



# Implementasi Algoritma K-Means Clustering untuk Pengelompokan Produk E-Commerce Berdasarkan Harga, Diskon, dan Total Revenue

Rinaldi Pasaribu\*, Saidi Ramadan Siregar

Fakultas Ilmu komputer dan Teknologi Informasi, Prodi Sistem Informasi, Universitas Budi Darma, Medan, Indonesia

Email: <sup>1,\*</sup>Rinaldipasaribu229a@gmail.com, <sup>2</sup>saidiramadan89@gmail.com

Email Penulis Korespondensi: Rinaldipasaribu229a@gmail.com

**Abstrak**—Perkembangan e-commerce yang pesat menghasilkan volume data transaksi yang besar, namun belum seluruhnya dimanfaatkan secara optimal untuk mendukung pengambilan keputusan strategis, khususnya dalam segmentasi produk. Permasalahan utama dalam penelitian ini adalah belum adanya pengelompokan produk yang sistematis berdasarkan atribut penting seperti harga, diskon, dan pendapatan, sehingga strategi penetapan harga dan promosi menjadi kurang tepat sasaran. Oleh karena itu, penelitian ini bertujuan untuk menganalisis pola penjualan produk serta mengelompokkan produk e-commerce berdasarkan karakteristik price, discount\_percent, dan total\_revenue. Dataset yang digunakan merupakan Amazon-style e-commerce dataset dengan jumlah 50.000 data transaksi dan 13 atribut, di mana penelitian ini memfokuskan pada tiga atribut utama tersebut sebagai dasar clustering. Metode yang digunakan adalah K-Means Clustering, dengan tahapan preprocessing, normalisasi menggunakan Min-Max Scaling, serta penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah 3 cluster, yang didukung oleh nilai Silhouette Score tertinggi sebesar 0.354 dan pola siku yang jelas pada grafik Elbow. Evaluasi tambahan menggunakan Davies-Bouldin Index sebesar 0.9335 menunjukkan kualitas cluster yang cukup baik meskipun belum optimal. Hasil clustering menghasilkan tiga kelompok utama, yaitu cluster produk premium (harga tinggi, diskon rendah, revenue tinggi), cluster produk diskon (harga sedang, diskon tinggi, revenue menengah), dan cluster produk low performance (harga rendah, diskon rendah, revenue rendah). Kesimpulan dari penelitian ini adalah bahwa algoritma K-Means mampu mengelompokkan produk e-commerce secara efektif berdasarkan atribut numerik yang relevan, serta menghasilkan insight yang dapat digunakan untuk mendukung strategi bisnis seperti penentuan harga dan promosi.

**Kata Kunci:** K-Means Clustering; E-Commerce; Segmentasi Produk; Business Intelligence; Elbow Method; Silhouette Score

**Abstract**—The rapid growth of e-commerce has generated a large volume of transactional data; however, this data has not been fully utilized to support strategic decision-making, particularly in product segmentation. The main problem addressed in this study is the absence of a systematic product grouping approach based on key attributes such as price, discount, and revenue, which leads to less effective pricing and promotional strategies. Therefore, this study aims to analyze product sales patterns and cluster e-commerce products based on the characteristics of price, discount\_percent, and total\_revenue. The dataset used is an Amazon-style e-commerce dataset consisting of 50,000 transaction records and 13 attributes, with the analysis focusing on the three main attributes as the basis for clustering. The method applied in this research is K-Means Clustering, which involves data preprocessing, normalization using Min-Max Scaling, and determining the optimal number of clusters using the Elbow Method and Silhouette Score. The results indicate that the optimal number of clusters is three clusters, supported by the highest Silhouette Score of 0.354 and a clear elbow pattern in the Elbow graph. Additional evaluation using the Davies-Bouldin Index of 0.9335 indicates that the clustering quality is fairly good, although not yet optimal. The clustering results produce three main groups: premium product cluster (high price, low discount, high revenue), discount product cluster (moderate price, high discount, moderate revenue), and low-performance product cluster (low price, low discount, low revenue). In conclusion, the K-Means algorithm is capable of effectively clustering e-commerce products based on relevant numerical attributes and generating insights that can support business strategies such as pricing and promotional decisions.

**Keywords:** K-Means Clustering; E-Commerce; Product Segmentation; Business Intelligence; Elbow Method; Silhouette Score

## 1. PENDAHULUAN

Perkembangan teknologi internet telah mendorong pertumbuhan pesat sektor e-commerce sehingga meningkatkan volume transaksi digital yang terjadi setiap hari di berbagai platform perdagangan online (Shen, 2025)(A'yun et al., 2024). Peningkatan aktivitas transaksi tersebut menunjukkan bahwa e-commerce telah menjadi salah satu ekosistem bisnis digital yang menghasilkan data transaksi dalam jumlah besar dan terus bertambah seiring dengan meningkatnya jumlah pengguna internet dan aktivitas belanja online (Caroline et al., 2023)(Munshi et al., 2023). Dalam praktiknya, setiap transaksi e-commerce menyimpan berbagai informasi penting yang dapat dimanfaatkan untuk analisis bisnis, seperti harga produk, tingkat diskon, jumlah produk terjual, hingga total pendapatan yang dihasilkan dari penjualan tersebut. Informasi tersebut membentuk kumpulan data transaksi yang sangat bernilai karena dapat menggambarkan perilaku konsumen dan dinamika pasar digital secara lebih komprehensif (Qu, 2025). Selain itu, data transaksi juga sering digunakan oleh perusahaan untuk memahami hubungan antar produk serta pola pembelian konsumen dalam aktivitas belanja online (Hendra et al., 2024). Dengan memanfaatkan teknik analisis data, perusahaan dapat mengidentifikasi pola penjualan produk yang sebelumnya tersembunyi di dalam kumpulan data transaksi yang besar dan kompleks (Yuliarnis et al., 2020). Proses analisis tersebut memungkinkan organisasi memperoleh wawasan strategis yang dapat digunakan untuk meningkatkan efektivitas pemasaran, pengelolaan inventori, serta strategi promosi produk (Siregar & Khalidy, 2024). Selain itu, pemanfaatan teknik data mining dan analisis data juga mampu membantu perusahaan dalam mengoptimalkan strategi bisnis melalui pemahaman yang lebih baik terhadap pola permintaan pasar dan performa penjualan produk. Oleh karena itu, analisis data transaksi e-commerce menjadi salah satu pendekatan



penting dalam mendukung pengambilan keputusan berbasis data pada perusahaan digital modern (Caroline et al., 2023). Dengan demikian, pemanfaatan analisis data pada transaksi e-commerce tidak hanya membantu memahami pola penjualan produk, tetapi juga memberikan dasar analitis yang kuat bagi perusahaan dalam merancang strategi bisnis yang lebih efektif dan kompetitif (Qu, 2025).

Banyak perusahaan saat ini memiliki volume data yang sangat besar yang berasal dari aktivitas transaksi digital, namun sebagian besar data tersebut belum dimanfaatkan secara optimal untuk melakukan segmentasi produk atau pelanggan yang dapat mendukung pengambilan keputusan bisnis secara strategis (Kannan & Khan, 2025). Kondisi ini sering terjadi karena organisasi hanya menyimpan data sebagai arsip operasional tanpa melakukan analisis lanjutan untuk menemukan pola atau informasi yang tersembunyi di dalamnya (Zulkurnain & Eryanto, 2025)(Sholeh et al., 2025). Padahal, data yang dihasilkan dari aktivitas bisnis seperti transaksi penjualan, harga produk, maupun interaksi konsumen dapat memberikan wawasan penting mengenai karakteristik pasar dan perilaku pembelian (Sun, 2024). Dalam konteks pemasaran modern, segmentasi menjadi pendekatan penting untuk membagi pasar atau produk ke dalam kelompok yang memiliki karakteristik serupa sehingga perusahaan dapat memahami kebutuhan setiap kelompok secara lebih spesifik (Suhairi et al., 2023). Tanpa adanya proses segmentasi, perusahaan akan kesulitan mengidentifikasi kelompok produk atau pelanggan yang paling berpotensi menghasilkan keuntungan sehingga strategi pemasaran menjadi kurang efektif (Yoseph et al., 2020). Penelitian dalam bidang data mining menunjukkan bahwa teknik analisis seperti clustering mampu membantu perusahaan mengekstraksi pola dari dataset besar sehingga informasi yang sebelumnya tersembunyi dapat dimanfaatkan untuk perencanaan bisnis (Tjia et al., 2025). Selain itu, penerapan segmentasi juga memungkinkan perusahaan merancang strategi pemasaran yang lebih terarah dengan menyesuaikan penawaran produk pada karakteristik kelompok tertentu (Suh, 2025). Tanpa pemahaman terhadap segmentasi pasar atau produk, strategi harga dan promosi yang diterapkan cenderung bersifat umum sehingga kurang mampu menjangkau target konsumen secara tepat (Moodley et al., 2020). Hal ini dapat menyebabkan sumber daya pemasaran tidak dimanfaatkan secara optimal karena promosi diberikan kepada segmen yang kurang relevan dengan karakteristik produk yang ditawarkan. Oleh karena itu, pemanfaatan analisis data dan teknik segmentasi menjadi langkah penting bagi perusahaan untuk meningkatkan efektivitas strategi harga, promosi, dan pengelolaan produk dalam lingkungan bisnis digital yang semakin kompetitif.

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan objek atau data berdasarkan tingkat kemiripan karakteristik sehingga objek yang memiliki karakteristik serupa berada dalam satu kelompok yang sama (Amalina et al., 2022)(Azzahra & Yasir, 2024). Metode ini banyak digunakan dalam analisis data bisnis karena mampu mengekstraksi pola tersembunyi dari dataset yang besar tanpa memerlukan label data sebelumnya (Manarung et al., 2025). Dalam konteks e-commerce, teknik clustering dapat dimanfaatkan untuk mengidentifikasi pola performa produk berdasarkan atribut penjualan sehingga perusahaan dapat memahami karakteristik produk yang memiliki tingkat permintaan tinggi maupun rendah (Noval et al., 2025). Penerapan clustering pada data penjualan juga memungkinkan perusahaan menganalisis hubungan antara harga, jumlah penjualan, dan strategi promosi yang mempengaruhi performa produk di pasar digital (Apriyanto & Sitio, 2025). Salah satu algoritma clustering yang paling banyak digunakan dalam penelitian analisis data adalah K-Means karena memiliki mekanisme yang sederhana dan mampu memproses dataset berukuran besar secara efisien (Falih et al., 2025). Algoritma K-Means bekerja dengan membagi data ke dalam sejumlah cluster tertentu dengan cara meminimalkan jarak antar data terhadap pusat cluster atau centroid yang terbentuk (Sakinah & Awaliyah, 2025).

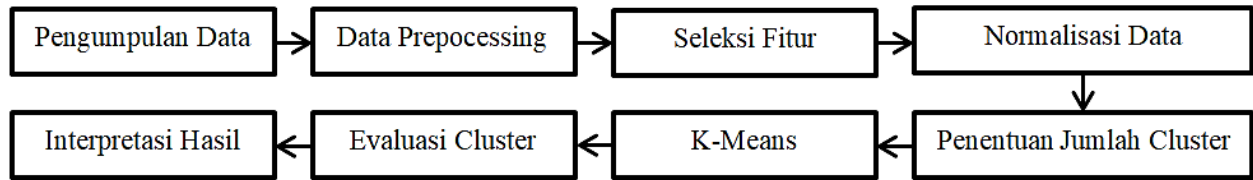
Beberapa penelitian sebelumnya telah menerapkan K-Means untuk mengelompokkan data penjualan produk guna membantu perusahaan dalam menentukan strategi pemasaran dan pengelolaan stok secara lebih efektif (Almaripat et al., 2025). Penelitian lain juga menunjukkan bahwa teknik clustering dapat membantu mengidentifikasi pola pembelian pelanggan serta preferensi terhadap produk tertentu pada platform e-commerce (Effendy et al., 2026). Namun sebagian besar penelitian yang ada masih berfokus pada segmentasi pelanggan atau customer segmentation dibandingkan dengan segmentasi produk berdasarkan performa penjualan (Mado & Hendry, 2025). Oleh karena itu, penelitian ini berfokus pada analisis segmentasi produk dengan menggunakan atribut harga, diskon, dan total revenue untuk mengidentifikasi pola performa penjualan produk e-commerce secara lebih komprehensif.

## **2. METODOLOGI PENELITIAN**

Metodologi penelitian merupakan tahapan sistematis yang digunakan untuk menjelaskan proses penelitian mulai dari pengumpulan data hingga analisis hasil sehingga penelitian dapat dilakukan secara terstruktur dan dapat direplikasi oleh peneliti lain (Pandiangan & Albina, 2025). Pendekatan metodologis yang jelas sangat penting dalam penelitian berbasis data karena membantu memastikan bahwa proses pengolahan dan analisis data dilakukan secara konsisten dan valid. Dalam penelitian ini, metodologi mencakup tahapan pengumpulan dataset, preprocessing data, pemilihan fitur, penerapan algoritma analisis, serta evaluasi hasil yang diperoleh dari proses pengolahan data. Penerapan metodologi yang sistematis juga memungkinkan peneliti mengidentifikasi hubungan antar variabel dalam dataset sehingga dapat menghasilkan temuan yang relevan untuk mendukung pengambilan keputusan berbasis data. Oleh karena itu, penelitian ini menggunakan pendekatan metodologis yang terstruktur dengan memanfaatkan dataset transaksi e-commerce untuk dianalisis menggunakan teknik clustering guna mengidentifikasi pola performa produk.

## 2.1 Tahapan penelitian

Tahapan penelitian pada studi ini dirancang untuk mengolah data transaksi e-commerce sehingga dapat menghasilkan pengelompokan produk berdasarkan harga, diskon, dan total revenue menggunakan algoritma K-Means Clustering. Setiap tahapan dilakukan secara berurutan mulai dari pengumpulan data hingga interpretasi hasil clustering sehingga pola performa produk pada dataset dapat dianalisis secara sistematis.



**Gambar 1.** Tahapan Penelitian

Berdasarkan Gambar 1 diatas, Tahapan penelitian yang dilakukan dimulai dari pengumpulan data transaksi e-commerce yang memuat informasi harga, diskon, dan total revenue sebagai dasar analisis performa produk, kemudian dilakukan preprocessing untuk membersihkan data agar tidak terdapat noise atau inkonsistensi yang dapat mempengaruhi hasil clustering. Selanjutnya dilakukan seleksi fitur dengan memfokuskan pada atribut price, discount\_percent, dan total\_revenue karena ketiganya secara langsung merepresentasikan strategi penetapan harga dan kontribusi pendapatan produk. Data kemudian dinormalisasi untuk menyamakan skala antar atribut sehingga proses perhitungan jarak pada algoritma K-Means menjadi lebih akurat. Setelah itu ditentukan jumlah cluster optimal menggunakan metode seperti Elbow agar hasil pengelompokan produk lebih representatif. Proses K-Means kemudian dilakukan untuk mengelompokkan produk berdasarkan kemiripan karakteristik harga, diskon, dan revenue, yang selanjutnya dievaluasi untuk memastikan kualitas cluster sebelum akhirnya diinterpretasikan guna mengidentifikasi pola performa produk sebagai dasar rekomendasi strategi bisnis e-commerce.

## 2.2 Dataset

Dataset yang digunakan dalam penelitian ini merupakan Amazon-style E-Commerce Sales Dataset yang berisi data transaksi penjualan produk pada platform e-commerce. Dataset ini mencatat berbagai informasi transaksi yang mencerminkan aktivitas penjualan, perilaku konsumen, serta performa produk yang dijual pada platform digital. Penggunaan dataset transaksi e-commerce dalam penelitian data mining sering dilakukan karena data tersebut memiliki atribut yang kaya dan mampu menggambarkan dinamika aktivitas perdagangan digital secara komprehensif. Informasi yang tersedia dalam dataset ini dapat dimanfaatkan untuk berbagai analisis seperti analisis penjualan, segmentasi produk, maupun analisis perilaku konsumen.

Struktur atribut yang terdapat dalam dataset penelitian ini disajikan pada Tabel 1 yang menjelaskan nama atribut serta deskripsi dari setiap variabel yang tersedia dalam dataset.

**Tabel 1.** Atribut Dataset

No	Nama Atribut	Deskripsi
1	order_id	ID unik yang menunjukkan setiap transaksi pemesanan
2	order_date	Tanggal terjadinya transaksi pembelian
3	product_id	ID unik yang mengidentifikasi setiap produk
4	product_category	Kategori produk yang dijual
5	price	Harga asli produk sebelum diskon
6	discount_percent	Persentase diskon yang diberikan pada produk
7	quantity_sold	Jumlah unit produk yang terjual
8	customer_region	Wilayah asal pelanggan yang melakukan pembelian
9	payment_method	Metode pembayaran yang digunakan pelanggan
10	rating	Nilai penilaian produk yang diberikan pelanggan
11	review_count	Jumlah ulasan yang diberikan pelanggan terhadap produk
12	discounted_price	Harga produk setelah diberikan diskon
13	total_revenue	Total pendapatan yang diperoleh dari penjualan produk

Berdasarkan Tabel 1, dapat diketahui bahwa dataset penelitian ini terdiri dari 13 atribut yang mencakup informasi transaksi, karakteristik produk, serta interaksi pelanggan terhadap produk. Meskipun dataset memiliki berbagai atribut, penelitian ini hanya menggunakan tiga atribut utama sebagai variabel analisis clustering yaitu price, discount\_percent, dan total\_revenue. Ketiga atribut tersebut dipilih karena secara langsung merepresentasikan strategi harga, tingkat promosi, serta performa pendapatan produk pada platform e-commerce. Dengan memanfaatkan atribut tersebut, proses clustering diharapkan mampu mengelompokkan produk berdasarkan karakteristik harga dan kontribusinya terhadap pendapatan penjualan sehingga dapat memberikan gambaran segmentasi produk yang lebih jelas bagi perusahaan dalam merumuskan strategi pemasaran dan penjualan.



### 2.3 Data Preprocessing

Tahap data preprocessing pada penelitian ini berperan penting dalam menyiapkan dataset transaksi e-commerce agar layak digunakan pada proses clustering berbasis K-Means, khususnya karena analisis difokuskan pada atribut price, discount\_percent, dan total\_revenue. Proses diawali dengan penghapusan data duplikat, yaitu mengidentifikasi dan menghilangkan baris data yang memiliki nilai identik pada atribut tertentu seperti order\_id atau kombinasi atribut lainnya, sehingga tidak terjadi bias akibat penggandaan informasi transaksi yang dapat mempengaruhi pembentukan cluster. Selanjutnya dilakukan penanganan missing value, di mana data yang memiliki nilai kosong pada atribut penting seperti harga atau revenue akan ditangani menggunakan metode tertentu, misalnya penghapusan data jika jumlahnya kecil atau imputasi menggunakan nilai rata-rata (mean) agar tidak mengganggu distribusi data. Tahap berikutnya adalah transformasi atau perubahan format data numerik, seperti memastikan seluruh atribut numerik (price, discount\_percent, total\_revenue) berada dalam tipe data yang sesuai (float atau integer) serta menghilangkan karakter non-numerik yang dapat menyebabkan error pada proses komputasi. Terakhir, dilakukan seleksi atribut dengan hanya mengambil variabel yang relevan dengan tujuan penelitian, yaitu atribut yang secara langsung merepresentasikan performa produk dari sisi harga, strategi diskon, dan pendapatan, sehingga proses clustering menjadi lebih fokus, efisien, dan menghasilkan segmentasi produk yang lebih akurat serta mudah diinterpretasikan dalam konteks strategi bisnis e-commerce.

### 2.4 Normalisasi Data

Normalisasi data dalam penelitian ini dilakukan untuk mengatasi perbedaan skala nilai antar atribut price, discount\_percent, dan total\_revenue, karena ketiga variabel tersebut memiliki rentang nilai yang sangat berbeda sehingga dapat menyebabkan bias pada perhitungan jarak dalam algoritma K-Means. Tanpa normalisasi, atribut dengan nilai lebih besar seperti total\_revenue akan mendominasi proses clustering dibandingkan atribut lain, sehingga hasil pengelompokan menjadi tidak representatif terhadap keseluruhan karakteristik data. Oleh karena itu, digunakan metode Min-Max Scaling yang berfungsi untuk mentransformasikan nilai data ke dalam rentang tertentu, umumnya antara 0 hingga 1, sehingga setiap atribut memiliki kontribusi yang seimbang dalam proses perhitungan jarak (Allorerung et al., 2024) (Kusnaldi et al., 2022). Adapun rumus normalisasi Min-Max Scaling yang digunakan adalah sebagai berikut:

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

Dimana  $x$  merupakan nilai asli,  $x_{min}$  adalah nilai minimum dari atribut, dan  $x_{max}$  adalah nilai maksimum dari atribut tersebut. Dengan menerapkan normalisasi ini, data menjadi berada pada skala yang seragam sehingga algoritma K-Means dapat mengelompokkan produk secara lebih akurat berdasarkan kemiripan relatif antara harga, diskon, dan total revenue, yang pada akhirnya menghasilkan segmentasi produk yang lebih valid untuk analisis strategi bisnis e-commerce.

### 2.5 K-Means Clustering

Algoritma K-Means Clustering pada penelitian ini digunakan untuk mengelompokkan produk e-commerce berdasarkan kemiripan nilai price, discount\_percent, dan total\_revenue, sehingga dapat diidentifikasi pola performa produk secara objektif. Tahapan pertama adalah menentukan jumlah cluster (K) yang merepresentasikan jumlah kelompok produk yang ingin dibentuk, di mana nilai K biasanya diperoleh melalui metode seperti Elbow untuk memastikan jumlah cluster optimal dan tidak terlalu sedikit atau berlebihan. Penentuan nilai K sangat berpengaruh terhadap kualitas hasil clustering karena akan menentukan tingkat granularitas segmentasi produk berdasarkan karakteristik harga dan pendapatan (Salman & Alomary, 2024).

Tahap berikutnya adalah menentukan centroid awal, yaitu titik pusat awal dari setiap cluster yang biasanya dipilih secara acak dari data yang tersedia (Wan et al., 2024). Centroid ini akan menjadi acuan awal dalam proses pengelompokan data, sehingga pemilihan centroid awal dapat mempengaruhi hasil akhir clustering. Setelah centroid ditentukan, dilakukan proses perhitungan jarak antara setiap data dengan centroid, yang dalam penelitian ini umumnya menggunakan jarak Euclidean untuk mengukur kedekatan antar data dalam ruang multidimensi.

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

Dimana  $x$  adalah data,  $y$  adalah centroid, dan  $n$  adalah jumlah atribut (dalam penelitian ini terdiri dari price, discount, dan revenue). Berdasarkan hasil perhitungan jarak tersebut, setiap data kemudian dikelompokkan ke dalam cluster terdekat, yaitu cluster yang memiliki jarak minimum terhadap centroid. Proses ini menghasilkan pembagian awal data ke dalam beberapa kelompok berdasarkan kemiripan karakteristiknya.

Setelah proses pengelompokan awal, dilakukan tahap update centroid, yaitu menghitung ulang posisi centroid baru berdasarkan rata-rata nilai seluruh data yang berada dalam masing-masing cluster. Perhitungan centroid baru dilakukan dengan rumus rata-rata sebagai berikut (Minh et al., 2022):

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i \quad (3)$$

Dimana  $C_j$  adalah centroid cluster ke- $j$ , dan  $n_j$  adalah jumlah data dalam cluster tersebut. Tahapan ini bertujuan untuk memperbarui posisi pusat cluster agar lebih merepresentasikan anggota cluster yang sebenarnya.



Proses kemudian dilanjutkan dengan iterasi berulang, yaitu mengulangi langkah perhitungan jarak, pengelompokan data, dan pembaruan centroid hingga kondisi konvergen tercapai, yaitu ketika tidak ada perubahan signifikan pada posisi centroid atau anggota cluster. Dengan demikian, algoritma K-Means dalam penelitian ini mampu menghasilkan pengelompokan produk e-commerce yang stabil dan optimal berdasarkan kombinasi atribut harga, diskon, dan total revenue, sehingga dapat digunakan untuk analisis segmentasi produk dalam mendukung strategi bisnis yang lebih tepat sasaran.

## 2.6 Penentuan Jumlah Cluster

Penentuan jumlah cluster (K) merupakan tahap krusial dalam penerapan algoritma K-Means pada penelitian ini, karena jumlah cluster akan menentukan bagaimana produk e-commerce dikelompokkan berdasarkan atribut price, discount\_percent, dan total\_revenue. Pemilihan nilai K yang tidak tepat dapat menyebabkan hasil clustering menjadi kurang representatif, misalnya terlalu sedikit cluster akan menggabungkan produk dengan karakteristik berbeda, sedangkan terlalu banyak cluster akan menghasilkan segmentasi yang terlalu spesifik dan sulit diinterpretasikan. Oleh karena itu, penelitian ini menggunakan dua pendekatan umum yaitu Elbow Method dan Silhouette Score untuk memperoleh jumlah cluster yang optimal dan mampu merepresentasikan pola performa produk secara akurat.

Metode pertama adalah Elbow Method, yang digunakan untuk menentukan jumlah cluster dengan melihat nilai Sum of Squared Errors (SSE) atau total jarak kuadrat antara data dengan centroid dalam setiap cluster. Prinsip metode ini adalah mencari titik “siku” (elbow) pada grafik antara jumlah cluster (K) dan nilai SSE, di mana penurunan SSE mulai melambat secara signifikan. Secara matematis, SSE dapat dirumuskan sebagai berikut (Heidari et al., 2024) (Wu, 2024):

$$SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} \|x_i - C_j\|^2 \quad (4)$$

Dimana  $x_i$  adalah data ke- $i$ ,  $C_j$  adalah centroid cluster ke- $j$ , dan  $n_j$  adalah jumlah data dalam cluster tersebut. Dalam konteks penelitian ini, nilai SSE dihitung berdasarkan jarak antara produk dengan centroid cluster yang dibentuk dari atribut harga, diskon, dan revenue. Nilai K optimal ditentukan pada titik di mana penurunan SSE tidak lagi signifikan, sehingga jumlah cluster tersebut dianggap paling efisien dalam merepresentasikan data.

Metode kedua adalah Silhouette Score, yang digunakan untuk mengevaluasi kualitas cluster berdasarkan tingkat kedekatan data dalam satu cluster (cohesion) dan pemisahan antar cluster (separation). Nilai Silhouette berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa data berada dalam cluster yang tepat dan terpisah dengan baik dari cluster lain. Rumus Silhouette Score untuk setiap data adalah sebagai berikut (Gwak et al., 2025) (Faizi & Adnan, 2024):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

Dimana  $a(i)$  adalah rata-rata jarak data ke semua data dalam cluster yang sama, sedangkan  $b(i)$  adalah rata-rata jarak data ke cluster terdekat lainnya. Dalam penelitian ini, Silhouette Score digunakan untuk mengevaluasi hasil clustering produk berdasarkan kesamaan karakteristik harga, diskon, dan total revenue. Nilai rata-rata Silhouette yang lebih tinggi menunjukkan bahwa struktur cluster yang terbentuk semakin baik.

Dengan menggabungkan Elbow Method dan Silhouette Score, penelitian ini dapat menentukan jumlah cluster yang tidak hanya optimal secara matematis tetapi juga memiliki kualitas pengelompokan yang baik, sehingga hasil segmentasi produk e-commerce menjadi lebih akurat dan dapat digunakan sebagai dasar dalam analisis strategi harga dan promosi.

## 3. HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan dalam penelitian ini menyajikan secara sistematis keluaran dari proses pengolahan data e-commerce yang telah melalui tahapan preprocessing, normalisasi, hingga penerapan algoritma K-Means Clustering, dengan fokus pada atribut price, discount\_percent, dan total\_revenue sebagai variabel utama dalam segmentasi produk. Pada bagian ini, dilakukan analisis deskriptif untuk memahami karakteristik dataset, dilanjutkan dengan penentuan jumlah cluster optimal menggunakan metode Elbow dan Silhouette, serta evaluasi kualitas cluster melalui metrik seperti Davies-Bouldin Index dan Calinski-Harabasz Index. Hasil clustering kemudian divisualisasikan dan diinterpretasikan untuk mengidentifikasi pola performa produk berdasarkan kombinasi harga, tingkat diskon, dan kontribusi terhadap pendapatan. Selain itu, pembahasan juga mencakup perbandingan model clustering serta analisis sensitivitas jumlah cluster terhadap insight bisnis yang dihasilkan, sehingga mampu memberikan pemahaman yang lebih komprehensif dan aplikatif dalam mendukung pengambilan keputusan strategis pada platform e-commerce.

### 3.1 Deskripsi Dataset

Pada bagian ini disajikan gambaran umum karakteristik dataset yang digunakan dalam penelitian, khususnya pada atribut utama yaitu price, discount\_percent, dan total\_revenue. Dataset yang dianalisis terdiri dari 50.000 data transaksi dengan 13 fitur, sehingga dinilai cukup representatif untuk menggambarkan pola penjualan pada platform e-commerce.

Ringkasan statistik deskriptif dari ketiga atribut utama tersebut ditampilkan pada Gambar 2 untuk memberikan pemahaman awal terkait distribusi data sebelum dilakukan proses clustering.

	price	discount_percent	total_revenue
count	50000.000000	50000.000000	50000.000000
mean	252.507260	13.340700	657.331475
std	143.025544	9.850694	526.223968
min	5.010000	0.000000	3.840000
25%	127.840000	5.000000	240.320000
50%	252.970000	10.000000	505.410000
75%	376.335000	20.000000	968.970000
max	499.990000	30.000000	2499.550000

**Gambar 2.** Statistik Deskriptif Dataset

Berdasarkan Gambar 2, dapat dianalisis bahwa rata-rata harga produk (price) sebesar 252.51 dengan nilai minimum 5.01 dan maksimum 499.99, yang menunjukkan rentang harga produk cukup luas dari kategori murah hingga premium. Sementara itu, nilai rata-rata discount\_percent sebesar 13.34% dengan nilai maksimum 30% mengindikasikan bahwa strategi diskon cukup bervariasi, meskipun sebagian besar produk cenderung berada pada rentang diskon rendah hingga menengah, terlihat dari nilai median sebesar 10%. Pada atribut total\_revenue, rata-rata pendapatan sebesar 657.33 dengan nilai maksimum mencapai 2499.55 menunjukkan adanya perbedaan kontribusi revenue yang signifikan antar produk, yang kemungkinan dipengaruhi oleh kombinasi harga dan jumlah penjualan. Selain itu, nilai standar deviasi yang relatif tinggi pada ketiga variabel mengindikasikan adanya penyebaran data yang cukup besar, sehingga memperkuat kebutuhan penggunaan metode clustering untuk mengelompokkan produk berdasarkan karakteristik yang serupa. Secara keseluruhan, Gambar 5.1 menunjukkan bahwa dataset memiliki variasi yang kompleks dan relevan untuk dianalisis lebih lanjut dalam proses segmentasi produk e-commerce.

### 3.2 Hasil Normalisasi

Pada tahap ini dilakukan proses normalisasi terhadap atribut price, discount\_percent, dan total\_revenue menggunakan metode Min-Max Scaling dengan tujuan menyetarakan rentang nilai antar fitur sebelum dilakukan proses clustering. Normalisasi ini penting karena perbedaan skala yang signifikan antar variabel dapat menyebabkan bias dalam perhitungan jarak pada algoritma K-Means. Hasil normalisasi ditampilkan pada Gambar 3 sebagai perbandingan langsung antara nilai asli dan nilai setelah normalisasi.

	price_asli	discount_asli	revenue_asli	price_normalisasi	discount_normalisasi	revenue_normalisasi
0	128.75	10	463.52	0.249990	0.333333	0.184188
1	302.60	20	1210.40	0.601216	0.666667	0.483454
2	495.80	20	793.28	0.991535	0.666667	0.316319
3	371.95	15	1264.64	0.741323	0.500000	0.505187
4	201.68	0	806.72	0.397329	0.000000	0.321704

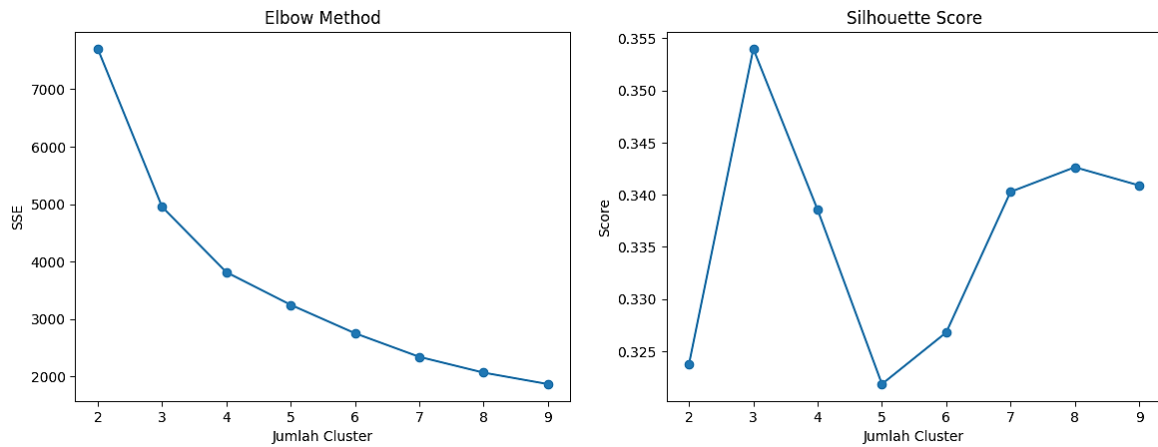
**Gambar 3.** Perbandingan Data Sebelum dan Setelah Normalisasi

Berdasarkan Gambar 3, terlihat bahwa nilai atribut yang semula memiliki rentang berbeda telah berhasil ditransformasikan ke dalam skala 0 hingga 1. Sebagai contoh, nilai price\_asli sebesar 128.75 berubah menjadi 0.249990, sedangkan nilai maksimum seperti 495.80 mendekati 1 yaitu 0.991535. Hal serupa juga terjadi pada atribut discount\_percent, di mana nilai diskon 0% menjadi 0.000000 dan diskon 20% menjadi 0.666667, menunjukkan proses normalisasi berjalan secara proporsional. Pada atribut total\_revenue, nilai 463.52 dinormalisasi menjadi 0.184188, sementara nilai yang lebih tinggi seperti 1264.64 menjadi 0.505187, yang mengindikasikan bahwa perbedaan skala telah dikompresi tanpa menghilangkan pola distribusi data. Dengan demikian, Gambar 3 menunjukkan bahwa proses normalisasi berhasil menghilangkan dominasi variabel dengan skala besar seperti revenue, sehingga setiap fitur memiliki kontribusi yang seimbang dalam proses clustering. Hal ini menjadi langkah krusial untuk memastikan bahwa hasil pengelompokan produk yang dihasilkan oleh algoritma K-Means lebih akurat dan tidak bias terhadap variabel tertentu.

### 3.3 Penentuan Jumlah Cluster

Pada tahap ini dilakukan penentuan jumlah cluster optimal sebagai parameter utama dalam algoritma K-Means Clustering. Penentuan nilai K dilakukan menggunakan dua pendekatan evaluasi, yaitu Elbow Method dan Silhouette Score, guna memastikan bahwa jumlah cluster yang dipilih tidak hanya optimal secara matematis tetapi juga memiliki

kualitas pemisahan yang baik. Hasil visualisasi kedua metode tersebut ditampilkan pada Gambar 4 sebagai dasar dalam menentukan jumlah cluster terbaik.



**Gambar 4.** Grafik Elbow Method dan Silhouette Score

Berdasarkan Gambar 4, pada grafik Elbow Method terlihat bahwa nilai Sum of Squared Errors (SSE) mengalami penurunan yang sangat signifikan dari K=2 ke K=3, kemudian mulai melandai setelah K=3 hingga K berikutnya. Titik ini membentuk pola “siku” (elbow) yang jelas pada K=3, yang mengindikasikan bahwa penambahan jumlah cluster setelah nilai tersebut tidak memberikan peningkatan yang signifikan dalam mengurangi error. Hal ini menunjukkan bahwa secara struktural, K=3 merupakan jumlah cluster yang optimal. Sementara itu, pada grafik Silhouette Score, nilai tertinggi juga diperoleh pada K=3 dengan skor sekitar 0.35, yang menunjukkan bahwa pada jumlah cluster tersebut, tingkat kohesi dalam cluster dan separasi antar cluster berada pada kondisi terbaik dibandingkan dengan nilai K lainnya. Nilai silhouette yang menurun setelah K=3 mengindikasikan bahwa kualitas pengelompokan menjadi kurang optimal ketika jumlah cluster ditambah. Dengan demikian, berdasarkan konsistensi hasil dari kedua metode evaluasi tersebut, dapat disimpulkan bahwa jumlah cluster optimal yang digunakan dalam penelitian ini adalah 3 cluster, yang selanjutnya digunakan dalam proses analisis clustering dan interpretasi pola produk e-commerce.

### 3.4 Hasil Clustering dengan K-Means

Pada tahap ini disajikan hasil pengelompokan produk e-commerce menggunakan algoritma K-Means dengan jumlah cluster optimal sebanyak 3 cluster. Hasil clustering ditunjukkan dalam bentuk rata-rata nilai setiap atribut (price, discount\_percent, dan total\_revenue) pada masing-masing cluster, sehingga dapat digunakan untuk mengidentifikasi karakteristik tiap kelompok produk. Ringkasan hasil clustering tersebut ditampilkan pada Gambar 5 sebagai dasar dalam analisis segmentasi produk.

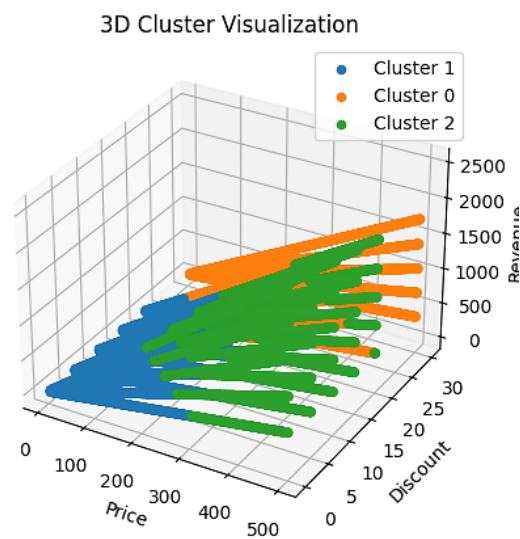
	price	discount_percent	total_revenue
cluster			
0	253.890598	25.943944	523.918572
1	125.632949	8.590791	307.943667
2	383.570016	8.506706	1124.746551

**Gambar 5.** Rata-rata Nilai Setiap Cluster

Berdasarkan Gambar 5, terlihat bahwa setiap cluster memiliki karakteristik yang berbeda secara signifikan. Cluster 2 memiliki nilai rata-rata price tertinggi (383.57) dan total\_revenue tertinggi (1124.75) dengan tingkat diskon relatif rendah (8.51%), sehingga dapat dikategorikan sebagai Cluster Produk Premium, yaitu produk dengan harga tinggi, diskon rendah, dan menghasilkan revenue tinggi. Selanjutnya, Cluster 0 memiliki rata-rata harga sedang (253.89) dengan tingkat diskon paling tinggi (25.94%) serta revenue menengah (523.92), sehingga dapat dikategorikan sebagai Cluster Produk Diskon, yaitu produk dengan strategi promosi aktif untuk meningkatkan penjualan. Sementara itu, Cluster 1 menunjukkan rata-rata harga paling rendah (125.63), diskon rendah (8.59%), dan revenue terendah (307.94), sehingga dikategorikan sebagai Cluster Produk Low Performance, yaitu produk dengan kontribusi penjualan yang relatif rendah. Dengan demikian, Gambar 5.4 menunjukkan bahwa algoritma K-Means berhasil mengelompokkan produk berdasarkan pola harga, diskon, dan pendapatan secara jelas, sehingga dapat digunakan sebagai dasar dalam perumusan strategi bisnis yang lebih tepat sasaran pada masing-masing segmen produk.

Pada tahap ini juga dilakukan visualisasi hasil clustering untuk memberikan gambaran yang lebih intuitif terhadap pola pengelompokan produk berdasarkan tiga atribut utama, yaitu price, discount\_percent, dan total\_revenue. Visualisasi ini disajikan dalam bentuk grafik tiga dimensi (3D) sehingga distribusi dan pemisahan antar cluster dapat

diamati secara lebih jelas. Hasil visualisasi cluster ditampilkan pada Gambar 6 sebagai representasi dari hasil pengelompokan yang telah dilakukan menggunakan algoritma K-Means dengan jumlah cluster optimal yang sebelumnya ditentukan melalui metode Elbow dan Silhouette.



**Gambar 6.** Visualisasi 3D Hasil Clustering

Berdasarkan Gambar 6, terlihat bahwa data terbagi ke dalam tiga kelompok cluster yang memiliki pola distribusi berbeda pada ruang tiga dimensi. Cluster dengan warna tertentu menunjukkan kecenderungan produk dengan harga tinggi dan revenue besar yang terkonsentrasi pada area tertentu, sementara cluster lainnya menunjukkan distribusi produk dengan harga dan revenue yang lebih rendah. Meskipun terdapat beberapa titik yang saling berdekatan, secara umum terlihat adanya pemisahan antar cluster yang cukup jelas, yang menandakan bahwa algoritma K-Means mampu mengelompokkan data berdasarkan kemiripan karakteristik. Selain itu, distribusi data juga menunjukkan bahwa variabel price dan total\_revenue memiliki hubungan yang cenderung searah, di mana peningkatan harga sering diikuti dengan peningkatan revenue, meskipun dipengaruhi oleh faktor diskon.

Jika dikaitkan dengan hasil Elbow Method, pemilihan jumlah cluster sebanyak 3 terbukti cukup representatif karena mampu membentuk kelompok yang tidak terlalu umum (underfitting) maupun terlalu spesifik (overfitting). Hal ini terlihat dari distribusi cluster yang masih dapat dipisahkan secara visual tanpa terjadi fragmentasi berlebihan. Dengan demikian, Gambar 6 memperkuat hasil analisis sebelumnya bahwa penggunaan tiga cluster sudah optimal dan mampu merepresentasikan segmentasi produk e-commerce secara efektif, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan bisnis yang lebih terarah.

Kualitas cluster yang dihasilkan dinilai berdasarkan Davies-Bouldin Index (DBI) yaitu 0.9335 menunjukkan kategori cukup baik, tetapi belum optimal. Secara konsep, DBI mengukur rasio antara kedekatan antar data dalam satu cluster (intra-cluster similarity) dan jarak antar cluster (inter-cluster separation), di mana nilai yang semakin kecil (mendekati 0) menandakan cluster semakin kompak dan semakin terpisah dengan jelas. Jika dilihat, nilai 0.93 masih tergolong sedang, dimana Cluster yang terbentuk sudah memiliki struktur, tetapi masih terdapat overlap atau kemiripan antar cluster. Beberapa produk kemungkinan memiliki karakteristik harga, diskon, dan revenue yang mirip, sehingga tidak sepenuhnya terpisah dengan tegas. Kualitas pemisahan cluster belum maksimal, terutama jika dibandingkan dengan nilai DBI yang ideal ( $< 0.5$ ). Namun, hasil ini tidak buruk dan masih dapat diterima dalam konteks data bisnis seperti e-commerce, karena Data bersifat variatif dan realistis (tidak terdistribusi sempurna). Atribut yang digunakan hanya 3 variabel utama, sehingga pemisahan alami memang terbatas. Jika dikombinasikan dengan hasil Silhouette Score yang cukup baik ( $\sim 0.35$ ), maka model masih layak digunakan untuk analisis segmentasi.

### 3.5 Pembahasan

Hasil penelitian ini menunjukkan bahwa penerapan algoritma K-Means Clustering pada data produk e-commerce dengan atribut price, discount\_percent, dan total\_revenue mampu menghasilkan segmentasi produk yang jelas menjadi tiga kelompok utama, yaitu produk premium, produk berbasis diskon, dan produk dengan performa rendah. Temuan ini sejalan dengan penelitian yang dilakukan oleh (Praditya et al., 2024) yang menyatakan bahwa K-Means efektif dalam mengelompokkan data transaksi untuk menghasilkan segmentasi yang dapat digunakan dalam strategi pemasaran yang lebih terarah. Selain itu, penelitian oleh (Sakinah & Awaliyah, 2025) juga menunjukkan bahwa penggunaan fitur seperti harga dan volume penjualan dalam K-Means mampu menghasilkan segmentasi e-commerce yang optimal ketika jumlah cluster ditentukan menggunakan Silhouette Score.

Lebih lanjut, hasil penelitian ini juga konsisten dengan studi oleh (Almaripat et al., 2025) yang menyatakan bahwa clustering data penjualan dapat membantu perusahaan dalam mengidentifikasi pola produk dan mendukung pengambilan keputusan strategis berbasis data. Namun, perbedaan utama dari penelitian ini dibandingkan penelitian



sebelumnya terletak pada fokus segmentasi, di mana sebagian besar penelitian terdahulu lebih menitikberatkan pada customer segmentation, seperti pada penelitian (Putri et al., 2024) yang menggunakan pendekatan RFM untuk mengelompokkan pelanggan. Sebaliknya, penelitian ini berfokus pada product segmentation, yang memberikan perspektif berbeda dalam analisis bisnis e-commerce, khususnya dalam pengelolaan harga dan strategi promosi produk.

Selain itu, dibandingkan dengan penelitian lain yang mengombinasikan beberapa metode clustering seperti K-Means dan DBSCAN, penelitian ini menunjukkan bahwa K-Means masih relevan dan efektif untuk data numerik terstruktur, meskipun terdapat keterbatasan dalam menangani noise atau data yang tidak terdistribusi secara jelas (Johan, 2025). Nilai evaluasi seperti Silhouette Score (~0.35) dan Davies-Bouldin Index (~0.93) dalam penelitian ini menunjukkan kualitas cluster yang cukup baik, meskipun belum optimal, yang juga sejalan dengan temuan penelitian sebelumnya bahwa performa clustering sangat dipengaruhi oleh pemilihan fitur dan karakteristik data (Gustriansyah et al., 2024).

Dengan demikian, dapat disimpulkan bahwa penelitian ini tidak hanya memperkuat temuan dari studi sebelumnya mengenai efektivitas K-Means dalam analisis data e-commerce, tetapi juga memberikan kontribusi baru dengan menekankan pentingnya segmentasi produk berbasis atribut harga, diskon, dan revenue sebagai dasar dalam perumusan strategi bisnis yang lebih spesifik dan aplikatif.

## 4. KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan produk e-commerce berdasarkan atribut price, discount\_percent, dan total\_revenue menggunakan algoritma K-Means Clustering guna mengidentifikasi pola performa produk dan mendukung pengambilan keputusan bisnis. Berdasarkan hasil analisis, jumlah cluster optimal yang diperoleh adalah 3 cluster, yang ditentukan secara konsisten menggunakan metode Elbow dan Silhouette Score. Hasil clustering menunjukkan adanya tiga karakteristik utama, yaitu cluster produk premium dengan harga tinggi dan revenue tinggi, cluster produk diskon dengan tingkat diskon tinggi dan revenue menengah, serta cluster produk low performance dengan harga dan revenue rendah. Evaluasi model menggunakan Silhouette Score (~0.35) dan Davies-Bouldin Index (~0.93) menunjukkan bahwa kualitas cluster tergolong cukup baik meskipun masih terdapat sedikit overlap antar cluster. Secara keseluruhan, penelitian ini berhasil membuktikan bahwa K-Means efektif dalam melakukan segmentasi produk berbasis data numerik dan mampu menghasilkan insight yang relevan untuk strategi penetapan harga dan promosi. Adapun saran untuk penelitian selanjutnya adalah dengan menambahkan variabel lain seperti quantity\_sold, rating, dan review\_count agar segmentasi menjadi lebih komprehensif dan akurat. Selain itu, penggunaan metode clustering lain seperti Hierarchical Clustering atau DBSCAN dapat dilakukan untuk membandingkan performa model secara lebih mendalam. Penelitian berikutnya juga disarankan untuk melakukan feature engineering serta uji statistik lanjutan guna meningkatkan kualitas cluster. Dari sisi implementasi, hasil penelitian ini dapat dimanfaatkan oleh pelaku e-commerce sebagai dasar dalam menyusun strategi pemasaran yang lebih spesifik, seperti optimalisasi promosi pada cluster diskon dan peningkatan kualitas produk pada cluster premium, sehingga diharapkan mampu meningkatkan total revenue secara keseluruhan.

## REFERENCES

- A'yun, I. Q., Anggraini, L., Asmara, G. D., & Khoirunnisa, R. M. (2024). Analysis of the Development of E-Commerce Transactions in the 6 Highest Transaction Countries in Southeast Asia. *Journal of Economics Research and Social Sciences*, 8(2), 207–221. <https://doi.org/10.18196/jerss.v8i2.22033>
- Allorerung, P. P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 178–191. <https://doi.org/10.14421/jiska.2024.9.3.178-191>
- Almaripat, M., Faqih, A., & Rinaldy, A. R. (2025). Sales Data Classterization Analysis Using K-Means Method for Marketing Strategy Development. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 972–976. <https://doi.org/10.59934/jaiea.v4i2.792>
- Amalina, T., Pramana, D. B. A., & Sari, B. N. (2022). Metode K-Means clustering dalam pengelompokan penjualan produk frozen food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574–583. <https://doi.org/10.5281/zenodo.7052276>
- Apriyanto, B., & Sitio, S. L. M. (2025). Penerapan k-means dalam menganalisis pola pembelian pelanggan pada data transaksi e-commerce. *Bit-Tech*, 7(3), 790–797. <https://doi.org/10.32877/bt.v7i3.2195>
- Azzahra, L., & Yasir, A. (2024). A Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 3(1), 1–10. <https://doi.org/10.70340/jirsi.v3i1.88>
- Caroline, Yuswardi, & Rofi'i, Y. U. (2023). Analysis of e-commerce purchase patterns using big data: An integrative approach to understanding consumer behavior. *International Journal Software Engineering and Computer Science (IJSECS)*, 3(3), 352–364. <https://doi.org/10.35870/ijsecs.v3i3.1840>
- Effendy, I. R., Manurung, Y. A., & Hujaifah, M. I. (2026). Implementasi Algoritma K-means Clustering untuk Mencari Preferensi Pelanggan Toko Online Tazee Clothes. *Jurnal Media Informatika*, 7(1), 69–76. <https://doi.org/10.55338/jumin.v7i1.8054>
- Faizi, M. I., & Adnan, S. M. (2024). Improved segmentation model for melanoma lesion detection using normalized cross-correlation-based k-means clustering. *IEEE Access*, 12, 20753–20766.



<https://doi.org/10.1109/ACCESS.2024.3360223>

- Falih, A. R. F., Kurniawan, R., Wijaya, Y. A., & Anwar, S. (2025). Algoritma K-Mean Untuk Optimalisasi Model Clustering Data Penjualan Toko Online Di Tiktok Shop Dalam Strategi Pemasaran. *Jurnal Sistem Informasi Kaputama (JSIK)*, 9(1), 1–11. <https://doi.org/10.59697/jsik.v9i1.929>
- Gustriansyah, R., Alie, J., & Suhandi, N. (2024). A hybrid machine learning model for market clustering. *Engineering, Technology & Applied Science Research*, 14(6), 18824–18828. <https://doi.org/10.48084/etasr.9259>
- Gwak, G., Hwang, U., & Kim, J. (2025). Clustering of shoulder movement patterns using K-means algorithm based on the shoulder range of motion. *Journal of Bodywork and Movement Therapies*, 41, 164–170. <https://doi.org/https://doi.org/10.1016/j.jbmt.2024.11.034>
- Heidari, J., Daneshpour, N., & Zangeneh, A. (2024). A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers. *Pattern Recognition*, 155, 1–12. <https://doi.org/https://doi.org/10.1016/j.patcog.2024.110639>
- Hendra, H., Hermawan, A., & Edy, E. (2024). Smart product recommendations in web e-commerce: leveraging apriori algorithm for market basket analysis. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 18(3). <https://doi.org/10.22146/ijccs.89075>
- Johan, M. E. (2025). Implementation of Customer Segmentation Model using K-Means and DBSCAN for Fashion Industry Product Transaction. *JOIV: International Journal on Informatics Visualization*, 9(6), 2559–2568. <https://doi.org/10.62527/joiv.9.6.2978>
- Kannan, M. K. J., & Khan, A. (2025). Predictive Analysis and Data-Driven Strategies for Turning Data into Dollars to Visualize ROI Using Retail Intelligence 2.0. *International Journal for Multidisciplinary Research (IJFMR)*, 7(2), 1–17. <https://doi.org/10.36948/ijfmr.2025.v07i02.40721>
- Kusnaldi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 330–338. <https://doi.org/10.47065/josyc.v3i4.2112>
- Mado, P. M. K., & Hendry, H. (2025). Implementasi algoritma clustering K-Means untuk segmentasi pelanggan di e-commerce. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 6(3), 1680–1686. <https://doi.org/10.63447/jimik.v6i3.1563>
- Manarung, R. I., Widodo, E., & Rifai, A. M. (2025). Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs. *International Journal Software Engineering and Computer Science (IJSECS)*, 5(1), 226–234. <https://doi.org/10.35870/ijsecs.v5i1.4090>
- Minh, H.-L., Sang-To, T., Wahab, M. A., & Cuong-Le, T. (2022). A new metaheuristic optimization based on K-means clustering algorithm and its application to structural damage identification. *Knowledge-Based Systems*, 251, 109189. <https://doi.org/https://doi.org/10.1016/j.knosys.2022.109189>
- Moodley, R., Chiclana, F., Caraffini, F., & Carter, J. (2020). A product-centric data mining algorithm for targeted promotions. *Journal of Retailing and Consumer Services*, 54, 101940. <https://doi.org/10.1016/j.jretconser.2019.101940>
- Munshi, A., Alhindi, A., Qadah, T. M., & Alqurashi, A. (2023). An Electronic Commerce Big Data Analytics Architecture and Platform. In *Applied Sciences* (Vol. 13, Issue 19, p. 10962). <https://doi.org/10.3390/app131910962>
- Noval, M., Windarsyah, W., & Marleny, F. D. (2025). Implementasi Algoritma K-Means Untuk Analisis Pola Penjualan Pada Toko Monisa. *Jurnal Media Informatika*, 6(3), 1996–2002. <https://doi.org/10.55338/jumin.v6i3.6237>
- Pandiangan, D. F., & Albina, M. (2025). Model dan Tahapan Penelitian Kuantitatif: Pendekatan Teoretis dan Praktis dalam Kajian Pendidikan. *IHSAN: Jurnal Pendidikan Islam*, 3(3), 724–730. <https://doi.org/10.61104/ihsan.v3i3.1494>
- Praditya, R. G., Sembodo, G., & Heikal, J. (2024). Market segmentation analysis to find out products and services that suit customer needs using the python Kmeans clustering method (Case study: Superindo Tambun Area, Bekasi). *Jurnal Teknik Industri Terintegrasi*, 7(4), 2072–2081. <https://doi.org/10.31004/jutin.v7i4.35889>
- Putri, Y., Aldo, D., & Ilham, W. (2024). Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering. *SINKRON: Jurnal Dan Penelitian Teknik Informatika*, 8(1), 20–28. <https://doi.org/10.33395/sinkron.v8i4.14000>
- Qu, Y. (2025). Research on Purchasing Behavior Pattern of E-commerce Platform Consumers Based on Big Data Analysis. *Advances in Economics, Management and Political Sciences*, 177, 187–191. <https://doi.org/10.54254/2754-1169/2025.22481>
- Sakinah, A., & Awaliyah, D. S. (2025). Optimization E-Commerce Consumer Segmentation Based On K-Means Clustering And Machine Learning. *Journal of Mathematics, Computations and Statistics*, 8(2), 606–619. <https://doi.org/10.35580/jmathcos.v8i2.9548>
- Salman, Z., & Alomary, A. (2024). Performance of the K-means and fuzzy C-means algorithms in big data analytics. *International Journal of Information Technology*, 16(1), 465–470. <https://doi.org/https://doi.org/10.1007/s41870-023-01436-y>
- Shen, W. (2025). Analysis of E-commerce Customer Consumption Data and Traffic Risk Detection Based on User Recommendation Algorithm. *Results in Engineering*, 27, 106255. <https://doi.org/10.1016/j.rineng.2025.106255>



- Sholeh, M., La'i, H. N., Arif, R., Fahtezi, N. A., & Darojah, Z. (2025). Pengelolaan Arsip yang Efektif Terhadap Akurasi Data dan Pengambilan Keputusan di SDN Siwalankerto I. *Jurnal Administrasi Pendidikan Islam*, 7(1), 85–99. <https://doi.org/10.15642/japi.2025.7.1.85-99>
- Siregar, M. N. H., & Khalidy, F. (2024). Utilization of Sales Data Analysis for Product Recommendation Systems in E-Commerce Using the Apriori Algorithm. *Journal of Computer Science Artificial Intelligence and Communications*, 1(2), 41–45. <https://doi.org/10.64803/jocsaic.v1i2.17>
- Suh, Y. (2025). Discovering customer segments through interaction behaviors for home appliance business. *Journal of Big Data*, 12(57), 1–39. <https://doi.org/10.1186/s40537-025-01111-y>
- Suhairi, S., Siregar, M. M., Ningrum, L. D., Bintang, R., & Mutiara, A. (2023). Strategi segmentasi, targeting, dan positioning dalam pasar global: Pendekatan untuk keberhasilan bisnis internasional. *Innovative: Journal Of Social Science Research*, 3(6), 5120–5131. <https://j-innovative.org/index.php/Innovative/article/view/6606>
- Sun, C. (2024). Data Analysis of Customer Segmentation and Personalized Strategy in the Era of Big Data. *Advances in Economics, Management and Political Sciences*, 92, 46–52. <https://doi.org/10.54254/2754-1169/92/20231411>
- Tjia, T. E., Yasir, F. N., & Ekawati, S. (2025). Implementation of Data Mining for Analyzing Consumer Purchasing Patterns at TeTa Ino Café. *Brilliance: Research of Artificial Intelligence*, 5(2), 751–760. <https://doi.org/10.47709/brilliance.v5i2.6767>
- Wan, B., Huang, W., Pierre, B., Cheng, Y., & Zhou, S. (2024). K-Means algorithm based on multi-feature-induced order. *Granular Computing*, 9(2), 45. <https://doi.org/https://doi.org/10.1007/s41066-024-00470-w>
- Wu, R. (2024). Behavioral analysis of electricity consumption characteristics for customer groups using the k-means algorithm. *Systems and Soft Computing*, 6, 200143. <https://doi.org/https://doi.org/10.1016/j.sasc.2024.200143>
- Yoseph, F., Ahamed Hassain Malim, N. H., Heikkilä, M., Brezulianu, A., Geman, O., & Paskhal Rostam, N. A. (2020). The impact of big data market segmentation using data mining and clustering techniques. *Journal of Intelligent & Fuzzy Systems*, 38(5), 6159–6173. <https://doi.org/10.3233/JIFS-179698>
- Yuliarnis, S. K., Hendriyani, Y., Kurniadi, D., & Giatman, M. (2020). Application of data mining for analysis of consumer purchase data on sales transaction data at halal mart hni hpai dharmasraya. *Jurnal Pendidikan Teknologi Kejuruan*, 3(1), 68–75. <https://doi.org/10.24036/jptk.v3i1.6923>
- Zulkurnain, D. R., & Eryanto, H. (2025). Analisis Sistem Pengelolaan Arsip Inaktif Pada Lembaga Sosial XYZ. *Musytari: Jurnal Manajemen, Akuntansi, Dan Ekonomi*, 22(10), 141–150. <https://doi.org/10.2324/jy3q6085>