Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume. 6 Nomor. 2 Juli 2026
E-ISSN :2827-7953; P-ISSN : 2827-8135, Hal. 564-579
DOI: https://doi.org/10.55606/juisik.v6i2.2532
Tersedia: https://journal.sinov.id/index.php/juisik

Analisis Pengguna *E-Commerce* Menggunakan Algoritma *K-Means* dan *Hierarchical Clustering*

Atiqah Noor Zhaafirah[1*], Adhisa Nanda Kurnia[2], Pungkas Subarkah[3], Agus Pramono[4]
[1-4]Informatika, Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia
Email: viraira99@gmail.com[1], adhisanandak@gmail.com[2], subarkah@amikompurwokerto.ac.id[3], agus@amikompurwokerto.ac.id[4]
**Penulis Korespondensi: viraira99@gmail.com*

Abstract. *This study aims to analyze E-commerce user behavior through customer segmentation using K-Means and Hierarchical Clustering algorithms. The research applies a quantitative Data Mining approach using the E-commerce Customer Behavior Dataset obtained from Kaggle, consisting of 5,000 customer transaction records. The analysis process was conducted using Orange Data Mining 3.40.0 through several stages, including data collection, attribute selection, preprocessing, Clustering, evaluation, and interpretation of results. The attributes used include customer demographics, transaction characteristics, product information, session activities, and customer ratings. The Clustering process generated three user groups using both algorithms. Results show that K-Means produced clearer cluster separation and easier interpretation through scatter plot visualization and statistical evaluation. Cluster distribution analysis indicates that most users belong to groups with relatively similar transaction characteristics, while smaller clusters represent users with significantly higher transaction values. Meanwhile, Hierarchical Clustering provides additional insight through dendrogram structures that illustrate hierarchical relationships among customer groups. Overall, both algorithms successfully identified user segments based on transaction behavior, although K-Means produced more interpretable Clustering results for E-commerce customer segmentation.*

Keywords*: Customer Segmentation; Data Mining; E-Commerce; Hierarchical Clustering; K-Means Clustering.*

Abstrak. Penelitian ini bertujuan untuk menganalisis perilaku pengguna *E-commerce* melalui segmentasi pelanggan menggunakan algoritma *K-Means* dan *Hierarchical Clustering*. Penelitian ini menerapkan pendekatan kuantitatif berbasis *Data Mining* dengan menggunakan *E-commerce* Customer Behavior Dataset yang diperoleh dari Kaggle, yang terdiri dari 5.000 data transaksi pelanggan. Proses analisis dilakukan menggunakan Orange *Data Mining* 3.40.0 melalui beberapa tahapan, yaitu pengumpulan data, pemilihan atribut, pra-pemrosesan data, *Clustering*, evaluasi, dan interpretasi hasil. Atribut yang digunakan meliputi demografi pelanggan, karakteristik transaksi, informasi produk, aktivitas sesi, dan penilaian pelanggan. Proses *Clustering* menghasilkan tiga kelompok pengguna menggunakan kedua algoritma tersebut. Hasil penelitian menunjukkan bahwa algoritma *K-Means* menghasilkan pemisahan cluster yang lebih jelas dan lebih mudah diinterpretasikan melalui visualisasi scatter plot serta evaluasi statistik. Analisis distribusi cluster menunjukkan bahwa sebagian besar pengguna berada dalam kelompok dengan karakteristik transaksi yang relatif serupa, sementara cluster yang lebih kecil merepresentasikan pengguna dengan nilai transaksi yang jauh lebih tinggi. Sementara itu, *Hierarchical Clustering* memberikan wawasan tambahan melalui struktur dendrogram yang menggambarkan hubungan hierarkis antar kelompok pelanggan. Secara keseluruhan, kedua algoritma berhasil mengidentifikasi segmen pengguna berdasarkan perilaku transaksi, meskipun *K-Means* menghasilkan hasil *Clustering* yang lebih mudah diinterpretasikan untuk segmentasi pelanggan *e-commerce*.

Kata Kunci: *Data Mining*; *E-Commerce*; *Hierarchical Clustering*; *K-Means Clustering*; Segmentasi Pelanggan.

1. LATAR BELAKANG

Persaingan bisnis untuk menarik, mempertahankan, dan memahami pelanggan semakin ketat di era digital yang ditandai dengan pesatnya perkembangan industri *e-commerce*(Sypriansyah & Hasibuan, 2025). Perdagangan *E-commerce* mudah dan praktis karena dapat dilakukan hanya dengan menggunakan perangkat elektronik seperti komputer, laptop, atau smartphone yang terhubung ke internet sebagai media perantara (Ikhwan Syarif et al., 2023).Dengan pertumbuhan *E-commerce* yang semakin pesat, platform ini telah menjadi

Naskah Masuk: 15 April 2026; Revisi:28 Mei 2026; Diterima: 02 Juni 2026; Terbit: 13 Juli 2026

salah satu pilihan utama bagi orang-orang untuk melakukan pembelian online (Kartika & Romansyah, 2024). Oleh karena itu, bisnis harus mampu menggunakan data yang mereka miliki untuk melihat perilaku pelanggan dan pola belanja mereka sehingga mereka dapat memberikan layanan yang lebih baik dan menerapkan strategi bisnis yang lebih efisien (Sudirjo et al., 2018). Kehadiran *E-commerce* juga menjadikan proses jual beli lebih mudah, cepat, dan efisien sehingga banyak digunakan oleh masyarakat dalam memenuhi kebutuhan sehari-hari (Sari, 2021). Dalam penggunaan *e-commerce*, setiap pengguna memiliki perilaku yang berbeda-beda. Ada pengguna yang aktif melakukan pembelian, ada yang hanya sesekali bertransaksi, bahkan ada pula pengguna yang hanya melihat produk tanpa melakukan pembelian (Mauludin et al., 2022). Perbedaan perilaku tersebut menunjukkan bahwa setiap pengguna memiliki karakteristik dan pola belanja yang unik (Oktaviani et al., 2025). Oleh karena itu, diperlukan analisis lebih lanjut untuk memahami pola perilaku pengguna tersebut.

Meningkatnya jumlah pengguna *E-commerce* menyebabkan volume data yang dihasilkan juga semakin besar (Br jabat, Tarigan, 2024). Data tersebut meliputi riwayat transaksi, frekuensi pembelian, jenis produk yang diminati, hingga aktivitas pengguna dalam aplikasi. Besarnya volume dan kompleksitas data membuat proses analisis menjadi sulit apabila dilakukan secara manual (*E-commerce* et al., 2025). Selain membutuhkan waktu yang lama, analisis manual juga berpotensi menghasilkan informasi yang kurang akurat. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data dalam jumlah besar secara otomatis dan efisien. Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah *Data Mining*. *Data Mining* merupakan teknik yang digunakan untuk mengekstraksi informasi atau pola tersembunyi dari sekumpulan data dalam jumlah besar (*K-Means* et al., 2023).

Salah satu metode dalam *Data Mining* yang sering digunakan adalah *Clustering*, yaitu teknik untuk mengelompokkan data berdasarkan kemiripan karakteristik tertentu (A-priori, n.d.). Dalam konteks *e-commerce*, *Clustering* dapat digunakan untuk mengelompokkan pengguna berdasarkan perilaku dan pola belanja mereka sehingga dapat membantu dalam memahami segmentasi pengguna (Fahrillah Fahrillah & Zaehol Fatah, 2025). Dalam penelitian ini, kedua metode mengelompokkan yang digunakan adalah *Hierarchical Clustering* dan *K-Means*. *K-Means* adalah metode mengelompokkan yang sederhana dan memiliki kecepatan yang baik dalam mengelompokkan data berdasarkan kedekatan antar data (Zuhal, 2022). Sementara itu, *Hierarchical Clustering* mengelompokkan memiliki keunggulan dalam membuat struktur *Hierarchical Clustering*, yang membuat hubungan antar kelompok data lebih jelas dalam dendrogram (Kumar et al., 2025). Penelitian sebelumnya juga menunjukkan

bahwa kombinasi *K-Means* dan *Hierarchical Clustering* dapat menghasilkan segmentasi pelanggan yang lebih optimal sesuai karakteristik data transaksi *E-commerce* (Yosua et al., 2025). Oleh karena itu, kedua metode tersebut dipilih untuk membandingkan hasil pengelompokan pengguna *E-commerce* sehingga diperoleh cluster yang lebih sesuai dan mudah dianalisis (Rahayu et al., 2024).

Berdasarkan kesamaan perilaku transaksi dan pola pembelian, penelitian ini mampu mengidentifikasi beberapa cluster pengguna berdasarkan penerapan metode *K-Means* dan *Hierarchical Clustering* terhadap dataset pengguna *E-commerce* yang diperoleh melalui Kaggle. Proses *Clustering* memungkinkan data pelanggan dikelompokkan sesuai dengan fitur transaksi yang sebanding, dan menghasilkan segmentasi yang lebih terstruktur (Dan et al., 2025). Perusahaan dapat memperoleh pemahaman yang lebih baik tentang perilaku pelanggan, seperti frekuensi pembelian, nilai transaksi, dan kecenderungan terhadap produk tertentu, melalui segmen ini (Digital et al., 2026). Selain itu, penelitian sebelumnya menunjukkan bahwa temuan yang berkaitan dengan pengelompokan pelanggan dapat digunakan untuk membantu dalam pembuatan strategi pemasaran yang lebih tepat sasaran, meningkatkan loyalitas pelanggan, dan membantu bisnis meningkatkan kualitas layanan yang diberikan oleh masing-masing kelompok (Sakinah & Awaliyah, 2025). Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengelompokkan pengguna *E-commerce* ke dalam beberapa cluster berdasarkan pola perilaku mereka serta mengetahui karakteristik dari masing-masing kelompok pengguna. Hasil penelitian diharapkan dapat membantu perusahaan *E-commerce* dalam memahami konsumennya sehingga dapat digunakan sebagai dasar dalam menentukan strategi pemasaran yang lebih efektif dan tepat sasaran.

2. KAJIAN TEORITIS

Bagian ini menguraikan teori-teori yang mendasari penelitian mengenai analisis perilaku pengguna *E-commerce* menggunakan teknik *Data Mining*, khususnya metode *Clustering* dengan algoritma *K-Means* dan *Hierarchical Clustering*. Selain itu, bagian ini juga memaparkan beberapa penelitian terdahulu yang relevan sebagai landasan dalam pelaksanaan penelitian.

E-commerce

E-commerce merupakan kegiatan jual beli barang atau jasa yang dilakukan melalui media elektronik dan jaringan internet (Nursania Dasopang, 2023). Perkembangan *E-commerce* telah mengubah pola transaksi masyarakat menjadi lebih mudah, cepat, dan efisien. Setiap aktivitas pengguna dalam platform *E-commerce* menghasilkan data yang dapat

dimanfaatkan untuk memahami perilaku pelanggan, seperti riwayat transaksi, frekuensi pembelian, kategori produk yang diminati, serta aktivitas pengguna selama mengakses platform (Nursania Dasopang, 2023). Informasi tersebut dapat digunakan sebagai dasar dalam pengambilan keputusan bisnis dan penyusunan strategi pemasaran yang lebih tepat sasaran.

Perilaku Pengguna *E-commerce*

Perilaku pengguna *E-commerce* merupakan pola tindakan dan aktivitas pelanggan ketika berinteraksi dengan platform belanja online (Sutisna & Sutrisna, 2023). Perilaku tersebut dapat dilihat dari berbagai aspek, seperti frekuensi pembelian, nilai transaksi, durasi penggunaan aplikasi, jumlah halaman yang dikunjungi, serta tingkat kepuasan pelanggan terhadap layanan yang diberikan. Analisis perilaku pengguna sangat penting karena dapat membantu perusahaan memahami kebutuhan pelanggan dan mengidentifikasi kelompok pengguna dengan karakteristik yang berbeda (Apriana & Yuliansyah, 2024).

Data Mining

Data Mining adalah proses menemukan pola, hubungan, atau informasi yang tersembunyi dari kumpulan data dalam jumlah besar menggunakan teknik statistik, machine learning, dan analisis basis data (Pujiono et al., 2024). *Data Mining* digunakan untuk mengubah data mentah menjadi informasi yang bermanfaat dalam mendukung pengambilan keputusan. Dalam bidang *e-commerce*, *Data Mining* banyak dimanfaatkan untuk analisis pelanggan, rekomendasi produk, prediksi perilaku konsumen, dan segmentasi pelanggan (Nanda et al., 2025).

Clustering

Clustering merupakan salah satu teknik dalam *Data Mining* yang bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik yang dimiliki (Adji et al., 2025). Data yang memiliki karakteristik serupa akan ditempatkan dalam kelompok yang sama (cluster), sedangkan data yang memiliki karakteristik berbeda akan ditempatkan pada cluster yang berbeda. Metode *Clustering* termasuk ke dalam teknik unsupervised learning karena tidak memerlukan data berlabel dalam proses pembelajarannya (Aldi et al., 2025). Pada penelitian ini, *Clustering* digunakan untuk mengelompokkan pengguna *E-commerce* berdasarkan karakteristik transaksi dan aktivitas pelanggan.

Algoritma *K-Means*

K-Means merupakan salah satu algoritma *Clustering* yang paling banyak digunakan karena memiliki proses yang sederhana dan efisien. Algoritma ini bekerja dengan membagi data ke dalam sejumlah cluster yang telah ditentukan sebelumnya berdasarkan kedekatan data terhadap titik pusat cluster (*centroid*). Proses *K-Means* dilakukan secara iteratif hingga posisi

centroid stabil dan tidak mengalami perubahan yang signifikan. Keunggulan *K-Means* terletak pada kemampuannya dalam mengelompokkan data berukuran besar dengan waktu komputasi yang relatif cepat serta menghasilkan cluster yang mudah diinterpretasikan (Anggraini & Hasibuan, 2025).

Hierarchical Clustering

Hierarchical Clustering adalah metode *Clustering* yang membentuk struktur pengelompokan secara bertingkat berdasarkan tingkat kemiripan antar data (Fahmi et al., 2026). Hasil pengelompokan divisualisasikan dalam bentuk dendrogram yang menunjukkan hubungan hierarkis antar objek. Metode ini tidak memerlukan penentuan jumlah cluster pada tahap awal dan dapat memberikan informasi mengenai kedekatan antar kelompok data (Suharti et al., 2022). Oleh karena itu, *Hierarchical Clustering* sering digunakan untuk memperoleh gambaran struktur hubungan antar data yang lebih rinci.

Orange *Data Mining*

Orange *Data Mining* merupakan perangkat lunak analisis data berbasis visual yang menyediakan berbagai fitur untuk pengolahan data, visualisasi, machine learning, dan *Data Mining*. Orange memungkinkan pengguna membangun alur kerja analisis melalui widget yang saling terhubung sehingga mempermudah proses eksplorasi data, preprocessing, *Clustering*, evaluasi, dan interpretasi hasil (Fathiarahma et al., 2021). Dalam penelitian ini, Orange digunakan untuk menjalankan seluruh proses analisis mulai dari pengolahan dataset hingga evaluasi hasil *Clustering*.

Penelitian Terdahulu

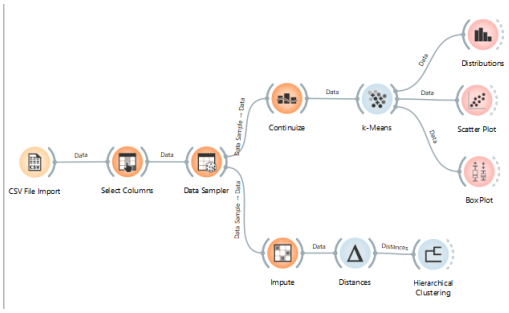
Beberapa penelitian sebelumnya menunjukkan bahwa metode *Clustering* dapat digunakan secara efektif untuk melakukan segmentasi pelanggan berdasarkan perilaku transaksi (Safitri et al., 2025). Penelitian yang dilakukan oleh beberapa peneliti dalam bidang *E-commerce* menunjukkan bahwa algoritma *K-Means* mampu menghasilkan kelompok pelanggan yang jelas berdasarkan karakteristik transaksi dan nilai pembelian (Chandra & Darmayanti, 2019). Sementara itu, penelitian lain menunjukkan bahwa *Hierarchical Clustering* mampu memberikan informasi tambahan mengenai hubungan antar kelompok pelanggan melalui struktur dendrogram yang dihasilkan (Ramiro Firjatullah et al., 2025).

Penelitian mengenai segmentasi pelanggan menggunakan kombinasi *K-Means* dan *Hierarchical Clustering* juga telah dilakukan pada berbagai dataset transaksi. Hasil penelitian tersebut menunjukkan bahwa kedua algoritma mampu mengidentifikasi pola perilaku pelanggan dengan baik, meskipun *K-Means* cenderung menghasilkan cluster yang lebih mudah diinterpretasikan, sedangkan *Hierarchical Clustering* memberikan gambaran hubungan antar

kelompok secara lebih rinci. Berdasarkan hasil penelitian terdahulu tersebut, penggunaan kedua algoritma dalam penelitian ini diharapkan dapat memberikan hasil segmentasi pengguna *E-commerce* yang lebih komprehensif dan mendukung pengambilan keputusan bisnis yang lebih efektif (Nuryansyah & Rasyid, 2026).

3. METODE PENELITIAN

Penelitian ini menerapkan metode kuantitatif dengan pendekatan *Data Mining* untuk menganalisis perilaku pengguna *E-commerce* berdasarkan data transaksi dan aktivitas pelanggan. Pendekatan kuantitatif digunakan karena penelitian berfokus pada pengolahan data numerik yang dapat diukur dan dianalisis secara statistik untuk menemukan pola perilaku pengguna. Seluruh proses eksperimen dilakukan menggunakan perangkat lunak Orange versi 3.40.0, yang menyediakan antarmuka visual berbasis alur kerja (workflow) untuk membangun pipa analisis data yang modular (Kaylista et al., 2023). Alur penelitian terdiri atas beberapa tahapan, yaitu pengumpulan dataset, eksplorasi dan seleksi data, preprocessing data, proses *Clustering* menggunakan algoritma *K-Means* dan *Hierarchical Clustering*, evaluasi hasil *Clustering*, serta analisis dan interpretasi hasil. Setiap tahapan saling berkaitan untuk menghasilkan pengelompokan pengguna yang optimal serta memungkinkan perbandingan kinerja kedua algoritma dilakukan secara konsisten. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian menggunakan Orange *Data Mining.*

Pengumpulan Data

Tahap pertama adalah pengumpulan dataset. Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan topik perilaku pelanggan *E-commerce* (*E-commerce* Customer Behavior Dataset). Dataset ini dipilih karena memuat data transaksi dan aktivitas pelanggan yang relevan untuk analisis perilaku pengguna, seperti usia pelanggan, kategori produk, harga produk, jumlah pembelian, total transaksi, metode pembayaran, durasi sesi, jumlah halaman yang dilihat, status pelanggan tetap, waktu pengiriman, dan rating pelanggan. Dataset diunduh dalam format CSV dan selanjutnya digunakan sebagai sumber data utama

dalam proses analisis serta pengelompokkan pengguna menggunakan algoritma *K-Means* dan *Hierarchical Clustering*.

Eksplorasi dan seleksi data

Tahap kedua adalah eksplorasi dan seleksi data awal. Pada tahap ini, widget Tabel Data digunakan untuk memeriksa dataset yang telah diimpor ke Orange *Data Mining* untuk mengetahui struktur dan jenis atribut data. Ini juga dilakukan untuk memastikan bahwa proses impor data tidak mengandung kesalahan. Hasil pemeriksaan menunjukkan bahwa dataset terdiri dari 5.000 data transaksi pelanggan yang diidentifikasi melalui berbagai fitur yang menunjukkan fitur dan aktivitas pengguna *e-commerce*.

Selanjutnya dilakukan seleksi atribut menggunakan widget Select Columns untuk menentukan variabel yang akan digunakan dalam proses *Clustering*. Atribut yang dipilih sebagai fitur meliputi Age, Product_Category, Unit_Price, Quantity, Discount_Amount, Total_Amount, Payment_Method, Device_Type, Session_Duration_Minutes, Pages_Viewed, Is_Returning_Customer, Delivery_Time_Days, dan Customer_Rating. Sementara itu, atribut Order_ID, Customer_ID, Date, City, dan Gender ditempatkan sebagai metadata karena tidak digunakan secara langsung dalam proses pengelompokan. Tahap ini bertujuan untuk memastikan bahwa hanya atribut yang relevan yang digunakan sehingga hasil *Clustering* dapat menggambarkan perilaku pengguna *E-commerce* secara lebih akurat.

Preprocessing Data

Tahap ketiga adalah preprocessing data yang bertujuan untuk mempersiapkan dataset sebelum dilakukan proses *Clustering*. Pada tahap ini digunakan widget Data Sampler untuk mengambil sampel data yang akan digunakan dalam analisis sehingga proses komputasi dapat dilakukan dengan lebih efisien. Selanjutnya, data diproses menggunakan widget Continuize untuk mengubah atribut kategorikal menjadi bentuk numerik agar dapat diolah algoritma *K-Means*. Untuk proses *Hierarchical Clustering*, dilakukan penanganan data menggunakan widget Impute guna memastikan tidak terdapat nilai yang hilang pada dataset. Hasil preprocessing kemudian digunakan sebagai masukan pada tahap *Clustering* untuk memperoleh kelompok pengguna berdasarkan karakteristik dan perilaku transaksi yang serupa.

Proses *Clustering*

Tahap keempat adalah proses *Clustering* menggunakan algoritma *K-Means* dan *Hierarchical Clustering*. Algoritma *K-Means* digunakan untuk mengelompokkan data berdasarkan kedekatan terhadap pusat cluster (*centroid*) (Widya et al., n.d.), sedangkan *Hierarchical Clustering* digunakan untuk membentuk struktur pengelompokan secara bertingkat berdasarkan tingkat kemiripan antar data (Alamtaha et al., 2023). Penggunaan kedua

algoritma ini bertujuan untuk membandingkan hasil pengelompokan pengguna *E-commerce* sehingga dapat diperoleh gambaran karakteristik setiap kelompok pelanggan (*Clustering* & Belajar, 2022).

Evaluasi Hasil *Clustering*

Tahap kelima adalah evaluasi yang dilakukan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh algoritma *K-Means* dan *Hierarchical Clustering*. Proses evaluasi dilakukan melalui analisis visual menggunakan widget Scatter Plot, Box Plot, dan Distributions pada Orange *Data Mining*. Evaluasi ini bertujuan untuk mengamati pola persebaran data, tingkat pemisahan antar cluster, serta karakteristik setiap cluster yang terbentuk sehingga dapat diketahui efektivitas proses pengelompokan yang dilakukan.

Analisis dan Interpretasi Hasil

Tahap keenam adalah analisis dan interpretasi hasil dilakukan untuk mengidentifikasi karakteristik dari setiap cluster yang terbentuk berdasarkan atribut yang digunakan dalam proses *Clustering*. Analisis dilakukan dengan membandingkan nilai dan pola pada masing-masing cluster sehingga dapat diketahui perbedaan karakteristik antar kelompok pengguna. Hasil interpretasi digunakan untuk memahami perilaku pelanggan *e-commerce*, seperti pola pembelian, tingkat aktivitas pengguna, dan kecenderungan transaksi yang dapat dimiliki setiap segmen pelanggan. Selain itu, hasil *Clustering* juga dapat digunakan untuk mendukung penyusunan strategi pemasaran yang lebih tepat sasaran berdasarkan karakteristik masing-masing kelompok pelanggan. Pada penelitian ini, dilakukan pula perbandingan hasil pengelompokan yang diperoleh dari algoritma *K-Means* dan *Hierarchical Clustering* untuk menentukan metode yang menghasilkan cluster yang lebih representatif terhadap karakteristik data.

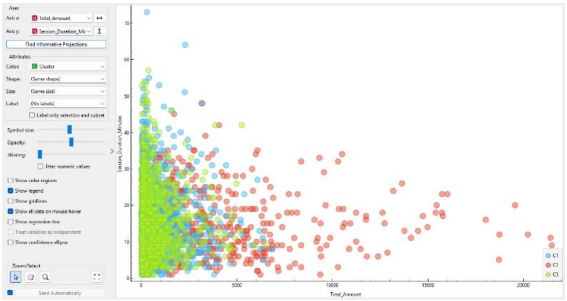
4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengelompokan pengguna *E-commerce* menggunakan algoritma *K-Means* dan *Hierarchical Clustering* beserta interpretasi dari setiap cluster yang terbentuk. Proses *Clustering* dilakukan terhadap data pelanggan yang telah melalui tahap preprocessing, meliputi seleksi atribut, transformasi data, dan penanganan nilai yang hilang. Atribut yang digunakan dalam proses pengelompokan mencakup karakteristik pelanggan dan aktivitas transaksi, seperti usia, kategori produk, harga produk, jumlah pembelian, total transaksi, durasi sesi, jumlah halaman yang dilihat, status pelanggan tetap, waktu pengiriman, dan rating pelanggan. Hasil *Clustering* kemudian dievaluasi melalui visualisasi menggunakan Scatter Plot, Distributions, dan Box Plot untuk mengamati pola persebaran data, distribusi

anggota cluster, serta karakteristik masing-masing cluster. Selanjutnya, dilakukan analisis dan perbandingan hasil pengelompokan yang dihasilkan oleh algoritma *K-Means* dan *Hierarchical Clustering* untuk memperoleh pemahaman yang lebih mendalam mengenai karakteristik pengguna *E-commerce* berdasarkan tingkat kemiripan atribut yang dimiliki.

Hasil *Clustering* Menggunakan *K-Means*

Proses pengelompokan data pengguna *E-commerce* dilakukan menggunakan algoritma *K-Means* dengan jumlah cluster sebanyak tiga kelompok (k = 3). Hasil pengelompokan divisualisasikan menggunakan Scatter Plot dengan atribut Total_Amount sebagai sumbu X dan Session_Duration_Minutes sebagai sumbu Y. Visualisasi hasil *Clustering* ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi hasil *Clustering* menggunakan algoritma *K-Means.*

Berdasarkan hasil visualisasi pada Gambar 2, data pengguna berhasil dikelompokkan menjadi tiga cluster, yaitu C1, C2, dan C3 menggunakan algoritma *K-Means*. Persebaran data menunjukkan bahwa sebagian besar pengguna berada pada rentang nilai transaksi yang rendah hingga sedang, sedangkan sejumlah pengguna dengan nilai transaksi yang lebih tinggi membentuk kelompok yang berbeda. Selain itu, variasi durasi sesi pengguna menunjukkan adanya perbedaan tingkat aktivitas dalam menggunakan platform *e-commerce*. Hasil ini mengindikasikan bahwa proses *Clustering* mampu mengidentifikasi kelompok pengguna berdasarkan pola transaksi dan aktivitas yang dimiliki.

Tabel 1. Distribusi anggota cluster.

Cluster	Jumlah Anggota	Persentase
C1	2.100	51,2%
C2	500	12,2%
C3	1.500	36,6%
Total	4.100	100%

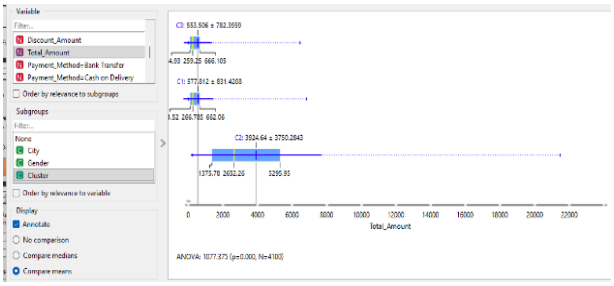
Tabel 1 menunjukkan distribusi jumlah anggota pada setiap cluster hasil pengelompokan menggunakan algoritma *K-Means*. Cluster 1 (C1) memiliki jumlah anggota terbanyak dengan lebih dari setengah total data, sedangkan Cluster 2 (C2) memiliki jumlah

anggota paling sedikit. Sementara itu, Cluster 3 (C3) menempati urutan kedua dalam jumlah anggota. Distribusi ini menunjukkan bahwa sebagian besar pengguna memiliki karakteristik yang cenderung serupa dengan anggota pada Cluster 1, sedangkan Cluster 2 merepresentasikan kelompok pengguna yang lebih spesifik dengan jumlah anggota yang relatif.

Hasil dari distribusi tersebut sejalan dengan visualisasi Scatter Plot dan evaluasi Box Plot yang menunjukkan adanya perbedaan karakteristik antar cluster. Meskipun jumlah anggota Cluster 2 relatif lebih sedikit, cluster ini memiliki nilai transaksi yang lebih tinggi dibandingkan cluster lainnya. Sebaliknya, Cluster 1 dan Cluster 3 mencakup sebagian besar pengguna dengan karakteristik transaksi yang lebih umum. Dengan demikian, hasil *Clustering* tidak hanya mampu memisahkan pengguna berdasarkan kemiripan karakteristik, tetapi juga menunjukkan proporsi anggota pada masing-masing kelompok yang terbentuk.

Evaluasi Hasil *Clustering* Menggunakan Box Plot

Untuk mengevaluasi karakteristik setiap cluster yang terbentuk, dilakukan analisis menggunakan Box Plot terhadap atribut Total_Amount. Visualisasi ini digunakan untuk mengamati persebaran data, nilai rata-rata, serta perbedaan tingkat transaksi pada masing-masing cluster. Hasil visualisasi ditunjukkan pada Gambar 3.



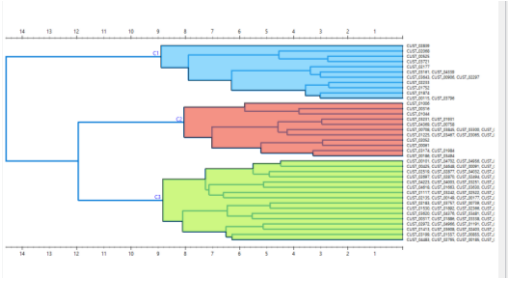
Gambar 3. Distribusi nilai Total_Amount pada setiap cluster hasil *K-Means.*

Berdasarkan Gambar 3, terdapat perbedaan yang cukup jelas pada nilai Total_Amount di antara ketiga cluster yang terbentuk. Cluster 2 (C2) memiliki rata-rata nilai transaksi tertinggi sebesar 3924,64, sedangkan Cluster 1 (C1) dan Cluster 3 (C3) memiliki rata-rata yang jauh lebih rendah, masing-masing sebesar 577,81 dan 553,51. Hasil ini menunjukkan bahwa pengguna pada Cluster 2 cenderung memiliki tingkat pengeluaran yang lebih tinggi dibandingkan pengguna pada cluster lainnya. Selain itu, Cluster 2 juga memiliki rentang persebaran data yang lebih luas dibandingkan Cluster 1 dan Cluster 3.

Kondisi ini mengindikasikan bahwa nilai transaksi pada Cluster 2 lebih bervariasi, termasuk adanya sejumlah pengguna dengan nilai transaksi yang sangat tinggi. Sebaliknya, Cluster 1 dan Cluster 3 menunjukkan persebaran data yang lebih sempit sehingga karakteristik transaksinya relatif lebih seragam. Hasil uji ANOVA menghasilkan nilai F sebesar 1077,375

dengan tingkat signifikansi p = 0,000. Nilai tersebut menunjukkan bahwa terdapat perbedaan rata-rata Total_Amount yang signifikan antar cluster. Dengan demikian, hasil evaluasi menggunakan Box Plot membuktikan bahwa algoritma *K-Means* mampu membentuk kelompok pengguna yang memiliki karakteristik transaksi yang berbeda secara statistik.

Hasil *Clustering* Menggunakan *Hierarchical Clustering*



Gambar 4. Hasil pengelompokan pengguna *E-commerce* menggunakan *Hierarchical Clustering.*

Berdasarkan Gambar 4, proses pengelompokan menggunakan algoritma *Hierarchical Clustering* menghasilkan tiga cluster utama, yaitu C1, C2, dan C3. Dendrogram menunjukkan hubungan hierarki antar data berdasarkan tingkat kemiripan karakteristik yang dimiliki oleh setiap pengguna. Penggabungan data dilakukan secara bertahap, dimulai dari objek yang memiliki jarak paling dekat hingga membentuk kelompok yang lebih besar. Selain itu, Cluster 2 juga memiliki rentang persebaran data yang lebih luas dibandingkan Cluster 1 dan Cluster 3. Kondisi ini mengindikasikan bahwa nilai transaksi pada Cluster 2 lebih bervariasi, termasuk adanya sejumlah pengguna dengan nilai transaksi yang sangat tinggi. Sebaliknya, Cluster 1 dan Cluster 3 menunjukkan persebaran data yang lebih sempit sehingga karakteristik transaksinya relatif lebih seragam. Secara keseluruhan, *Hierarchical Clustering* mampu mengidentifikasi struktur hubungan antar pengguna dan membentuk tiga kelompok utama berdasarkan kemiripan karakteristik transaksi dan aktivitas pelanggan. Hasil ini memberikan gambaran yang lebih rinci mengenai keterkaitan antar kelompok dibandingkan metode *K-Means* yang langsung membagi data ke dalam sejumlah cluster tertentu.

Perbandingan *K-Means* dan *Hierarchical Clustering*

Pada penelitian ini, algoritma *K-Means* dan *Hierarchical Clustering* digunakan untuk mengelompokkan pengguna *E-commerce* berdasarkan karakteristik transaksi dan aktivitas pelanggan. Kedua metode menghasilkan tiga cluster utama, namun memiliki pendekatan yang berbeda dalam proses pembentukan cluster. Berdasarkan hasil visualisasi Scatter Plot dan evaluasi menggunakan Box Plot, algoritma *K-Means* mampu menghasilkan pemisahan cluster yang cukup jelas. Perbedaan karakteristik antar cluster terlihat dari nilai Total_Amount yang

menunjukkan variasi rata-rata transaksi yang signifikan. Selain itu, distribusi anggota cluster menunjukkan bahwa sebagian besar data berada pada Cluster 1, sedangkan Cluster 2 memiliki jumlah anggota paling sedikit namun memiliki nilai transaksi yang lebih tinggi dibandingkan cluster lainnya.

Sementara itu, *Hierarchical Clustering* menghasilkan struktur pengelompokan dalam bentuk dendrogram yang menunjukkan hubungan kedekatan antar cluster. Hasil dendrogram memperlihatkan bahwa beberapa cluster memiliki tingkat kemiripan yang lebih tinggi dan bergabung terlebih dahulu sebelum membentuk kelompok yang lebih besar. Informasi ini memberikan gambaran mengenai hubungan hierarkis antar kelompok pengguna yang tidak ditampilkan secara langsung oleh algoritma *K-Means*. Secara keseluruhan, kedua algoritma berhasil mengelompokkan data pengguna berdasarkan tingkat kemiripan karakteristik yang dimiliki. Namun, *K-Means* menghasilkan cluster yang lebih mudah diinterpretasikan karena didukung oleh visualisasi persebaran data dan evaluasi statistik yang menunjukkan perbedaan karakteristik antar cluster. Di sisi lain, *Hierarchical Clustering* memberikan informasi tambahan mengenai struktur hubungan antar kelompok melalui dendrogram yang terbentuk.

5. KESIMPULAN DAN SARAN

Studi ini berhasil menerapkan algoritma *K-Means* dan *Hierarchical Clustering* untuk melakukan segmentasi pengguna *E-commerce* berdasarkan karakteristik transaksi dan aktivitas pelanggan dengan memanfaatkan dataset Perilaku Pelanggan *E-commerce*. Langkah-langkah analisis dilakukan menggunakan Orange *Data Mining* yang mencakup pengumpulan data, preprocessing, *Clustering*, evaluasi, dan interpretasi hasil. Temuan dari penelitian menunjukkan bahwa kedua algoritma tersebut efektif dalam mengelompokkan pengguna ke dalam tiga cluster utama yang berdasarkan pada kesamaan karakteristik yang ada. Algoritma *K-Means* menghasilkan pembagian cluster yang lebih jelas serta lebih mudah untuk diinterpretasikan melalui visualisasi scatter plot, di samping evaluasi statistik yang dilakukan dengan Box Plot dan ANOVA. Lebih jauh, *K-Means* menunjukkan performa yang lebih unggul dalam membedakan kelompok pengguna berdasarkan nilai transaksi serta pola aktivitas pelanggan. Sementara itu, *Hierarchical Clustering* memberikan informasi tambahan terkait struktur hubungan antara cluster melalui dendrogram, yang memungkinkan untuk analisis hubungan hierarkis di antara kelompok pengguna. Meskipun begitu, hasil *Clustering* yang diperoleh dari *K-Means* dianggap lebih representatif dan lebih mudah digunakan untuk interpretasi segmentasi pelanggan *e-commerce*. Secara keseluruhan, studi ini mengindikasikan bahwa penerapan metode *Clustering* dapat membantu dalam mengidentifikasi pola perilaku

pengguna *E-commerce* dan hal ini dapat digunakan sebagai dasar untuk merumuskan strategi pemasaran, meningkatkan layanan pelanggan, serta mendukung pengambilan keputusan bisnis yang lebih efektif dan tepat sasaran.

Berdasarkan temuan penelitian yang telah dilaksanakan, terdapat sejumlah rekomendasi yang bisa dipertimbangkan untuk penelitian mendatang. Pertama, studi berikutnya bisa memanfaatkan dataset yang lebih besar dan diperoleh dari platform *E-commerce* lain untuk membuat hasil segmentasi pelanggan lebih representatif. Kedua, penelitian bisa memasukkan atribut tambahan yang lebih terperinci, seperti frekuensi pembelian, waktu transaksi, tingkat loyalitas konsumen, serta riwayat pengembalian barang guna mendapatkan pemahaman yang lebih mendalam tentang karakteristik pelanggan. Di samping itu, studi selanjutnya dapat membandingkan algoritma pengelompokan lain, seperti DBSCAN, Fuzzy C-Means, atau K-Medoids untuk menemukan metode yang memberikan kualitas cluster yang lebih baik. Penilaian hasil pengelompokan juga bisa ditingkatkan dengan memanfaatkan metrik validasi seperti Silhouette Score, Davies-Bouldin Index, atau Calinski-Harabasz Index supaya kualitas pengelompokan dapat dinilai dengan lebih objektif. Penelitian berikutnya juga bisa mengintegrasikan metode *Clustering* dengan teknik machine learning lainnya, seperti klasifikasi atau prediksi pengunduran pelanggan, sehingga hasil segmentasi tidak hanya berfungsi untuk mengelompokkan pelanggan tetapi juga untuk mendukung pengambilan keputusan bisnis yang lebih strategis. Oleh karena itu, hasil studi dapat memberikan keuntungan yang lebih signifikan bagi perusahaan *E-commerce* dalam merancang strategi pemasaran dan meningkatkan kesetiaan pelanggan.

DAFTAR REFERENSI

A-priori, A. (n.d.). *Model Pengujian Akurasi Berbasis Empiris Pada*. *1*.

Adji, D. R., Lutfina, E., Ferdianto, B. E., Prashanti, E., Anwarri, K. A. P., & Prayogo, S. R. (2025). Metode dan Algoritma Dalam Data *Clustering*: Systematic Literature Review. *Science Technology and Management Journal*, *5*(1), 9–15. https://doi.org/10.53416/stmj.v5i1.326.

Alamtaha, Z., Djakaria, I., & Yahya, N. I. (2023). *Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokkan Pengguna Media Sosial*. *4*(1), 33–43. https://doi.org/10.20956/ejsa.vi.24830

Aldi, M. A., Fatah, Z., Informasi, T., Ibrahimy, U., Timur, S. J., Informasi, S., Ibrahimy, U., & Timur, S. J. (2025). *Implementasi K-Means Cluster Ing Dalam Pengelompokan Data Kunjungan*. *2*(1), 13–19 https://doi.org/10.69714/3hhfj353.

Anggraini, D., & Hasibuan, M. S. (2025). Initial *Centroid* Pada Algoritma *K-Means* Dan K-Medoids. *Jatilima : Jurnal Multimedia Dan Teknologi Informasi , 7*(3), 487–500.

Apriana, D., & Yuliansyah, C. (2024). Mengoptimalkan Penjualan Online Melalui Teknik *Data Mining* (Studi Kasus *E-commerce*). *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, *4*(02), 514–527. https://doi.org/10.37680/almikraj.v4i02.4774.

Br jabat, Tarigan, P. (2024). Pemanfaatan Platform *E-commerce* Melalui Marketplace Sebagai Upaya Peningkatan Penjualan. *Jurnal Abdimas Sosek*, *4*(2), 16–21.

Chandra, A. A., & Darmayanti, N. P. A. (2019). **肖沉** 1, 2, 孙莉 1, 2Δ, **曹杉杉** 1, 2, **梁浩** 1, 2, **程焱** 1, 2. *Tjyybjb.Ac.Cn*, *27*(2), 58–66.

Clustering, A. K., & Belajar, G. (2022). *Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar*. *12*. https://doi.org/10.34010/jati.v12i1.

Dan, A. K., Kaggle, D., Informatika, T., & Asahan, U. (2025). *Clustering Data Konsumen E-commerce Menggunakan*. *04*(2), 96–109.

Digital, B., Bisnis, P., & Indonesia, D. (2026). *Analisis Segmentasi Pelanggan E-commerce Menggunakan Metode Clustering Berbasis RFM*. *V*(1).

E-commerce, T., Apriyanto, B., Lina, S., & Sitio, M. (2025). *Penerapan K-Means dalam Menganalisis Pola Pembelian Pelanggan Pada Data*. *7*(3). https://doi.org/10.32877/bt.v7i3.2195.

Fahmi, H., Sari, W. K., Tania, K. D., Informasi, S., Sriwijaya, U., Indah, I., Ilir, K. O., Selatan, S., Behavior, S., & Method, P. (2026). *Analisis Komparasi Algoritma K-Means Dan Hierarchical Clustering Untuk Segmentasi Pelanggan Dalam Menggunakan Metode Pembayaran*. *10*(3), 4389–4397https://doi.org/10.36040/jati.v10i3.18173.

Fahrillah Fahrillah, & Zaehol Fatah. (2025). Pengelompokan Data Nilai Siswa Madrasah Ta'Hiliyah Menggunakan Metode *K-Means Clustering*. *Jurnal Riset Sistem Informasi*, *2*(1), 53–59. https://doi.org/10.69714/0v1pkz05.

Fathiarahma, A., Voutama, A., Ridwan, T., & Heryana, N. (2021). Analisis_Text_Mining_Klasifikasi_Kegiatan_Keluarga. *Jurnal Teknologi Terpadu*, *Vol. 9 No. 1 2023*(1), 35–41https://doi.org/10.54914/jtt.v9i1.606.

Ikhwan Syarif, M., Hannum, M., Wahyuni, S., Ekonomi, F., Islam Universitas Islam Negeri Sumatra Utara, B., & Artikel, H. (2023). Potensi Perkembangan *E-commerce* Dalam Menunjang Bisnis di Indonesia I N F O A R T I K E L A B S T R A K. *Jcbd Journal of Computers and Digital Business*, *2*(1), 11–14https://doi.org/10.56427/jcbd.v2i1.30.

K-Means, P. M. A., Nurani, S., Syahra, Y., & Calam, A. (2023). *Penerapan Data Mining Dalam Clustering Pencapaian Target*. *2*, 355–363 https://doi.org/10.53513/jursi.v2i3.6552.

Kartika, E. D., & Romansyah, N. (2024). Strategi Pemasaran Digital dan Pengaruhnya terhadap Loyalitas Pelanggan di Era *E-commerce*. *FISKAL: Jurnal Ekonomi, Bisnis, Dan Manajemen*, *1*(01), 1–7https://doi.org/10.70716/emis.v1i1.293.

Kaylista, N. N. K., Khoirunnisaa, N., F, G. V. E. N., & Pratama, A. Y. (2023). *Implementasi Algoritma K-Means Clustering Menggunakan Aplikasi Orange Untuk Mengetahui Pola Indeks Pembangunan Manusia Tahun 2022*. *4*(1), 65–76.

Kumar, S., Rani, R., Pippal, S. K., & Agrawal, R. (2025). *Customer segmentation in e-commerce : K-Means vs Hierarchical Clustering*. *23*(1), 119–128.

https://doi.org/10.12928/telkomnika.v23i1.26384.

Mauludin, M. S., Saputra, A. D., Sari, A. Z., & Munawaroh, I. (2022). *Analisis Perilaku Konsumen Dalam Transaksi Di e-commerce*. *1*(1).

Nanda, S. G., Shafira, H. I., & Zahira, T. (2025). Implementasi *Data Mining* di Sektor *E-commerce* 1*. *Jurnal Ilmu Komputer (JILKOMP)*, *1*(1), 13–16.

Nursania Dasopang. (2023). Jiemas E – Commerce Bisnis Dan Internet. *Jurnal Ilmiah Ekonomi, Manajemen Dan Syariah JIEMAS*, *2*, 129–135https://doi.org/10.55883/jiemas.v2i3.26.

Nuryansyah, B. A., & Rasyid, O. (2026). *Analisis Segmentasi Pelanggan dan Prediksi Churn pada E-commerce Menggunakan K-Means Clustering dan Random Forest : Studi Kasus Olist Brazilian E-commerce*. *02*(03), 795–801.

Oktaviani, M., Sari, I. P., & Miftah, Z. (2025). Pengaruh Tren Perilaku Konsumtif Dari Konsumen Dalam *E-commerce* Di Era Digital. *Jurnal Media Akademik (JMA)*, *3*(4), 1–1.

Pujiono, S., Astuti, R., & Muhamad Basysyar, F. (2024). Implementasi *Data Mining* Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma *K-Means Clustering*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 615–620. https://doi.org/10.36040/jati.v8i1.8360

Rahayu, N. D., Anshor, A. H., & Afriantoro, I. (2024). Penerapan *Data Mining* untuk Pemetaan Siswa Berprestasi menggunakan Metode *Clustering K-Means*. *JUKI : Jurnal Komputer Dan Informatika*, *6*(1), 71–83. https://doi.org/10.53842/juki.v6i1.474.

Ramiro Firjatullah, F., Faqih, A., & Dwilestari, G. (2025). Optimalisasi Strategi Pemasaran Menggunakan Algotima Agglomerative *Hierarchical Clustering* Pada Hasil Penjualan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(3), 4348–4352. https://doi.org/10.36040/jati.v9i3.13662.

Safitri, H., Putri Lenggo Geni, S., Merry, F., & Wati, M. (2025). Penerapan *K-Means Clustering* untuk Segmentasi Konsumen *E-commerce* Penerapan *K-Means Clustering* untuk Segmentasi Konsumen *E-commerce* Berdasarkan Pola Pembelian. *JUKI : Jurnal Komputer Dan Informatika*, *7*(1), 89–99https://doi.org/10.53842/juki.v7i1.1049.

Sakinah, A. M., & Awaliyah, D. S. (2025). *Optimization Of E-commerce Consumer Segmentation Based On K-Means Clustering And Machine Learning*. *8*(2), 606–619https://doi.org/10.35580/jmathcos.v8i2.9548.

Sari, R. (2021). *Pengaruh Penggunaan Paylater Terhadap Perilaku Impulse Buying Pengguna E-commerce di Indonesia*. *7*(1), 44–57https://doi.org/10.35313/jrbi.v7i1.2058.

Sudirjo, F., Purwati, T., Unggul Budiman, Y., & Manuhutu, M. (2018). Analisis Dampak Strategi Pemasaran Digital dalam Meningkatkan Loyalitas Pelanggan: Perspektif Industri *E-commerce*. *Jurnal Pendidikan Tambusa*, *7*(August), 7524–7532.

Suharti, P. H., Suryandari, A. S., & Amalia, R. N. (2022). Analisis Kinerja Modul Pengendali Tekanan Udara Pct-14 Berbasis Plc Dengan Berbagai Metoda Tuning. *Sebatik*, *26*(2), 420–427. https://doi.org/10.46984/sebatik.v26i2.2134.

Sutisna, N., & Sutrisna. (2023). Implementasikan Sistem Informasi dalam Mendukung Perilaku Pembelian Terhadap Keputusan Pembelian *E-commerce*. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, *2*(1), 20–30. https://doi.org/10.33050/mentari.v2i1.343.

Sypriansyah, D., & Hasibuan, A. (2025). Strategi Membangun Bisnis *E-commerce* Yang Sukses Di Era Digital Strategies for Building a Successful *E-commerce* Business in the Digital Era. *Variable Research Journal*, *02*(01), 1.

Widya, P., Septi, A., Wulandari, R. F., & Daniati, E. (n.d.). *Analisis Preferensi Tren Produk Berkelanjutan dalam E-commerce Menggunakan Metode K-Means Clustering*. *9*, 642–650.

Yosua, I. K., Agung, M., Bagus, I., & Sarasvananda, G. (2025). *Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering*. *3*(M), 917–924.

Zuhal, N. K. (2022). Study Comparison *K-Means Clustering* dengan Algoritma *Hierarchical Clustering*. *Prosiding Seminar Nasional Teknologi Dan Sains*, *1*, 200–205.