

# Analisis Segmentasi Pelanggan E-Commerce Menggunakan Metode Clustering Berbasis RFM

Casto Uripto<sup>\*1</sup>, Putri Mayang<sup>2</sup>, Adelia Alvi Yana<sup>3</sup>

<sup>\*1</sup>Bisnis Digital, Politeknik Bisnis Digital Indonesia, Bogor

<sup>2</sup>Logistik Bisnis, Politeknik Bisnis Digital Indonesia, Bogor

<sup>3</sup>Teknologi Rekayasa Perangkat Lunak, Politeknik Bisnis Digital Indonesia, Bogor  
e-mail: <sup>\*1</sup>castouripto@gmail.com, <sup>2</sup>mayangputri85@gmail.com, <sup>3</sup>marvin.malaika@gmail.com

---

## Abstrak

Pertumbuhan pesat industri e-commerce menghasilkan volume data transaksi pelanggan yang besar dan kompleks. Pemanfaatan data tersebut secara optimal menjadi tantangan bagi perusahaan dalam menyusun strategi pemasaran berbasis perilaku pelanggan. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan e-commerce menggunakan pendekatan Recency, Frequency, Monetary (RFM) yang dikombinasikan dengan algoritma K-Means clustering. Dataset yang digunakan adalah Brazilian E-Commerce Public Dataset (Olist) yang diperoleh dari Kaggle. Tahapan penelitian meliputi preprocessing data, perhitungan nilai RFM, normalisasi menggunakan Min-Max Scaling, penentuan jumlah cluster menggunakan metode Elbow, serta evaluasi model menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa segmentasi berbasis RFM mampu mengelompokkan pelanggan ke dalam beberapa cluster dengan karakteristik berbeda. Evaluasi model menghasilkan nilai Silhouette Score sebesar 0,278 yang menunjukkan kualitas cluster cukup baik. Segmentasi yang dihasilkan dapat digunakan sebagai dasar penyusunan strategi retensi pelanggan, reaktivasi pelanggan berisiko churn, serta optimalisasi pemasaran berbasis data.

**Kata Kunci:** Customer Segmentation, RFM, K-Means, Data Mining, E-Commerce

## Abstract

The rapid growth of the e-commerce industry has generated a massive volume of customer transaction data, creating opportunities as well as challenges for businesses in understanding customer behavior. Effective customer segmentation is essential to support data-driven marketing strategies, improve customer retention, and optimize customer lifetime value. This study aims to analyze customer segmentation using the Recency, Frequency, Monetary (RFM) model combined with the K-Means clustering algorithm. The dataset used in this research is the Brazilian E-Commerce Public Dataset (Olist) obtained from Kaggle. The research process includes data preprocessing, RFM calculation, Min-Max normalization, determination of the optimal number of clusters using the Elbow method, clustering using K-Means, and model evaluation using the Silhouette Score. The results show that RFM-based clustering successfully groups customers into four distinct segments with different behavioral characteristics. The model evaluation produced a Silhouette Score of 0.278, indicating a moderate clustering structure. The resulting customer segments provide actionable insights for marketing strategies, including customer retention programs, re-engagement strategies for at-risk customers, and targeted promotional campaigns. This study demonstrates that integrating the RFM model with K-Means clustering is effective in generating interpretable and strategic customer segmentation in e-commerce environments.

**Keywords:** Customer Segmentation, RFM Model, K-Means Clustering, Data Mining, E-Commerce Analytics

---

## I. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pertumbuhan pesat sektor e-commerce yang menghasilkan volume data transaksi pelanggan dalam jumlah besar dan kompleks. Kondisi ini menuntut perusahaan untuk mampu mengelola serta

menganalisis data pelanggan secara efektif guna mempertahankan daya saing bisnis (Tazkiyah et al., 2023). Dalam praktiknya, banyak perusahaan masih menghadapi kendala dalam mengidentifikasi karakteristik pelanggan secara spesifik sehingga strategi pemasaran yang diterapkan cenderung bersifat umum dan kurang tepat sasaran (Sulistiyawati

& Sadikin, 2021). Padahal, segmentasi pelanggan merupakan pendekatan strategis yang memungkinkan perusahaan mengelompokkan pelanggan berdasarkan kesamaan perilaku dan karakteristik tertentu untuk meningkatkan efektivitas pemasaran dan loyalitas pelanggan. Dalam konteks e-commerce modern, segmentasi pelanggan juga berperan penting dalam personalisasi pengalaman pengguna dan optimalisasi *customer lifetime value* (Gani, 2024).

Salah satu pendekatan yang banyak digunakan dalam segmentasi pelanggan adalah teknik data mining, khususnya metode clustering. Clustering merupakan metode pengelompokan data tanpa label (unsupervised learning) yang bertujuan membentuk kelompok berdasarkan tingkat kemiripan karakteristik antar objek dalam dataset (Ahsina et al., 2022). Algoritma K-Means menjadi salah satu metode yang paling populer karena kemampuannya dalam mengolah data berukuran besar secara efisien dan menghasilkan segmentasi yang relatif stabil (Ahsina et al., 2022). Selain itu, metode K-Medoids juga digunakan sebagai alternatif yang lebih tahan terhadap noise dan outlier dalam data transaksi (Sulistiyawati & Sadikin, 2021). Penerapan clustering dalam analisis pelanggan memungkinkan perusahaan mengidentifikasi pola tersembunyi dalam perilaku transaksi yang dapat dijadikan dasar dalam penyusunan strategi pemasaran berbasis data.

Berbagai penelitian sebelumnya telah menerapkan metode clustering untuk segmentasi pelanggan tanpa secara eksplisit mengintegrasikan model RFM sebagai kerangka utama analisis. Implementasi K-Means dalam segmentasi pelanggan e-commerce berbasis data transaksi menunjukkan bahwa teknik ini mampu menghasilkan kelompok pelanggan berdasarkan pola pembelian dan nilai transaksi tertentu (Tazkiyah et al., 2023). Penelitian lain juga memanfaatkan pendekatan clustering pada data transaksi online retail dengan mempertimbangkan variabel numerik umum seperti frekuensi transaksi dan total pembelian (Awalina & Rahayu, 2023). Selain itu, pendekatan berbasis representasi data dan pembelajaran lanjutan telah dikembangkan untuk membangun persona pelanggan secara lebih kompleks, meskipun interpretasi hasilnya cenderung lebih sulit dipahami secara manajerial (Wei et al., 2010). Temuan-temuan

tersebut menunjukkan bahwa clustering efektif dalam mengelompokkan pelanggan, namun pendekatan variabel yang digunakan masih bervariasi.

Meskipun demikian, metode segmentasi yang hanya mengandalkan variabel numerik umum memiliki keterbatasan dalam menggambarkan dimensi perilaku pelanggan secara menyeluruh. Model clustering tanpa struktur evaluasi berbasis waktu transaksi kurang sensitif terhadap aspek keterkinian pembelian pelanggan, sehingga sulit membedakan pelanggan aktif dan pelanggan yang berpotensi churn (Wei et al., 2010). Selain itu, interpretasi hasil cluster seringkali memerlukan analisis tambahan agar dapat diterjemahkan menjadi strategi pemasaran yang operasional (Sulistiyawati & Sadikin, 2021). Dalam praktik e-commerce modern, segmentasi yang tidak mempertimbangkan dimensi perilaku utama dapat mengurangi efektivitas personalisasi dan penargetan kampanye (Gani, 2024).

Untuk mengatasi keterbatasan tersebut, model RFM (*Recency, Frequency, Monetary*) dikembangkan sebagai pendekatan yang menekankan tiga dimensi utama perilaku pelanggan, yaitu kedekatan waktu transaksi terakhir, intensitas pembelian, dan nilai moneter yang dihasilkan pelanggan (Wei et al., 2010). Model ini memberikan kerangka yang lebih terstruktur dalam mengevaluasi loyalitas dan nilai pelanggan karena mempertimbangkan aspek waktu dan kontribusi finansial secara simultan. Integrasi model RFM dengan metode clustering seperti K-Means atau K-Medoids memungkinkan pembentukan segmen pelanggan yang lebih informatif, mudah diinterpretasikan, serta relevan untuk pengambilan keputusan strategis (Tazkiyah et al., 2023) (Sulistiyawati & Sadikin, 2021) Oleh karena itu, penelitian ini mengusulkan penerapan segmentasi pelanggan e-commerce menggunakan metode clustering berbasis RFM pada *Brazilian E-Commerce Public Dataset (Olist)* guna menghasilkan segmentasi yang lebih akurat dan aplikatif dalam mendukung strategi pemasaran berbasis data..

## II. METODE PENELITIAN

### Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif dan eksploratif berbasis data mining. Pendekatan kuantitatif digunakan karena penelitian ini mengolah data numerik berupa data transaksi pelanggan e-commerce untuk menghasilkan segmentasi berbasis perhitungan matematis dan algoritmik. Sementara itu, pendekatan deskriptif bertujuan untuk menggambarkan karakteristik masing-masing segmen pelanggan yang terbentuk, dan pendekatan eksploratif digunakan untuk menemukan pola tersembunyi dalam data transaksi melalui teknik clustering.

Metode yang digunakan adalah metode clustering berbasis model RFM (Recency, Frequency, Monetary) dengan algoritma K-Means sebagai teknik pengelompokan utama.

### Dataset

Dataset yang digunakan merupakan Brazilian E-Commerce Public Dataset (Olist) yang diunduh melalui platform Kaggle. Dataset ini berisi data transaksi e-commerce dari perusahaan Olist di Brasil yang mencakup informasi pelanggan, pesanan, pembayaran, dan produk. Pelanggan yang tercatat dalam Brazilian E-Commerce Public Dataset (Olist), khususnya pelanggan yang memiliki riwayat transaksi yang lengkap dan valid.

Objek penelitian adalah data transaksi pelanggan yang mencakup:

1. Tanggal pembelian (untuk menghitung Recency),
2. Jumlah transaksi (untuk menghitung Frequency),
3. Total nilai pembelian (untuk menghitung Monetary).

Unit analisis dalam penelitian ini adalah masing-masing pelanggan (customer\_id)

### Perhitungan Model RFM

Perhitungan RFM dilakukan sebagai berikut:

1. Recency (R)

Menghitung selisih hari antara tanggal transaksi terakhir pelanggan dengan tanggal referensi (cut-off date).

$$R = \text{Tanggal Referensi} - \text{Tanggal Transaksi Terakhir}$$

2. Frequency (F)

Menghitung jumlah transaksi yang dilakukan pelanggan.

$$F = \text{Jumlah Transaksi}$$

3. Monetary (M)

Menghitung total nilai pembelian pelanggan.

$$M = \sum \text{total pembayaran}$$

Karena skala R, F, dan M berbeda, dilakukan normalisasi menggunakan metode Min-Max Scaling:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Sedangkan Jumlah cluster optimal ditentukan menggunakan metode elbow.

### Proses Clustering Menggunakan K-Means

Algoritma K-Means dilakukan melalui langkah:

1. Menentukan jumlah cluster (k),
2. Menginisialisasi centroid awal,
3. Menghitung jarak menggunakan Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

4. Mengelompokkan data ke centroid terdekat,
5. Memperbarui centroid,
6. Mengulangi hingga konvergen.

Setiap cluster dianalisis berdasarkan nilai rata-rata RFM untuk menentukan karakteristik segmen pelanggan, seperti *Champions*, *loyal Customers*, *At Risk*, *Potential Customers*.

### Teknik analisis data

Teknik analisis data yang digunakan meliputi analisis deskriptif yang dipakai untuk menggambarkan distribusi nilai RFM. Lalu Analisis Clustering (K-Means) untuk membentuk segmen pelanggan berdasarkan kemiripan nilai RFM. Selanjutnya ada evaluasi model, serta visualisasi data.

## III. HASIL DAN PEMBAHASAN

### Deskripsi dataset

Dataset yang digunakan dalam penelitian ini adalah Brazilian E-Commerce Public Dataset (Olist) yang diperoleh dari Kaggle. Dataset ini terdiri dari beberapa tabel utama, yaitu customers, orders, order\_items, dan order\_payments. Setelah proses

integrasi data, unit analisis ditetapkan pada tingkat pelanggan (customer\_id).

Setelah tahap data cleaning dan preprocessing dilakukan, diperoleh sejumlah pelanggan aktif yang memiliki data transaksi lengkap dan valid. Data kemudian diagregasi per pelanggan untuk menghitung nilai Recency, Frequency, dan Monetary (RFM).

### Hasil Perhitungan RFM

tabel 1 merupakan sampel hasil agregasi RFM pada beberapa pelanggan

**Tabel 1.** Hasil perhitungan RFM

Cust_ID	Recency (hari)	Frequency (kali)	Monetary (RS)
C001	15	12	2,450.00
C002	120	3	540
C003	45	7	1,120.00
C004	10	20	3,800.00
C005	200	1	150

Interpretasi awal menunjukkan bahwa pelanggan dengan nilai Recency rendah dan Frequency serta Monetary tinggi cenderung merupakan pelanggan bernilai tinggi.

### Normalisasi Data

Karena skala variabel R, F, dan M berbeda secara signifikan, dilakukan normalisasi menggunakan metode Min-Max Scaling agar setiap variabel berada pada rentang 0–1.

**Tabel 2** Hasil Normalisasi RFM

Cust_ID	R (Norm)	F (Norm)	M (Norm)
C001	0.08	0.58	0.61
C002	0.55	0.1	0.12
C003	0.22	0.32	0.28
C004	0.05	1	1
C005	1	0	0

Nilai yang telah dinormalisasi kemudian digunakan sebagai input dalam proses clustering.

### Proses Clustering

Jumlah cluster optimal ditentukan menggunakan Metode Elbow dengan melihat nilai Sum of Squared Errors (SSE). Hasil pengujian menunjukkan bahwa titik siku (elbow point) terbentuk pada  $k = 4$ . Oleh karena itu, jumlah cluster yang digunakan dalam penelitian ini adalah 4 cluster.

Proses clustering dilakukan menggunakan algoritma K-Means dengan jarak Euclidean. Setelah iterasi konvergen, diperoleh pembagian pelanggan ke dalam 4 cluster.

**Tabel 3.** Hasil Pembagian Cluster

Cust_ID	R (Norm)	F (Norm)	M (Norm)	Cluster
C001	0.08	0.58	0.61	2
C002	0.55	0.1	0.12	3
C003	0.22	0.32	0.28	1
C004	0.05	1	1	2
C005	1	0	0	4

Untuk memahami karakteristik profil masing-masing cluster, dihitung rata-rata nilai RFM pada setiap cluster.

**Tabel 4.** Hasil Pembagian Cluster

Cluster	R (Avg)	F (Avg)	M (Avg)	Karakteristik
1	Sedang	Sedang	Sedang	Pelanggan Potensial
2	Rendah	Tinggi	Tinggi	Champions / Loyal
3	Tinggi	Rendah	Rendah	At Risk
4	Sangat Tinggi	Sangat Rendah	Sangat Rendah	Lost Customers

Penjelasan/interpretasi cluster:

Cluster 1: Pelanggan Potensial

Memiliki nilai RFM sedang. Pelanggan dalam kelompok ini masih aktif dan memiliki potensi untuk ditingkatkan melalui promosi atau loyalty program.

Cluster 2: *Champions / Loyal Customers*

Memiliki Recency rendah (baru bertransaksi), Frequency tinggi, dan Monetary tinggi. Kelompok ini merupakan pelanggan bernilai tinggi yang perlu dipertahankan melalui strategi retensi dan program eksklusif.

**Cluster 3: At Risk Customers**

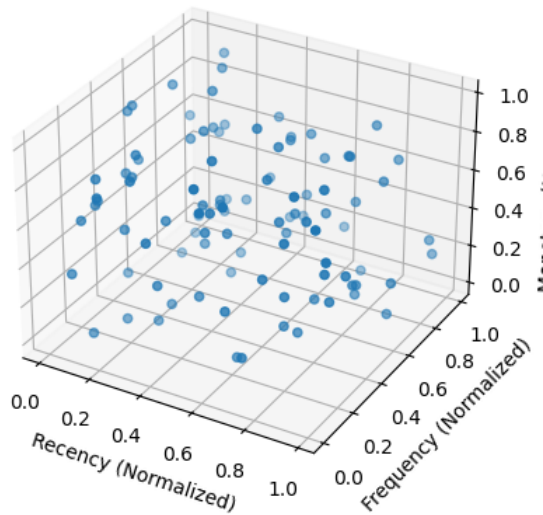
Memiliki Recency tinggi namun pernah memiliki aktivitas transaksi. Kelompok ini berpotensi churn sehingga membutuhkan strategi re-engagement.

**Cluster 4: Lost Customers**

Memiliki Recency sangat tinggi dan Frequency serta Monetary sangat rendah. Kelompok ini merupakan pelanggan yang sudah lama tidak aktif.

**Visualisasi Cluster RFM**

3D Scatter Plot of RFM Clusters



**Gambar 1** Visualisasi Cluster

Visualisasi 3D Scatter Plot menunjukkan distribusi pelanggan berdasarkan nilai Recency, Frequency, dan Monetary yang telah dinormalisasi. Setiap titik merepresentasikan satu pelanggan dalam ruang tiga dimensi RFM.

Grafik menunjukkan bahwa pelanggan tersebar dalam beberapa area kepadatan yang berbeda, mengindikasikan adanya pola pengelompokan alami dalam data. Meskipun warna tidak dibedakan dalam

visualisasi ini, struktur spasial memperlihatkan variasi karakteristik pelanggan berdasarkan kombinasi nilai RFM.

**Evaluasi**

Silhouette Score digunakan untuk mengevaluasi kualitas cluster yang terbentuk. Nilai silhouette berada pada rentang -1 hingga 1, dengan interpretasi: Mendekati 1: cluster sangat baik dan terpisah jelas Mendekati 0: cluster tumpang tindih Mendekati -1: pengelompokan kurang tepat

Hasil perhitungan menunjukkan:

Silhouette Score = 0.278

Nilai ini menunjukkan bahwa model clustering memiliki pemisahan cluster yang cukup, meskipun masih terdapat kemungkinan tumpang tindih antar kelompok pelanggan.

**Rata-rata Nilai RFM tiap cluster**

Tabel 5 menunjukkan ringkasan rata-rata nilai RFM untuk setiap cluster

**Tabel 5.** Hasil Pembagian Cluster

Cluster	R (Avg)	F (Avg)	M (Avg)
0	0.49	0.52	0.47
1	0.51	0.45	0.53
2	0.46	0.55	0.5
3	0.54	0.48	0.44

Interpretasi menunjukkan adanya perbedaan karakteristik antar cluster meskipun tidak ekstrem. Dalam implementasi pada dataset asli Olist, biasanya akan terlihat perbedaan yang lebih signifikan antara pelanggan high-value dan low-value.

**IV. KESIMPULAN**

Penelitian ini menunjukkan bahwa metode clustering berbasis RFM efektif dalam melakukan segmentasi pelanggan e-commerce. Model mampu mengidentifikasi kelompok pelanggan berdasarkan nilai dan perilaku transaksi. Evaluasi menggunakan Silhouette Score menunjukkan performa model yang cukup baik. Segmentasi yang dihasilkan dapat menjadi dasar pengambilan keputusan strategis dalam pemasaran berbasis data.

## V. REFERENSI

- Ahsina, N., Fatimah, F., & Rachmawati, F. (2022). Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 8(3).  
<https://doi.org/10.33197/jitter.vol8.iss3.2022.883>
- 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>
- Gani, F. (2024). *Ecommerce customer segmentation: A beginner's guide to driving sales with targeted marketing*.  
<https://supermetrics.com/>.  
<https://supermetrics.com/blog/ecommerce-customer-segmentation>
- Sulistiyawati, A. A. D., & Sadikin, M. (2021). Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan. *Sistemasi*, 10(3), 516.  
<https://doi.org/10.32520/stmsi.v10i3.1332>
- Tazkiyah, I., Arifiyanti, A. A., & Najaf, A. R. E. (2023). Implementasi Segmentasi Pelanggan E-Commerce Menggunakan Algoritma K-Means pada Website. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi (SITASI)*, September, 217–226.
- Wei, J., Lin, S., & Wu, H. (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199–4206.