

## Analisis Segmentasi Karakteristik Pelanggan pada Platform Olist Brazilian E-commerce dengan Metode RFM dan K-Means Clustering

<sup>1\*</sup>Farhan Damar Wildan, <sup>2</sup>Nur Azizah, <sup>3</sup>Aldiyudha Pratama, <sup>4</sup>Mufidah Karimah

<sup>1234</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

<sup>1\*</sup>[farhandamar20@gmail.com](mailto:farhandamar20@gmail.com), <sup>2</sup>[nurazizah.na207@gmail.com](mailto:nurazizah.na207@gmail.com), <sup>3</sup>[aldiyudha0905@gmail.com](mailto:aldiyudha0905@gmail.com), <sup>4</sup>[dosen02829@unpam.ac.id](mailto:dosen02829@unpam.ac.id)

### Abstract

Customer segmentation is a key strategy for understanding consumer characteristics and facilitating better marketing decisions. The objective of this study is to use the K-Means Clustering algorithm and the Recency, Frequency, Monetary (RFM) technique to segment customers based on their transaction behavior. The CRISP-DM research methodology steps include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The Brazilian Public E-Commerce dataset by Olist, obtained from Kaggle, was used. Data integration, data cleaning, RFM variable development, data normalization, and K-Means clustering are the steps in the analytical process. Using the Elbow Method, the ideal number of clusters was determined to be four. Based on the findings of this study, consumers can be divided into four categories: Hibernating, At-Risk Customers, Loyal Customers, and Best Customers. A result of 0.4886, indicating sufficiently adequate cluster quality, was obtained from the model evaluation using the Silhouette Score. These segmentation findings can serve as a starting point for developing more specific marketing plans based on the characteristics of each client.

**Keywords:** Data Mining, Customer Segmentation, RFM, K-Means Clustering, CRISP-DM.

### Abstrak

Segmentasi pelanggan merupakan taktik penting untuk memahami karakteristik konsumen dan memfasilitasi pengambilan keputusan pemasaran yang lebih baik. Tujuan penelitian ini adalah menggunakan algoritma K-Means Clustering dan teknik Recency, Frequency, Monetary (RFM) untuk menyegmentasikan pelanggan berdasarkan perilaku transaksinya. Langkah-langkah metodologi penelitian CRISP-DM meliputi Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Penerapan. Dataset E-Commerce Publik Brasil oleh Olist, yang diperoleh dari Kaggle, adalah dataset yang digunakan. Penggabungan data, pembersihan data, pengembangan variabel RFM, normalisasi data, dan pengelompokan K-Means merupakan langkah-langkah dalam proses analitis. Dengan menggunakan Metode Elbow, jumlah kluster yang ideal ditemukan sebanyak empat. Berdasarkan temuan studi ini, konsumen dapat dibagi menjadi empat kategori: Hibernating, At-Risk Customers, Loyal Customers, dan Best Customers. Hasil 0,4886, yang menunjukkan kualitas kluster yang cukup memadai, diperoleh dari evaluasi model menggunakan Skor Silhouette. Temuan segmentasi ini dapat digunakan sebagai titik awal untuk menyusun rencana pemasaran yang lebih spesifik berdasarkan karakteristik masing-masing klien.

**Kata Kunci:** Data Mining, Segmentasi Pelanggan, RFM, K-Means Clustering, CRISP-DM.

### A. PENDAHULUAN

Perkembangan teknologi informasi yang sangat pesat pada dekade terakhir telah memicu transformasi digital yang signifikan di berbagai sektor industri secara global. Perubahan ini menuntut organisasi untuk beralih dari pola konvensional menuju ekosistem digital yang lebih terintegrasi guna mempertahankan daya saing di pasar yang semakin kompetitif. Inovasi teknologi tidak hanya mengubah cara perusahaan beroperasi, tetapi juga mendefinisikan ulang pola interaksi antara penyedia layanan dengan konsumen secara menyeluruh. Hal ini menciptakan tantangan sekaligus peluang besar bagi organisasi dalam mengelola aliran data yang masuk secara masif setiap harinya. Transformasi digital menjadi fondasi

utama bagi perusahaan untuk dapat bertahan dan berkembang di tengah ketidakpastian kondisi ekonomi global saat ini.

Sistem informasi memegang peran yang sangat krusial dalam sebuah organisasi sebagai pendukung utama dalam proses pengambilan keputusan strategis yang berbasis data. Melalui implementasi sistem informasi yang efektif, perusahaan dapat mengumpulkan, menyimpan, serta mengolah data mentah menjadi informasi berharga yang memiliki nilai guna tinggi bagi manajemen. Keandalan sistem ini memungkinkan organisasi untuk meningkatkan efisiensi operasional serta meminimalisir risiko kesalahan manusia dalam aktivitas bisnis sehari-hari. Selain itu, sistem informasi berfungsi sebagai alat pemantauan kinerja

organisasi secara *real-time* yang memberikan gambaran objektif mengenai kondisi perusahaan saat ini. Pemanfaatan teknologi informasi yang optimal pada akhirnya akan memberikan keunggulan kompetitif yang berkelanjutan bagi organisasi dalam menghadapi dinamika pasar yang terus berubah (Wijayanto et al., 2026).

Salah satu representasi nyata dari keberhasilan transformasi digital di sektor perdagangan adalah platform Olist Brazilian E-commerce yang merupakan marketplace terbesar di wilayah Brasil. Olist berfungsi sebagai integrator yang menghubungkan ribuan penjual kecil dengan berbagai platform penjualan daring besar untuk menjangkau pasar yang lebih luas secara efektif. Platform ini memiliki struktur data relasional yang sangat kompleks, mencakup informasi transaksi, profil pelanggan, ulasan produk, hingga rincian pembayaran. Dataset Olist mencerminkan dinamika transaksi riil yang sangat dinamis dengan volume data mencapai ratusan ribu baris yang mencakup periode waktu selama beberapa tahun terakhir. Sebagai objek penelitian, Olist menyediakan sumber daya data yang sangat kaya untuk dieksplorasi guna memahami karakteristik perilaku konsumen dalam ekosistem e-commerce.

Meskipun memiliki volume transaksi yang sangat besar, platform Olist saat ini menghadapi permasalahan strategis terkait kesulitan dalam membedakan karakteristik pelanggan secara spesifik. Manajemen perusahaan seringkali mengalami kendala dalam mengidentifikasi mana pelanggan yang memiliki loyalitas tinggi dan mana pelanggan yang berisiko berhenti berlangganan atau *churn*. Ketiadaan pengelompokan yang jelas menyebabkan perusahaan memperlakukan seluruh basis pelanggan dengan pendekatan pemasaran yang bersifat umum atau seragam. Permasalahan ini muncul akibat besarnya volume data transaksi harian yang tidak diimbangi dengan proses pengolahan data mining yang mendalam untuk menghasilkan profil pelanggan yang akurat. Kondisi tersebut menciptakan kesenjangan antara ketersediaan data mentah dengan pengetahuan yang dapat diterapkan untuk mendukung strategi bisnis perusahaan.

Dampak dari permasalahan ketidakmampuan melakukan segmentasi ini sangat berpengaruh terhadap efisiensi biaya pemasaran yang dikeluarkan oleh perusahaan. Strategi pemasaran yang bersifat general berisiko tinggi menyebabkan pemborosan sumber daya karena promosi yang ditawarkan seringkali tidak sesuai dengan kebutuhan atau minat spesifik pelanggan. Selain itu, perusahaan berisiko kehilangan pelanggan bernilai tinggi (*high-value customers*) karena kurangnya program retensi yang dipersonalisasi untuk menjaga loyalitas mereka. Dampak jangka panjangnya adalah penurunan tingkat kepuasan pelanggan secara keseluruhan yang dapat merusak citra merek platform di mata konsumen. Jika tidak segera diatasi, penurunan retensi pelanggan ini akan berdampak langsung pada berkurangnya pendapatan perusahaan serta melemahnya posisi tawar Olist dibandingkan dengan kompetitor sejenis.

Penelitian ini menggunakan algoritma Pengelompokan K-Means yang dipadukan dengan teknik Recency, Frequency, Monetary (RFM) untuk mengatasi masalah ini. Pendekatan RFM dipilih karena mampu mengukur perilaku konsumen berdasarkan kontribusi finansial, frekuensi pembelian, dan transaksi terbaru. Sementara itu, teknik K-Means Clustering digunakan karena keefektifannya dalam mengklasifikasikan data yang luas secara otomatis ke dalam banyak kluster sesuai dengan kesamaan atribut konsumen. Diharapkan bahwa penggabungan kedua metode ini akan menghasilkan model segmentasi dengan landasan matematis yang kokoh, objektivitas, dan akurasi. Tujuan utama penelitian ini adalah menghasilkan laporan segmentasi karakteristik pelanggan yang dapat dimanfaatkan oleh manajemen Olist sebagai landasan untuk mengembangkan strategi bisnis yang lebih efisien, disesuaikan, dan terfokus.

## B. TINJAUAN PUSTAKA

### 1) Data Mining

Data mining merupakan proses analisis terhadap data dalam jumlah besar yang berasal dari berbagai sumber, seperti basis data relasional, data berorientasi objek, dan data transaksi, untuk menemukan informasi, pola, atau pengetahuan baru yang bermanfaat. Proses ini dilakukan secara interaktif guna mengidentifikasi tren atau hubungan tertentu yang dapat digunakan untuk mendukung pengambilan keputusan di masa mendatang. Pola-pola yang ditemukan kemudian dapat dianalisis lebih lanjut sehingga menghasilkan informasi yang bernilai bagi berbagai bidang, termasuk analisis transaksi dan strategi bisnis (Andini et al., 2022). Pemanfaatan data mining pada data transaksi juga telah diterapkan untuk menemukan pola penjualan yang dapat digunakan sebagai dasar dalam mendukung strategi bisnis dan pengambilan keputusan yang lebih efektif (Karimah & Pahira, 2022).

### 2) Analisis RFM

Analisis Recency, Frequency, dan Monetary dikenal sebagai analisis RFM. RFM adalah alat analisis pemasaran yang digunakan untuk mengidentifikasi secara kuantitatif pelanggan mana yang paling berharga bagi suatu bisnis dengan menganalisis dan menelaah data mengenai frekuensi transaksi pelanggan, waktu pembelian terakhir mereka (recency), serta total jumlah yang telah mereka belanjakan (monetary). (Tivano et al., 2023)

### 3) Clustering

Clustering adalah metode pembelajaran mesin yang digunakan untuk mengelompokkan data menjadi beberapa kelompok atau klaster yang sesuai (Rafi Nahjan et al., 2023). Clustering digunakan dalam berbagai bidang, termasuk segmentasi pelanggan, analisis data genetik, pemrosesan gambar dan video, serta pengelompokan dokumen. Data yang kompleks dapat disederhanakan menjadi kelompok-kelompok yang lebih teratur melalui metode pengelompokan, yang mempermudah pemahaman mengenai hubungan antar data dan mendukung

pengambilan keputusan yang lebih baik.(Hendrastuty, 2024).

#### 4) *K-Means Clustering*

Menurut (Mardiansa et al., 2023) pengelompokan K-means adalah teknik penambangan data yang umum digunakan untuk menemukan kelompok-kelompok alami dalam suatu kumpulan data dengan mengelompokkan titik-titik data yang memiliki kesamaan. Untuk meminimalkan jarak rata-rata antar kelompok data, K-means membagi data tersebut menjadi beberapa kelompok. (Pujiono et al., 2024).

#### 5) *Silhouette Score*

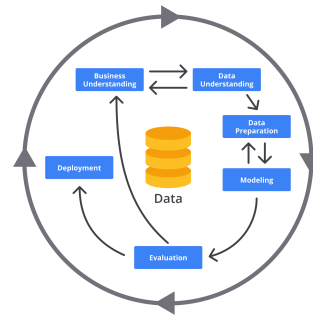
Metode untuk menggambarkan seberapa erat suatu objek selaras dengan klusternya dibandingkan dengan kluster-kluster lain disebut analisis skor siluet. Literatur menyatakan bahwa nilai rata-rata yang dihasilkan dari analisis siluet menunjukkan seberapa optimal jumlah kluster yang dihasilkan. Jumlah kluster yang optimal ( $k$ ) ditandai dengan nilai rata-rata yang lebih besar, dan disarankan untuk membagi korpus menjadi kluster-kluster pada nilai-nilai tersebut. (Sudrajat et al., 2025).

#### 6) *Elbow Method*

Dengan menghitung *Within-Cluster Sum of Squares* (WCSS), Metode Elbow merupakan metode heuristik untuk menentukan jumlah kluster yang ideal dalam algoritma K-Means. Secara teknis, pendekatan ini menggambarkan hubungan antara jumlah kluster dan jumlah kuadrat total jarak antara titik data dan pusat klusternya; titik “siku” menandakan titik di mana penurunan WCSS mulai mendatar secara drastis. Memilih titik ini memastikan bahwa pemisahan data didasarkan pada optimalitas varians internal dan validitas matematis yang kuat dengan memberikan penjelasan ilmiah mengapa penambahan kluster tidak lagi secara signifikan meningkatkan efisiensi model. (Onumanyi et al., 2022).

### C. METODE

Proses Standar Lintas Industri untuk Penambangan Data, atau CRISP-DM, adalah metodologi yang digunakan dalam penelitian ini. Metodologi ini terdiri dari langkah-langkah berikut: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (Nugraha et al., 2025) dapat dilihat pada gambar 1 berikut ini.



Gambar 1 CRISP-DM

#### 1) *Business Understanding*

Mengidentifikasi permasalahan manajemen dalam mengelola basis pelanggan yang sangat besar, di mana ketiadaan segmentasi yang akurat menyebabkan strategi pemasaran menjadi kurang efektif dan biaya promosi menjadi tidak efisien.

#### 2) *Data Understanding*

Mengeksplorasi struktur dan kualitas informasi yang tersedia guna memastikan keselarasan data dengan tujuan penelitian. analisis dilakukan terhadap kumpulan data mentah yang mencakup lebih dari seratus ribu catatan transaksi untuk mengidentifikasi variabel-variabel krusial yang merepresentasikan perilaku konsumen. Proses pemahaman ini mencakup peninjauan terhadap tipe data, distribusi nilai pada atribut kunci seperti waktu pembelian dan nominal transaksi, serta deteksi awal terhadap anomali data.

#### 3) *Data Preparation*

Pada tahap data preparation dilakukan proses pembersihan data, penanganan missing value, serta penggabungan beberapa dataset untuk memperoleh data pelanggan dan transaksi yang lengkap. Selanjutnya dilakukan perhitungan nilai Recency, Frequency, dan Monetary (RFM) sebagai representasi perilaku pelanggan. Setelah nilai RFM diperoleh, dilakukan standarisasi data menggunakan StandardScaler untuk menyamakan skala antar atribut sebelum proses clustering dilakukan.

Rumus *Recency*

$$R = T_{ref} - T_{last} \quad (1)$$

Keterangan:

- $R$ = Recency
- $T_{ref}$ = tanggal referensi
- $T_{last}$ = tanggal transaksi terakhir pelanggan

Rumus *Frequency*

$$F = \sum \text{Transaksi} \quad (2)$$

Keterangan:

- $F$ = Frequency

- Menunjukkan jumlah transaksi yang dilakukan pelanggan.

Rumus *Monetary*

$$M = \sum \text{Nilai Transaksi} \quad (3)$$

Keterangan:

- $M$ = Monetary
- Menunjukkan total nilai transaksi pelanggan.

Rumus Standarisasi Data

$$Z = \frac{x - \mu}{\sigma} \quad (4)$$

Keterangan:

- $x$ = nilai data
- $\mu$ = rata-rata data
- $\sigma$ = standar deviasi data

#### 4) Modelling

Selama tahap pemodelan, pelanggan dikelompokkan berdasarkan nilai RFM mereka menggunakan teknik pengelompokan K-Means. Metode Elbow yang didasarkan pada Within-Cluster Sum of Squares (WCSS) digunakan untuk menentukan jumlah kelompok yang ideal sebelum proses pengelompokan dilakukan. Sebagai hasil dari tahap ini, pelanggan dibagi ke dalam beberapa kategori dengan karakteristik yang unik..

Fungsi Objektif K-Means

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i - c_j\|^2 \quad (5)$$

Keterangan:

- $J$ = fungsi objektif K-Means
- $k$ = jumlah kluster
- $x_i$ = data ke- $i$
- $c_j$ = centroid kluster ke- $j$

#### 5) Evaluation

Tujuan dari tahap evaluasi adalah untuk menentukan seberapa baik kualitas kluster akhir tersebut. Tingkat kemiripan di dalam suatu kluster dan tingkat perbedaan antar kluster diukur dalam penelitian ini menggunakan metode Silhouette Score. Kualitas kluster yang dihasilkan meningkat seiring dengan meningkatnya nilai Silhouette Score..

Rumus Silhouette Score:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

#### 6) Deployment

Tahapan deployment merupakan fase final dalam metodologi *data mining* yang berfokus pada

pengintegrasian hasil analisis ke dalam lingkungan operasional untuk menghasilkan nilai manfaat yang nyata bagi organisasi. Pada konteks segmentasi pelanggan ini, proses tersebut melibatkan transformasi temuan strategis dari model klusterisasi menjadi instrumen praktis, seperti pembuatan dashboard interaktif yang mampu memonitor pergeseran perilaku segmen pelanggan secara *real-time*.

#### Deskripsi Dataset

Penelitian ini menggunakan dataset *Brazilian E-Commerce Public Dataset by Olist* yang diperoleh secara resmi melalui platform *Kaggle*. Dataset ini merupakan kumpulan data transaksi dari *marketplace* Olist di Brasil yang mencakup sekitar 100.000 riwayat pesanan pelanggan antara tahun 2016 hingga 2018. Data tersebut mencatat aktivitas belanja, mulai dari waktu transaksi hingga nominal pembayaran, yang kemudian diolah untuk menghitung nilai *Recency*, *Frequency*, dan *Monetary* (RFM) sebagai basis pengelompokan pelanggan.

### D. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis data dan pembahasan mendalam mengenai segmentasi pelanggan Olist berdasarkan metodologi CRISP-DM

#### 1) Data Preparation

Tahap data preparation diawali dengan proses penggabungan dataset pelanggan, pesanan, dan pembayaran untuk memperoleh data transaksi yang lengkap. Hasil penggabungan menghasilkan sebanyak 103.886 baris data yang selanjutnya digunakan dalam proses analisis. Setelah proses integrasi data dilakukan, tahap berikutnya adalah pemeriksaan kualitas data melalui identifikasi missing value dan pembersihan data pada atribut yang digunakan dalam perhitungan RFM. Data yang telah bersih kemudian digunakan untuk membentuk variabel *Recency*, *Frequency*, dan *Monetary* sebagai representasi perilaku pelanggan sebelum dilakukan proses standarisasi data.

**Tabel 1 Hasil Penggabungan Dataset**

Keterangan	Nilai
Jumlah Baris	103.886
Jumlah Kolom	16

Berdasarkan Tabel 1, proses penggabungan dataset berhasil menghasilkan 103.886 baris data transaksi yang terdiri dari 16 atribut. Data tersebut merupakan hasil integrasi data pelanggan, pesanan, dan pembayaran yang selanjutnya digunakan sebagai dasar dalam proses analisis segmentasi pelanggan.

**Tabel 2 Struktur Dataset Hasil Penggabungan**

No	Nama Kolom	Tipe Data
1	order_id	object
2	customer_id	object
3	order_status	object
4	order_purchase_timestamp	object
5	order_approved_at	object
6	order_delivered_carrier_date	object

No	Nama Kolom	Tipe Data
7	order_delivered_customer_date	object
8	order_estimated_delivery_date	object
9	payment_sequential	int64
10	payment_type	object
11	payment_installments	int64
12	payment_value	float64
13	customer_unique_id	object
14	customer_zip_code_prefix	int64
15	customer_city	object
16	customer_state	object

Hasil eksplorasi dataset menunjukkan bahwa data terdiri atas atribut yang berkaitan dengan informasi pelanggan, transaksi, dan pembayaran. Atribut `customer_unique_id`, `order_purchase_timestamp`, dan `payment_value` digunakan sebagai variabel utama dalam perhitungan RFM.

**Tabel 3 Hasil Pemeriksaan Missing Value**

Nama Kolom	Missing Value
order_id	0
customer_id	0
order_status	0
order_purchase_timestamp	0
order_approved_at	175
order_delivered_carrier_date	1888
order_delivered_customer_date	3132
order_estimated_delivery_date	0
payment_sequential	0
payment_type	0
payment_installments	0
payment_value	0
customer_unique_id	0
customer_zip_code_prefix	0
customer_city	0
customer_state	0

Berdasarkan hasil pemeriksaan missing value, ditemukan beberapa atribut yang memiliki nilai kosong, terutama pada atribut yang berkaitan dengan proses pengiriman pesanan. Namun, atribut yang digunakan dalam pembentukan variabel RFM tidak memiliki missing value sehingga tetap dapat digunakan untuk proses analisis.

**Tabel 4 Hasil Pembersihan Data**

Keterangan	Jumlah Data
Sebelum Pembersihan	103.886
Setelah Pembersihan	103.886

Berdasarkan hasil pembersihan data, jumlah data sebelum dan sesudah proses cleaning tetap sebanyak 103.886 data. Hal ini disebabkan atribut yang digunakan dalam pembentukan variabel RFM, yaitu `customer_unique_id`, `order_purchase_timestamp`, dan `payment_value`, tidak memiliki nilai yang hilang (missing value). Oleh karena itu,

seluruh data transaksi dapat digunakan pada tahap analisis selanjutnya.

**Tabel 5 Hasil Perhitungan RFM**

Customer Unique ID	Recency	Frequency	Monetary
0000366f3b9a7992bf8c76cfd3221e2	161	1	141.90
0000b849f77a49e4a4ce2b2a4ca5be3f	164	1	27.19
0000f46a3911fa3c0805444483337064	586	1	86.22
0000f6ccb0745a6a4b88665a16c9f078	370	1	43.62
0004aac84e0df4da2b147fca70cf8255	337	1	196.89

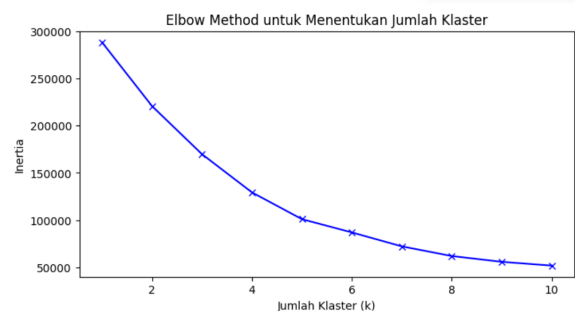
Berdasarkan hasil analisis RFM, setiap konsumen memiliki karakteristik transaksi yang unik. Skor Recency menunjukkan berapa hari yang telah berlalu sejak transaksi terakhir pelanggan, skor Frequency menunjukkan berapa kali pelanggan telah melakukan transaksi, dan skor Monetary menunjukkan nilai total transaksi pelanggan selama periode pengamatan.

Setelah memperoleh nilai-nilai Recency, Frequency, dan Monetary (RFM), metode StandardScaler digunakan untuk menstandarkan data. Untuk mencegah agar tidak ada satu variabel pun yang mendominasi proses pengelompokan, tahap ini bertujuan untuk menstandarkan skala di seluruh atribut. Seluruh 96.095 catatan pelanggan telah diproses dengan sukses dan siap digunakan pada tahap pemodelan dengan memanfaatkan algoritma K-Means, sesuai dengan hasil standarisasi.

## 2) Modelling

Selama tahap pemodelan, teknik pengelompokan K-means digunakan untuk mengelompokkan klien berdasarkan nilai recency, frequency, dan monetary (RFM) mereka. Selanjutnya, Metode Elbow digunakan untuk memperkirakan jumlah kelompok yang ideal sebelum memulai proses pengelompokan, yang memungkinkan kami memperoleh hasil segmentasi yang lebih representatif.

### a. Menentukan Jumlah Cluster



**Gambar 2 Elbow Method**

Hasil penerapan Metode Elbow untuk menentukan jumlah kluster ditampilkan pada Gambar 2. Grafik tersebut menunjukkan bahwa titik siku berada pada  $k = 4$ , yang berarti penambahan kluster setelah titik tersebut tidak secara signifikan menurunkan nilai inersia. Oleh karena itu, dalam prosedur pengelompokan K-Means pada penelitian ini digunakan empat kluster.

b. Hasil K-Means Clustering

**Tabel 6** Tabel Hasil K-Means Clustering

Customer ID	Recenc	Frequenc	Monetar	Cluste
	y	y	y	r
0000366f3b9a7992bf8c76cfd3221e2	161	1	141,90	1
0000b849f77a49e4a4ce2b2a4ca5be3f	164	1	27,19	1
0000f46a3911fa3c0805444483337064	586	1	86,22	0
0000f6ccb0745a6a4b88665a16c9f078	370	1	43,62	0
0004aac84e0df4da2b147fca70cf8255	337	1	196,89	0

Hasil clustering menunjukkan bahwa setiap pelanggan berhasil dikelompokkan ke dalam salah satu dari empat klaster berdasarkan karakteristik nilai RFM yang dimiliki. Pengelompokan ini menjadi dasar untuk menganalisis perilaku pelanggan pada masing-masing segmen.

c. Distribusi Pelanggan per Cluster

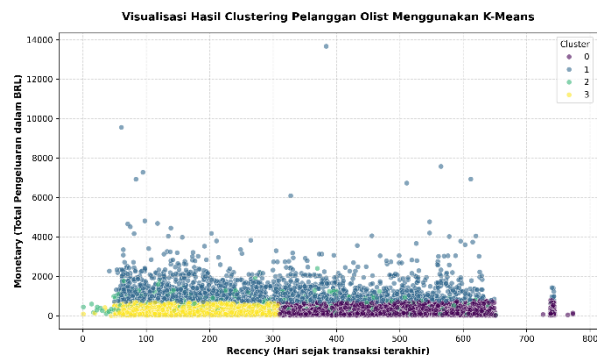
**Tabel 7** Distribusi Pelanggan per Cluster

Cluster	Jumlah Pelanggan
0	38.655
1	2.422
2	2.962
3	52.056
Total	96.095

Berdasarkan hasil clustering menggunakan algoritma K-Means, diperoleh empat segmen pelanggan dengan jumlah anggota yang berbeda. Segmen Best Customers (Cluster 3) memiliki jumlah anggota terbanyak yaitu 52.056 pelanggan atau sekitar 54,17% dari total pelanggan. Selanjutnya, segmen Hibernating (Cluster 0) memiliki 38.655 pelanggan atau sekitar 40,23%.

Sementara itu, segmen Loyal Customers (Cluster 2) dan At Risk Customers (Cluster 1) memiliki jumlah anggota yang relatif lebih sedikit, masing-masing sebanyak 2.962 pelanggan (3,08%) dan 2.422 pelanggan (2,52%). Hasil tersebut menunjukkan bahwa sebagian besar pelanggan Olist berada pada segmen Best Customers dan Hibernating, sedangkan pelanggan yang termasuk dalam kategori Loyal Customers maupun At Risk Customers hanya mencakup sebagian kecil dari keseluruhan pelanggan.

d. Visualisasi Hasil Clustering



**Gambar 3** Hasil Visualisasi Clustering

Hasil visualisasi pengelompokan pelanggan menggunakan pendekatan K-Means dengan empat kelompok ditampilkan pada Gambar 3. Sifat-sifat Recency dan Monetary digunakan dalam visualisasi tersebut untuk menampilkan distribusi konsumen berdasarkan total pengeluaran mereka dan waktu transaksi terakhir mereka. Terlihat jelas bahwa meskipun hanya sebagian kecil pelanggan yang memiliki nilai pengeluaran sangat tinggi, mayoritas pelanggan berada dalam kisaran pengeluaran sedang hingga menengah. Metode K-Means berhasil mengelompokkan pelanggan berdasarkan karakteristik transaksi mereka, sebagaimana dibuktikan oleh variasi pola distribusi di antara kluster.

e. Karakteristik Cluster

**Tabel 8** Karakteristik Cluster

Cluster	Recency	Frequency	Monetary
0	438.80	1.00	134.93
1	289.49	1.02	1196.23
2	269.31	2.12	290.31
3	178.36	1.00	135.16

Berdasarkan rata-rata nilai Recency, Frequency, dan Monetary pada masing-masing cluster, dilakukan interpretasi segmen pelanggan untuk memudahkan penyusunan strategi bisnis..

Cluster 0 dikategorikan sebagai *Hibernating* karena memiliki nilai Recency tertinggi sebesar 438,80 hari serta nilai Frequency dan Monetary yang relatif rendah. Hal ini menunjukkan bahwa pelanggan pada cluster tersebut sudah lama tidak melakukan transaksi.

Cluster 1 dikategorikan sebagai *At Risk Customers* karena memiliki nilai Monetary yang tinggi sebesar 1.196,23 namun menunjukkan kecenderungan penurunan aktivitas transaksi berdasarkan nilai Recency yang cukup besar. Pelanggan pada segmen ini perlu mendapatkan perhatian khusus agar tidak berpindah atau berhenti bertransaksi.

Cluster 2 dikategorikan sebagai *Loyal Customers* karena memiliki nilai Frequency tertinggi sebesar 2,12 transaksi, yang menunjukkan tingkat loyalitas pelanggan lebih baik dibandingkan cluster lainnya.

Cluster 3 dikategorikan sebagai *Best Customers* karena memiliki nilai Recency terendah sebesar 178,36 hari yang menunjukkan pelanggan masih aktif bertransaksi. Segmen ini merupakan kelompok pelanggan yang perlu dipertahankan melalui program loyalitas dan promosi eksklusif.

### 3) Evaluation

Evaluasi model dilakukan menggunakan Silhouette Score untuk mengukur kualitas hasil clustering yang terbentuk. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan kualitas cluster yang semakin baik.

**Tabel 9 Hasil Evaluasi Clustering**

Parameter	Nilai
Jumlah Cluster	4
Silhouette Score	0,4886

Model K-Means dengan empat kluster menghasilkan Skor Silhouette sebesar 0,4886 berdasarkan hasil evaluasi. Mengingat adanya perbedaan yang mencolok antar kluster dan bahwa anggota setiap kluster memiliki karakteristik yang relatif serupa, nilai ini menunjukkan bahwa kluster yang dihasilkan memiliki kualitas yang cukup tinggi. Sehingga, hasil segmentasi pelanggan diyakini mampu menggambarkan pola perilaku konsumen berdasarkan karakteristik Recency, Frequency, dan Monetary (RFM). Dengan demikian, hasil ini dapat menjadi landasan untuk merancang strategi pemasaran yang lebih efektif.

### 4) Deployment

Pelanggan pada segmen *Hibernating* direkomendasikan untuk diberikan program reaktivasi seperti voucher diskon, promo khusus, atau kampanye pemasaran melalui email dan media sosial guna mendorong pelanggan kembali melakukan transaksi. Pada segmen *At Risk Customers*, perusahaan perlu menerapkan strategi retensi pelanggan melalui penawaran personal, pemberian reward, serta promosi yang disesuaikan dengan riwayat pembelian pelanggan agar tidak berpindah ke kompetitor.

Untuk segmen *Loyal Customers*, perusahaan dapat meningkatkan loyalitas melalui program membership, sistem poin, atau penawaran eksklusif guna mempertahankan frekuensi transaksi yang tinggi. Sementara itu, pada segmen *Best Customers*, perusahaan perlu memberikan apresiasi khusus seperti layanan prioritas, promo eksklusif, maupun program loyalitas premium karena segmen ini memberikan kontribusi bisnis yang paling penting bagi perusahaan.

Dengan menerapkan strategi yang berbeda pada setiap segmen pelanggan, perusahaan dapat meningkatkan efektivitas pemasaran, mempertahankan pelanggan potensial, serta mengoptimalkan pendapatan secara lebih efisien.

## E. PENUTUP

### Simpulan

Berdasarkan hasil penelitian, segmentasi pelanggan pada dataset Olist berhasil dilakukan secara efektif dengan menggunakan pendekatan RFM dan algoritma pengelompokan K-Means. Temuan Metode Elbow menunjukkan bahwa empat kelompok—Pelanggan yang Tidak Aktif, Pelanggan Berisiko, Pelanggan Setia, dan Pelanggan Terbaik—merupakan jumlah yang ideal. Evaluasi Skor Silhouette menghasilkan nilai 0,4886, yang menunjukkan bahwa model pengelompokan tersebut berhasil mengklasifikasikan pelanggan berdasarkan kriteria transaksi dengan cukup baik.

### Saran

Untuk memperoleh hasil segmentasi yang lebih baik, penelitian selanjutnya dapat membandingkan pendekatan K-Means dengan teknik pengelompokan alternatif. Untuk membantu perusahaan mengidentifikasi strategi pemasaran yang lebih efektif, hasil segmentasi konsumen juga dapat dimanfaatkan sebagai dasbor atau sistem pendukung keputusan.

## E. DAFTAR PUSTAKA

- Andini, Y., Hardinata, J. T., & Purba, Y. P. (2022). Penerapan Data Mining Terhadap Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar Menggunakan Metode Apriori. *Jurnal TIMES*, 11(1), 9–15. <https://doi.org/10.51351/jtm.11.1.2022661>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Karimah, M., & Pahira, W. (2022). Penerapan Algoritma Apriori Pada Prediksi Penjualan Tanaman Hias Bromelia (Studi Kasus: Karimah Flora). *Maret*, 7(1), 2622–4615. <http://openjournal.unpam.ac.id/index.php/informatika68>
- Mardiansa, M., Sari, H. L., & Prahasti, P. (2023). Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering. *Jurnal Multidisiplin Dehasen (MUDE)*, 2(4), 693–702. <https://doi.org/10.37676/mude.v2i4.4749>
- Nugraha, D. S., Thoib, I., Sururi, N., Bayu F., & Candra, B. P. (2025). Segmentasi Pelanggan Berbasis Analisis RFM Menggunakan Algoritma K-Means Clustering. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, 4(2), 1361–1369. <https://doi.org/10.62712/juktisi.v4i2.641>
- Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An Automatic Elbow Detection Method for Estimating the Number

- of Clusters in a Dataset. *Applied Sciences (Switzerland)*, 12(15).  
<https://doi.org/10.3390/app12157515>
- Pujiono, S., Astuti, R., & Muhamad Basysyar, F. (2024). Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 615–620.  
<https://doi.org/10.36040/jati.v8i1.8360>
- Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104.  
<https://doi.org/10.36040/jati.v7i1.6094>
- Sudrajat, R., Hadiana, A. I., & Melina, M. (2025). Evaluasi Kualitas Klaster Wilayah Rawan Bencana Menggunakan K-Means dengan Silhouette dan Elbow Method. *Jurnal Algoritma*, 22(2), 127–139.  
<https://doi.org/10.33364/algoritma/v.22-2.2379>
- Tivano, D., Ruslianto, I., & Prawira, D. (2023). Sistem Segmentasi Pelanggan Potensial Menggunakan Metode K-Means Dan Analisis Rfm. *Coding Jurnal Komputer Dan Aplikasi*, 11(3), 325.  
<https://doi.org/10.26418/coding.v11i03.53238>
- Wijayanto, G., Widayati, T., Alamsyah, N., & Aminudin, R. (2026). Importance of Digitalization and Information System in Business: A Literature Review. *International Journal of Business, Law, and Education*, 7(1), 210–218.