JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Implementasi Algoritma *K-means Clustering* untuk Mencari Preferensi Pelanggan Toko Online Tazeee *Clothes*

Ikram Ramadhan Effendy[1], Yehezkiel Albert Manurung[2], Muhamad Imam Hujaifah[3]

[1,2,3]Teknologi Informasi, Unversitas Bina Sarana Informatika, Depok, Indonesia
Email: [1]17230332@bsi.ac.id, [2]17230336@bsi.ac.id, [3]17230356@bsi.ac.id
Email Penulis Korespondensi: [1]17230332@bsi.ac.id

Abstrak−Pertumbuhan transaksi pada platform e-commerce mendorong pelaku bisnis untuk memahami preferensi pelanggan secara lebih tepat. Tazeee Clothes, sebagai toko fashion berbasis online, memiliki volume data penjualan yang besar namun belum dimanfaatkan secara maksimal dalam melakukan segmentasi produk. Penelitian ini bertujuan mengelompokkan produk berdasarkan performa penjualannya untuk mengidentifikasi pola pembelian pelanggan. Metode yang diterapkan adalah algoritma K-Means Clustering dengan pendekatan kuantitatif menggunakan data penjualan Januari–Desember 2024, meliputi variabel harga, jumlah pesanan, dan total berat produk. Proses pengelompokan dilakukan melalui iterasi hingga diperoleh centroid yang konvergen, kemudian divalidasi menggunakan Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Hasil analisis menghasilkan tiga klaster utama, yaitu produk dengan penjualan tinggi, penjualan sedang, dan penjualan rendah. Nilai evaluasi Silhouette Score 0.5782, Davies-Bouldin Index 0.4361, dan Calinski-Harabasz Index 3450.61 mengindikasikan kualitas klaster yang baik. Temuan ini memberikan landasan strategis bagi Tazeee Clothes dalam menyusun keputusan pemasaran, manajemen stok, serta pengembangan produk.

Kata Kunci: K-Means, Clustering, Segmentasi Pelanggan, Data Mining, Penjualan Produk.

Abstract− The rapid growth of e-commerce transactions encourages businesses to better understand customer preferences. Tazeee Clothes, as an online-based fashion store, has a large volume of sales data that has not been fully utilized for product segmentation. This study aims to group products based on their sales performance in order to identify customer purchasing patterns. The applied method is the K-Means Clustering algorithm with a quantitative approach using sales data from January to December 2024, including variables such as price, number of orders, and total product weight. The clustering process was performed iteratively until the centroids converged, followed by validation using Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. The results indicate three main clusters: products with high, medium, and low sales performance. The evaluation values—Silhouette Score of 0.5782, Davies-Bouldin Index of 0.4361, and Calinski-Harabasz Index of 3450.61—demonstrate good clustering quality. These findings provide a strategic foundation for Tazeee Clothes in making marketing decisions, inventory management, and product development.

Keywords: K-Means, Clustering, Customer Segmentation, Data Mining, Product Sales.

1. PENDAHULUAN

Kemajuan teknologi digital telah mendorong transformasi mendalam dalam pola interaksi konsumen dengan layanan perdagangan, khususnya pada ekosistem e-commerce yang berkembang pesat [1]. Platform belanja daring kini tidak hanya berfungsi sebagai media transaksi, tetapi juga berevolusi menjadi pusat penyimpanan informasi strategis yang mencatat beragam aktivitas pengguna, mulai dari pola pencarian produk, intensitas kunjungan, respons terhadap promosi, hingga rekam jejak pembelian secara keseluruhan. Beragam data ini menjadi sumber penting bagi perusahaan untuk memahami kecenderungan perilaku pelanggan serta dinamika permintaan pasar. Melalui proses segmentasi, perusahaan dapat mengelompokkan pelanggan berdasarkan kesamaan karakteristik pada variabel tertentu sehingga perilaku konsumen dapat dianalisis secara lebih sistematis, terukur, dan mudah diprediksi [2]. Pemahaman yang lebih tajam mengenai hasil segmentasi tersebut memungkinkan pelaku bisnis merancang keputusan yang lebih tepat dalam bidang pemasaran, pengelolaan produk, maupun strategi layanan.

Dalam konteks Tazeee Clothes sebagai usaha fashion berbasis online, tantangan utama terletak pada kemampuan perusahaan dalam mengolah dan menafsirkan akumulasi data pelanggan yang terus meningkat. Pertumbuhan transaksi, keberagaman produk, serta intensitas aktivitas pengguna membuat struktur data semakin kompleks dan sulit diinterpretasikan tanpa pendekatan analitis yang memadai. Tanpa dukungan analitik yang kuat, strategi pemasaran berisiko menjadi terlalu generik dan tidak mampu menyesuaikan kebutuhan tiap segmen pelanggan secara khusus. Kondisi ini dapat menurunkan efektivitas kampanye promosi, memicu ketidaktepatan dalam pengelolaan persediaan, dan berdampak pada penurunan kualitas pengalaman pelanggan. Oleh sebab itu, penggunaan metode analitik yang dapat menggali pola tersembunyi dalam data menjadi kebutuhan penting agar keputusan bisnis dapat diambil secara berbasis bukti. Permasalahan utama dalam penelitian ini adalah bagaimana membangun struktur kluster pelanggan berdasarkan kesamaan pola belanja dan preferensi mereka, mengingat data transaksi tidak secara langsung menunjukkan batasan antar segmen. Diperlukan pendekatan analitik untuk mengekstraksi pola laten sehingga perusahaan mampu mengidentifikasi kelompok pelanggan yang benar-benar berbeda dari segi perilaku maupun kontribusi bisnisnya [3]. Tantangan ini semakin signifikan mengingat tingginya tingkat persaingan di industri fashion online, yang menuntut strategi yang lebih fokus, efisien, dan mampu memberikan nilai diferensiasi. Ketidaktepatan

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

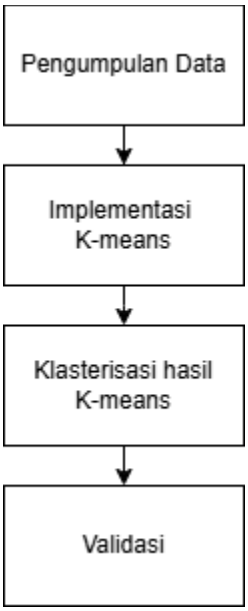
segmentasi dapat mengakibatkan rekomendasi produk yang kurang relevan, efektivitas promosi yang menurun, hingga berkurangnya retensi pelanggan karena layanan tidak sesuai preferensi pengguna.

Penggunaan algoritma K-Means memberikan pendekatan komputasional yang efisien dalam mengelompokkan pelanggan berdasarkan kemiripan atribut. Melalui metode ini, Tazeee Clothes dapat mengidentifikasi kelompok pelanggan dengan karakteristik yang beragam, seperti konsumen dengan frekuensi pembelian tinggi, pelanggan yang sensitif terhadap harga, pengguna yang loyal pada kategori produk tertentu, atau mereka yang hanya bertransaksi pada periode tertentu. Selain itu, metode ini juga dapat diterapkan untuk mengelompokkan produk berdasarkan performa penjualannya tinggi, menengah, maupun rendah yang dapat menjadi dasar evaluasi portofolio produk [4]. Temuan ini tidak hanya memberikan gambaran yang lebih jelas mengenai struktur pelanggan, tetapi juga mendukung perusahaan dalam merancang pendekatan pemasaran yang lebih personal, menyusun strategi penawaran yang relevan, serta meningkatkan kualitas pengalaman pengguna. Dengan segmentasi yang lebih akurat, penelitian ini diharapkan memberikan kontribusi praktis bagi Tazeee Clothes dalam meningkatkan efisiensi operasional, mempertahankan pelanggan bernilai tinggi, dan memperkuat daya saing perusahaan di pasar e-commerce yang semakin kompetitif.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan data yang memiliki karakteristik kategorikal. K-Means merupakan algoritma penting dalam bidang data mining dan banyak digunakan karena prosedur implementasinya yang sederhana serta efisiensi kinerjanya [5]. Pendekatan ini kerap dipilih karena mampu memberikan proses komputasi yang cepat dan beradaptasi dengan berbagai variasi dataset, sehingga mendukung pembentukan kluster yang lebih stabil dan konsisten [6]. Metode penelitian menggunakan pendekatan kuantitatif dengan memanfaatkan data berbentuk numerik yang dapat diukur secara objektif [7]. Data kuantitatif memungkinkan pelaksanaan perhitungan, perbandingan, dan analisis statistik untuk memperoleh pemahaman mengenai parameter seperti besaran, durasi, frekuensi, maupun nilai-nilai numerik lain yang relevan untuk dianalisis secara matematis. Secara keseluruhan, rangkaian tahapan algoritma K-Means dalam penelitian ini mengikuti proses yang ditunjukkan pada Gambar 1.



Gambar 1. Alur metode penelitian

Tahapan penelitian dengan pendekatan kuantitatif pada Gambar 1 dapat dijelaskan sebagai berikut:

1) Pengumpulan Data
   Dataset dibangun dari riwayat transaksi pembelian pelanggan pada platform Tazeee *Clothes*. Informasi yang dihimpun meliputi nama produk, harga, jumlah pesanan, serta total berat setiap produk.
2) Implementasi K-means
   Pada tahap ini, algoritma K-Means Clustering diterapkan untuk mengelompokkan data ke dalam sejumlah kluster berdasarkan tingkat kemiripannya. Setiap record dialokasikan ke kluster dengan centroid yang memiliki jarak minimum.
3) Proses Klasterisasi

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Fase klasterisasi merupakan tahap akhir dari algoritma, di mana data terbagi ke dalam kluster yang stabil sesuai pola kesamaannya. Setelah iterasi mencapai konvergensi dan perubahan centroid tidak lagi signifikan, setiap data memperoleh label kluster secara final.

4) Validasi Hasil Klasterisasi
Pengujian kualitas kluster dilakukan untuk menilai sejauh mana K-Means menghasilkan pengelompokan yang representatif. Karena K-Means merupakan metode *unsupervised* tanpa label awal, evaluasi dilakukan melalui validasi internal berbasis pola jarak antardata. Penelitian ini memanfaatkan tiga metrik utama *Silhouette Score*, *Calinski Harabasz Index*, dan *Davies Bouldin Index* untuk mengukur tingkat kepadatan dan keterpisahan kluster yang terbentuk.

2.2 Sumber Data

Data yang digunakan dalam penelitian ini bersumber dari riwayat penjualan pada platform e-commerce Tazeee Clothes. Dataset mencakup transaksi yang berlangsung dari Januari hingga Desember 2024 dan berisi beberapa atribut utama, meliputi nama produk, harga, jumlah pesanan, serta total berat produk. Secara keseluruhan, terdapat 639 entri dengan tiga variabel numerik utama yang menjadi fokus analisis. Struktur lengkap dari dataset tersebut disajikan pada Tabel 1.

Tabel 1 Data Penjualan

Nomor	Nama Produk	Harga	Jumlah Dipesan	Total berat (gr)
1	atasan cewek merk remix jeans	30000	1	200
2	kaos anak kecil merk CROCODILE	25000	2	500
3	kemeja lengan pendek cowok	25000	2	500
4	kemeja lengan panjang cowok	30000	1	300
5	blazer cewek	30000	1	300
6	kemeja lengan panjang cewek	30000	4	1000
7	celana panjang cewek	30000	4	1000
…	…	…	…	…
637	Baju dan Celana merk ZARA BASIC	75000	1	1000
638	rok katun cewek merk PRISCILLA	30000	2	600
639	Rok cewek merk PER UNA	30000	2	600

2.3 *Clustering*

Clustering dan klasifikasi merupakan pendekatan yang digunakan untuk mengorganisasi data ke dalam sejumlah kategori berdasarkan kemiripan karakteristik yang telah ditetapkan sebelumnya [8]. Prosedur *clustering* bertujuan memetakan himpunan data atau sinyal ke dalam kelompok tertentu melalui proses pengaitan hipotesis indeks secara terstruktur. Dengan demikian, setiap kluster yang dihasilkan mencerminkan keseragaman karakteristik internal sekaligus menunjukkan perbedaan yang tegas dibandingkan dengan kluster lainnya.

2.4. Implementasi Algoritma K-means

Algoritma K-Means merupakan teknik pengelompokan non-hierarki (partisi) yang dikembangkan untuk membagi data ke dalam dua atau lebih kluster. Metode ini bekerja dengan mengelompokkan data yang memiliki karakteristik serupa ke dalam satu kluster, sementara data dengan sifat berbeda dipisahkan ke kluster lainnya. Inti dari algoritma ini adalah meminimalkan fungsi objektif dengan menekan variasi internal kluster dan memperbesar jarak antar kluster [9]. Penggunaannya terbatas pada data dengan atribut numerik, sehingga setiap variabel harus dapat dinyatakan dalam bentuk nilai yang dapat diolah secara matematis [10]. Dalam analisis penjualan, K-Means berperan dalam mengidentifikasi kategori produk yang termasuk tingkat penjualan tinggi, sedang, maupun rendah [11].

Proses klasterisasi pada K-Means berlangsung secara iteratif, dimulai dari pemilihan centroid awal secara acak. Setelah itu, jarak antara setiap data dan masing-masing centroid dihitung untuk menentukan kedekatan relatif terhadap kluster tertentu. Ketika seluruh data telah dialokasikan, posisi centroid diperbarui berdasarkan nilai rata-rata anggota kluster. Tahapan ini terus diulang sampai perubahan posisi centroid tidak lagi signifikan sebagai tanda bahwa kondisi konvergen telah dicapai. Berkat struktur langkah kerja yang sederhana dan efisiensi komputasi yang tinggi, K-Means menjadi salah satu pendekatan klasterisasi yang paling banyak digunakan dalam analisis data modern. Adapun tahapan implementasi K-Means meliputi:

1. Penentuan Jumlah Kluster (K)
   Tahap awal menetapkan jumlah kluster K yang diperlukan sesuai tujuan analisis.
2. Inisialisasi Pusat Kluster
   Pusat kluster dapat ditentukan melalui berbagai metode, namun pendekatan yang paling umum adalah pemilihan nilai awal centroid secara acak.
3. Perhitungan Jarak Menggunakan Euclidean Distance

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Setelah centroid awal ditetapkan, jarak setiap data terhadap seluruh centroid dihitung menggunakan Euclidean Distance sebagaimana ditunjukkan pada persamaan (1). Langkah ini menentukan kluster terdekat bagi setiap data.

$$d\,(x_i, y_j) = \sqrt{\sum_{i=1}^{n} (x_i - y_j)^2} \quad (1)$$

Keterangan :
d : jarak data ke pusat kluster
x_i : objek ke-i
y_i : *centroid* ke-j
n : jumlah objek

4. Pengelompokan Data ke Kluster
Setiap data dialokasikan ke kluster dengan jarak minimum terhadap centroid sebagaimana ditunjukkan pada persamaan (2).

$$Cluster(x_i) = arg\ min\ j\ \ d(x_i, y_j) \quad (2)$$

5. Pembaruan Centroid
Setelah data terkelompokkan, centroid diperbarui berdasarkan nilai rata-rata seluruh data dalam satu kluster sesuai persamaan (3).

$$y_j(t+1) = \frac{1}{NS_j} \sum_{j \in S_j} x_j \quad (3)$$

Keterangan :
$y_j(t+1)$: *centroid* baru pada iterasi ke-1
NSj : banyak data pada kluster S_j

6. Iterasi Hingga Konvergensi
Langkah 3 hingga 5 diulangi hingga perubahan centroid tidak lagi signifikan atau batas maksimum iterasi tercapai [12].

2.5 Alat yang digunakan

Penerapan metode clustering dalam penelitian ini memanfaatkan berbagai perangkat dan teknologi untuk memastikan proses analisis berjalan efektif serta menghasilkan keluaran yang presisi. Pengolahan dan analisis data dilakukan menggunakan JavaScript. Pada tahap pembentukan kelompok, algoritma K-Means digunakan dengan jumlah klaster yang ditentukan melalui pendekatan kuantitatif. Setelah proses klasterisasi selesai, hasilnya divisualisasikan melalui scatter plot dua dan tiga dimensi untuk memperlihatkan pola persebaran tiap klaster. Selain itu, kualitas pengelompokan dievaluasi menggunakan Silhouette Score, Davies Bouldin Index, dan Calinski Harabasz Index, kemudian dikombinasikan dengan visualisasi scatter plot 2D dan 3D guna memperkuat interpretasi struktur klaster. Pendekatan implementatif ini juga telah diterapkan secara luas pada riset-riset sebelumnya terkait analisis pengelompokan data. Dengan metode tersebut, penelitian ini mampu menyajikan pemahaman yang lebih menyeluruh mengenai preferensi pelanggan berdasarkan hasil analisis menggunakan algoritma K-Means.

2.6 *Calinski–Harabasz Indeks*

Indeks Calinski–Harabasz menilai mutu pengelompokan dengan mengukur perbandingan antara variasi antar klaster (SSB), yang menunjukkan tingkat keterpisahan, dan variasi dalam klaster (SSW), yang merepresentasikan tingkat kohesi. Perbandingan ini kemudian disesuaikan melalui faktor normalisasi yang diperoleh dari selisih jumlah observasi dan jumlah klaster, dibagi dengan jumlah klaster yang dikurangi satu. Semakin tinggi skor CH, semakin kuat indikasi bahwa struktur klaster yang terbentuk memiliki pemisahan yang lebih tegas dan konsistensi internal yang lebih stabil [13].

2.7 *Davies-Bouldin Index*

Davies–Bouldin Index (DBI) merupakan indikator yang digunakan untuk menilai mutu hasil pengelompokan dengan mengukur tingkat konsistensi internal suatu kluster serta tingkat pemisahannya dari kluster lain [14]. Perhitungan indeks ini didasarkan pada rasio antara rerata jarak setiap elemen dalam kluster terhadap centroid nya dan jarak antar centroid. Nilai DBI yang lebih rendah mendekati nol mengindikasikan kualitas kluster yang lebih baik, karena menunjukkan bahwa data dalam kluster berada dalam struktur yang lebih kompak dan antar kluster memiliki pemisahan yang jelas.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

2.8 *Silhoutte Score*

Silhouette Score merupakan indikator yang digunakan untuk mengevaluasi kedekatan suatu data terhadap klaster tempatnya berada dibandingkan dengan kedekatannya terhadap klaster lain [15]. Metrik ini membantu menilai ketepatan penempatan data di dalam klusternya serta tingkat keterpisahannya dari klaster berbeda. Nilai *Silhouette Score* berada pada rentang −1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa data sesuai dengan klaster asal dan memiliki pemisahan yang jelas dari klaster lain. Nilai mendekati 0 menandakan posisi data berada di daerah batas antarklaster, sedangkan nilai negatif mengindikasikan potensi mis-klasifikasi karena objek lebih dekat dengan klaster berbeda. Oleh karena itu, *Silhouette Score* digunakan untuk menilai kesesuaian jumlah klaster yang dipilih dan kualitas struktur klaster secara keseluruhan. Nilai *Silhouette Coefficient* (s) untuk tiap objek dihitung secara individual sesuai persamaan (4).

$$s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (4)$$

Keterangan :

$a(i)$ = Jarak rata-rata antara data i dengan semua data lain yang berada dalam kluster yang sama

$b(i)$ = Jarak rata-rata antara data i dengan data-data yang berada pada kluster terdekat (kluster lain yang paling dekat jaraknya)

3. HASIL DAN PEMBAHASAN

3.1 Data Hasil Iterasi

Setelah proses penerapan algoritma k-means sebagaimana diuraikan pada Bab 2, analisis menghasilkan data akhir setelah melalui 13 kali iterasi, yang dirangkum pada Tabel 2 serta divisualisasikan pada Gambar 2.

Tabel 2 Data Hasil Iterasi

Nomor	Nama Produk	Jarak C1	Jarak C2	Jarak C3	Cluster
1	atasan cewek merk remix jeans	24057.67	970000.33	4610.08	C3
2	kaos anak kecil merk CROCODILE	29052.18	975000.13	515.44	C3
3	kemeja lengan pendek cowok	29052.18	975000.13	515.44	C3
4	kemeja lengan panjang cowok	24055.51	970000.25	4598.40	C3
5	blazer cewek	24055.51	970000.25	4598.40	C3
6	kemeja lengan panjang cewek	24052.03	970000.00	4577.36	C3
7	celana panjang cewek	24052.03	970000.00	4577.36	C3
…	…	…	…	…	…
637	Baju dan Celana merk ZARA BASIC	20950.34	925000.00	49572.90	C1
638	rok katun cewek merk PRISCILLA	24051.52	970000.08	4576.31	C3
639	Rok cewek merk PER UNA	24051.52	970000.08	4576.31	C3



Gambar 2. Hasil Iterasi ke-13

3.2 Hasil Implementasi Algoritma *K-Means Clustering*

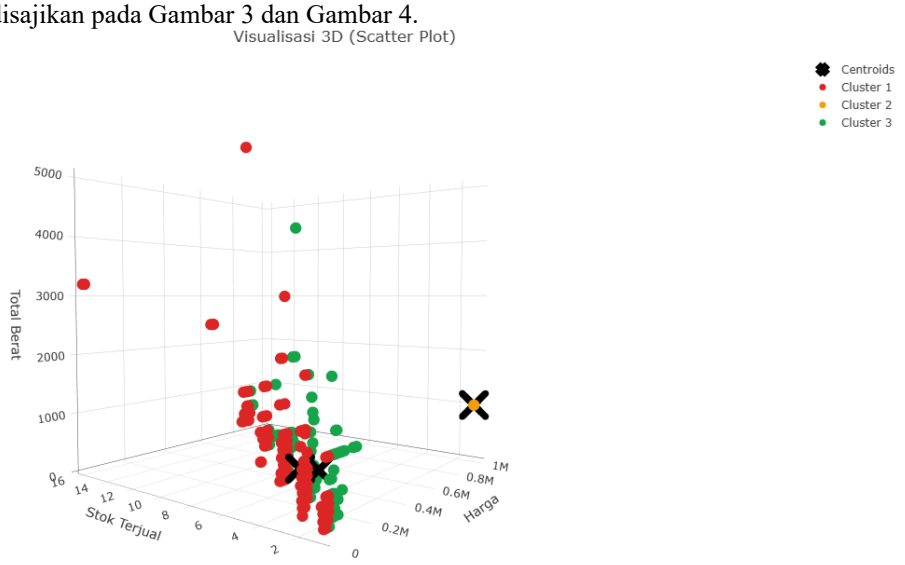
Visualisasi hasil klasterisasi ditampilkan dalam bentuk scatter plot 2D dan 3D dengan penggunaan warna kontras yang berbeda pada setiap klaster untuk memperjelas batas pemisahan antar kelompok data. Klaster pertama ditandai dengan warna merah, klaster kedua dengan warna kuning, dan klaster ketiga dengan warna hijau. Perbedaan warna yang tegas ini bertujuan untuk membantu pembaca mengidentifikasi pola sebaran data serta tingkat keterpisahan antar klaster

This is an open access article under the CC–BY-SA license

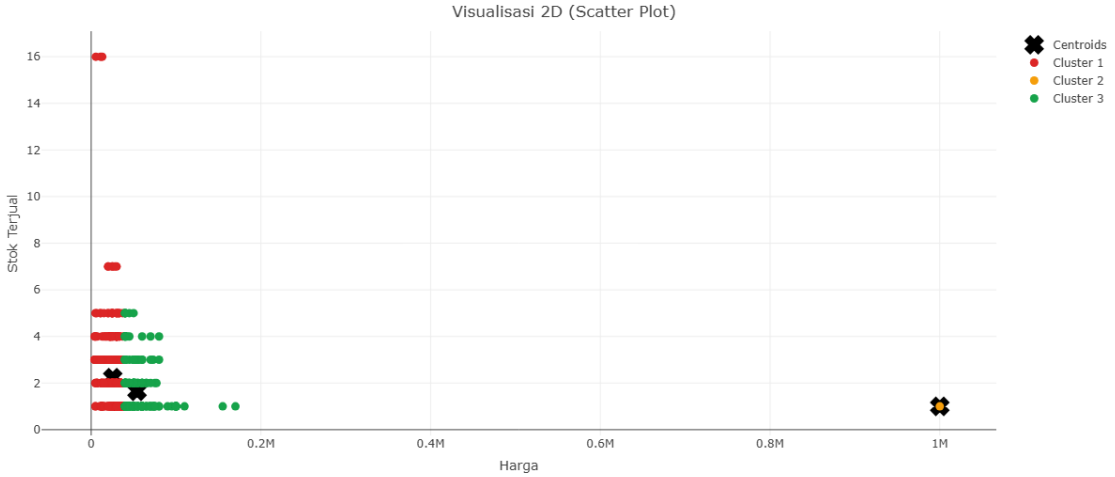
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

secara lebih intuitif. Berdasarkan visualisasi tersebut, terlihat bahwa masing-masing klaster membentuk kelompok yang relatif terpisah, yang mengindikasikan bahwa algoritma K-Means mampu mengelompokkan data penjualan produk dengan baik, sebagaimana disajikan pada Gambar 3 dan Gambar 4.



Gambar 3. Visualisasi Scatter Plot 3D hasil *Clustering*



Gambar 4. Visualisasi Scatter Plot 2D hasil *Clustering*

Metode K-Means digunakan dengan penentuan jumlah klaster optimal sebanyak tiga kelompok. Setiap klaster merepresentasikan pola karakteristik produk berdasarkan variabel harga, jumlah pemesanan, dan total berat (g).

a) Klaster 1 (Merah) terdiri dari 421 produk yang tergolong sebagai item dengan tingkat penjualan tertinggi.
b) Klaster 2 (Kuning) mencakup 1 produk yang memiliki performa penjualan relatif rendah dibandingkan kelompok lainnya.
c) Klaster 3 (Hijau) memuat 216 produk menunjukkan tingkat penjualan pada kategori normal pada toko online Tazeee Clothes.

3.3 Evaluasi dan Validasi Hasil *Clustering*

Kualitas klasterisasi dievaluasi menggunakan tiga metrik yang saling melengkapi, yakni *Silhouette Score* untuk menilai kedekatan titik data dalam kluster yang sama, *Davies-Bouldin Index* untuk mengukur tingkat kemiripan antar kluster, serta *Calinski-Harabasz Index* untuk menilai kepadatan dan keterpisahan kluster. Ketiga indikator tersebut dimanfaatkan berdasarkan hasil visualisasi yang disajikan pada Gambar 4**.**

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 1 Februari 2026, Page 127-134
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



EVALUASI KUALITAS CLUSTERING

Analisis validitas cluster menggunakan metode Silhouette, DBI, dan CHI.

Metode Evaluasi	Skor	Interpretasi
Silhouette Score Rentang: -1 s.d 1	0.5782	**Bagus.** *Semakin mendekati 1, semakin baik pemisahannya.*
Davies-Bouldin (DBI) Rentang: 0 s.d ∞	0.4361	**Bagus (Padat).** *Semakin KECIL nilainya, semakin baik clusternya.*
Calinski-Harabasz (CHI) Rentang: 0 s.d ∞	3450.61	**Higher is Better.** *Nilai tinggi menandakan cluster padat & terpisah jelas.*

Gambar 5. Evaluasi Kualitas *Clustering*

Jika dibandingkan dengan penelitian sebelumnya, hasil klasterisasi pada penelitian ini menunjukkan performa yang kompetitif. Penelitian Pamungkas dan Februariyanti [1] melaporkan nilai Silhouette Score sebesar 0.52 pada pengelompokan data e-commerce Lazada, sementara penelitian Wahyuni et al. [3] memperoleh nilai Davies-Bouldin Index sebesar 0.48 pada segmentasi pelanggan berbasis RFM. Dalam penelitian ini, nilai Silhouette Score sebesar 0.5782 dan Davies-Bouldin Index sebesar 0.4361 menunjukkan kualitas klaster yang lebih baik atau setidaknya setara dibandingkan penelitian terdahulu. Hal ini mengindikasikan bahwa pemilihan variabel harga, jumlah pesanan, dan total berat produk cukup efektif dalam membentuk struktur klaster yang jelas pada kasus penjualan Tazeee Clothes.

Dengan demikian, penerapan algoritma K-Means pada penelitian ini tidak hanya konsisten dengan temuan penelitian sebelumnya, tetapi juga memperkuat bukti bahwa metode ini efektif digunakan untuk analisis segmentasi produk pada sektor fashion berbasis e-commerce.

4. KESIMPULAN

Secara komprehensif, hasil evaluasi klasterisasi yang diperoleh menunjukkan bahwa model pengelompokan yang dibangun memiliki tingkat akurasi dan stabilitas yang mampu diterapkan dalam konteks pengambilan keputusan strategi. Nilai Silhouette Score yang mencapai 0.5782 mengindikasikan bahwa sebagian besar data telah terkelompok dengan baik, di mana jarak antar data dalam klaster relatif lebih dekat dibandingkan jaraknya dengan klaster lain. Kondisi ini mencerminkan adanya pola yang konsisten dalam perilaku penjualan produk. Selain itu, nilai Davies-Bouldin Index sebesar 0.4361 yang tergolong rendah menunjukkan bahwa tingkat kesamaan antar klaster berada pada level minimal, sehingga masing-masing klaster mampu merepresentasikan karakteristik produk yang berbeda secara jelas dan tidak saling tumpang tindih. Nilai Calinski-Harabasz Index yang mencapai 3450.61 semakin memperkuat kualitas hasil klasterisasi karena menunjukkan perhitungan yang tinggi antara variasi antar klaster dan variasi di dalam.

Berdasarkan temuan tersebut, dapat disimpulkan bahwa penerapan algoritma K-Means pada data penjualan Tazeee Clothes efektif dalam mengidentifikasi segmentasi produk yang relevan dan bermakna. Informasi klaster yang dihasilkan dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran yang lebih terarah, seperti penentuan prioritas promosi pada kelompok produk tertentu serta penyesuaian strategi harga sesuai dengan karakteristik masing-masing klaster. Selain itu, hasil klasterisasi ini juga berpotensi meningkatkan efisiensi pengelolaan persediaan dengan membantu perusahaan dalam memprediksi kebutuhan stok secara lebih akurat. Dengan pendekatan berbasis data ini, Tazeee Clothes diharapkan mampu mengembangkan kebijakan penjualan yang lebih adaptif dan berkelanjutan, sekaligus meningkatkan daya saing perusahaan di tengah persaingan industri.

UCAPAN TERIMAKASIH

Kami segenap penulis mengucapkan terima kasih kepada Tazeee Clothes telah memberikan ketersediaan data selama penelitian ini berlangsung.

REFERENCES

[1] M. D. Pamungkas and H. Februariyanti, “Penerapan Algoritma K-Means Clustering Untuk Mengelompokan Data Review Barang Pada E-Commerce Lazada,” *semanTIK*, vol. 8, no. 2, p. 99, 2022, doi: 10.55679/semantik.v8i2.29058.

[2] D. A. Imanuel and G. Alfian, “Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 283–292, 2025, doi: 10.25126/jtiik.2025128619.

[3] S. Wahyuni, T. T. Wulansari, and F. Fahrullah, “Segmentasi Pelanggan Berdasarkan Analisis Recency,

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**

Frequency, Monetary Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group," *J. Rekayasa Teknol. Inf.*, vol. 7, no. 2, p. 180, 2023, doi: 10.30872/jurti.v7i2.8748.

[4] N. S. Niko, A. Rahman, D. Marini Umi Atmaja, and A. Basri, "Klasterisasi Stok Produk Retail Untuk Menetukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 306–312, 2023, doi: 10.47065/bit.v4i3.736.

[5] I. Pii, N. Suarna, and N. Rahaningsih, "Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 423–430, 2023, doi: 10.36040/jati.v7i1.6336.

[6] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.

[7] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.

[8] S. A. Rahmah, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2020.

[9] A. Yahya1* and R. Kurniawan2, "Implementation of K-Means Algorithm for Clustering Sales Data Based on Sales Patterns," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 05, no. 1, pp. 350–358, 2025.

[10] F. Hadi and Y. Diana, "Pengklusteran Penjualan Bahan Bangunan Menggunakan Algoritma K-Means," *JOISIE (Journal Inf. Syst. Informatics Eng.*, vol. 4, no. 1, p. 22, 2020, doi: 10.35145/joisie.v4i1.629.

[11] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1–7, 2022.

[12] N. Noviati, M. Mulyawan, D. A. Kurnia, and A. R. Rinaldi, "Clustering Data Penjualan Produk Makanan pada Toko Toserba Yogya Siliwangi dengan Menggunakan Metode K-Means," *MEANS (Media Inf. Anal. dan Sist.*, vol. 7, no. 1, pp. 77–84, 2022, doi: 10.54367/means.v7i1.1850.

[13] P. Algoritma *et al.*, "Comparison of K - Means and K - Medoids Algorithms in Clustering Islamic Commercial Banks in Indonesia Using Davies - Bouldin Index , Calinski - Harabasz Index , and Silhouette Coefficient sektor perbankan syariah . Bank (ROA), yang biasanya disebut Ind," *J. Kelitbangan*, vol. 13, no. 1, pp. 1–12, 2022.

[14] B. Kristanto, A. Turmudi Zy, and M. Fatchan, "Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 112–120, 2023, doi: 10.47065/bit.v4i1.521.

[15] R. Anggari, M. Ifandi, A. Firdaus, M. Wati, and H. Haviluddin, "Penerapan Algoritma K-Means untuk Pengelompokan Negara di Dunia Berdasarkan Indikator Ekonomi," *J. Rekayasa Teknol. Inf.*, vol. 8, no. 2, p. 195, 2024, doi: 10.30872/jurti.v8i2.19745.

This is an open access article under the CC–BY-SA license

Submitted: **19/12/2025**; Accepted: **05/01/2026**; Published: **15/01/2026**