

Evaluasi Klasterisasi Produk Ritel Berdasarkan Kategori dan Harga Menggunakan Algoritma K-Means

Maha Valne Datin Mahfujah Tambunan¹, Siska Mayasari Rambe^{2*}, Rahma Syahri³

^{1,2,3} Program Studi Magister Teknologi Informasi, Universitas Pembangunan Panca Budi Medan

ABSTRAK

Persaingan bisnis ritel yang kompetitif menuntut strategi pemasaran yang tepat berdasarkan karakteristik produk. Variasi produk dengan kategori dan harga berbeda membuat pengelompokan manual kurang efektif dan memakan waktu. Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan produk ritel berdasarkan kategori dan harga eceran sebagai pendukung strategi pemasaran. Penelitian dilakukan di Supermarket Indobali, Tanjung Balai Karimun, menggunakan data harga jual produk tahun 2024 sebanyak 4.245 data dengan 73 kategori. Setelah pembersihan data, diperoleh 4.226 data yang diproses melalui preprocessing, Label Encoding, normalisasi Min-Max, penentuan jumlah cluster dengan Elbow Method, klasterisasi K-Means, serta evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index. Meskipun Silhouette Score tertinggi diperoleh pada K=2 (0,638), jumlah cluster ditetapkan tiga berdasarkan Elbow Method dan pertimbangan interpretasi praktis. Hasil menunjukkan K-Means berhasil mengelompokkan produk ke tiga cluster dengan Silhouette Score 0,578 dan Davies-Bouldin Index 0,534, mengindikasikan kualitas pengelompokan cukup baik. Pembentukan cluster lebih dipengaruhi Label Encoding kategori dibanding harga, sehingga pelabelan segmen ekonomis, menengah, dan premium bersifat interpretatif, bukan pemisahan rentang harga tegas, akibat tumpang tindih harga antar-cluster. Meski demikian, hasil pengelompokan tetap memberi informasi awal bermanfaat bagi pengelola ritel dalam menentukan strategi pemasaran, promosi, pengelolaan stok, dan prioritas produk.

ABSTRACT

Increasingly competitive retail business conditions require appropriate marketing strategies based on product characteristics. The variety of products across different categories and price levels makes manual product grouping inefficient and time-consuming. This study applies the K-Means Clustering algorithm to segment retail products based on product category and retail price to support marketing strategy formulation. The research was conducted at Supermarket Indobali, Tanjung Balai Karimun, using 2024 retail price data comprising 4,245 records across 73 categories. After data cleaning, 4,226 records were processed through preprocessing, Label Encoding of the category attribute, Min-Max normalization, determination of the optimal number of clusters via the Elbow Method, K-Means clustering, and evaluation using the Silhouette Score and Davies-Bouldin Index. Although the highest Silhouette Score was obtained at K=2 (0.638), three clusters were selected based on the Elbow Method and practical interpretability considerations. Results show that K-Means successfully segmented products into three clusters with a Silhouette Score of 0.578 and a Davies-Bouldin Index of 0.534, indicating reasonably good clustering quality. Cluster formation was more strongly influenced by the categorical Label Encoding results than by price, meaning the economical, moderate, and premium segment labels are interpretive rather than strict price-range divisions, due to price overlap across clusters. Nevertheless, the clustering results provide useful preliminary insights for retail managers in formulating marketing strategies, promotions, inventory management, and product prioritization.

Kata Kunci: K-Means Clustering; segmentasi produk ritel; Label Encoding; Elbow Method; Silhouette Score
Email: *mayasarisiska5@gmail.com

DOI: <http://dx.doi.org/10.30829/jistech.v11i1.31085>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong transformasi besar dalam berbagai sektor bisnis, termasuk industri ritel [1]. Digitalisasi proses bisnis memungkinkan perusahaan mengumpulkan data produk dan transaksi

dalam jumlah yang sangat besar setiap hari [2]. Data tersebut tidak hanya berfungsi sebagai arsip administratif, tetapi juga menjadi sumber informasi yang bernilai dalam mendukung pengambilan keputusan bisnis [3]. Melalui pemanfaatan teknik analisis data yang tepat, perusahaan dapat mengidentifikasi karakteristik produk serta peluang pemasaran yang sebelumnya sulit diketahui melalui analisis konvensional [4].

Salah satu jenis data yang paling banyak dihasilkan oleh sistem ritel adalah data produk yang memuat informasi seperti nama barang, kategori produk, harga, stok, maupun transaksi penjualan. Seiring bertambahnya jumlah produk yang dikelola, proses identifikasi karakteristik produk secara manual menjadi semakin sulit dilakukan. Produk dengan kategori yang berbeda dapat memiliki rentang harga yang hampir sama, sedangkan produk dalam kategori yang sama sering kali memiliki variasi harga yang cukup besar. Kondisi tersebut menyebabkan proses pengelompokan produk berdasarkan pengamatan manual menjadi kurang efisien dan berpotensi menghasilkan interpretasi yang subjektif [5]. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengelompokkan produk berdasarkan karakteristik yang dimiliki sehingga struktur data dapat dipahami secara lebih sistematis.

Industri ritel merupakan salah satu sektor yang memiliki tingkat persaingan tinggi karena menawarkan ribuan produk dengan kategori dan harga jual yang beragam [6]. Setiap produk memiliki tingkat harga dan posisi pasar yang berbeda-beda, sehingga perusahaan sering mengalami kesulitan dalam menentukan strategi pemasaran yang sesuai untuk setiap kelompok produk apabila proses analisis masih dilakukan secara manual [7]. Pengelompokan produk berdasarkan pengalaman atau intuisi sering kali menghasilkan keputusan yang kurang objektif sehingga berdampak pada efektivitas promosi, pengelolaan persediaan, maupun penempatan produk di area penjualan [8].

Salah satu pendekatan yang banyak digunakan dalam mengolah data berskala besar adalah data mining [9]. Data mining merupakan proses menemukan pola, hubungan, maupun informasi tersembunyi dari sekumpulan data yang berukuran besar sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan [10]. Di antara berbagai teknik data mining, Clustering menjadi salah satu metode yang efektif untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label kelas sebelumnya (*unsupervised learning*) [11].

Algoritma K-Means Clustering merupakan salah satu algoritma Clustering yang paling banyak digunakan karena memiliki proses komputasi yang relatif sederhana, cepat, dan mampu menangani data dalam jumlah besar [10]. Algoritma ini bekerja dengan cara membagi data ke dalam sejumlah cluster berdasarkan kedekatan jarak antar data terhadap pusat cluster (centroid) [12]. Setiap iterasi dilakukan hingga posisi centroid tidak lagi mengalami perubahan sehingga diperoleh kelompok data yang memiliki tingkat homogenitas tinggi di dalam cluster dan heterogenitas tinggi antar cluster [13].

Penelitian ini menggunakan dataset berupa daftar harga jual produk ritel yang terdiri atas 4.245 data produk dengan atribut nomor urut, nama produk, kategori produk, dan harga eceran, yang mencakup 73 kategori produk berbeda. Kebaruan penelitian ini terletak pada penerapan kombinasi atribut kategori produk melalui Label Encoding dan harga eceran melalui normalisasi Min-Max dalam proses Clustering, disertai interpretasi karakteristik setiap cluster sebagai dasar penyusunan strategi pemasaran. Validasi hasil Clustering menggunakan Elbow Method, Silhouette Score, dan Davies-Bouldin Index sehingga kualitas cluster dapat dievaluasi secara lebih menyeluruh [14]. Hasil pengelompokan yang diperoleh selanjutnya diinterpretasikan sebagai segmen produk yang dapat digunakan sebagai referensi pendukung dalam penyusunan kebijakan pemasaran ritel, khususnya dalam hal diferensiasi perlakuan promosi, pengelompokan stok, dan penetapan prioritas produk berdasarkan segmen harga.

Berbagai penelitian telah menerapkan algoritma K-Means untuk melakukan pengelompokan produk pada sektor ritel. Hidayat et al. (2023) dalam penelitian *Klasterisasi Stok Produk Retail Menggunakan Algoritma K-Means* menggunakan atribut stok produk untuk mengelompokkan pergerakan barang menjadi beberapa kategori. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu membantu proses pengelolaan persediaan produk secara lebih terstruktur. Namun, penelitian tersebut hanya menggunakan atribut numerik sehingga belum mengakomodasi atribut kategorikal, seperti kategori produk, yang umum dijumpai pada data ritel [15].

Penelitian lain dilakukan oleh Pratama et al. (2024) melalui penelitian *Analisis Produk Terlaris Menggunakan Metode K-Means Clustering*. Penelitian tersebut mengelompokkan produk berdasarkan data penjualan untuk mengidentifikasi kelompok produk terlaris dan kurang diminati. Hasil penelitian menunjukkan bahwa K-Means mampu membentuk kelompok produk sesuai karakteristik penjualan. Akan tetapi, penelitian tersebut masih menggunakan atribut numerik dan belum mengevaluasi penggunaan data campuran yang terdiri atas atribut kategorikal dan numerik maupun pengaruh representasi atribut kategorikal terhadap kualitas hasil klasterisasi [16].

Berdasarkan kedua penelitian tersebut, dapat diketahui bahwa penelitian klasterisasi produk ritel di Indonesia masih didominasi oleh penggunaan algoritma K-Means pada data numerik, seperti stok dan penjualan. Penelitian yang secara khusus mengevaluasi segmentasi produk menggunakan data campuran, yaitu atribut kategorikal (kategori produk) dan numerik (harga eceran), masih sangat terbatas. Selain itu, belum banyak penelitian yang mengkaji dampak penggunaan representasi atribut kategorikal terhadap kualitas hasil klasterisasi. Oleh karena itu, penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk mengelompokkan produk ritel berdasarkan kategori dan harga eceran sehingga dapat memberikan informasi yang mendukung penyusunan strategi pemasaran produk ritel secara lebih efektif, efisien, dan berbasis data.

2. Metodologi Penelitian

2.1 Tahapan Penelitian

Penelitian ini menerapkan pendekatan Knowledge Discovery in Database (KDD) sebagai kerangka kerja dalam proses pengolahan data hingga diperoleh informasi yang dapat digunakan sebagai dasar pengambilan keputusan [17]. Tahapan penelitian terdiri atas seleksi data (data selection), preprocessing, transformasi data, proses Clustering menggunakan algoritma K-Means, evaluasi hasil Clustering, serta interpretasi hasil sebagai rekomendasi strategi pemasaran produk ritel [18].

2.2 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data daftar harga jual (price list) produk ritel yang bersumber dari sistem pencatatan harga sebuah usaha ritel. Dataset terdiri atas 4.245 baris data dengan 4 atribut, yaitu nomor urut, nama barang, kategori produk, dan harga eceran, yang mencakup 73 kategori produk berbeda, seperti Battery, Biskuit, Parfum, Rambut Shampoo, dan lain sebagainya. Deskripsi atribut dataset ditunjukkan pada Tabel 1.

Tabel 1. Deskripsi Dataset

Atribut	Tipe Data	Keterangan
No	Integer	Nomor urut baris data
Item Barang	Kategori (Teks)	Nama produk yang dijual
Kategori	Kategori (Teks)	Kelompok/jenis produk (73 kategori)
Harga Eceran	Numerik	Harga jual eceran produk (dalam Rupiah)

Hasil data observasi sebelum diproses ditunjukkan pada Tabel 2.

Tabel 2. Data Observasi

No	Item Barang	Kategori	Harga Eceran (Rp)
1	ABC Baterai Biru Besar 2S	Battery	14.000
2	ABC Baterai Biru Kecil 4S	Battery	11.000
3	ABC Baterai Biru Kecil 4S	Battery	10.000
4	ABC Baterai Hijau Besar 2S	Battery	15.000
5	ABC Baterai Hijau Kecil 4S	Battery	9.500
...	...	...	...
4.245	Zwitsal Minyak Telon 60ml	Minyak Kesehatan	26.800

Berdasarkan hasil eksplorasi data, atribut Harga Eceran memiliki rentang nilai yang sangat lebar, yaitu dari Rp1.000 hingga Rp2.000.000, dengan rata-rata sebesar Rp26.494 dan nilai tengah (median) sebesar Rp15.500. Rentang nilai yang lebar tersebut mengindikasikan adanya sebaran harga yang tidak seragam antar kategori produk, sehingga diperlukan proses normalisasi sebelum data digunakan dalam proses Clustering.

2.3 Preprocessing Data

Tahap preprocessing bertujuan meningkatkan kualitas data sebelum dilakukan proses Clustering, meliputi pemeriksaan missing value, data tidak valid, data duplikat, transformasi Label Encoding, dan normalisasi Min-Max [19].

2.3.1 Pemeriksaan Missing Value, Data Tidak Valid, dan Data Duplikat

Pemeriksaan dilakukan terhadap atribut Kategori dan Harga Eceran untuk memastikan tidak terdapat nilai kosong. Berdasarkan hasil pemeriksaan, kedua atribut tersebut tidak memiliki missing value. Selanjutnya dilakukan pemeriksaan data dengan nilai harga yang tidak valid (Harga Eceran kurang dari atau sama dengan 0), dan ditemukan sebanyak 19 baris data yang tidak valid sehingga dikeluarkan dari dataset. Pemeriksaan data duplikat berdasarkan kombinasi nama produk, kategori, dan harga eceran tidak menemukan adanya baris yang terduplikasi. Dengan demikian, jumlah data akhir yang digunakan dalam proses Clustering adalah 4.226 baris data. Ringkasan hasil pemeriksaan ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pemeriksaan Kualitas Data

Pemeriksaan	Jumlah
Total data awal	4.245
Missing value (Kategori/Harga Eceran)	0
Data dengan harga tidak valid (≤ 0)	19
Data duplikat	0
Data valid digunakan untuk Clustering	4.226

2.3.2 Label Encoding

Algoritma K-Means hanya dapat memproses data numerik, sehingga atribut Kategori yang bertipe teks dikonversi menjadi kode numerik menggunakan Label Encoding [20]. Proses ini menghasilkan 73 kode kategori (0 sampai dengan 72) yang diurutkan secara alfabetis. Contoh hasil Label Encoding ditunjukkan pada Tabel 4.

Tabel 4. Hasil Label Encoding	
Kategori Produk	Kode
Alat Cukur	0
Bahan Kue	1
Battery	2
Biskuit	3
Biskuit Cake	4
...	...
Tissue Kering	72

2.3.3 Normalisasi Min-Max

Normalisasi dilakukan menggunakan metode Min-Max Normalization dengan persamaan:

$$x' = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

Normalisasi diterapkan pada atribut Kategori_Kode (hasil Label Encoding) dan Harga Eceran agar kedua atribut berada pada rentang 0 hingga 1, sehingga tidak terjadi dominasi atribut tertentu dalam pembentukan cluster akibat perbedaan skala nilai. Sebagai contoh, data produk "ABC Baterai Biru Besar 2S" dengan kategori Battery (kode = 2) dan harga eceran Rp14.000 dinormalisasi sebagai berikut.

$$\begin{aligned} \text{Kategori Kode}' &= \frac{(2 - 0)}{(72 - 0)} = 0,0278 \\ \text{Harga Eceran} &= \frac{(14.000 - 1.000)}{(2.000.000 - 1.000)} = 0,0065 \end{aligned}$$

Proses yang sama diterapkan pada seluruh 4.226 data produk sehingga seluruh nilai atribut berada pada rentang 0–1 dan siap digunakan pada tahap Clustering.

2.4 Algoritma K-Means Clustering

Algoritma K-Means merupakan metode unsupervised learning yang mengelompokkan data berdasarkan tingkat kemiripan karakteristik [21]. Tahapannya meliputi: (1) menentukan jumlah cluster K; (2) menentukan centroid awal menggunakan metode K-Means++; (3) menghitung jarak setiap data ke centroid menggunakan Euclidean Distance; (4) menempatkan setiap data pada cluster dengan jarak terkecil; (5) menghitung centroid baru; dan (6) mengulang proses hingga centroid konvergen.

Persamaan Euclidean Distance yang digunakan adalah:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2} \quad (2)$$

Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method melalui perhitungan Within Cluster Sum of Squares (WCSS):

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} (x - \mu_i)^2 \quad (3)$$

2.5 Evaluasi Clustering

Evaluasi kualitas Clustering dilakukan menggunakan dua metode. Pertama, Silhouette Score mengukur seberapa baik setiap data berada di dalam clusternya dengan rentang nilai -1 hingga 1, di mana nilai mendekati 1 menunjukkan kualitas yang baik. Kedua, Davies-Bouldin Index (DBI) mengukur tingkat pemisahan antar cluster, di mana nilai yang semakin kecil menunjukkan kualitas Clustering yang semakin baik.

3. Hasil dan Pembahasan

3.1 Hasil Preprocessing Data

Berdasarkan pemeriksaan yang dilakukan terhadap 4.245 data produk, tidak ditemukan missing value pada atribut Kategori maupun Harga Eceran. Ditemukan 19 baris data dengan harga tidak valid (Harga Eceran kurang dari atau sama dengan 0) yang kemudian dikeluarkan dari dataset, serta tidak ditemukan data duplikat, sehingga diperoleh 4.226 data valid yang digunakan pada tahap Clustering (lihat Tabel 3). Selanjutnya, atribut Kategori dikonversi menggunakan Label Encoding menjadi 73 kode kategori (lihat Tabel 4), dan seluruh atribut numerik (Kategori_Kode dan Harga Eceran) dinormalisasi menggunakan Min-Max Normalization sehingga berada pada rentang 0 hingga 1.

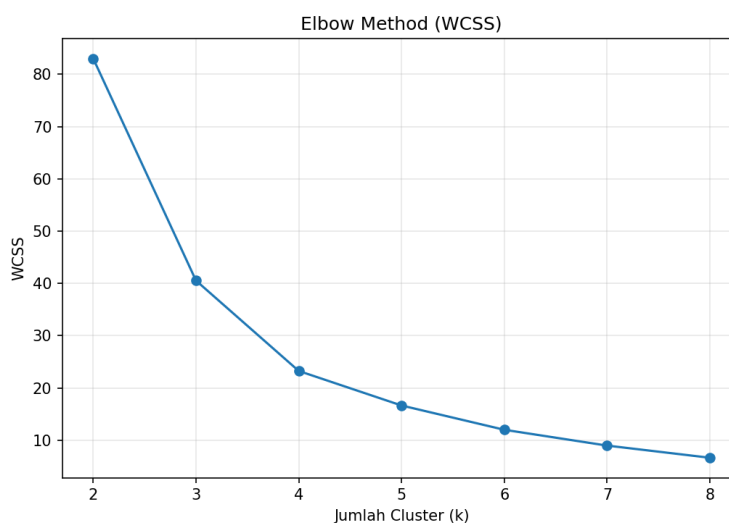
3.2 Penentuan Jumlah Cluster Optimal

Elbow Method diterapkan dengan menghitung nilai WCSS untuk K = 2 hingga K = 8. Hasil perhitungan ditunjukkan pada Tabel 5.

Tabel 5. Hasil Perhitungan WCSS, Silhouette Score, dan Davies-Bouldin Index

K	WCSS	Silhouette Score	Davies-Bouldin Index
2	83,03	0,638	0,496
3	40,54	0,578	0,534

4	23,24	0,594	0,504
5	16,63	0,587	0,506
6	11,99	0,577	0,510
7	8,96	0,612	0,496
8	6,63	0,616	0,477



Gambar 1. Grafik *Elbow Method* (WCSS)

Grafik *Elbow Method* pada Gambar 1 menunjukkan penurunan nilai WCSS yang tajam dari $K = 2$ ke $K = 3$, kemudian laju penurunan mulai melandai pada $K = 3$ hingga $K = 4$, sebelum kembali melandai secara bertahap hingga $K = 8$. Pola tersebut mengindikasikan bahwa titik siku (*elbow*) berada di sekitar $K = 3$ sampai $K = 4$.

Perlu dicatat bahwa nilai *Silhouette Score* tertinggi secara numerik diperoleh pada $K = 2$ (0,638). Meskipun demikian, penelitian ini menetapkan jumlah cluster optimal sebanyak $K = 3$ berdasarkan dua pertimbangan utama. Pertama, hasil *Elbow Method* menunjukkan titik siku yang paling signifikan pada $K = 3$, di mana penurunan WCSS dari $K = 2$ ke $K = 3$ (sebesar 42,49 poin) jauh lebih besar dibandingkan penurunan pada nilai K berikutnya. Kedua, segmentasi tiga tingkat harga yang terdiri atas segmen ekonomis, menengah, dan premium merupakan pendekatan yang lebih lazim dan aplikatif dalam praktik strategi pemasaran ritel dibandingkan segmentasi dua kelompok, sehingga memberikan granularitas interpretasi yang lebih bermakna bagi pengelola ritel. Nilai *Silhouette Score* sebesar 0,578 dan *Davies-Bouldin Index* sebesar 0,534 pada $K = 3$ masih berada pada kategori yang cukup baik, sebagaimana dibahas lebih lanjut pada Sub-bab 3.5.

3.3 Proses Clustering K-Means

Proses Clustering dilakukan menggunakan $K = 3$ dengan inisialisasi centroid menggunakan metode *K-Means++*. Nilai centroid akhir setelah proses konvergen (dalam data ternormalisasi) ditunjukkan pada Tabel 6.

Tabel 6. Nilai Centroid Akhir (Data Ternormalisasi)

Cluster	Kategori	Kode	Harga Eceran
Cluster 1	0,1721		0,0077
Cluster 2	0,5281		0,0132
Cluster 3	0,8145		0,0163

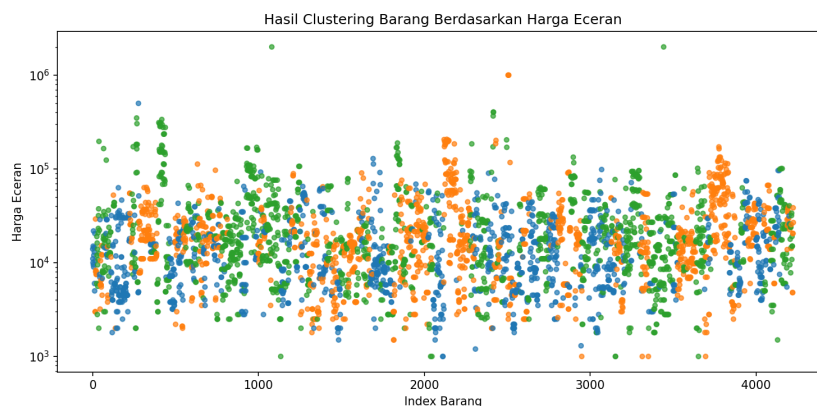
Sebagai ilustrasi proses perhitungan jarak, data produk "ABC Baterai Biru Besar 2S" (*Kategori_Kode* = 0,0278; *Harga Eceran* = 0,0065) dihitung jaraknya terhadap ketiga centroid menggunakan Euclidean Distance, dengan hasil: $d(\text{Cluster 1}) = 0,1443$; $d(\text{Cluster 2}) = 0,5004$; $d(\text{Cluster 3}) = 0,7868$. Karena nilai terkecil adalah 0,1443, maka data tersebut ditempatkan pada Cluster 1. Proses iterasi yang sama dilakukan terhadap seluruh 4.226 data hingga posisi centroid konvergen dan tidak terjadi perubahan keanggotaan cluster.

3.4 Hasil Pengelompokan K-Means

Setelah proses iterasi selesai dan centroid konvergen, seluruh 4.226 data produk berhasil dikelompokkan ke dalam 3 cluster dengan distribusi keanggotaan sebagaimana ditunjukkan pada Tabel 7.

Tabel 7. Distribusi Anggota dan Karakteristik Harga Setiap Cluster

Cluster	Segmen Harga	Jumlah Anggota	Persentase	Harga Min (Rp)	Harga Rata-rata (Rp)	Harga Maks (Rp)
Cluster 1	Ekonomis	1.231	29,13%	1.000	16.457	500.000
Cluster 2	Menengah	1.451	34,34%	1.000	27.439	1.000.000
Cluster 3	Premium	1.544	36,54%	1.000	33.609	2.000.000



Gambar 2. Sebaran Harga Eceran Produk Berdasarkan Hasil Clustering

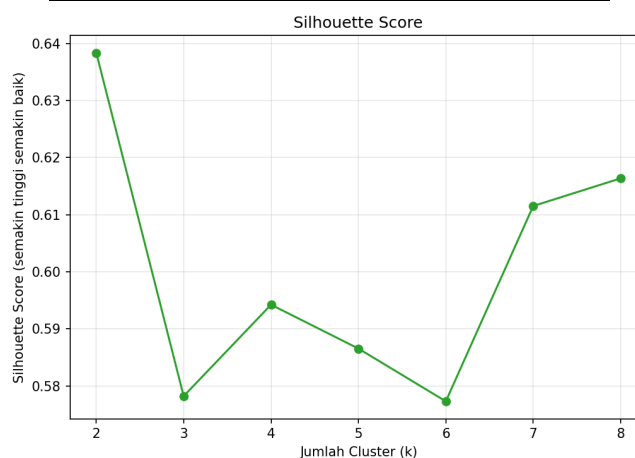
Cluster 1 (Ekonomis) terdiri atas 1.231 produk (29,13%) dengan harga rata-rata Rp16.457, merupakan kelompok produk kebutuhan harian dengan harga jual relatif rendah. Cluster 2 (Menengah) mencakup 1.451 produk (34,34%) dengan harga rata-rata Rp27.439. Cluster 3 (Premium) merupakan kelompok terbesar dengan 1.544 produk (36,54%) dan harga rata-rata tertinggi, yaitu Rp33.609, sekaligus mencakup produk dengan harga maksimum Rp2.000.000.

3.5 Evaluasi Kualitas Clustering

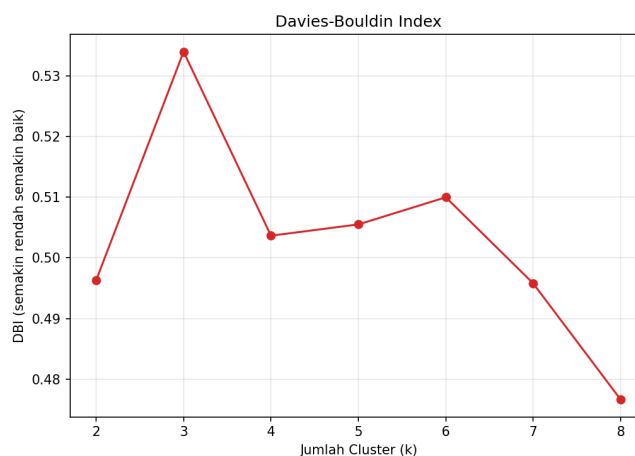
Evaluasi kualitas hasil Clustering pada $K = 3$ dilakukan menggunakan Silhouette Score dan Davies-Bouldin Index. Hasil evaluasi ditunjukkan pada Tabel 8.

Tabel 8. Hasil Evaluasi Clustering pada $K = 3$

Metrik Evaluasi	Nilai	Interpretasi
Silhouette Score	0,578	Cukup Baik
Davies-Bouldin Index	0,534	Cukup Baik



Gambar 3. Grafik Silhouette Score untuk $K = 2$ s.d. 8



Gambar 4. Grafik Davies-Bouldin Index untuk $K = 2$ s.d. 8

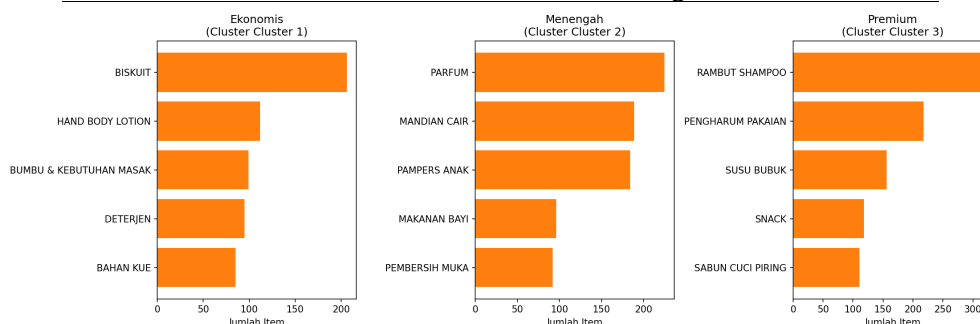
Silhouette Score sebesar 0,578 menunjukkan bahwa data dalam setiap cluster memiliki keterpaduan (cohesion) yang cukup baik dan berada pada rentang yang umum diinterpretasikan sebagai struktur cluster yang cukup kuat. Davies-Bouldin Index sebesar 0,534 menunjukkan tingkat pemisahan antar cluster yang cukup memadai, karena nilai di bawah 1,0 umumnya dianggap baik dalam penelitian Clustering. Secara keseluruhan, kedua metrik evaluasi menunjukkan bahwa hasil Clustering dengan $K = 3$ berada pada kategori cukup baik dan dapat digunakan sebagai dasar penyusunan strategi pemasaran.

3.6 Karakteristik dan Interpretasi Setiap Cluster

Untuk memperoleh interpretasi yang lebih mendalam, dilakukan analisis kategori produk yang paling dominan (memiliki jumlah item terbanyak) pada masing-masing cluster. Hasil analisis ditunjukkan pada Tabel 9 dan Gambar 5.

Tabel 9. Lima Kategori Produk Dominan per Cluster

Cluster	Segmen Harga	Kategori Produk Dominan	Jumlah Item
Cluster 1	Ekonomis	Biskuit	206
Cluster 1	Ekonomis	Hand Body Lotion	112
Cluster 1	Ekonomis	Bumbu & Kebutuhan Masak	99
Cluster 1	Ekonomis	Deterjen	95
Cluster 1	Ekonomis	Bahan Kue	85
Cluster 2	Menengah	Parfum	225
Cluster 2	Menengah	Mandian Cair	189
Cluster 2	Menengah	Pampers Anak	184
Cluster 2	Menengah	Makanan Bayi	96
Cluster 2	Menengah	Pembersih Muka	92
Cluster 3	Premium	Rambut Shampoo	316
Cluster 3	Premium	Pengharum Pakaian	218
Cluster 3	Premium	Susu Bubuk	156
Cluster 3	Premium	Snack	118
Cluster 3	Premium	Sabun Cuci Piring	111



Gambar 5. Kategori Produk Dominan pada Setiap Cluster

Cluster 1 (Ekonomis) didominasi oleh kategori kebutuhan harian dengan harga jual rendah, seperti Biskuit, Hand Body Lotion, serta Bumbu dan Kebutuhan Masak. Produk pada cluster ini umumnya dibeli secara rutin dengan nilai transaksi kecil per satuan, namun memiliki kontribusi volume penjualan yang signifikan. Cluster 2 (Menengah) didominasi oleh kategori kebersihan diri dan kebutuhan bayi, seperti Parfum, Mandian Cair, dan Pampers Anak, yang umumnya memiliki harga menengah dan tingkat pembelian yang cukup stabil. Cluster 3 (Premium) memiliki harga rata-rata tertinggi dan didominasi oleh kategori Rambut Shampoo, Pengharum Pakaian, dan Susu Bubuk, serta mencakup produk dengan harga tertinggi dalam dataset (hingga Rp2.000.000).

Sebagai catatan analitis, nilai komponen Harga Eceran pada seluruh centroid (Tabel 6) relatif kecil dan berdekatan (berkisar antara 0,0077 hingga 0,0163) dibandingkan dengan komponen Kategori_Kode (berkisar antara 0,1721 hingga 0,8145). Kondisi ini terjadi karena rentang harga pada dataset sangat lebar (Rp1.000 hingga Rp2.000.000), sehingga setelah normalisasi Min-Max, sebagian besar nilai harga produk terkompresi mendekati nol akibat pengaruh nilai ekstrem (outlier) pada batas atas. Akibatnya, pembentukan cluster pada penelitian ini lebih banyak dipengaruhi oleh perbedaan kode kategori dibandingkan murni oleh perbedaan harga, meskipun secara agregat ketiga cluster yang terbentuk tetap menunjukkan perbedaan rata-rata harga yang bermakna (Tabel 7). Penelitian selanjutnya disarankan mempertimbangkan transformasi logaritmik pada atribut harga sebelum normalisasi, atau penggunaan skema pembobotan atribut, agar kontribusi atribut harga terhadap pembentukan cluster menjadi lebih proporsional.

3.7 Rekomendasi Strategi Pemasaran Berdasarkan Cluster

Hasil pengelompokan produk ke dalam tiga cluster memberikan informasi segmentasi berbasis data yang dapat dijadikan referensi pendukung bagi pengelola ritel dalam menyusun kebijakan pemasaran yang lebih terarah. Perlu ditegaskan bahwa rekomendasi strategi yang dirumuskan berikut ini bersifat indikatif berdasarkan karakteristik segmen harga dan kategori dominan pada masing-masing cluster, sehingga perlu disesuaikan dengan kondisi

operasional toko, data penjualan aktual, margin produk, dan preferensi pelanggan yang dimiliki oleh pengelola ritel. Berdasarkan interpretasi karakteristik setiap cluster, strategi pemasaran yang dapat diterapkan ditunjukkan pada Tabel 10.

Tabel 10. Rekomendasi Strategi Pemasaran per Cluster

Cluster	Label	Karakteristik	Strategi Pemasaran
Cluster 1	Produk Dasar (Ekonomis)	Harga rendah, kebutuhan harian	Bundling produk, penempatan strategis di area kasir, promosi paket hemat
Cluster 2	Produk Kebersihan diri & Bayi (Menengah)	Harga menengah, kebersihan diri & kebutuhan bayi	Program loyalitas pelanggan, diskon kuantitas, cross-selling produk sejenis
Cluster 3	Produk Unggulan (Premium)	Harga tinggi, mencakup produk bernilai tertinggi	Promosi eksklusif, display khusus, program cicilan/kemudahan pembiayaan untuk produk bernilai tinggi

Strategi untuk Cluster 1 (Produk Dasar) difokuskan pada peningkatan frekuensi pembelian melalui program bundling dan penempatan produk yang strategis, mengingat produk pada cluster ini merupakan kebutuhan harian dengan volume penjualan tinggi. Strategi untuk Cluster 2 (Produk Kebersihan Diri dan Bayi) diarahkan pada program loyalitas dan cross-selling antar kategori kebersihan diri dan kebutuhan bayi. Adapun strategi untuk Cluster 3 (Produk Unggulan) berfokus pada promosi eksklusif dan kemudahan pembiayaan, mengingat cluster ini mencakup produk dengan rentang harga tertinggi dalam dataset. Seluruh rekomendasi strategi tersebut merupakan informasi pendukung yang dihasilkan dari analisis segmentasi berbasis data, dan bukan merupakan pengganti kajian pemasaran yang lebih komprehensif yang melibatkan data transaksi, perilaku konsumen, serta kondisi persaingan pasar.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma K-Means Clustering untuk mengelompokkan 4.226 data produk ritel (dari 4.245 data awal, setelah pembersihan 19 data harga tidak valid) ke dalam tiga cluster berdasarkan atribut kategori produk (Label Encoding) dan harga eceran (Min-Max Normalization). Penerapan Elbow Method menunjukkan penurunan WCSS yang tajam dari K=2 ke K=3, sehingga jumlah cluster final ditetapkan K=3 meskipun Silhouette Score tertinggi diperoleh pada K=2 (0,638); pertimbangan ini didasarkan pada keseimbangan antara hasil Elbow Method dan kebutuhan diferensiasi segmen pemasaran yang lebih terperinci. Hasil evaluasi menunjukkan Silhouette Score sebesar 0,578 dan Davies-Bouldin Index sebesar 0,534, yang mengindikasikan kualitas clustering berada pada kategori cukup baik dengan kohesivitas internal dan separabilitas antar cluster yang memadai.

Ketiga cluster yang terbentuk memiliki karakteristik berbeda: Cluster 1 (1.231 produk, 29,13%) merupakan segmen Ekonomis yang didominasi kebutuhan harian; Cluster 2 (1.451 produk, 34,34%) merupakan segmen Menengah yang didominasi produk kebersihan diri dan kebutuhan bayi; serta Cluster 3 (1.544 produk, 36,54%) merupakan segmen Premium dengan harga rata-rata tertinggi. Perlu dicatat bahwa pembentukan cluster lebih banyak dipengaruhi oleh perbedaan kode kategori dibandingkan harga, sebagai konsekuensi rentang harga dataset yang lebar sehingga nilai harga terkompresi setelah normalisasi Min-Max. Hasil pengelompokan ini dapat menjadi referensi bagi pengelola ritel dalam menyusun strategi pemasaran terdiferensiasi untuk masing-masing segmen—mencakup bundling, program loyalitas, promosi eksklusif, dan pengelolaan stok—namun bersifat indikatif dan perlu diintegrasikan dengan data operasional lain seperti data transaksi penjualan, margin produk, dan perilaku konsumen agar mendukung pengambilan keputusan pemasaran secara lebih komprehensif.

Daftar Pustaka

- [1] S. Aryza, M. S. Novelan, and M. R. Islam, "A Data-Driven Framework for Integrating Decision-Making and Operational Efficiency in Multi-Product Retail : A Case Study with Experimental Evaluation," vol. 9, no. 1, pp. 140–149, 2025, doi: 10.30829/zero.v9i1.24301.
- [2] A. D. Saputro and M. Ikaningtyas, "Pengaruh Digitalisasi terhadap Perkembangan Bisnis di Era Modern," vol. 1, no. 4, pp. 1199–1210, 2025.
- [3] M. Sholeh, H. Najma, R. Arif, N. A. Fahtezi, and Z. Darojah, "Pengelolaan Arsip yang Efektif Terhadap Akurasi Data dan Pengambilan Keputusan di SDN Siwalankerto I," vol. 07, no. 01, pp. 85–99, 2025, doi: 10.15642/JAPI.2025.7.1.85-99.
- [4] M. A. V. Fitrah, "Analisis potensi pasar dan strategi pemasaran untuk produk atau layanan inovatif terhadap usaha coffeeshop latera jambi," vol. 10, no. 3, pp. 1264–1276, 2025.
- [5] E. Nurarofah, R. Herdiana, and N. Dienwati Nuris, "Penerapan Asosiasi Menggunakan Algoritma Fp-Growth Pada Pola Transaksi Penjualan Di Toko Roti," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 353–359, 2023, doi: 10.36040/jati.v7i1.6299.
- [6] Y. C. Ramadhani and A. Budiarti, "Peran Strategi Manajemen Pemasaran, Kepuasan Pelanggan, dan Loyalitas Pelanggan dalam Membangun Keunggulan Kompetitif pada Industri Ritel," vol. 4, no. 2, pp. 210–234, 2025.
- [7] A. W. Ningsih, V. S. Sahliyyah, W. Aayatul, H. Tri, and Y. S. Shafrani, "STRATEGI PEMASARAN BERBASIS PORTOFOLIO PRODUK MENGGUNAKAN MATRISK BOSTON CONSULTING GROUP (BCG) PADA OUTLET

- MINUMAN NESCAFE DI PURWOKERTO,” *J. Media Akad.*, vol. 4, no. 6, 2026.
- [8] M. C. Gumilang and I. Tahyudin, “Optimasi Manajemen Stok melalui Pengelompokan Produk Ritel Berdasarkan Data Penjualan dengan K-Means,” pp. 644–655, 2025, doi: 10.33364/algorithm/v.22-2.2526.
 - [9] N. Asyifa and M. S. Novelan, “Performance Analysis of the K-Means Algorithm in Classifying Organic and Inorganic Waste Types,” pp. 6681–6690, 2025.
 - [10] Arifin, Z. Sitorus, M. Iqbal, A. P. Utama Siahaan, and S. Wahyuni, “ANALYSIS OF THE SALES POTENTIAL OF BUMDES PRODUCTS USING THE K-MEANS CLUSTERING,” vol. 4, no. 1, 2025.
 - [11] S. Sutiono, R. F. Wijaya, and M. S. Novelan, “ANALISIS PENGELOMPOKAN JADWAL MENGAJAR GURU DI SMKN 1 STABAT MENGGUNAKAN METODE K-MEANS CLUSTERING,” vol. 4307, no. August, pp. 3450–3456, 2025.
 - [12] C. A. Nugraha, M. A. Kesuma, O. I. Cahyani, and M. Wati, “Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means,” vol. 7, pp. 80–88, 2025.
 - [13] B. Yudha Putra, F. Yuanita Azzahra, and I. Alfrina Erlanda, “KLAUSTERISASI PENGUNJUNG MALL MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN PENDAPATAN,” vol. 11, no. 3, 2023.
 - [14] Y. Nurohmah *et al.*, “OPTIMALISASI PERFORMA K- MEANS CLUSTERING DENGAN PCA DALAM ANALISIS TINGKAT KEMISKINAN DI JAWA BARAT,” vol. 7, no. 3, pp. 1657–1665, 2023.
 - [15] N. S. Niko, A. Rahman, D. Marini Umi Atmaja, and A. Basri, “Klasterisasi Stok Produk Retail Untuk Menentukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means,” *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 306–312, 2023, doi: 10.47065/bit.v4i3.736.
 - [16] Z. V. Arla and T. Sutabri, “Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada ‘ Toko Hartati ’ Universitas Bina Darma , Indonesia untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan tertentu . dalam periode tertentu . Penjualan produk sering di,” vol. 2, no. 4, 2024.
 - [17] D. Lasandi, R. Astuti, K. Anam, A. P. Putra, and B. Nurhakim, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA POPULASI TERNAK DI KOTA BANDUNG BERBASIS DATA MINING,” *JUNRLA Inform. DAN Tek. ELEKTRO Terap.*, vol. 14, no. 2, 2026.
 - [18] B. D. Oktavian, I. Lumintu, and S. Amar, “Optimalisasi Strategi Product Bundling melalui Pemetaan Pola Peminatan dan Pola Penjualan Produk menggunakan K-Means Clustering dan Apriori Optimizing Product Bundling Strategy through Mapping Product Interest Patterns and Sales Patterns using K-Means Clu,” vol. 8, no. 1, pp. 26–41, 2025.
 - [19] A. Ichwani, R. I. Kesuma, A. Setiawan, and I. E. Wicaksono, “Preventing Data Leakage in Classification via Integrated Machine Learning Pipelines : Preprocessing , Feature Transformation , and Hyperparameter Tuning,” *J. Tek. Inform.*, vol. 7, no. 1, pp. 391–410, 2026.
 - [20] F. Nuraeni, M. Hasby, and M. Firdaus, “Model Klasterisasi Data Pertanian Menggunakan Algoritma K-Means,” *J. Algoritm.*, vol. 22, no. 2, pp. 890–898, 2025, doi: 10.33364/algorithm/v.22-2.2054.
 - [21] N. Umaira, W. Astuti, R. Ayu, P. Sari, S. R. Karlina, and I. S. Putri, “Analisa Klaster Pola Karakteristik Pengunjung Event di DKI Jakarta,” vol. 8, pp. 99–105, 2025.