

## SEGMENTASI PASAR TERHADAP GENERASI Z BERDASARKAN PERILAKU BELANJA ONLINE MENGGUNAKAN METODE K-MEANS CLUSTERING

<sup>1</sup>Meri Nova Marito<sup>✉</sup>, <sup>1</sup>Ratna Wati Simbolon, <sup>1</sup>Duma Lasmaria Siagian,  
<sup>1</sup>Bertha Nerpy Siahaan, <sup>2</sup>Ade Linhar P., <sup>1</sup>Anugrah Zai

<sup>1</sup>Universitas Mandiri Bina Prestasi, Medan, Indonesia

<sup>2</sup>Universitas Islam Negeri Syekh Ali Hasan Ahmad Addary, Padangsidempuan, Indonesia

Email: [merinovamarito@yahoo.com](mailto:merinovamarito@yahoo.com)

DOI: <https://doi.org/10.46880/jmika.Vol10No1.pp244-248>

### ABSTRACT

*The advancement of digital technology greatly affects the consumption patterns of society, especially generation Z, towards e-commerce platforms. generation Z has a high affinity for digital technology. This condition creates a need for more precise market segmentation so that marketing strategies can be tailored to consumer characteristics. This study aims to conduct market segmentation based on online shopping behavior using the K-Means Clustering method. The data applied for this study is the Online Retail Dataset obtained from the UCI Machine Learning Repository. The research stages include data preprocessing, attribute selection, data normalization, determination of the number of clusters using the Elbow Method, evaluation of cluster quality using the Silhouette Score, as well as the clustering process using the K-Means algorithm. The variables analyzed include transaction frequency, number of product purchases, customer transaction value, and the interval or time gap of transactions. The research results show that customer data can be grouped into several segments with different characteristics, namely customers with high purchase levels, customers with moderate transaction levels, and customers with low purchase activity. The evaluation of cluster quality using the Silhouette Score obtained a value of 0.8957, indicating that the quality of the resulting clusters is considered very good. This, the resulting customer segmentation can help business actors in understanding consumer characteristics and in developing more effective and targeted marketing strategies.*

**Keyword:** Market Segmentation, Generation Z, Online Shopping Behavior, Data Mining, K-Means Clustering.

### ABSTRAK

*Dengan kemajuan teknologi digital sangat mempengaruhi pola konsumsi masyarakat khususnya generasi Z terhadap platform E-commerce. Generasi Z memiliki kedekatan tinggi dengan teknologi digital. Kondisi ini menimbulkan kebutuhan akan segmentasi pasar yang lebih tepat agar strategi pemasaran dapat disesuaikan dengan karakteristik konsumen. Penelitian ini bertujuan untuk melakukan segmentasi pasar berdasarkan perilaku belanja online menggunakan metode K-Means Clustering. Data yang diterapkan untuk penelitian ini adalah Online Retail Dataset yang diperoleh dari repositori UCI Machine Learning Repository. Tahapan penelitian meliputi preprocessing data, pemilihan atribut, normalisasi data, penentuan jumlah cluster menggunakan Elbow Method, evaluasi kualitas cluster menggunakan Silhouette Score, serta proses clustering menggunakan algoritma K-Means. Variabel yang dianalisis meliputi frekuensi transaksi, jumlah pembelian produk, nilai transaksi pelanggan, dan interval atau jarak waktu transaksi. Hasil penelitian bahwa data pelanggan dapat dikelompokkan ke dalam beberapa segmen dengan karakteristik yang berbeda, yaitu pelanggan dengan tingkat pembelian tinggi, pelanggan dengan tingkat transaksi sedang, dan pelanggan dengan aktivitas pembelian rendah. Hasil evaluasi kualitas cluster menggunakan Silhouette Score memperoleh nilai sebesar 0,8957 yang menunjukkan bahwa kualitas cluster yang dihasilkan tergolong sangat baik. Dengan demikian, segmentasi pelanggan yang dihasilkan dapat membantu pelaku bisnis dalam memahami karakteristik konsumen serta menyusun strategi pemasaran yang lebih efektif dan tepat sasaran.*

**Kata Kunci:** Segmentasi Pasar, Generasi Z, Perilaku Belanja Online, Data Mining, K-Means Clustering.

### PENDAHULUAN

Perkembangan e-commerce mampu menghasilkan perubahan terhadap perilaku konsumen dalam melakukan transaksi digital (Afriani et al.,

2025). Kehadiran platform e-commerce nantinya akan memudahkan masyarakat dalam kegiatan online. Kemudahan tersebut mengakibatkan meningkatnya aktivitas belanja online dan menghasilkan data



transaksi dalam jumlah besar yang dapat dimanfaatkan sebagai sumber informasi dalam pengambilan keputusan bisnis (Sandora, 2020).

Perubahan perilaku konsumen dalam melakukan transaksi digital juga dipengaruhi oleh perkembangan gaya hidup masyarakat modern, terutama pada Generasi Z yang dikenal memiliki tingkat adaptasi tinggi terhadap teknologi dan internet. Generasi Z merupakan kelompok konsumen yang aktif menggunakan platform digital dalam aktivitas belanja online (Sativa, 2024). Karakteristik konsumen digital pada generasi ini umumnya dipengaruhi oleh kemudahan akses, kecepatan transaksi, promosi, tren media sosial, serta variasi produk yang tersedia pada platform e-commerce (Febriani et al., 2026). Kondisi tersebut menjadikan perilaku belanja online sebagai aspek penting untuk diketahui oleh pelaku bisnis dalam menentukan strategi pemasaran yang tepat sasaran.

Dalam persaingan bisnis digital yang semakin kompetitif, perusahaan tidak hanya dituntut untuk meningkatkan kualitas produk dan layanan, tetapi juga memahami karakteristik konsumennya secara lebih mendalam. Dan peningkatan volume data transaksi pelanggan pada platform e-commerce sering kali menjadi tantangan terhadap proses analisis perilaku konsumen (Arnold et al., 2022). Data transaksi yang besar dan kompleks memerlukan teknik pengolahan data yang mampu menghasilkan informasi secara efektif dan terstruktur. Jika data tersebut tidak dimanfaatkan dengan baik, maka potensi informasi yang dapat digunakan untuk mendukung strategi bisnis menjadi kurang maksimal.

Data mining merupakan salah satu pendekatan yang digunakan dalam analisis perilaku pelanggan berbasis data transaksi online (Soleh & Jonas, 2025). Teknik yang dapat digunakan untuk melakukan pengelompokan data adalah metode K-Means Clustering. Metode ini merupakan algoritma clustering yang bekerja dengan mengelompokkan data ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik antar data. K-Means Clustering sering digunakan dalam analisis segmentasi pelanggan karena memiliki proses perhitungan yang relatif sederhana, dan mampu menghasilkan pengelompokan data yang baik (Aprilianur & Hadisaputro, 2022).

Penelitian terdahulu mengatakan bahwa metode K-Means Clustering mampu mengelompokkan pelanggan berdasarkan pola transaksi online retail sehingga membantu perusahaan dalam menentukan strategi pemasaran yang lebih efektif (Awalina & Rahayu, 2023). Selain itu, penelitian lain mengatakan bahwa segmentasi pelanggan menggunakan pendekatan K-Means clustering bertujuan untuk

memahami karakteristik pelanggan berdasarkan perilaku transaksi dan tingkat pembelian pelanggan. Karakteristik pelanggan berdasarkan pola transaksi, frekuensi pembelian, maupun nilai transaksi yang dilakukan pelanggan (Awalina & Rahayu, 2023).

Dari ke-2 penelitian tersebut memiliki pembaharuan terhadap analisis segmentasi perilaku belanja online Generasi Z dengan menambahkan variabel interval atau jarak waktu transaksi untuk memperkuat identifikasi pola perilaku pelanggan pada transaksi e-commerce. Penambahan variabel tersebut bertujuan untuk menghasilkan segmentasi pelanggan yang lebih detail dalam menggambarkan intensitas aktivitas belanja online pelanggan yang merepresentasikan karakteristik konsumen digital Generasi Z menggunakan metode K-Means Clustering.

Untuk data yang digunakan dalam penelitian ini adalah Online Retail Dataset yang diperoleh dari UCI Machine Learning Repository (Chen, 2025). Dataset tersebut berisi data transaksi pelanggan pada platform retail online yang digunakan untuk menganalisis pola pembelian konsumen berdasarkan atribut atau variabel transaksi yaitu frekuensi transaksi, jumlah produk, total belanja dan interval atau jarak waktu transaksi. Penelitian ini tidak menggunakan data usia pelanggan secara langsung karena dataset tidak menyediakan atribut umur pelanggan. Oleh karena itu, representasi Generasi Z didasarkan pada karakteristik perilaku konsumen digital dalam transaksi belanja online.

Penelitian ini diharapkan dapat diperoleh kelompok pelanggan dengan karakteristik perilaku belanja online yang berbeda sehingga hasil segmentasi dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih efektif, efisien, dan sesuai dengan kebutuhan konsumen digital pada era modern.

## **KAJIAN PUSTAKA**

### **K-Means**

K-Means merupakan sebuah metode clustering non hirarki dimana kegiatan utama yang dilakukan ialah mengelompokkan data sesuai dengan kriteria yang sama dan dilakukan pengelompokkan ke dalam sebuah cluster data, untuk data yang berbeda dimasukkan kedalam sebuah cluster yang berbeda (Gustientiedina et al., 2019). Metode ini efektif digunakan contohnya untuk menghitung segmentasi pelanggan.

Alur algoritma ini, dengan menentukan jumlah K yang diinginkan lalu tentukan titik pusat atau centroid secara acak namun setelah tahap iterasi dilakukan rumus sebagai berikut:

1. Tentukan jumlah cluster (K) yang diinginkan
2. Tentukan titik centroid awal dengan acak

- Lakukan perhitungan nilai centroid dengan nilai rata-rata data pada cluster dengan rumus:

$$V_{ij} = \frac{\sum X_{kj}}{N_i}$$

Keterangan:

$V_{ij}$  = centroid/titik pusat cluster ke-i variabel ke-j

$N_i$  = jumlah anggota cluster ke-i

$X_{kj}$  = nilai data ke-k variabel ke-j

- Lakukan perhitungan jarak data dan centroid dengan rumus Euclidean Distance:

$$D_e = \sqrt{(x - s)^2 + (y - t)^2}$$

Keterangan:

$D_e$  = jarak antara data dengan centroid

$(x, y)$  = koordinat data

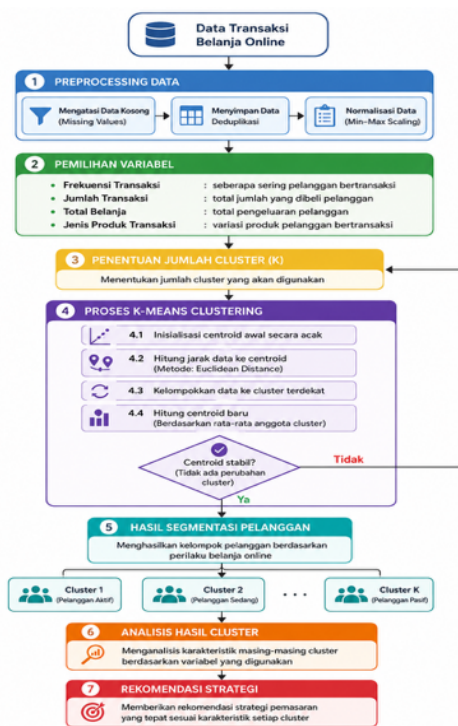
$(s, t)$  = koordinat centroid

- Lakukan penggabungan data dengan jarak terdekat terhadap centroid
- Lakukan iterasi dengan melakukan perhitungan Kembali terhadap centroid baru berdasarkan anggota cluster yang sudah terbentuk.

Dari proses iterasi dapat dilakukan proses berulang hingga nilai centroid stabil dan tidak terjadi perubahan anggota cluster, yang nantinya centroid bernilai optimal.

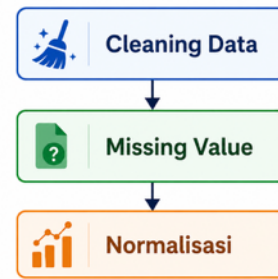
### METODE PENELITIAN

Dalam hal ini alur kerja terhadap metode yang diusulkan pada penelitian Segmentasi Pasar terhadap generasi Z berdasarkan perilaku belanja online ini dapat digambarkan pada gambar 1 sebagai berikut:



Gambar 1. Model yang Diusulkan

Dengan alur kerja diatas, dilakukan preprocessing data dengan alur seperti berikut:



Gambar 2. Preprocessing Data

Data awal yang digunakan pada penelitian ini adalah diperoleh dari UCI Machine Learning Repository yaitu online retail dataset (Chen, 2011) dengan 541.909 data dan 5 variabel aktif yaitu CustomerID, frekuensi transaksi, jumlah produk, total belanja dan interval atau jarak waktu transaksi. Ditahap ini dilakukan preprocessing untuk mengatasi missing value, outlier dan duplikasi di dalam dataset. Atribut CustomerID memiliki missing value atau null value sebesar 20%. Dengan dilakukan preprocessing data nilai ini dianggap nilai besar maka di hapus dari dataset.

Sehingga variabel yang digunakan untuk transaksi ini meliputi frekuensi transaksi, jumlah produk, total belanja dan interval atau jarak waktu transaksi.

Tabel 1. Kategori Variabel

Variabel	Fungsi
InvoiceNo (Frekuensi Transaksi)	Intensitas belanja online
Quantity (Jumlah Produk)	Pola pembelian
Total Belanja (Q*Unit Prize)	Tingkat pengeluaran
InvoiceDate (Jarak Waktu Transaksi)	Aktivitas pelanggan terbaru

### HASIL DAN PEMBAHASAN

Terdapat 4 variabel yang ada pada dataset retail ini. Dengan dataset sebanyak 541.909 item. Namun dipakai 2 variabel yaitu Quantity /Jumlah Produk (Sumbu X) dan Total Belanja/Q\*Unit Prize (Sumbu Y). Variabel tersebut digunakan untuk mengetahui pola perilaku belanja pelanggan berdasarkan jumlah produk yang dibeli dan total nilai transaksi pelanggan.

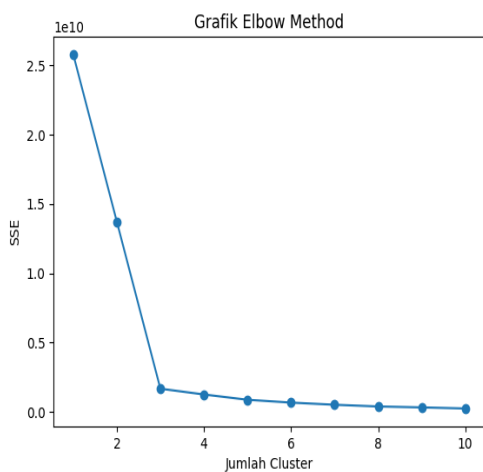
Tabel 2. Tabel Transaksi

	Sumbu X	Sumbu Y
0	6	15.30
1	6	20.34
2	8	22.00

3	6	20.34
4	6	20.34
5	2	15.30
6	6	25.50
7	6	11.10
8	6	11.10
9	32	54.08
10	6	12.60
11	6	12.60
12	8	30.00
13	6	9.90
14	6	25.50
15	3	14.85
16	2	19.90
17	3	17.85
18	3	17.85
19	4	31.80

Setelah dilakukan preprocessing data dengan data diatas, maka dilakukan penentuan jumlah cluster. Penentuan jumlah cluster menggunakan metode Elbow Method. Dimana metode ini digunakan untuk menentukan jumlah cluster optimal berdasarkan nilai Sum of Squared Error (SSE) pada setiap jumlah cluster yang diuji.

Pengujian dilakukan dengan membandingkan beberapa jumlah cluster mulai dari K=1 hingga K=10. Hasil pengujian Elbow Method ditunjukkan pada Gambar 3 seperti berikut ini:



Gambar 3. Grafik Elbow

Dari grafik ini, terlihat bahwa penurunan nilai SSE terjadi secara signifikan hingga cluster ke-3 dan mulai melandai setelahnya. Kondisi tersebut menunjukkan bahwa jumlah cluster optimal pada penelitian ini adalah sebanyak 3 cluster. Oleh karena itu, proses pengelompokan data pada metode K-Means

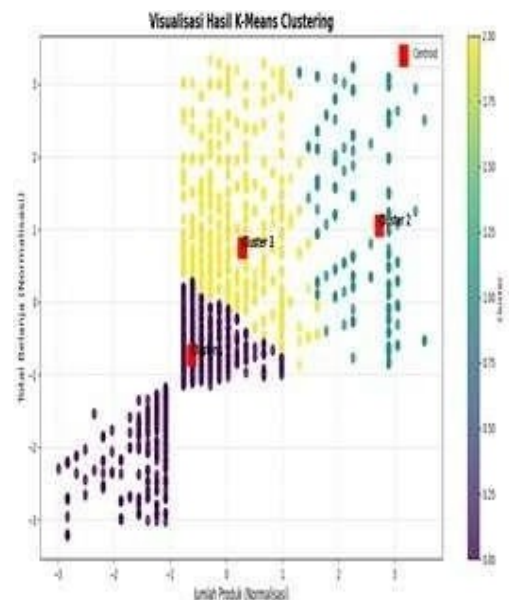
Clustering dilakukan menggunakan tiga cluster untuk menghasilkan segmentasi pelanggan yang lebih optimal.

Untuk Evaluasi kualitas cluster pada penelitian ini dilakukan secara kuantitatif menggunakan metode Silhouette Score. Nilai Silhouette Score berada pada nilai 0.8957424485014445, artinya nilai yang mendekati 1 menunjukkan kualitas cluster yang sangat baik. karena data dalam setiap cluster memiliki tingkat kemiripan yang tinggi dan perbedaan antar cluster yang cukup jelas. Perhitungan evaluasi sebagai berikut :

```
[3]: data = pd.read_excel("Online_Retail.xlsx")
[4]: X = data[['Quantity']]
[9]: X_sample = X.sample(1000, random_state=42)
[10]: from sklearn.cluster import KMeans
      from sklearn.metrics import silhouette_score
[11]: kmeans = KMeans(n_clusters=3, random_state=42)
      labels = kmeans.fit_predict(X_sample)
[12]: score = silhouette_score(X_sample, labels)
      print(score)
0.8957424485014445
```

Gambar 4. Perhitungan Evaluasi Cluster

Hasil proses clustering menghasilkan tiga kelompok pelanggan dengan karakteristik perilaku belanja online yang berbeda. Untuk hasil visualisasi cluster dapat dilihat dengan gambar sebagai berikut :



Gambar 5. Visualisasi Hasil K-Means

Berdasarkan hasil visualisasi K-Means Clustering, data pelanggan berhasil dikelompokkan menjadi tiga cluster berdasarkan variabel jumlah

produk dan total belanja yang telah dinormalisasi. Dengan penjelasan sebagai berikut :

1. Cluster 1

Bahwa pelanggan dengan jumlah pembelian dan total belanja relatif rendah. Kelompok ini merupakan pelanggan dengan aktivitas transaksi rendah sehingga strategi yang dapat diterapkan adalah promosi diskon dan peningkatan loyalitas pelanggan.

2. Cluster 2

Bahwa terdiri dari pelanggan dengan jumlah produk dan total belanja tinggi. Cluster ini termasuk pelanggan potensial atau pelanggan prioritas karena memberikan kontribusi transaksi yang besar. Strategi yang dapat diterapkan yaitu pemberian layanan premium, membership, dan promo eksklusif.

3. Cluster 3

Bahwa pelanggan dengan tingkat pembelian sedang hingga tinggi. Kelompok ini berpotensi ditingkatkan menjadi pelanggan prioritas.

Hasil clustering menunjukkan bahwa metode K-Means mampu memisahkan karakteristik pelanggan dengan cukup baik sehingga bermanfaat bagi perusahaan dalam memberikan keputusan agar strategi pemasaran tepat sasaran.

## KESIMPULAN

Hasil clustering menginformasikan bahwa pelanggan dapat dikelompokkan menjadi beberapa segmen, yaitu pelanggan dengan aktivitas pembelian tinggi, sedang, dan rendah. Segmentasi tersebut nantinya dapat menentukan strategi pemasaran, promosi, dan pelayanan yang lebih tepat sasaran berdasarkan karakteristik masing-masing pelanggan

## DAFTAR PUSTAKA

- Afriani, W. N., Hermawan, A., & Jasman, J. (2025). Dampak E-commerce terhadap Perilaku Konsumen: Studi Perbandingan Studi Kasus Belanja Online dan Offline di Kabupaten Bima. *MUQADDIMAH: Jurnal Ekonomi, Manajemen, Akuntansi Dan Bisnis*, 3(1), 197–208. <https://doi.org/10.59246/muqaddimah.v3i1.1233>
- Aprilianur, G., & Hadisaputro, E. L. (2022). Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Myam Hijab Penajam. *JUPITER (Jurnal Penelitian Ilmu Dan Teknik Komputer)*, 14(1), 161–170.
- Arnold, D. R. S., Sukmawati, W. I., & Fitri, S. D. (2022). Dampak e-commerce terhadap perilaku konsumen dan strategi bisnis. *Jurnal Manajemen Pemasaran Internasional*, 1(2), 57–68.

- Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122–137. <https://doi.org/10.34010/jati.v13i2.10090>
- Chen, D. (2025). *Online Retail Dataset*.
- Febriani, T., Falakhi, A., Hapsari, M., & Riza, A. F. (2026). Karakteristik perilaku belanja online antar generasi Y, Z dan ALPHA. *Jurnal Ilmiah Ekonomi Dan Manajemen*, 4(1), 178–189.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/TEKNOSI.v5i1.2019.17-24>
- Sandora, M. (2020). Pengaruh Kepercayaan Dan Kemudahan Terhadap Keputusan Pembelian Belanja Secara Online. *Eko Dan Bisnis: Riau Economic and Business Review*, 11(3), 290–310.
- Sativa, O. (2024). *Budaya Berbelanja Online Generasi Z dan Generasi Milenial (Studi Kasus Produk Fashion Wanita di Online Shop Shopatcloth)* [Thesis]. Universitas Islam Indonesia.
- Soleh, O., & Jonas, D. (2025). Penggunaan Algoritma K-Means untuk Segmentasi Data Pelanggan pada Sistem Pemasaran Berbasis Data Mining. *Nusa: Journal of Science Studies*, 1(2), 63–69. <https://doi.org/10.59613/0e5cay32>