

# Implementasi Algoritma *Clustering K-Means* untuk Segmentasi Pelanggan di *E-Commerce*

Priscianus Mikael Kia Mado <sup>1\*</sup>, Hendry <sup>2</sup>

<sup>1\*,2</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

*Email:* priscianusmikaelkiamado@gmail.com <sup>1\*</sup>, hendry@uksw.edu d <sup>2</sup>

## Histori Artikel:

*Dikirim* 26 Juli 2025; *Diterima dalam bentuk revisi* 7 Agustus 2025; *Diterima* 10 Agustus 2025; *Diterbitkan* 10 September 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

## Abstrak

Dalam era digital yang semakin maju, persaingan di dunia e-commerce menuntut perusahaan untuk memahami perilaku pelanggan secara mendalam agar dapat mempertahankan loyalitas dan meningkatkan penjualan. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan e-commerce dengan menerapkan algoritma clustering K-means menggunakan analisis RFM (Recency, Frequency, Monetary). Data transaksi pelanggan diolah melalui tahap pra-pemrosesan seperti pembersihan data dan normalisasi, kemudian diterapkan algoritma K-means untuk mengelompokkan pelanggan ke dalam segmen-segmen yang homogen berdasarkan karakteristik perilaku pembelian mereka. Pengelompokan yang optimal didapatkan dengan menggunakan metrik evaluasi Silhouette Score, menghasilkan tiga segmen utama pelanggan. Hasil segmentasi ini dapat membantu perusahaan dalam merancang strategi pemasaran yang lebih efektif dan terfokus sesuai kebutuhan masing-masing segmen pelanggan.

**Kata Kunci:** Algoritma Clustering K-Means; Segmentasi Pelanggan; E-Commerce.

## Abstract

In the increasingly advanced digital era, competition in the e-commerce world requires companies to understand customer behavior in depth in order to maintain loyalty and increase sales. This study aims to segment e-commerce customers by applying the K-means clustering algorithm using RFM (Recency, Frequency, Monetary) analysis. Customer transaction data is processed through pre-processing stages such as data cleaning and normalization, then the K-means algorithm is applied to group customers into homogeneous segments based on their purchasing behavior characteristics. Optimal grouping is obtained using the Silhouette Score evaluation metric, resulting in three main customer segments. The results of this segmentation can help companies design more effective and focused marketing strategies according to the needs of each customer segment.

**Keyword:** K-Means Clustering Algorithm; Customer Segmentation; E-Commerce.

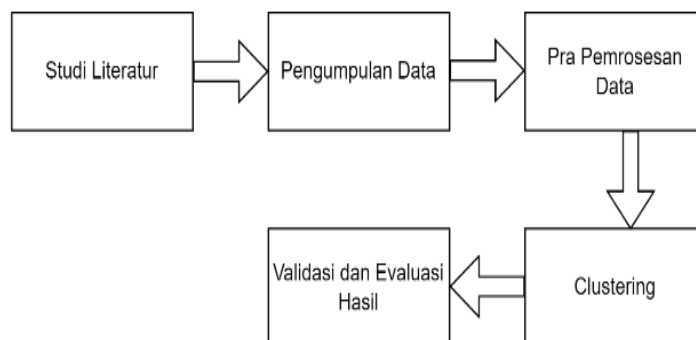
## 1. Pendahuluan

Dalam era digital yang semakin maju, dunia e-commerce menghadapi persaingan yang semakin ketat. Untuk tetap kompetitif, perusahaan perlu memahami perilaku pelanggan secara lebih mendalam untuk meningkatkan loyalitas dan kepuasan mereka. Proses ini sangat penting mengingat masyarakat telah memasuki fase revolusi industri 5.0, yang ditandai oleh integrasi teknologi canggih yang mempercepat berbagai aktivitas sehari-hari. Teknologi ini juga memudahkan akses ke dunia digital, sehingga mendorong terwujudnya konsep *Society 5.0* yang menekankan keseimbangan antara manusia dan teknologi modern. Di Indonesia, adopsi teknologi semakin meluas, terutama dengan peningkatan jumlah pengguna internet yang sangat signifikan, yang mencapai 185,3 juta pada tahun 2024 (F. Rizti, 2025). Pandemi Covid-19 menjadi faktor utama yang mempercepat pergeseran kegiatan ekonomi ke ranah daring, mengubah cara orang berinteraksi dan bertransaksi secara mendalam. Hal ini menunjukkan pentingnya pengembangan infrastruktur digital untuk mendukung kegiatan ekonomi masyarakat yang semakin bergantung pada teknologi. Dengan ini, bisnis daring atau e-commerce tumbuh dengan pesat, baik di Indonesia maupun di tingkat global. Platform-platform e-commerce seperti Shopee, Tokopedia, dan TikTok kini memiliki jutaan pengguna aktif, dengan Shopee tercatat memiliki 235,9 juta pengguna (Fatika, 2025) dan Tokopedia sekitar 100,3 juta pengguna (eDot, 2025). Namun, meskipun e-commerce berkembang pesat, tantangan yang dihadapi oleh pelaku e-commerce semakin beragam. Salah satu pendekatan strategis yang dapat digunakan untuk menghadapi tantangan ini adalah segmentasi pelanggan. Segmentasi memungkinkan perusahaan untuk mengelompokkan pelanggan berdasarkan perilaku atau karakteristik tertentu, seperti kebiasaan berbelanja atau kebutuhan spesifik. Dengan segmentasi yang tepat, perusahaan dapat merancang strategi pemasaran yang lebih terfokus dan sesuai dengan kebutuhan tiap segmen, yang pada gilirannya dapat meningkatkan efektivitas pemasaran dan memperkuat posisi perusahaan di pasar yang semakin kompetitif.

Salah satu metode yang digunakan untuk melakukan segmentasi pelanggan adalah algoritma clustering K-Means, yang sering dipadukan dengan analisis RFM (*Recency, Frequency, Monetary*). Pendekatan ini memungkinkan pengelompokan pelanggan berdasarkan karakteristik pembelian mereka, di mana K-Means mengelompokkan data yang serupa dalam satu kluster, sementara analisis RFM menganalisis tiga faktor utama waktu terakhir pembelian (*Recency*), frekuensi pembelian (*Frequency*), dan total pengeluaran (*Monetary*) untuk mengidentifikasi segmen pelanggan yang berbeda (Febri, Ningsih, & Lemantara, 2021; Collins *et al.*, 2021; Techtarget, 2025). Kombinasi kedua metode ini memberikan pemahaman yang lebih jelas tentang perilaku pelanggan, yang dapat membantu perusahaan dalam merancang strategi pemasaran yang lebih efisien dan terukur. Penerapan segmentasi menggunakan algoritma K-Means dan analisis RFM memungkinkan perusahaan untuk mengidentifikasi pola-pola pembelian yang berbeda di antara pelanggan mereka. Hasil dari pengelompokan ini dapat dijadikan dasar untuk merancang promosi yang lebih spesifik untuk tiap segmen, memperkuat upaya retensi pelanggan yang berisiko beralih ke pesaing, dan meningkatkan penggunaan sumber daya yang lebih efisien (Pramudiansyah, 2021; Siagian *et al.*, 2021). Selain itu, strategi berbasis data ini juga dapat mempercepat proses pengambilan keputusan strategis dalam menghadapi tantangan pasar yang dinamis. Dalam hal ini, segmentasi pelanggan tidak hanya membantu perusahaan menarik pelanggan baru, tetapi juga memperpanjang hubungan dengan pelanggan yang sudah ada, yang menjadi faktor penting dalam meningkatkan keberhasilan bisnis e-commerce di era digital. Seiring dengan kemajuan teknologi digital yang semakin pesat, penggunaan metode analisis data seperti K-Means clustering dan *RFM analysis* menjadi lebih relevan untuk memastikan daya saing bisnis tetap terjaga. Penggunaan metode-metode ini sangat strategis di tengah cepatnya adopsi teknologi digital di Indonesia dan global, yang mendorong perusahaan untuk selalu berinovasi dan memahami perilaku konsumen secara lebih tepat dan mendalam (Silamantha & Hadiono, 2024; Suharti *et al.*, 2022).

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma K-means Clustering untuk segmentasi pelanggan *e-commerce* berdasarkan analisis RFM (*Recency, Frequency, Monetary*). Tujuan utama penelitian ini adalah untuk memperoleh pengelompokan pelanggan yang homogen dan strategis. Tahapan-tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan-tahapan Penelitian

Langkah pertama dalam penelitian ini adalah melakukan studi literatur untuk merangkum teori-teori dan pendekatan-pendekatan yang relevan terkait dengan analisis RFM (*Recency, Frequency, Monetary*) serta penerapan algoritma *K-Means Clustering* dalam segmentasi pelanggan *e-commerce*. Setelah itu, data transaksi pelanggan dikumpulkan dari platform Kaggle, yang mencakup informasi mengenai nilai total transaksi selama periode waktu tertentu. Setelah data terkumpul, dilakukan tahap pra-pemrosesan untuk memastikan kualitas data yang digunakan. Tahap ini mencakup pembersihan data (*cleaning data*) untuk menghilangkan duplikat, nilai yang hilang, atau data yang tidak relevan (*outlier*), serta normalisasi data menggunakan metode Z-Score atau Min-Max Scaling agar variabel memiliki skala yang seragam. Jika diperlukan, tahap feature engineering juga dilakukan untuk membuat fitur baru, seperti menghitung skor RFM untuk setiap pelanggan. Setelah data siap, algoritma K-Means Clustering diterapkan untuk mengelompokkan pelanggan berdasarkan karakteristik RFM mereka. Proses *clustering* ini melibatkan beberapa langkah, seperti menentukan jumlah kluster optimal (*k*) menggunakan metode *Elbow*, *Silhouette Analysis*, serta beberapa indeks evaluasi lainnya, seperti *Calinski-Harabasz Index*, *Davies-Bouldin Index*, dan *Dunn Index*. Hasil *clustering* kemudian dievaluasi menggunakan metrik *Silhouette Score* untuk mengukur kualitas kluster berdasarkan kerapian dan jarak antar kluster.

## 3. Hasil dan Pembahasan

### 3.1 Hasil

Hasil penerapan algoritma K-Means clustering pada data RFM menunjukkan bahwa pelanggan berhasil dikelompokkan ke dalam empat segmen utama berdasarkan perilaku mereka. Setiap kluster memiliki karakteristik yang unik, yang tercermin dari nilai rata-rata *Recency, Frequency, dan Monetary*. Analisis lebih lanjut dilakukan terhadap masing-masing kluster untuk memberikan label yang sesuai, yang kemudian divisualisasikan pada Gambar 2.

Ringkasan Segmen:

Segment	JumlahPelanggan	Recency	Frequency	Monetary
At Risk	184	68.440217	18.293478	4.441504e+07
Best Customers	1085	19.344700	28.344700	7.282650e+07
Loyal Customers	410	31.565854	22.729268	5.687091e+07
Potential Loyalist	321	39.551402	20.439252	5.046065e+07

Gambar 2. Nilai rata-rata *Recency*, *Frequency*, dan *Monetary*

Visualisasi hasil clustering memperkuat temuan segmentasi dengan menunjukkan pemisahan yang jelas antar kelompok pelanggan. Segmentasi ini memberikan wawasan strategis yang penting bagi perusahaan untuk merancang kampanye pemasaran yang lebih tepat sasaran, meningkatkan loyalitas pelanggan, serta memaksimalkan nilai bisnis dari setiap segmen yang teridentifikasi.

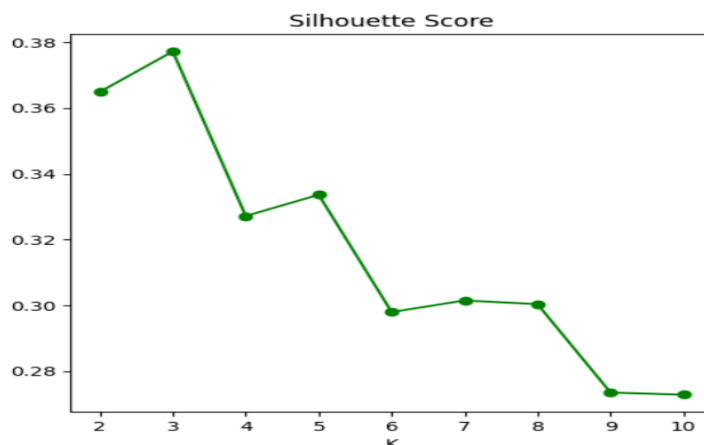
### 3.2 Pembahasan

Analisis RFM (*Recency*, *Frequency*, dan *Monetary*) adalah Teknik yang digunakan untuk menentukan peringkat dan pengelompokan pelanggan berdasarkan perilaku dalam melakukan transaksi.

	CustomerID	Recency	Frequency	Monetary
0	CUST00001	1	22	63882371
1	CUST00002	18	23	57227590
2	CUST00003	22	28	76991379
3	CUST00004	10	33	87579881
4	CUST00005	21	33	96718054

Gambar 3. Hasil analisis menggunakan RFM

Data pada Gambar 3 merupakan hasil analisis menggunakan RFM (*Recency*, *Frequency*, *Monetary*) untuk mengelompokkan pelanggan berdasarkan perilaku transaksi mereka. Dari analisis ini, terlihat bahwa setiap pelanggan memiliki perilaku transaksi yang sangat variatif. Berdasarkan Gambar 3, analisis ini bertujuan untuk menentukan seberapa sering pelanggan melakukan pembelian pada suatu produk, mengukur frekuensi pembelian produk oleh pelanggan, serta mengetahui total transaksi pembelian yang dilakukan oleh pelanggan pada produk tertentu. Dalam menentukan jumlah kluster optimal (K) pada algoritma K-Means, metode validasi yang digunakan adalah *silhouette score*, yang mengukur seberapa dekat suatu objek dengan kluster lain dibandingkan dengan kluster yang ditempatinya. Metode ini membantu dalam memilih jumlah kluster terbaik, di mana nilai skor siluet yang lebih tinggi mengindikasikan segmentasi pelanggan yang lebih jelas dan efektif. Dengan mengandalkan *silhouette score*, hasil pengelompokan pelanggan menjadi lebih akurat dan memberikan makna yang lebih dalam untuk analisis segmentasi.



Gambar 4. Silhouette Scores

Setelah menentukan jumlah kluster yang optimal, langkah selanjutnya adalah melakukan normalisasi data untuk mengelompokkan karakteristik pelanggan berdasarkan RFM (*Recency, Frequency, Monetary*) menggunakan algoritma K-Means. Metode yang dipilih untuk menentukan jumlah kluster adalah *Silhouette Index*, dengan jumlah kluster optimal sebanyak tiga, karena metode ini memberikan keseimbangan dan kejelasan dalam segmentasi pelanggan berdasarkan nilai RFM. Setelah jumlah kluster ditentukan, *K-Means Clustering* diterapkan untuk mengelompokkan pelanggan ke dalam segmen-segmen berdasarkan kemiripan nilai RFM yang telah dinormalisasi. Setiap pelanggan kemudian diberi label kluster yang dianalisis berdasarkan rata-rata nilai RFM masing-masing, sehingga menghasilkan segmen pelanggan yang bermakna, seperti pelanggan loyal, berisiko, atau potensial. Hasil segmentasi ini dapat dilihat pada Gambar 5.

```
Distribusi Segmentasi:
Segment
Best Customers          1085
Loyal Customers         410
Potential Loyalist      321
At Risk                  184
Name: count, dtype: int64
```

Gambar 5. Hasil Segmentasi

#### 4. Kesimpulan

Analisis RFM dan segmentasi pelanggan memberikan wawasan yang berharga mengenai perilaku konsumen serta nilai strategis dari setiap kelompok pelanggan. Dengan menerapkan analisis RFM dan algoritma K-Means clustering, perusahaan berhasil mengelompokkan pelanggan ke dalam segmen-segmen yang berbeda berdasarkan perilaku belanja mereka. Penentuan jumlah kluster optimal, yang dilakukan menggunakan metrik evaluasi seperti *Silhouette Score*, menghasilkan empat segmen utama yang merepresentasikan karakteristik pelanggan secara jelas. Hasil segmentasi ini menjadi dasar yang kuat bagi perusahaan untuk merancang strategi pemasaran yang lebih efektif dan efisien, seperti program loyalitas untuk pelanggan setia, promosi khusus untuk pelanggan potensial,

serta upaya re-engagement bagi pelanggan yang berisiko. Pendekatan berbasis data ini terbukti efektif dalam meningkatkan retensi pelanggan, mendorong pertumbuhan bisnis yang berkelanjutan, serta meningkatkan efisiensi pemasaran. Dengan demikian, perusahaan dapat mengembangkan hubungan yang lebih baik dengan setiap segmen pelanggan dan mencapai pertumbuhan bisnis yang optimal.

## 5. Daftar Pustaka

- Ardi, R. B., Nastiti, F. E., & Sumarlinda, S. (2023). Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus: Fashion Viral Solo). *INFOTECH journal*, 9(1), 124-131.
- 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 Informatika*, 13(2), 122-137. <https://doi.org/10.34010/jati.v13i2.10090>.
- Cuhwanto, Y. A., & DA, R. (2021). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Petir*, 15(1), 48-56.
- Fatika, R. A. (2025). negara dengan pengguna TikTok terbesar, Indonesia urutan berapa. *GoodStats Data. Diambil*, 30.
- Febri, A., Ningsih, N., & Lemantara, J. (2021). Aplikasi Analisis Segmentasi Pelanggan untuk Menentukan Strategi Pemasaran Menggunakan Kombinasi Metode K-Means dan Model RFM. *SISTEMASI*, 10(1), 139-151. <https://doi.org/10.32520/stmsi.v10i1.1123>.
- Fitriyani, R., Firmansyah, A. L., & Nurfadillah, L. A. (2024, October). Penerapan Algoritma Clustering untuk Segmentasi Pelanggan E-commerce berdasarkan Data Pembelian dan Aktivitas. In *SEMINAR TEKNOLOGI MAJALENGKA (STIMA)* (Vol. 8, pp. 372-379). <https://doi.org/10.31949/stima.v8i0.1129>.
- Gustrianda, R., & Mulyana, D. I. (2022). Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids. *J. Media Inform. Budidarma*, 6(1), 27.
- Hafidz Ardana, C., Khoyum, A. A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1-9.
- Pramudiansyah, A. P. (2021). Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Azyariyah Mandar*, 7(2), 06-19. <https://doi.org/10.35329/jiik.v7i2.201>.
- Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *Intecom*, 1(1), 72-77.
- Romadhona, W., Nugroho, B. I., & Murtopo, A. A. (2022). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Jurnal Minfo Polgan*, 11(2), 100-104. <https://doi.org/10.33395/jmp.v11i2.11797>.
- Siagian, R., Sirait, P. S. P., & Halima, A. (2021). E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model. *Journal of informatics and telecommunication engineering*, 5(1), 21-30. <https://doi.org/10.31289/jite.v5i1.5182>.

- Silamantha, W. A., & Hadiono, K. (2024). Analisis rfm dan k-means clustering untuk segmentasi pelanggan pada pt. sanutama bumi arto. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 5(3), 1297-1305. <https://doi.org/10.30645/kesatria.v5i3.448>.
- Suharti, P. H., Suryandari, A. S., & Amalia, R. N. (2022). Analisis kinerja modul pengendali tekanan udara PCT-14 Berbasis PLC dengan berbagai metoda tuning. *Sebatik*, 26(2), 420-427. <https://doi.org/10.46984/sebatik.v26i2.2134>.
- Tazkiyah, I., Arifiyanti, A. A., & Najaf, A. R. E. (2023, November). Implementasi Segmentasi Pelanggan E-Commerce Menggunakan Algoritma K-Means Pada Website. In *Prosiding Seminar Nasional Teknologi dan Sistem Informasi* (Vol. 3, No. 1, pp. 217-226). <https://doi.org/10.33005/sitasi.v3i1.651>.