

# DOKUMENTASI SEGMENT PASAR DENGAN K- MEANS CLUSTERING



**Dibuat oleh:**

Ricki Kosjono

**Untuk:**

*Pelatihan Fundamental dan Intermediate Data Sains*

Diselenggarakan oleh **Digitalent 2025**

2025

## Daftar Isi

<b>1. Tujuan.....</b>	<b>1</b>
<b>2. Data yang Digunakan .....</b>	<b>1</b>
<b>3. Eksplorasi Data (EDA) .....</b>	<b>2</b>
<b>4. Preprocessing.....</b>	<b>3</b>
<b>5. Menentukan Jumlah Cluster Optimal .....</b>	<b>3</b>
<b>6. Hasil Clustering (K=3).....</b>	<b>5</b>
<b>7. Visualisasi .....</b>	<b>6</b>
<b>8. Kesimpulan.....</b>	<b>6</b>

## Daftar Gambar

Gambar 1. Boxplot Membership vs Total Spend.....	2
Gambar 2. Grafik Elbow .....	4
Gambar 3. silhouette score.....	4
Gambar 4. Scatterplot Clustering.....	6


## 1. Tujuan

Tujuan dari analisis ini adalah untuk melakukan segmentasi pelanggan berdasarkan perilaku belanja mereka pada dataset E-commerce Customer Behavior. Melalui proses segmentasi, perusahaan dapat:

1. Mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa.
2. Memahami perbedaan pola belanja antar segmen.
3. Menyusun strategi pemasaran yang lebih tepat sasaran, baik untuk retensi, upselling, maupun aktivasi pelanggan.

Dengan demikian, hasil analisis ini diharapkan mampu mendukung pengambilan keputusan bisnis secara lebih efektif.

## 2. Data yang Digunakan

Analisis ini menggunakan dataset **E-commerce Customer Behavior** yang diambil dari Kaggle:  [E-commerce Customer Behavior Dataset – Kaggle](#). Analisis ini menggunakan dataset E-commerce Customer Behavior yang terdiri dari 350 baris data dengan 11 kolom. Setiap kolom merepresentasikan atribut yang berkaitan dengan profil maupun perilaku pelanggan, yaitu:

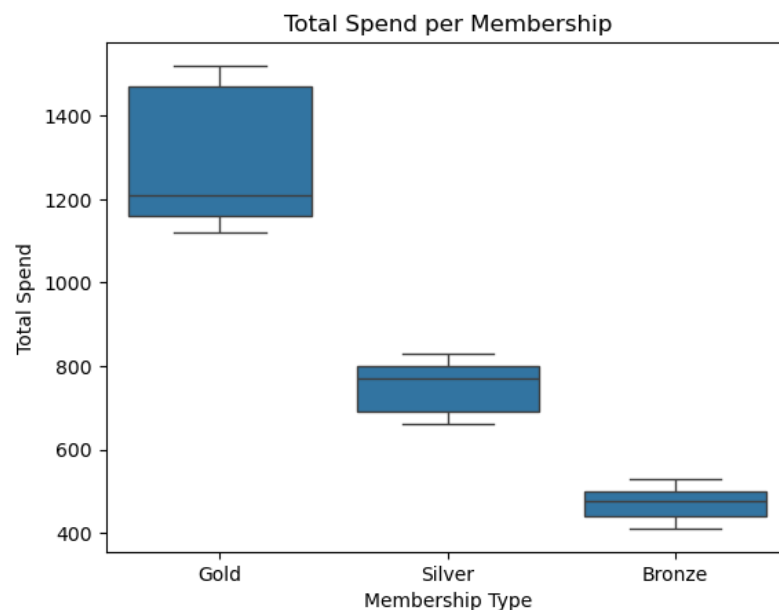
1. Customer ID : identitas unik setiap pelanggan.
2. Gender : jenis kelamin pelanggan (Male / Female).
3. Age : usia pelanggan.
4. City : kota asal pelanggan.
5. Membership Type : tipe keanggotaan (Bronze, Silver, Gold).
6. Total Spend : total pengeluaran pelanggan.
7. Items Purchased : jumlah barang yang dibeli.
8. Average Rating : rata-rata rating yang diberikan pelanggan.
9. Discount Applied : status penggunaan diskon (True / False).
10. Days Since Last Purchase : jumlah hari sejak transaksi terakhir.
11. Satisfaction Level : tingkat kepuasan pelanggan (Satisfied, Neutral, Unsatisfied).

Dataset ini memberikan gambaran menyeluruh mengenai profil pelanggan, pola belanja, serta tingkat kepuasan, sehingga relevan digunakan untuk segmentasi pelanggan dengan pendekatan clustering.

### 3. Eksplorasi Data (EDA)

Untuk memahami perbedaan perilaku konsumen berdasarkan level membership, dilakukan analisis menggunakan boxplot yang menggambarkan distribusi pengeluaran (Total Spend) pada tiap kategori membership. Visualisasi ini membantu melihat kecenderungan pola belanja serta perbedaan rata-rata pengeluaran antar kelompok.

Gambar 1. Boxplot Membership vs Total Spend



Gambar 1. Boxplot Membership vs Total Spend

Berdasarkan hasil analisis boxplot, dapat diperoleh beberapa temuan penting:

1. Kecenderungan Rata-Rata Pengeluaran

Terlihat bahwa semakin tinggi level membership (Gold → Silver → Bronze), semakin besar pula rata-rata pengeluaran (Total Spend).

2. Dominasi Membership Gold

Anggota dengan membership Gold memiliki tingkat pengeluaran yang secara signifikan lebih tinggi dibandingkan anggota dengan membership Silver maupun Bronze.

3. Perbedaan yang Jelas Antar Kelompok

Perbedaan distribusi ini menunjukkan adanya potensi segmentasi pelanggan berdasarkan membership, di mana kelompok Gold cenderung menjadi penyumbang pengeluaran terbesar.

#### **4. Preprocessing**

Sebelum proses clustering dilakukan, data perlu dipersiapkan melalui beberapa tahap pra-pemrosesan agar hasil analisis lebih akurat dan dapat dibandingkan secara adil antarvariabel. Langkah-langkah yang dilakukan antara lain:

1. Konversi Variabel Kategori

Variabel kategori seperti Gender, City, dan Membership Type diubah menjadi numerik menggunakan metode Label Encoding, sehingga dapat diproses oleh algoritma berbasis numerik.

2. Standardisasi Fitur

Seluruh fitur kemudian distandardisasi menggunakan StandardScaler agar berada pada skala yang sama. Hal ini penting dilakukan untuk mencegah perbedaan rentang nilai pada tiap variabel memengaruhi hasil clustering.

3. Fitur yang Digunakan

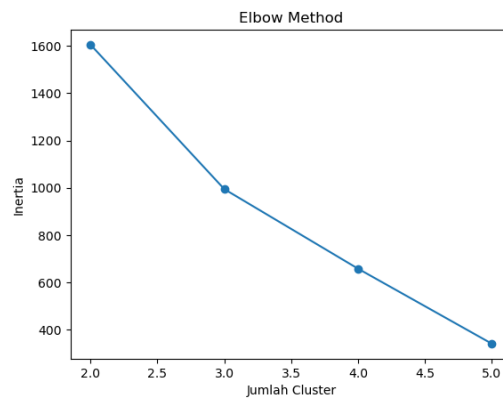
Setelah melalui tahap pra-pemrosesan, fitur yang digunakan dalam analisis clustering adalah sebagai berikut:

1. Age
2. Gender
3. City
4. Membership Type
5. Total Spend
6. Items Purchased
7. Average Rating
8. Discount Applied
9. Days Since Last Purchase

#### **5. Menentukan Jumlah Cluster Optimal**

Untuk menentukan jumlah cluster yang optimal dalam analisis K-Means, digunakan dua metode evaluasi, yaitu Elbow Method dan Silhouette Score. Kedua pendekatan ini membantu dalam menyeimbangkan aspek statistik dengan kebutuhan interpretasi bisnis.

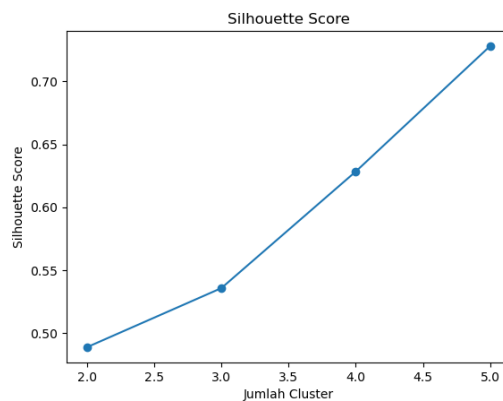
## 1. Elbow Method



Gambar 2. Grafik Elbow

Berdasarkan gambar 2, grafik Elbow memiliki titik siku yang terlihat pada  $k = 3$ . Meskipun nilai inertia terus menurun setelahnya, penurunan tersebut tidak lagi signifikan, sehingga  $k=3$  dianggap sebagai titik optimal.

## 2. Silhouette Score



Gambar 3. silhouette score

Pada gambar 3, hasil perhitungan silhouette score menunjukkan adanya peningkatan skor seiring bertambahnya jumlah cluster. Namun, apabila jumlah cluster terlalu banyak, interpretasi hasil menjadi kurang jelas dan tidak praktis untuk kebutuhan segmentasi pasar.

Dengan mempertimbangkan kedua metode tersebut, dipilih  $k = 3$  sebagai jumlah cluster yang optimal. Keputusan ini memberikan keseimbangan antara akurasi statistik (inertia dan silhouette score) dengan kemudahan interpretasi dalam konteks bisnis.

## 6. Hasil Clustering (K=3)

Setelah dilakukan proses K-Means dengan jumlah cluster optimal  $k = 3$ , diperoleh hasil segmentasi pelanggan seperti pada Tabel 1.1 berikut:

Table 1.1 Tabel Hasil Cluster

Cluster	Rata-rata Total Spend	Items Purchased	Days Since Last Purchase	Karakteristik
0	473	8	31	Usia lebih tua, belanja rendah, jarang transaksi
1	1311	18	18	Usia muda, belanja tinggi, aktif bertransaksi
2	748	12	30	Usia menengah, belanja sedang, aktivitas moderat

Berdasarkan tabel di atas, diperoleh interpretasi sebagai berikut:

1. Cluster 1 – High Value Customers (Gold-like)
  - Total belanja tinggi.
  - Banyak barang yang dibeli.
  - Sering bertransaksi (days since last purchase rendah).
  - Cocok untuk program loyalty premium atau exclusive offers.
2. Cluster 2 – Mid Value Customers (Silver-like)
  - Pengeluaran dan jumlah barang pada level menengah.
  - Aktivitas belanja cukup rutin.
  - Potensial untuk ditingkatkan menjadi pelanggan premium melalui promosi atau bundling produk.
3. Cluster 0 – Low Value Customers (Bronze-like)
  - Pengeluaran rendah.
  - Belanja jarang dan jumlah barang sedikit.
  - Dapat ditargetkan dengan promosi diskon atau kampanye retensi untuk meningkatkan keterlibatan.

## 7. Visualisasi

Untuk mendukung interpretasi hasil clustering, beberapa visualisasi digunakan sehingga pola data dapat terlihat lebih jelas. Visualisasi yang digunakan antara lain:

### 1. Boxplot Membership vs Total Spend

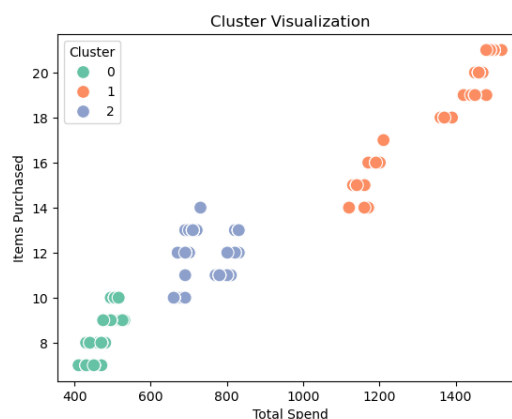
Menunjukkan perbedaan rata-rata pengeluaran antar membership. Dari boxplot terlihat bahwa membership Gold cenderung memiliki pengeluaran lebih tinggi dibandingkan Silver dan Bronze. Visualisasi ini dapat dilihat pada gambar 2.

### 2. Elbow Method & Silhouette Score

Digunakan untuk menentukan jumlah cluster yang optimal. Hasil menunjukkan bahwa  $k = 3$  merupakan pilihan terbaik karena memberikan keseimbangan antara nilai inertia dan kemudahan interpretasi bisnis. Visualisasi ini dapat dilihat pada gambar 3.

### 3. Scatterplot Clustering (Total Spend vs Items Purchased)

Memperlihatkan pemisahan cluster yang cukup jelas. Pola yang muncul sejalan dengan interpretasi segmentasi Gold – Silver – Bronze, di mana pelanggan dengan pengeluaran tinggi dan pembelian banyak terkelompok pada cluster bernilai tinggi (Gold-like), sementara cluster lainnya mewakili pelanggan menengah dan rendah. Scatterplot Clustering ini dapat dilihat pada gambar 4 sebagai berikut.



Gambar 4. Scatterplot Clustering

## 8. Kesimpulan

Berdasarkan hasil analisis clustering menggunakan metode K-Means, diperoleh beberapa temuan utama sebagai berikut:



1. Segmentasi pelanggan berhasil membagi konsumen menjadi tiga kelompok utama, yaitu High Value, Mid Value, dan Low Value Customers.
2. Hasil cluster memiliki keselarasan dengan kategori membership yang sudah ada (Gold, Silver, Bronze), sehingga memberikan validasi bahwa data bisnis mendukung hasil clustering.

Dari hasil tersebut, dapat ditarik beberapa implikasi strategis untuk mendukung keputusan pemasaran perusahaan:

1. Retensi pelanggan Gold (High Value) dengan memberikan benefit eksklusif, program loyalitas premium, atau layanan personalisasi.
2. Upselling pelanggan Silver (Mid Value) agar dapat ditingkatkan ke level Gold melalui penawaran menarik, bundling produk, atau promosi berjenjang.
3. Aktivasi pelanggan Bronze (Low Value) dengan strategi diskon, promosi khusus, atau kampanye retensi untuk meningkatkan frekuensi belanja.