

Segmentasi Pelanggan E-Commerce dengan menggunakan Algoritma CLARANS

E-Commerce Customer Segmentation using the CLARANS Algorithm

Elimiana Br Berutu¹, Andre Hasudungan Lubis*²

^{1,2}Fakultas Teknik, Universitas Medan Area

¹elimiana7@gmail.com.ac.id, ²andrehasudunganlubis@uma.ac.id

Received: April 02, 2026 | Revised: April 05, 2026 | Accepted: April 29, 2026

Abstrak

Segmentasi pelanggan merupakan salah satu langkah penting dalam mendukung strategi pemasaran pada platform E-Commerce. Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan karakteristik dan perilaku transaksi menggunakan algoritma CLARANS (Clustering Large Applications based upon Randomized Search). Data yang digunakan berasal dari dataset pelanggan E-Commerce yang mencakup beberapa atribut, seperti usia, nilai rata-rata transaksi, jumlah pesanan, loyalitas pelanggan, dan risiko churn. Tahapan penelitian meliputi pengumpulan data, pembersihan data, feature engineering, exploratory data analysis (EDA), implementasi algoritma, serta evaluasi hasil clustering. Evaluasi dilakukan menggunakan Silhouette Score, Davies–Bouldin Index, dan Calinski–Harabasz Index, kemudian dibandingkan dengan metode K-Means dan Hierarchical Clustering. Hasil penelitian menunjukkan bahwa algoritma CLARANS memberikan performa terbaik dengan nilai Silhouette Score sebesar 0,381991, Davies–Bouldin Index sebesar 1,061123, dan Calinski–Harabasz Index sebesar 3458,564. Hasil tersebut menunjukkan bahwa CLARANS mampu menghasilkan cluster yang lebih kompak dan terpisah dengan baik, sehingga efektif digunakan untuk segmentasi pelanggan pada data E-Commerce.

Kata kunci: Segmentasi Pelanggan, E-Commerce, CLARANS, Clustering, Evaluasi Cluster

Abstract

Customer segmentation is an important step in supporting marketing strategies on E-Commerce platforms. This study aims to cluster customers based on their characteristics and transaction behavior using the CLARANS (Clustering Large Applications based upon Randomized Search) algorithm. The dataset used consists of E-Commerce customer attributes, including age, average transaction value, total orders, customer loyalty, and churn risk. The research stages include data collection, data cleaning, feature engineering, exploratory data analysis (EDA), algorithm implementation, and clustering evaluation. The evaluation was conducted using Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index, and benchmarked against K-Means and Hierarchical Clustering methods. The results show that the CLARANS algorithm provides the best performance with a Silhouette Score of 0.381991, Davies–Bouldin Index of 1.061123, and Calinski–Harabasz Index of 3458.564. These findings indicate that CLARANS is capable of producing more compact and well-separated clusters, making it effective for customer segmentation in E-Commerce data.

Keywords: Customer Segmentation, E-Commerce, CLARANS, Clustering, Cluster Evaluation

1. PENDAHULUAN

Pelanggan merupakan istilah yang terkait dengan orang atau pihak yang selalu berkunjung atau berbelanja kepada pihak bisnis. Dimana, pelanggan melakukan transaksi berupa membeli atau memanfaatkan suatu produk usaha baik berupa barang ataupun jasa yang dijual atau dipasarkan, atau hanya sekedar melihat-lihat tempat usaha serta bertukar pikiran tentang usaha tersebut [1]. Transformasi pelanggan dari bisnis konvensional menjadi pelanggan *E-Commerce* adalah perubahan perilaku konsumen dari yang sebelumnya melakukan pembelian secara langsung di toko fisik menjadi melakukan transaksi melalui platform digital seperti *website*, *marketplace*, atau aplikasi belanja *online* [2].

Melalui *E-Commerce*, pelanggan dapat melihat katalog produk, membaca deskripsi dan ulasan, memilih barang, melakukan pembayaran secara *online*, serta memantau proses pengiriman barang hingga sampai ke tujuan [3]. Segmentasi pelanggan di *E-Commerce* sangat penting karena membantu pelaku bisnis memahami siapa pembelinya, lalu menyesuaikan strategi pemasaran, produk, dan layanan sehingga lebih tepat sasaran dan efisien [4]. Dalam bisnis *E-Commerce*, pelanggan memiliki pola belanja yang berbeda-beda, seperti frekuensi pembelian, jenis produk yang diminati, nilai transaksi, dan waktu terakhir melakukan pembelian [5]. Dengan melakukan segmentasi, perusahaan dapat mengelompokkan pelanggan ke dalam beberapa kategori berdasarkan kesamaan perilaku tersebut, misalnya pelanggan loyal, pelanggan baru, pelanggan potensial, atau pelanggan yang jarang bertransaksi [6].

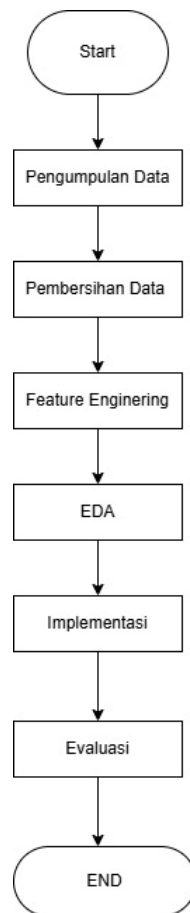
Penerapan kecerdasan buatan kini menjadi tren yang kerap di gunakan dalam bisnis [7]. Terdapat berbagai penelitian yang telah melakukan segmentasi pelanggan menggunakan beberapa metode clustering. Metode K-Means banyak digunakan karena prosesnya sederhana dan cepat, namun memiliki kelemahan yaitu harus menentukan jumlah cluster di awal serta sensitif terhadap pemilihan centroid awal. Selain itu, metode ini kurang optimal untuk data pelanggan yang memiliki pola kompleks dan tidak seragam [8]. Metode DBSCAN mampu mengelompokkan data berdasarkan kepadatan tanpa menentukan jumlah cluster terlebih dahulu serta dapat mendeteksi *noise*. Namun, metode ini sangat bergantung pada pemilihan parameter *epsilon* dan *MinPts*, sehingga hasil cluster dapat berubah jika parameter tidak ditentukan secara tepat [1].

Sementara itu, Hierarchical Clustering dapat membentuk struktur cluster secara bertingkat dan mudah diinterpretasikan, tetapi memiliki kelemahan pada kompleksitas komputasi yang tinggi sehingga kurang efisien untuk data pelanggan dalam jumlah besar [9]. Berdasarkan keterbatasan tersebut, diperlukan metode yang lebih efisien untuk menangani data skala besar, sehingga penelitian ini menggunakan algoritma CLARANS sebagai alternatif dalam segmentasi pelanggan E-Commerce.

CLARANS (*Clustering Large Applications based upon Randomized Search*) merupakan algoritma clustering yang dirancang untuk mengelompokkan data berukuran besar berdasarkan kesamaan karakteristik tertentu. Algoritma ini menggunakan pendekatan pencarian acak untuk menemukan solusi cluster yang optimal secara lebih efisien [10]. Secara umum, keunggulan CLARANS terletak pada kemampuannya dalam menangani data berukuran besar dengan proses pencarian cluster yang lebih efisien dibandingkan metode clustering tradisional. Melalui mekanisme pencarian acak pada ruang solusi, algoritma ini mampu mengurangi kompleksitas komputasi sekaligus tetap menghasilkan kualitas cluster yang baik [11]. Selain itu, CLARANS lebih fleksibel dalam menemukan pusat cluster yang optimal pada data dengan pola yang beragam [12]. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menerapkan algoritma CLARANS untuk melakukan segmentasi pelanggan E-Commerce guna menghasilkan kelompok pelanggan yang lebih jelas dan mendukung pengambilan keputusan strategi pemasaran yang lebih efektif.

2. METODE PENELITIAN

Dalam penelitian ini terdapat tahapan yang disusun secara sistematis untuk menghasilkan proses segmentasi pelanggan yang terstruktur dan akurat. Alur penelitian dimulai dari tahap pengumpulan data hingga evaluasi hasil clustering, sebagaimana ditunjukkan pada diagram alur penelitian seperti yang terlihat pada Gambar 1. Pada tahap pertama, pengumpulan data dilakukan berupa data transaksi pelanggan E-Commerce yang berasal dari tautan <https://www.kaggle.com/datasets/mirzayasirabdullah07/ecommerce-customer-dataset>.



Gambar 1. Tahapan Penelitian

Tahap selanjutnya adalah pembersihan data, yakni melakukan proses pengecekan terhadap data yang tidak lengkap, duplikasi data, nilai kosong (*missing values*), maupun data yang tidak konsisten. Tujuan dari tahap ini adalah memastikan bahwa data yang digunakan dalam proses analisis berada dalam kondisi bersih dan layak untuk diolah lebih lanjut.

Setelah data dibersihkan, penelitian dilanjutkan ke tahap *feature engineering* yang bertujuan untuk membentuk variabel atau fitur yang lebih representatif dari data asli, sehingga dapat meningkatkan kualitas proses *clustering*. Tahap berikutnya adalah *Exploratory Data Analysis* (EDA). Pada tahap ini dilakukan analisis awal untuk memahami karakteristik data, pola distribusi, hubungan antar variabel, serta mendeteksi kemungkinan adanya outlier. EDA

membantu peneliti memperoleh gambaran awal mengenai struktur data sebelum proses clustering dilakukan.

Selanjutnya adalah tahap implementasi, yaitu penerapan algoritma CLARANS pada data yang telah diproses. Pada tahap ini, data dikelompokkan ke dalam beberapa cluster berdasarkan kemiripan karakteristik pelanggan. Tahap terakhir adalah evaluasi, yang bertujuan untuk menilai kualitas cluster yang dihasilkan oleh algoritma CLARANS. Evaluasi dilakukan menggunakan metrik tertentu, seperti Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index.

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini berasal dari dataset publik yang diperoleh melalui platform Kaggle dengan judul Ecommerce Customer Dataset yang dipublikasikan oleh *Mirza Yasir Abdullah*. Dataset ini berisi informasi pelanggan dari platform E-Commerce yang digunakan untuk analisis perilaku pelanggan, segmentasi, serta pengambilan keputusan berbasis data. Dataset tersebut masih dapat diakses pada saat penelitian dilakukan dan diunduh pada tanggal 22 April 2026. Penggunaan dataset sekunder dari repositori publik seperti Kaggle umum digunakan dalam penelitian machine learning dan data mining karena mendukung replikasi penelitian serta transparansi eksperimen. Namun demikian, karena dataset publik dapat mengalami perubahan atau penghapusan sewaktu-waktu, informasi sumber dan tanggal akses dicantumkan secara eksplisit dalam penelitian ini.

Dataset penelitian memuat sebanyak 12 atribut yang merepresentasikan karakteristik pelanggan dan perilaku transaksi, seperti CustomerID, Gender, Age, Annual Income, Spending Score, serta riwayat pembelian pelanggan. Variabel-variabel tersebut digunakan sebagai dasar dalam proses segmentasi pelanggan karena mampu menggambarkan profil demografis dan pola pengeluaran pelanggan secara lebih jelas. Rangkuman dari data penelitian ini dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Rangkuman 5 Data Teratas pada Penelitian

Customer_id	Age	Gender	Country	Avg_order_value	...	Churn risk
CUST_8270	30	Female	Brazil	101.08	...	0.20
CUST_1860	53	Female	USA	90.39	...	0.34
CUST-6390	73	Male	Australia	83.28	...	0.05
CUST_6191	30	Other	Japan	109.90	...	0.19
CUST_6737	29	Female	Canada	269.38	...	0.15

2.2 Pembersihan Data

Pada tahap ini dilakukan proses pembersihan data untuk memastikan data yang digunakan berada dalam kondisi siap analisis. Proses pembersihan meliputi pengecekan nilai kosong (*missing values*), data duplikat, serta konsistensi tipe data pada setiap atribut. Berdasarkan dataset yang digunakan, atribut seperti CustomerID, Gender, Age, Annual Income, dan Spending Score diperiksa untuk memastikan tidak terdapat data yang hilang atau tidak valid. Selain itu, dilakukan penghapusan data duplikat dan penyesuaian format data agar sesuai dengan kebutuhan proses clustering. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga hasil segmentasi pelanggan menggunakan algoritma CLARANS dapat lebih akurat dan optimal.

2.3 Feature Engineering

Pada tahap ini dilakukan pemilihan dan transformasi atribut yang relevan agar proses

clustering dapat menghasilkan segmentasi yang lebih optimal. Fitur yang tidak digunakan dalam penelitian ini meliputi CustomerID, Customer_Since, Country, dan Preferred_Category. Karena fitur-fitur tersebut kurang mampu merepresentasikan karakteristik serta perilaku pelanggan dalam melakukan transaksi. Selain pemilihan fitur, atribut kategorikal seperti Gender terlebih dahulu diubah ke dalam bentuk numerik menggunakan teknik *label encoding* agar dapat diproses oleh algoritma clustering. Selanjutnya, dilakukan normalisasi data pada setiap fitur numerik untuk menyamakan skala nilai antar variabel, sehingga tidak ada atribut yang mendominasi proses pengelompokan. Tahap ini bertujuan untuk menghasilkan fitur yang lebih representatif dan sesuai dengan kebutuhan algoritma CLARANS dalam mengukur kemiripan antar data pelanggan. Adapun teknik normalisasi yang di gunakan adalah metode min-max dengan menggunakan persamaan berikut [13].

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

2.4 Exploratory Data Analysis (EDA)

Pada tahap Exploratory Data Analysis (EDA) dilakukan analisis awal terhadap data untuk memahami karakteristik dan pola distribusi setiap variabel yang digunakan dalam penelitian [14]. Analisis ini meliputi pemeriksaan statistik deskriptif, seperti nilai minimum, maksimum, rata-rata, dan standar deviasi pada fitur Age, Annual Income, dan Spending Score. Selain itu, dilakukan visualisasi data untuk melihat distribusi masing-masing fitur serta hubungan antar variabel, sehingga dapat diketahui pola awal perilaku pelanggan. Tahap ini juga bertujuan untuk mendeteksi adanya outlier atau data yang memiliki nilai ekstrem yang berpotensi memengaruhi hasil clustering..

2.5 Implementasi Algoritma CLARANS

Pada tahap ini dilakukan implementasi algoritma CLARANS untuk mengelompokkan data pelanggan berdasarkan kemiripan karakteristik yang telah dibentuk pada tahap sebelumnya. Pada algoritma CLARANS proses pengelompokan dilakukan dengan tujuan meminimalkan total jarak antara setiap objek data terhadap medoid terdekat pada masing-masing cluster [15]. Secara matematis, fungsi objektif yang digunakan dapat dinyatakan sebagai berikut.

$$J = \sum_{i=1}^n \min_{m_j \in M} d(x_i, m_j) \quad (2)$$

Tujuan utama algoritma adalah memperoleh nilai J sekecil mungkin, karena semakin kecil nilai fungsi objektif maka semakin baik kualitas cluster yang terbentuk. Apabila pengukuran jarak menggunakan Euclidean Distance, maka jarak antara objek data dan medoid dihitung menggunakan persamaan berikut:

$$d(x_i, m_j) = \sqrt{\sum_{l=1}^p (x_{il} - m_{jl})^2} \quad (3)$$

Dalam proses pencarian solusi, CLARANS melakukan pertukaran secara acak antara medoid dan non-medoid untuk membentuk solusi tetangga (*neighbor*). Perubahan kualitas cluster setelah pertukaran tersebut dihitung dengan persamaan:

$$\Delta J = J_{baru} - J_{lama} \quad (4)$$

Jika nilai $\Delta J < 0$, maka solusi baru dianggap lebih baik dan diterima sebagai medoid baru. Sebaliknya, jika $\Delta J \geq 0$, maka solusi ditolak dan proses pencarian tetangga dilanjutkan. Mekanisme ini memungkinkan CLARANS menemukan solusi optimum lokal secara efisien melalui pencarian acak. Selain itu, algoritma CLARANS memiliki dua parameter penting, yaitu numlocal dan maxneighbor. Parameter numlocal menyatakan jumlah pencarian optimum lokal yang dilakukan, sedangkan maxneighbor menyatakan jumlah maksimum tetangga yang diperiksa pada setiap pencarian. Secara matematis dapat dituliskan sebagai Semakin besar nilai numlocal, maka pencarian solusi akan semakin mendekati hasil optimal. Adapun konfigurasi parameter yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Konfigurasi Parameter CLARANS

Parameter	Nilai	Fungsi
numlocal	5	Menentukan jumlah pencarian local yang dilakukan algoritma
Maxneighbor	10	Menentukan jumlah maksimum tetangga yang di periksa pada setiap iterasi

2.6 Evaluasi

Pada tahapan ini yaitu meninjau hasil klusterisasi yang diperoleh dari algoritma CLARANS. Evaluasi bagaimana data terkuster dan apakah hasilnya sesuai dengan harapan. Menggunakan evaluasi metrik untuk menilai kualitas kluster dengan menggunakan *Davies Bouldin Index (DBI)*, *Silhouette Coefficient*, dan *Calinski-Harabasz Index*.

3. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil penelitian dan pembahasan yang diperoleh dari proses segmentasi pelanggan menggunakan algoritma CLARANS. Pembahasan diawali dengan hasil EDA untuk memberikan gambaran mengenai karakteristik, distribusi, dan pola awal data pelanggan yang digunakan dalam penelitian. Selanjutnya, dipaparkan hasil clustering menggunakan algoritma CLARANS, yang menunjukkan pembentukan beberapa kelompok pelanggan berdasarkan kemiripan karakteristik dan perilaku transaksi. Hasil cluster kemudian dianalisis untuk mengidentifikasi profil masing-masing kelompok pelanggan.

Untuk menilai kualitas hasil segmentasi, dilakukan evaluasi cluster menggunakan metode DBI dan *Silhouette Coefficient*. Kedua metrik ini digunakan untuk mengukur tingkat kekompakan cluster dan jarak antar cluster yang terbentuk. Sebagai bentuk validasi dan perbandingan, hasil clustering CLARANS juga akan dilakukan benchmark dengan algoritma K-Means dan Hierarchical Clustering, sehingga dapat diketahui performa dan kualitas cluster yang dihasilkan oleh masing-masing metode.

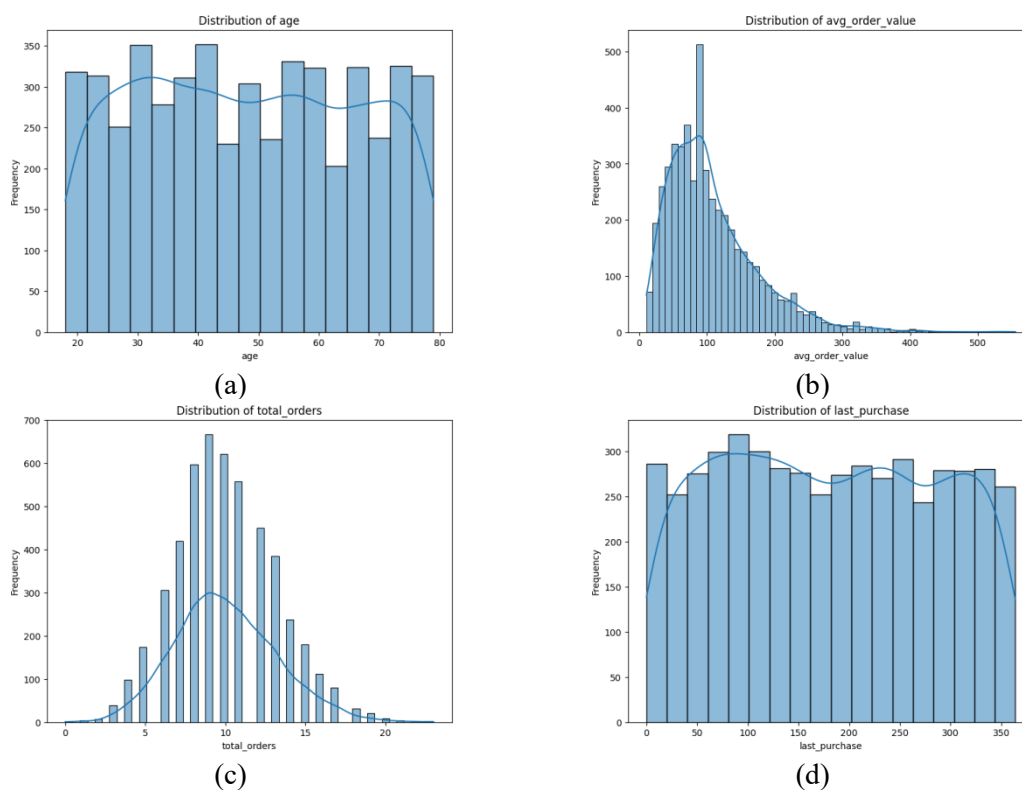
3.1 EDA

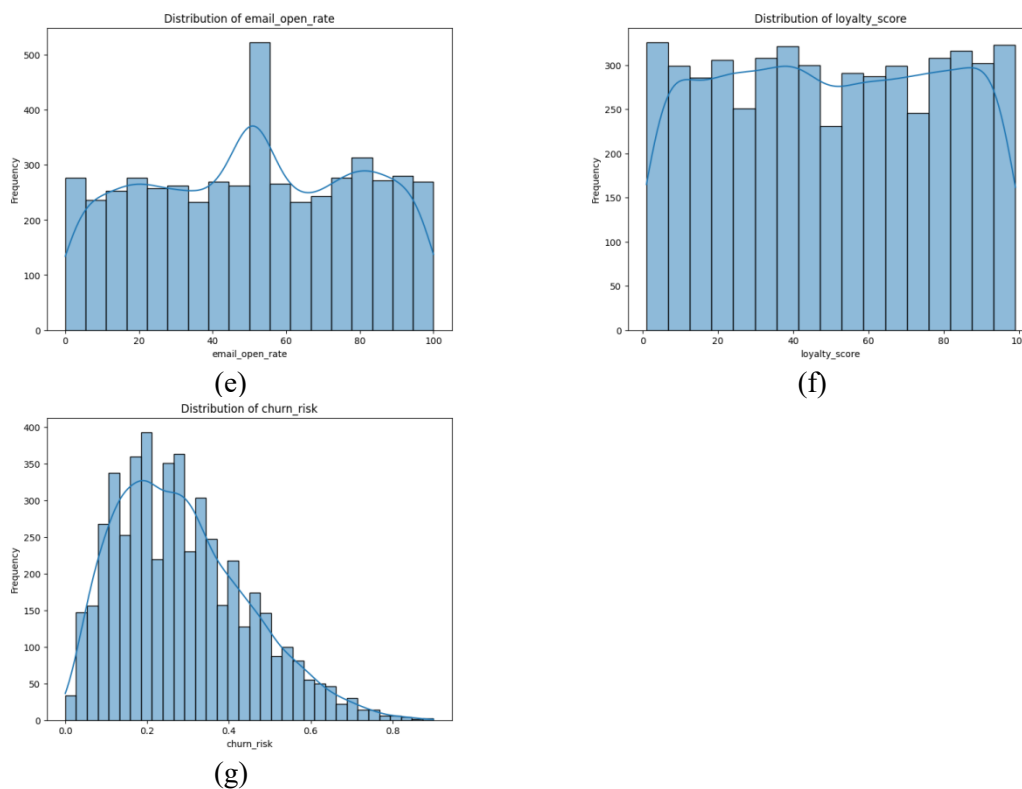
Pada tahap Exploratory Data Analysis (EDA), dilakukan analisis awal untuk memahami karakteristik data pelanggan yang digunakan dalam penelitian. Tahap ini bertujuan untuk melihat gambaran umum distribusi data, kecenderungan nilai, serta variasi pada setiap fitur sebelum proses clustering dilakukan.

Tabel 3. Statistik Deskriptif

Fitur	Mean	Std	Min	Max
Age	48,163	17,880	18,000	79,000
avg_order_value	107,680	67,593	10,660	555,460
total_orders	10,0270	3,1638	0,0000	23,000
last_purchase	180,073	104,926	0,0000	364,000
email_open_rate	50,7266	28,3618	0,0000	100,000
loyalty_score	50,0394	28,832	1,0000	99,000
churn_risk	0,2844	0,1596	0,0000	0,9000

Berdasarkan Tabel 3, dapat dilihat bahwa rata-rata usia pelanggan sebesar 48,163 tahun dengan rentang usia antara 18 hingga 79 tahun. Nilai rata-rata avg_order_value sebesar 107,680 menunjukkan rata-rata nilai transaksi pelanggan, sedangkan total_orders memiliki rata-rata 10,0270, yang menunjukkan frekuensi pemesanan pelanggan selama periode pengamatan. Selain itu, fitur last_purchase memiliki rata-rata 180,073, yang mengindikasikan rata-rata jarak waktu sejak transaksi terakhir pelanggan. Fitur lain juga menunjukkan variasi yang cukup besar berdasarkan nilai standar deviasi, sehingga berpotensi memberikan kontribusi yang baik dalam proses segmentasi pelanggan. Untuk memperjelas distribusi dan pola sebaran data pada masing-masing fitur, visualisasi statistik deskriptif ditampilkan pada Gambar 2.

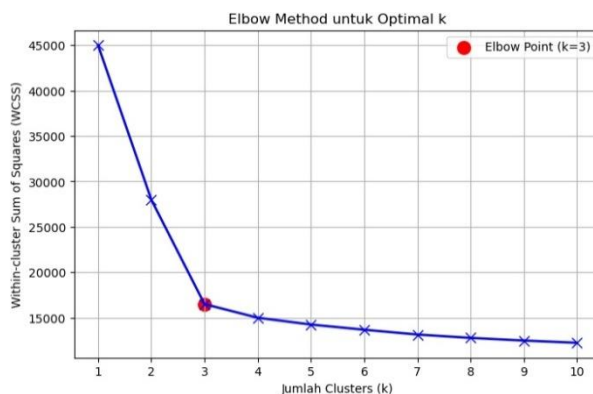




Gambar 2. Visualisasi EDA

3.2 Implementasi CLARANS

Pada tahap ini, algoritma CLARANS diterapkan untuk mengelompokkan data pelanggan berdasarkan kemiripan karakteristik yang telah diproses pada tahap sebelumnya. Proses clustering dilakukan dengan menentukan jumlah cluster optimal menggunakan metode Elbow Method berdasarkan nilai Within-cluster Sum of Squares (WCSS). Metode ini digunakan untuk melihat titik penurunan signifikan pada nilai WCSS yang menunjukkan jumlah cluster paling optimal. Hasil Elbow Method dapat dilihat pada Gambar 3 di bawah ini.



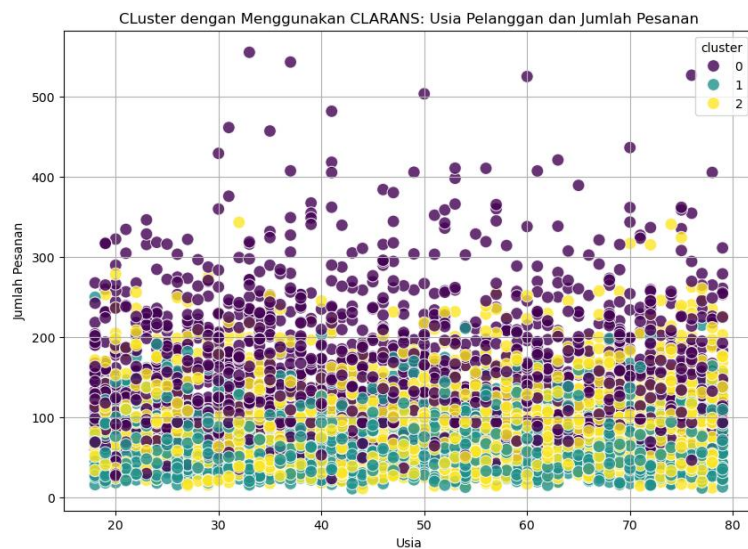
Gambar 3. Hasil Elbow Method untuk Penentuan Jumlah Cluster

Berdasarkan hasil pengujian pada Gambar 3, terlihat bahwa penurunan nilai WCSS terjadi secara signifikan dari $k=1$ hingga $k=3$, kemudian mulai melandai setelahnya. Kondisi ini menunjukkan bahwa titik elbow berada pada nilai $k=3$ sehingga jumlah cluster optimal yang digunakan dalam penelitian ini adalah tiga cluster. Pemilihan jumlah cluster tersebut diharapkan mampu menghasilkan segmentasi pelanggan yang lebih representatif dan memiliki kualitas cluster yang baik. Hasil clustering bertujuan untuk membentuk kelompok pelanggan yang memiliki pola perilaku dan profil yang serupa seperti yang di tampilkan pada Tabel 4.

Tabel 4. Hasil Cluster dengan algoritma CLARANS

Cluster	Jumlah
0	1559
1	2074
2	1367

Berdasarkan Tabel 4, hasil clustering menunjukkan bahwa data terbagi ke dalam tiga cluster, yaitu cluster 0 sebanyak 1559 data, cluster 1 sebanyak 2074 data, dan cluster 2 sebanyak 1367 data. Hasil ini menunjukkan bahwa cluster 1 memiliki jumlah anggota terbanyak dibandingkan cluster lainnya, sedangkan cluster 2 memiliki jumlah anggota paling sedikit. Untuk memperjelas pola persebaran antar cluster berdasarkan variabel usia pelanggan dan jumlah pesanan, visualisasi hasil clustering ditampilkan pada Gambar 4.



Gambar 4. Visualisasi Hasil Cluster dengan algoritma CLARANS

Berdasarkan Gambar 4, terlihat bahwa ketiga cluster memiliki pola persebaran yang berbeda pada variabel usia pelanggan dan jumlah pesanan. Cluster 1 cenderung terkonsentrasi pada jumlah pesanan yang relatif rendah, cluster 2 berada pada rentang jumlah pesanan menengah, sedangkan cluster 0 memiliki persebaran jumlah pesanan yang lebih tinggi dan lebih luas. Hal ini menunjukkan bahwa algoritma CLARANS mampu memisahkan pelanggan berdasarkan intensitas jumlah pesanan meskipun rentang usia pelanggan tersebar cukup merata. Secara umum, hasil ini mengindikasikan bahwa jumlah pesanan menjadi salah satu faktor pembeda utama dalam pembentukan cluster pelanggan, sementara variabel usia menunjukkan distribusi yang lebih menyebar di setiap cluster.

3.3 Evaluasi Hasil

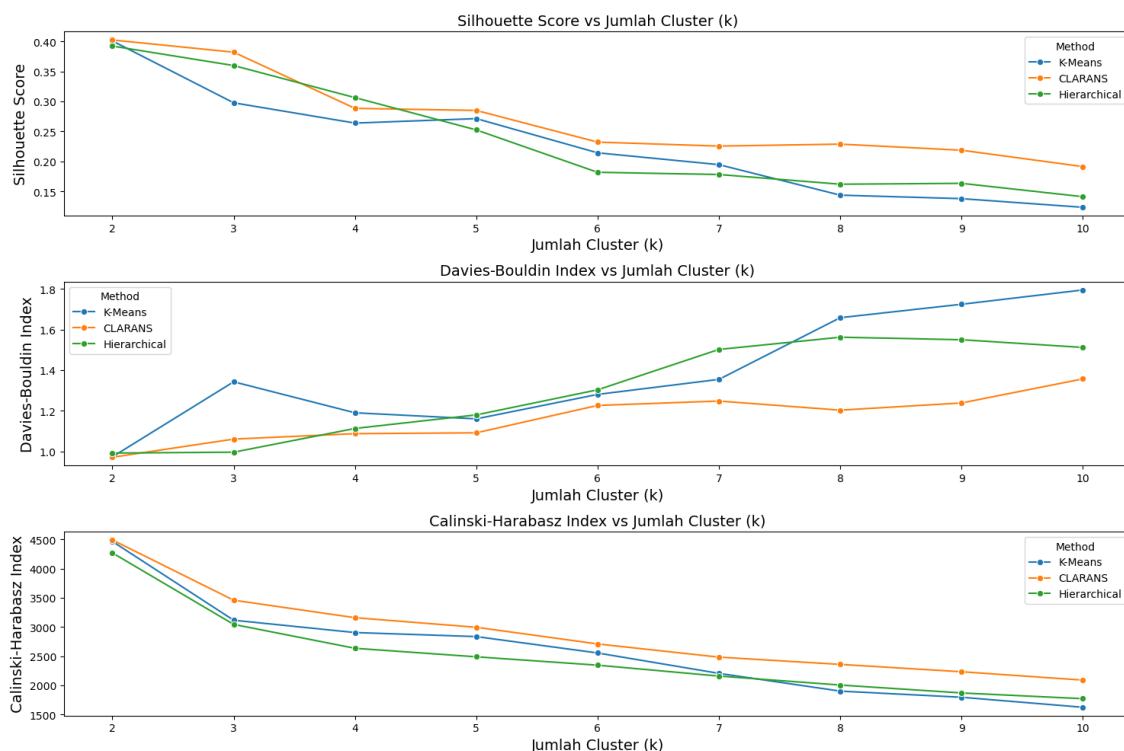
Pada tahap evaluasi, kualitas hasil clustering dibandingkan menggunakan tiga metrik, yaitu Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Evaluasi ini dilakukan untuk menilai tingkat kekompakan cluster, jarak antar cluster, serta kualitas pemisahan kelompok yang dihasilkan oleh masing-masing metode.

Tabel 5. Perbandingan Performa Clustering

Metode	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
CLARANS	0,381991	1,061123	3458,564
K-Means	0,297585	1,342858	3115,298
Hierarchical	0,359717	0,996654	3043,757

Berdasarkan Tabel 5, metode CLARANS memperoleh nilai Silhouette Score sebesar 0,381991, yang lebih tinggi dibandingkan K-Means (0,297585) dan sedikit lebih tinggi dibandingkan Hierarchical (0,359717). Nilai ini menunjukkan bahwa hasil cluster CLARANS memiliki tingkat kemiripan internal yang lebih baik. Pada metrik Davies–Bouldin Index, CLARANS memperoleh nilai 1,061123, yang lebih rendah dibandingkan K-Means (1,342858) dan mendekati Hierarchical (0,996654). Sementara itu, pada Calinski–Harabasz Index, CLARANS menunjukkan nilai tertinggi yaitu 3458,564, dibandingkan K-Means (3115,298) dan Hierarchical (3043,757). Hal ini mengindikasikan bahwa CLARANS menghasilkan cluster yang lebih kompak dan terpisah dengan baik. Meskipun metode Hierarchical memperoleh nilai Davies–Bouldin Index yang lebih rendah dibandingkan CLARANS, selisih nilai keduanya relatif kecil sehingga performa kedua metode masih dapat dianggap kompetitif. Namun demikian, CLARANS tetap dipertimbangkan lebih unggul secara keseluruhan karena menunjukkan hasil yang lebih baik pada metrik evaluasi lainnya, seperti Silhouette Score dan distribusi cluster yang lebih stabil. Oleh karena itu, penilaian performa tidak hanya didasarkan pada satu metrik, melainkan mempertimbangkan keseluruhan hasil evaluasi clustering.

Berdasarkan Gambar 5, terlihat bahwa metode CLARANS menunjukkan performa yang lebih stabil pada variasi jumlah cluster dari $k=2$ hingga $k=10$. Pada grafik Silhouette Score, CLARANS cenderung mempertahankan nilai yang lebih tinggi dibandingkan metode lainnya. Pada grafik Davies–Bouldin Index, nilai CLARANS relatif lebih rendah dan stabil, yang menunjukkan kualitas pemisahan cluster yang baik. Selain itu, pada grafik Calinski–Harabasz Index, CLARANS secara konsisten berada pada nilai yang lebih tinggi dibandingkan K-Means dan Hierarchical untuk sebagian besar variasi k . Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa CLARANS merupakan metode terbaik pada penelitian ini. Hal ini didukung oleh nilai Silhouette Score dan Calinski–Harabasz Index yang paling tinggi, serta nilai Davies–Bouldin Index yang kompetitif dan stabil. Keunggulan ini menunjukkan bahwa CLARANS lebih mampu membentuk cluster yang kompak, terpisah dengan baik, dan konsisten pada berbagai jumlah cluster.



Gambar 5. Visualisasi Perbandingan Kinerja antar Metode dengan Variasi k

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma CLARANS mampu melakukan segmentasi pelanggan dengan kualitas cluster yang lebih baik dibandingkan metode K-Means dan Hierarchical Clustering. Keunggulan ini terlihat dari hasil evaluasi yang menunjukkan nilai Silhouette Score dan Calinski-Harabasz Index tertinggi. Hal ini mengindikasikan bahwa CLARANS mampu menghasilkan cluster yang memiliki tingkat homogenitas internal yang baik sekaligus pemisahan antar cluster yang lebih jelas. Pendekatan pencarian acak yang digunakan CLARANS juga memberikan fleksibilitas dalam menemukan medoid yang lebih optimal dibandingkan metode berbasis centroid.

Selain itu, hasil clustering menunjukkan bahwa data pelanggan dapat dikelompokkan ke dalam beberapa segmen yang memiliki pola jumlah pesanan dan karakteristik usia yang berbeda. Segmentasi ini dapat digunakan untuk mendukung strategi bisnis, seperti identifikasi pelanggan aktif, pelanggan potensial, dan pelanggan dengan tingkat aktivitas rendah. Dengan demikian, penerapan CLARANS pada data pelanggan E-Commerce memberikan hasil yang tidak hanya baik secara metrik evaluasi, tetapi juga relevan untuk kebutuhan analisis bisnis dan pengambilan keputusan pemasaran.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma CLARANS terbukti mampu melakukan segmentasi pelanggan pada data E-Commerce dengan hasil yang baik. Proses clustering berhasil membagi data pelanggan ke dalam tiga cluster berdasarkan kemiripan karakteristik dan perilaku transaksi. Hasil evaluasi menunjukkan bahwa CLARANS memiliki performa yang lebih baik dibandingkan metode K-Means dan Hierarchical Clustering, yang

ditunjukkan oleh nilai Silhouette Score dan Calinski–Harabasz Index yang lebih tinggi, serta nilai Davies–Bouldin Index yang kompetitif. Dengan demikian, CLARANS dapat dinyatakan sebagai metode yang efektif untuk segmentasi pelanggan dan dapat digunakan sebagai dasar dalam mendukung strategi pemasaran yang lebih tepat sasaran pada platform E-Commerce. Penelitian selanjutnya dapat dilakukan dengan menggunakan dataset yang lebih besar dan lebih beragam untuk menguji konsistensi performa algoritma CLARANS pada skala data yang berbeda. Selain itu, penelitian berikutnya juga dapat menggabungkan CLARANS dengan metode optimasi atau teknik seleksi fitur untuk meningkatkan kualitas clustering. Validasi pada domain E-Commerce yang berbeda maupun penggunaan metrik evaluasi tambahan juga dapat dilakukan agar hasil penelitian menjadi lebih komprehensif dan generalisasi model dapat ditingkatkan.

DAFTAR PUSTAKA

- [1] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM: Comparison of the DBSCAN and K-MEANS Algorithms in Segmenting Customers Using Public Transportation of Transjakarta Using the RFM Method,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024.
- [2] M. Yusuf, S. Sutrisno, P. A. A. N. Putri, M. Asir, and P. A. Cakranegara, “Prospek Penggunaan E-Commerce Terhadap Profitabilitas Dan Kemudahan Pelayanan Konsumen: Literature Review,” *J. Darma Agung*, vol. 30, no. 3, pp. 786–801, 2024.
- [3] I. S. Riphah, *Pajak E-Commerce: Sebuah Pengertian Awal*. PT Elex Media Komputindo, 2022.
- [4] G. B. Sulistyono, N. Hasan, S. Kiswati, E. Muningsih, and others, “Segmentasi pelanggan dan optimalisasi penjualan pada data retail online berbasis model RFM,” *CONTEN Comput. Netw. Technol.*, vol. 5, no. 1, pp. 16–22, 2025.
- [5] 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, 2025.
- [6] S. Wulansari, “Analisis Segmentasi Pelanggan di Tiga E-Commerce Yang Paling Banyak Dikunjungi di Indonesia Pada Tahun 2023 Menggunakan RFM Model dan Teknik Clustering,” Universitas Bakrie, 2024.
- [7] A. J. Wijaya, “Peran Dan Implementasi Teknologi Kecerdasan Buatan Dalam Pengalaman Konsumen E-Commerce: Sebuah Tinjauan Sistematis,” *AIRA (Artificial Intell. Res. Appl. Learn.)*, vol. 1, no. 1, pp. 18–32, 2022.
- [8] E. F. L. Awalina and W. I. Rahayu, “Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-means clustering pada transaksi online retail,” *J. Teknol. Dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023.
- [9] M. A. Pakpahan, S. Sahid, M. M. F. Simanullang, and R. P. Winanda, “Analisis Hierarchical Clustering untuk Segmentasi Pelanggan pada Dataset Mall Customers,” *J. Sist. Inf. Dan Inform.*, vol. 4, no. 1, pp. 58–63, 2026.
- [10] N. H. Darmawan, “Segmentasi Loyalitas Pelanggan dengan Model RFM Menggunakan Algoritma Clarans,” *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, pp. 362–368, 2022.
- [11] P. Sarang, “CLARANS: Clustering Large Datasets with Randomized Search,” in *Thinking Data Science: A Data Science Practitioner’s Guide*, Springer, 2023, pp. 237–242.
- [12] R. Narasati, R. Lestari, R. Herdiana, R. Hamonangan, and S. Anawar, “Penerapan Algoritma K-Medoid Pada Penjualan Kerudung E-Commerce Shopee: Fifau Hijab,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1815–1819, 2024.

-
- [13] M. Mazziotta and A. Pareto, "Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min--Max method," *Environmetrics*, vol. 33, no. 5, p. e2730, 2022.
- [14] N. Ekbote, P. Dhanshetti, and S. Sakhrekar, "Techniques of exploratory data analysis," *Madhya Pradesh J. Soc. Sci.*, vol. 28, no. 2, pp. 10–14, 2023.
- [15] A. H. Lubis and E. Ramayana, "A review on appropriateness of partitional clustering algorithms in handling transactional data," *Int J Res Rev*, vol. 10, no. 9, pp. 162–169, 2023.