</i> tinggi, seperti 61,7% dan 61,5%, mengindikasikan hubungan kuat antara pertanyaan antecedent dan consequent. Nilai <i>lift</i> di atas 1 menunjukkan korelasi positif, artinya kemunculan satu item meningkatkan kemungkinan item lainnya. Secara umum, responden menunjukkan pola konsumsi yang konsisten, terutama dalam preferensi terhadap produk tertentu atau produk lokal.</p>
2.	Visualisasi Heatmap untuk Menunjukkan Hubungan Antar Produk		<p>Berdasarkan heatmap yang ditampilkan, dapat disimpulkan bahwa terdapat beberapa pola hubungan yang cukup kuat antar produk kecantikan. Hubungan paling kuat terlihat antara produk <i>maskara</i> dan <i>eyeliner</i>, serta sebaliknya, dengan nilai asosiasi sebesar 3.33. Ini menunjukkan bahwa konsumen yang membeli <i>maskara</i> cenderung juga membeli <i>eyeliner</i>, dan demikian pula sebaliknya. Selain itu, ada pula hubungan antara <i>lipstik</i> dengan <i>bedak</i>, serta <i>bedak</i> dengan <i>lipstik</i>, masing-masing dengan nilai asosiasi sebesar 2.22. Pola ini menunjukkan adanya kecenderungan pembelian produk secara berpasangan yang bisa dimanfaatkan untuk strategi bundling</p>

			atau promosi silang antar produk.
--	--	--	-----------------------------------

3.	Experimen dengan K Means Clustering (Umur, Frekuensi Belanja)		<p>Visualisasi klaster berdasarkan umur dan frekuensi belanja online menunjukkan perbedaan perilaku antar kelompok usia. Klaster 0 didominasi oleh usia muda hingga paruh baya dengan frekuensi belanja sedang (2–4). Klaster 1 berisi pengguna muda dengan frekuensi rendah hingga sedang (1–3), menunjukkan aktivitas belanja online yang masih rendah. Klaster 2 mencakup rentang usia lebih luas, termasuk yang lebih tua, dan menunjukkan frekuensi belanja tinggi (4–5), menandakan bahwa pengguna lebih tua cenderung lebih aktif berbelanja online.</p>
4.	Reduksi Dimensi dengan PCA & Visualisasi Biplot	 	<p>Scree Plot menunjukkan bahwa tiga komponen utama pertama menjelaskan sekitar 80% variasi data, sehingga cukup representatif. Biplot PC1 vs PC2 menggambarkan distribusi responden serta kontribusi variabel kuesioner terhadap komponen utama. Panjang dan arah panah (loading) menunjukkan pengaruh masing-masing pertanyaan, di mana topik terkait pola makan, gaya hidup sehat, dan keberlanjutan mendominasi pembentukan PC1 dan PC2.</p>
5.	Deteksi Outlier Menggunakan PCA dan Visualisasi Boxplot	 	<p>Berdasarkan visualisasi data dengan PCA, terlihat bahwa data tersebar cukup merata pada ruang dua dimensi (PCA1 dan PCA2), yang menunjukkan bahwa reduksi dimensi berhasil menangkap pola umum dari data. Sementara itu, boxplot untuk deteksi outlier menunjukkan adanya satu titik pencilan (outlier) yang menonjol pada dimensi PCA1. Hal ini mengindikasikan bahwa sebagian besar data berada dalam rentang nilai yang wajar, namun terdapat satu responden dengan karakteristik yang sangat berbeda dari yang lain berdasarkan hasil transformasi PCA.</p>

6.	Radar Chart untuk Preferensi Tiap Klaster		<p>Radar chart menunjukkan perbedaan preferensi antar klaster dalam berbelanja online. Klaster 0 unggul di semua aspek, terutama Brand Terkenal, Pengiriman Cepat, dan Review Positif. Klaster 1 memiliki preferensi terendah, menunjukkan minimnya perhatian pada faktor eksternal. Klaster 2 menonjol pada Harga Murah dan Diskon, menandakan sensitivitas tinggi terhadap faktor ekonomis.</p>
7.	Network Graph untuk Aturan Asosiasi		<p>Network graph menunjukkan keterkaitan kuat antar pertanyaan dalam kelompok tertentu. Q1, Q2, dan Q3 membentuk jaringan erat dengan pola jawaban konsisten. Q7 berperan sebagai simpul sentral, menghubungkan beberapa pertanyaan lain (Q4, Q5, Q6, Q12, Q16), menunjukkan perannya dalam preferensi responden. Terdapat pula subgrup terpisah seperti Q8–Q9–Q14 dan Q15, yang mencerminkan pola asosiasi spesifik dan lebih terisolasi.</p>
8.	Evaluasi Konsistensi Klaster (Unsupervised K-Fold CV)		<p>Boxplot "Distribusi Silhouette Score per Fold" menunjukkan bahwa nilai silhouette score untuk proses klasterisasi berada pada kisaran antara sekitar 0.305 hingga 0.37. Nilai median berada di atas 0.36, mengindikasikan bahwa sebagian besar fold menghasilkan kualitas klasterisasi yang cukup baik dan konsisten. Secara keseluruhan, model klasterisasi yang digunakan menunjukkan tingkat kohesi dan pemisahan antar klaster yang memadai.</p>
9.	Davies-Bouldin Index untuk validasi klaster tambahan	<p>Silhouette Score: 0.5136698968812698 Davies-Bouldin Index: 0.48679434038650116</p>	<p>Hasil evaluasi klasterisasi menunjukkan nilai Silhouette Score sebesar 0.514 dan Davies-Bouldin Index sebesar 0.487. Nilai Silhouette Score yang mendekati 1 menandakan bahwa klaster yang terbentuk memiliki jarak yang cukup baik antar klaster dan kepadatan dalam klaster yang memadai. Sementara itu, Davies-Bouldin Index yang lebih rendah dari 1 menunjukkan bahwa pemisahan antar klaster cukup optimal.</p>

10	Evaluasi K-Means Menggunakan Inertia (Elbow Method)	 <p>Elbow Method untuk Menentukan Jumlah Cluster Optimal</p> <p>The graph shows Inertia (y-axis, 0 to 8) versus Jumlah Cluster (x-axis, 2 to 5). The curve starts at (2, 8) and decreases sharply to (3, 2), then levels off with a slight decrease to (4, 1.5) and (5, 1). A vertical line is drawn at x=3, indicating the optimal number of clusters.</p> <table border="1"> <thead> <tr> <th>Jumlah Cluster</th> <th>Inertia</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>8</td> </tr> <tr> <td>3</td> <td>2</td> </tr> <tr> <td>4</td> <td>1.5</td> </tr> <tr> <td>5</td> <td>1</td> </tr> </tbody> </table>	Jumlah Cluster	Inertia	2	8	3	2	4	1.5	5	1	Grafik Elbow Method tersebut menunjukkan bahwa penurunan nilai <i>inertia</i> paling signifikan terjadi antara jumlah klaster 2 ke 3. Setelah titik tersebut, penurunan <i>inertia</i> menjadi lebih landai, yang menandakan bahwa menambah jumlah klaster setelah 3 tidak memberikan peningkatan signifikan terhadap kualitas pemisahan data. Maka, jumlah klaster yang optimal dalam analisis ini adalah 3 klaster .
Jumlah Cluster	Inertia												
2	8												
3	2												
4	1.5												
5	1												

IV. PENUTUP

4.1 Kesimpulan

Berdasarkan analisis data transaksi di platform e-commerce dengan menggunakan algoritma K-Means dan Apriori, dapat disimpulkan bahwa kedua metode ini sangat membantu dalam mengelompokkan konsumen serta menemukan pola produk yang sering dibeli bersamaan. K-Means memudahkan kita untuk membagi pelanggan ke dalam kelompok berdasarkan kebiasaan belanja mereka, sementara Apriori berguna untuk mengetahui produk-produk apa saja yang sering dibeli secara bersamaan dan bisa dijadikan rekomendasi. Dengan menggabungkan kedua algoritma ini, pelaku bisnis bisa menyusun strategi pemasaran yang lebih tepat sasaran dan berpeluang besar meningkatkan penjualan.

4.2 Saran

1. Penggunaan algoritma K-Means dan Apriori dapat dikembangkan lebih lanjut dengan menggabungkan teknik preprocessing yang lebih mendalam agar segmentasi dan aturan asosiasi menjadi lebih akurat.
2. Disarankan bagi pelaku e-commerce untuk memanfaatkan hasil segmentasi konsumen dalam merancang strategi pemasaran yang lebih personal dan tepat sasaran, seperti promosi bundling produk berdasarkan pola asosiasi yang ditemukan.
3. Pengembangan antarmuka pengguna (UI/UX) yang responsif dan mudah digunakan sangat penting untuk meningkatkan keterlibatan konsumen dalam platform e-commerce.
4. Penelitian selanjutnya sebaiknya menggunakan dataset dengan fitur yang lebih lengkap dan beragam agar hasil analisis menjadi lebih komprehensif.
5. Evaluasi berkala terhadap model segmentasi dan aturan asosiasi perlu dilakukan agar tetap relevan dengan perubahan tren dan perilaku konsumen.

DAFTAR PUSTAKA

- [1] B. Apriyanto and S. L. M. Sitio, “Penerapan K-Means dalam Menganalisis Pola Pembelian Pelanggan Pada Data Transaksi E-Commerce,” *bit-Tech*, vol. 7, no. 3, pp. 790–797, Apr. 2025, doi: [10.32877/bt.v7i3.2195](https://doi.org/10.32877/bt.v7i3.2195).
- [2] F. Maulana Alja, E. Dianiati, and A. Ristyawan, “PERANCANGAN UI/UX E-COMMERCE MENGGUNAKAN METODE USER CENTERED DESIGN (UCD),” 2024.
- [3] O. D. Kurnia *et al.*, “Analisis Perbandingan Algoritma Naïve Bayes dengan K-Nearest

Neighbor (KNN) Pada Dataset Mobile Price Classification Penulis Korespondensi: Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 1174,” Online, 2024.

- [4] A. S. Aji and F. Sulianta, “Mining Data Perilaku Belanja Online Konsumen Menggunakan Algoritma Apriori untuk Menghasilkan Aturan Asosiasi Sebagai Dasar Menentukan Aksi Bisnis pada Web E-Commerce.”
- [5] F. Naufal, Y. Herry Chrisnanto, and A. Kania Ningsih, “Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering,” Cimahi, Jul. 2022. [Online]. Available: <https://e-journal.unper.ac.id/index.php/informatics>
- [6] D. Selvia and A. Ramadhanu, “IMPLEMENTASI METODE K-MEANS CLUSTERING DENGAN TEKNIK PENGOLAHAN CITRA UNTUK MENGIDENTIFIKASI JENIS SEPATU,” *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 1, pp. 361–366, 2025.
- [7] D. Adistyan Pranata, R. Indriati, and A. Nugroho, “Desain UI/UX E-Commerce Menggunakan Aplikasi Figma,” Online, 2024.
- [8] N. F. Adani, A. F. Boy, S. Kom, M. Kom, and R. Syahputra, “Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan STMIK Triguna Dharma Program Studi Sistem Informasi, STMIK Triguna Dharma Program Studi Sistem Informasi, STMIK Triguna Dharma,” *Jurnal CyberTech*, vol. x. No.x, [Online]. Available: <https://ojs.trigunadharma.ac.id>
- [9] A. Ristyawan, A. Nugroho, and T. K. Amarya, “Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke,” vol. 12, no. 1, pp. 29–44, 2025.
- [10] A. N. dan K. N. A. S. Styawati, “Analisis pola transaksi pelanggan menggunakan algoritme Apriori,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 619–626, Sep. 2021.
- [11] V. E. Putri and H. D. Purnomo, “INTEGRASI ALGORITMA APRIORI DAN K-MEANS DALAM ANALISIS POLA PEMBELIAN UNTUK MENINGKATKAN STRATEGI PEMASARAN,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 409–423, Jan. 2025, doi: [10.29100/jipi.v10i1.5768](https://doi.org/10.29100/jipi.v10i1.5768).
- [12] Samsudin and U. hayati, “OPTIMALISASI STOK BARANG MELALUI ALGORITMA K-MEANS CLUSTERING ANALISIS UNTUK MANAJEMEN PERSEDIAAN DALAM KONTEKS BISNIS MODERN,” 2024.
- [13] M. Nara Bagdja, R. Taufiq Subagio, and V. Dwi Kartika, “PENERAPAN METODE K-MEANS UNTUK MENGKLASIFIKASI MINAT KONSUMEN TERHADAP PRODUK TOKO ONLINE,” 2024.
- [14] D. Apriana and C. Yuliansyah, “Mengoptimalkan Penjualan Online Melalui Teknik Data Mining (Studi Kasus E-Commerce),” *AL-MIKRAJ Jurnal Studi Islam dan Humaniora (E-ISSN 2745-4584)*, vol. 4, no. 02, pp. 514–527, Feb. 2024, doi: [10.37680/almikraj.v4i02.4774](https://doi.org/10.37680/almikraj.v4i02.4774).
- [15] S. P. Raudini¹, L. Purwopeni², D. Brilliant³, F. Azzahra⁴, N. Syahputri⁵, and S. Informasi, “PENERAPAN DATA MINING TERHADAP PENJUALAN KACAMATA DI OPTIK SURYA MEDAN MENGGUNAKAN ALGORITMA APRIORI,” *VARIABLE RESEARCH JOURNAL*, vol. 01, 2024.
- [16] F. A. Sianturi, “PENERAPAN ALGORITMA APRIORI UNTUK PENENTUAN TINGKAT PESANAN,” *Jurnal Mantik Penusa*, vol. 2, no. 1, pp. 50–57, 2018, [Online]. Available: <http://bowmasbow.blogspot.com/20>