

## Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian

Hersa Safitri<sup>1</sup>, Siti Putri Lenggo Geni<sup>2</sup>, Felisitas Merry<sup>3</sup>, Masna Wati<sup>4</sup>, Havaluddin<sup>5</sup>

<sup>1,2,3,4,5</sup>Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman,  
Jl. Sambaliung No. 9, Samarinda, Kalimantan Timur.

e-mail: [1hersasafitri7@gmail.com](mailto:hersasafitri7@gmail.com), [2lenggo.geni0305@gmail.com](mailto:lenggo.geni0305@gmail.com), [3felisitasmerry17@gmail.com](mailto:felisitasmerry17@gmail.com),  
[4masnawati@fkti.unmul.ac.id](mailto:masnawati@fkti.unmul.ac.id), [5havaluddin@unmul.ac.id](mailto:havaluddin@unmul.ac.id)

**Abstrak.** Pertumbuhan pesat e-commerce di Indonesia mendorong perusahaan untuk memahami karakteristik konsumen secara lebih mendalam guna meningkatkan strategi pemasaran. Penelitian ini bertujuan untuk mengelompokkan konsumen e-commerce berdasarkan pola pembelian menggunakan metode K-Means Clustering. Segmentasi dilakukan dengan mempertimbangkan tiga variabel utama, yaitu frekuensi pembelian, nilai transaksi, dan waktu pembelian. Data transaksi konsumen diolah dan dianalisis untuk menemukan pola-pola yang tersembunyi, kemudian ditentukan jumlah kluster optimal menggunakan metode Elbow. Hasil penelitian menunjukkan bahwa konsumen dapat dikelompokkan ke dalam lima segmen yang memiliki karakteristik perilaku belanja yang berbeda-beda, mulai dari konsumen bernilai tinggi dan aktif, hingga konsumen pasif dengan nilai transaksi rendah. Segmentasi ini memberikan wawasan penting bagi pelaku e-commerce dalam menyusun strategi pemasaran yang lebih tepat sasaran sesuai dengan karakter masing-masing segmen. Kesimpulan dari penelitian ini menunjukkan bahwa penerapan K-Means Clustering efektif dalam membantu memahami perilaku konsumen dan dapat dijadikan dasar pengambilan keputusan bisnis yang lebih strategis.

Kata Kunci : K-Means Clustering, Segmentasi Konsumen, E-Commerce

**Abstract.** *The rapid growth of e-commerce in Indonesia has pushed companies to gain a deeper understanding of consumer characteristics to improve their marketing strategies. This study aims to cluster e-commerce consumers based on their purchasing patterns using the K-Means Clustering method. Segmentation is performed by considering three key variables: purchase frequency, transaction value, and purchase time. Consumer transaction data is processed and analyzed to identify hidden patterns, and the optimal number of clusters is determined using the Elbow method. The results show that consumers can be grouped into five segments with distinct shopping behavior characteristics, ranging from high-value and active consumers to passive consumers with low transaction values. This segmentation provides valuable insights for e-commerce players in developing more targeted marketing strategies tailored to each segment's characteristics. The conclusion of this study indicates that the application of K-Means Clustering is effective in helping to understand consumer behavior and can serve as a basis for making more strategic business decisions.*

**Keywords:** *K-Means Clustering, Consumer Segmentation, E-Commerce*

### PENDAHULUAN

Pertumbuhan pesat industri e-commerce telah menyebabkan meningkatnya persaingan antar pelaku usaha dalam memperebutkan loyalitas konsumen. Dalam menghadapi tantangan ini, perusahaan dituntut untuk memahami perilaku dan karakteristik konsumen secara lebih mendalam agar mampu menyusun strategi pemasaran yang efektif dan tepat sasaran [1]. Salah satu pendekatan yang terbukti bermanfaat dalam menganalisis perilaku konsumen adalah segmentasi pelanggan berbasis data historis transaksi.

Model Recency, Frequency, Monetary (RFM) menjadi salah satu teknik yang paling umum digunakan dalam segmentasi pelanggan karena kemampuannya dalam mengukur loyalitas dan nilai pelanggan secara kuantitatif [1], [2]. RFM menilai pelanggan berdasarkan waktu transaksi terakhir (recency), frekuensi pembelian (frequency), dan total pengeluaran (monetary), yang selanjutnya dapat digunakan untuk



membedakan antara pelanggan bernilai tinggi dan rendah [3].

Untuk mendukung proses segmentasi, metode data mining seperti clustering sangat berperan. Algoritma K-Means, sebagai salah satu metode clustering non-hierarki yang paling populer, digunakan untuk mengelompokkan pelanggan ke dalam beberapa kluster berdasarkan kesamaan karakteristik perilaku mereka [4], [5]. Keunggulan K-Means terletak pada kesederhanaan dan efisiensinya dalam mengolah data skala besar, meskipun pemilihan jumlah kluster yang optimal menjadi tantangan tersendiri [6].

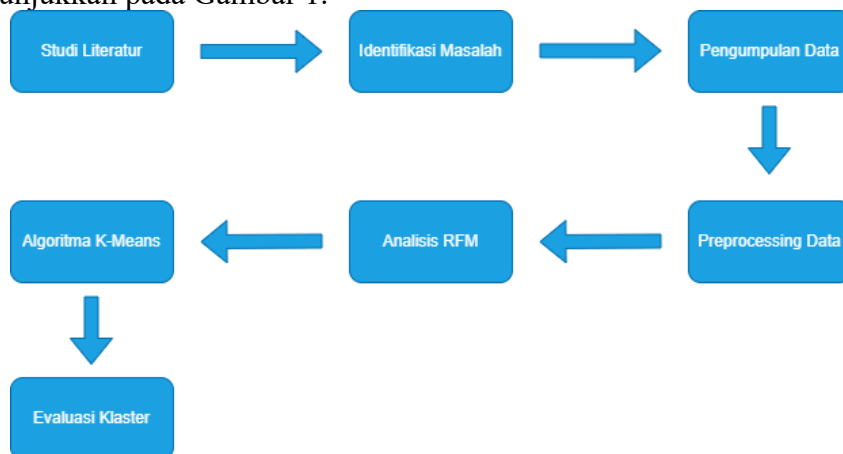
Berbagai penelitian sebelumnya telah membuktikan efektivitas kombinasi model RFM dan algoritma K-Means dalam berbagai sektor. Misalnya, pada PT Coversuper Indonesia Global, penerapan K-Means dan RFM berhasil mengelompokkan pelanggan ke dalam empat segmen berdasarkan perilaku belanja yang berbeda-beda [1]. Penelitian di Ova Gaming E-Sports Arena menunjukkan bahwa melalui segmentasi pelanggan, perusahaan dapat merancang strategi retensi berbasis prioritas nilai pelanggan [7]. Di sisi lain, penelitian [2] dalam konteks marketplace Ralali.com juga menegaskan bahwa RFM dapat membantu dalam mengidentifikasi pelanggan potensial dan setia.

Selain sektor e-commerce, penerapan K-Means juga digunakan secara luas dalam sektor lain, seperti asuransi [5], promosi pendidikan [8], dan bahkan pengelompokan wilayah rawan bencana [6], membuktikan fleksibilitas dan kemampuan metode ini dalam pengolahan data berbasis segmentasi.

Penelitian ini bertujuan untuk menerapkan metode K-Means Clustering dalam mengelompokkan pelanggan e-commerce berdasarkan atribut RFM untuk mendapatkan insight mengenai perilaku konsumen. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam penyusunan strategi pemasaran yang lebih personal dan berbasis data.

## METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif untuk mengidentifikasi dan mengelompokkan pelanggan *e-commerce* berdasarkan perilaku pembelian mereka. Tujuan utama dari penelitian ini adalah untuk menerapkan algoritma *unsupervised learning*, khususnya K-Means Clustering, dalam membentuk segmentasi pelanggan. Segmentasi dilakukan berdasarkan tiga indikator utama yang dikenal sebagai RFM, yang merupakan pendekatan populer dalam analisis perilaku konsumen. Dengan metode ini, diharapkan dapat diperoleh pemahaman yang lebih baik terhadap pola pembelian pelanggan, sehingga perusahaan dapat menyusun strategi pemasaran yang lebih efektif dan tepat sasaran. Adapun tahapan yang dilakukan untuk menyelesaikan penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Diagram Tahapan Penelitian

### 2.1. Studi Literatur

Segmentasi pelanggan merupakan pendekatan strategis yang penting dalam *e-commerce* untuk memahami perilaku konsumen dan menyusun strategi pemasaran yang tepat sasaran. Salah satu teknik yang paling umum digunakan adalah kombinasi analisis RFM dengan algoritma K-Means Clustering.

Penelitian [2] membahas pemanfaatan analisis RFM (Recency, Frequency, Monetary) sebagai strategi pendekatan terhadap konsumen marketplace Ralali.com. Analisis ini mampu mengidentifikasi pelanggan loyal dan potensial berdasarkan data transaksi, serta memberikan landasan untuk merancang program promosi yang lebih efektif, seperti diskon atau hadiah kepada pelanggan setia.

Penelitian [4] menerapkan metode clustering seperti K-Means, K-Medoids, dan X-Means yang digabungkan dengan model RFM untuk segmentasi pelanggan toko retail UD. XYZ. Penelitian ini menghasilkan lima segmen pelanggan dengan validasi terbaik diberikan oleh metode K-Medoids berdasarkan nilai Davies Bouldin Index (DBI), yang menunjukkan keunggulan dalam membedakan karakteristik pelanggan.

Penelitian [3] mengkaji implementasi K-Medoids clustering bersama analisis RFM untuk segmentasi pelanggan pada Toko Kosmetik Rahmadani. Penggunaan teknik ini mempermudah pemilik usaha dalam memahami karakteristik pelanggan dan mengembangkan strategi pemasaran yang sesuai. Hasilnya, empat cluster pelanggan terbentuk dengan validitas baik menggunakan metode silhouette coefficient.

Penelitian [8] menggunakan kombinasi RFM dan K-Means Clustering untuk mengelompokkan lokasi promosi dalam penerimaan mahasiswa baru. Segmentasi ini membantu institusi pendidikan dalam menentukan lokasi promosi yang lebih efektif berdasarkan perilaku dan kontribusi pendaftar sebelumnya, yang dikelompokkan menggunakan parameter transaksi.

Penelitian [9] melakukan analisis terhadap lebih dari 60.000 transaksi ritel untuk mengidentifikasi segmen pelanggan seperti Champions, Loyal Customers, dan At Risk melalui model RFM. Studi ini membuktikan bahwa penerapan RFM berbasis Python sangat berguna untuk meningkatkan strategi pemasaran berbasis perilaku pelanggan dan ROI.

Penelitian [10] menerapkan metode Naïve Bayes untuk memprediksi pelanggan loyal berdasarkan hasil segmentasi RFM. Penelitian ini menunjukkan akurasi tinggi sebesar 97,27%, dengan precision mencapai 100%, menjadikan pendekatan ini efektif untuk mendukung pengambilan keputusan berbasis data pelanggan dalam e-commerce.

Penelitian [5] menunjukkan implementasi algoritma K-Means Clustering untuk analisis data pelanggan pada perusahaan asuransi. Segmentasi pelanggan dilakukan berdasarkan usia, jenis kelamin, dan riwayat klaim untuk mendukung strategi bisnis yang lebih personal dan efisien.

Penelitian [11] menyajikan tinjauan literatur tentang implementasi K-Means dalam data mining. Mereka menekankan efektivitas algoritma ini dalam mengidentifikasi pola data besar meskipun memiliki kelemahan pada inisialisasi centroid dan sensitivitas terhadap outlier. Beberapa optimasi telah diusulkan untuk meningkatkan akurasi algoritma ini.

Penelitian [6]. memanfaatkan algoritma K-Means Clustering untuk mengelompokkan daerah rawan kebakaran hutan dan lahan di Indonesia. Dengan menggunakan model CRISP-DM dan indeks kejadian tahunan, penelitian ini memberikan visualisasi daerah berisiko tinggi, yang penting untuk perencanaan mitigasi bencana.

Penelitian [12] di Ova Gaming E-Sports Arena menggunakan model RFM dan algoritma K-Means untuk mengelompokkan pelanggan ke dalam empat segmen: *superstar*, *everyday*, *occasional*, dan *dormant*. Dengan dukungan metode Elbow dan Silhouette Coefficient, segmentasi yang dihasilkan dipakai untuk menyusun strategi retensi pelanggan, didukung oleh perhitungan Customer Lifetime Value (CLV) menggunakan pembobotan AHP.

Penelitian [13] juga memanfaatkan data transaksi selama tiga tahun dan menerapkannya pada model RFM dan algoritma K-Means untuk perusahaan jasa penjualan mobil Jepang. Segmentasi menghasilkan enam *cluster* dan hasil tersebut digunakan untuk menyusun strategi pemasaran baru berdasarkan karakteristik pelanggan dalam tiap *cluster*.

Penelitian [1] pada PT. Coversuper Indonesia Global menunjukkan efektivitas kombinasi RFM dan K-Means dalam mengidentifikasi empat segmen pelanggan. Segmentasi ini mendukung



pengambilan keputusan perusahaan untuk memprioritaskan pelanggan potensial dan mengalokasikan sumber daya secara optimal.

Penelitian [14] mengembangkan pendekatan segmentasi pelanggan tidak hanya berdasarkan data RFM tetapi juga atribut demografis seperti usia dan pendapatan. Dalam studi pada Lantikya Store, hasil segmentasi digunakan untuk menyusun strategi pemasaran berbasis AHP dan CLV, dengan hasil berupa empat kelompok pelanggan yang berbeda secara demografis dan perilaku belanja.

Sementara itu, [15] mengusulkan optimasi algoritma K-Means dengan algoritma Firefly untuk mengatasi kelemahan K-Means yang sensitif terhadap inisialisasi *centroid*. Pendekatan hybrid ini diterapkan pada data transaksi *e-commerce* berskala besar (>500.000 entri), dan menunjukkan peningkatan akurasi serta stabilitas klasterisasi berdasarkan nilai Silhouette Score > 0.9. Penelitian ini menggarisbawahi pentingnya pemilihan metode optimasi dalam meningkatkan kualitas segmentasi pelanggan.

### 2.2. Identifikasi Masalah

Perkembangan pesat industri *e-commerce* telah menyebabkan meningkatnya jumlah pelanggan dengan karakteristik dan perilaku belanja yang sangat beragam. Dalam menghadapi kondisi ini, perusahaan *e-commerce* sering kali menghadapi tantangan dalam memahami perilaku konsumen secara menyeluruh. Data transaksi yang melimpah belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan strategis dalam pemasaran. Perlakuan yang seragam terhadap seluruh konsumen tanpa memahami nilai dan kebiasaan belanja masing-masing pelanggan dapat menyebabkan tidak efektifnya strategi pemasaran, rendahnya tingkat retensi pelanggan, dan menurunnya loyalitas konsumen.

Selain itu, belum tersedianya sistem segmentasi yang berbasis data untuk mengelompokkan konsumen berdasarkan pola pembelian mengakibatkan perusahaan sulit dalam menentukan target pasar potensial dan dalam menyusun strategi pemasaran yang tepat sasaran. Algoritma K-Means merupakan salah satu metode klasterisasi yang dapat digunakan untuk mengatasi hal tersebut, namun pemanfaatannya dalam konteks *e-commerce* masih perlu ditingkatkan, terutama dalam hal identifikasi segmen yang akurat dan dapat ditindaklanjuti secara strategis.

### 2.3. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data transaksi pelanggan dari platform *e-commerce*, yang disimpan dalam format data “.csv”. Dataset ini mencakup informasi penting terkait aktivitas pembelian pelanggan seperti “InvoiceNo”, “CustomerID”, “InvoiceDate”, “Quantity”, “UnitPrice”, dan “Country”. Data tersebut merepresentasikan riwayat transaksi pelanggan selama kurun waktu tertentu dan bersifat sekunder karena diperoleh dari sistem informasi perusahaan.

Dataset ini adalah kumpulan data transaksi yang terjadi antara tanggal 1 Desember 2010 hingga 9 Desember 2011 untuk sebuah perusahaan ritel *online* yang berbasis dan terdaftar di Inggris. Perusahaan ini menjual produk berupa hadiah unik untuk berbagai acara dan momen khusus. Mayoritas pelanggan perusahaan ini merupakan pedagang grosir, yang melakukan pembelian dalam jumlah besar untuk dijual kembali. Adapun jumlah data yang didapat setelah pengumpulan data dalam waktu 1 tahun adalah 541.909 data.

### 2.4. Preprocessing Data

Preprocessing Data adalah proses mengolah data mentah menjadi data yang bersih dan terstruktur agar dapat digunakan dalam analisis atau oleh algoritma machine learning, melalui tahapan seperti pembersihan, transformasi, dan normalisasi data.

#### 2.4.1. Pembersihan Data (Data Cleaning)

Beberapa proses pembersihan data yang dilakukan meliputi:

- a. Menghapus transaksi yang tidak memiliki “CustomerID”



- b. Menghapus transaksi dengan  $Quantity \leq 0$  atau  $UnitPrice \leq 0$  (karena dapat dianggap tidak valid)
- c. Menghapus nilai duplikat jika ada
- d. Mengubah kolom InvoiceDate ke format “datetime” untuk pemrosesan selanjutnya
- e. Mengecek apakah terdapat *missing values* dan duplikasi

#### 2.4.3. Normalisasi Data

Nilai Recency, Frequency, dan Monetary memiliki skala yang berbeda, maka data harus dinormalisasi sebelum dimasukkan ke algoritma K-Means. Normalisasi dilakukan menggunakan Min-Max Scaling, sehingga seluruh nilai berada dalam rentang [0,1].

### 2.5. Analisis RFM

#### 2.5.1. Pembuatan Variabel RFM

Setelah data bersih, dilakukan perhitungan tiga metrik penting dari RFM:

- a. Recency (R): Mengukur jumlah hari sejak transaksi terakhir pelanggan hingga tanggal referensi (*cut-off date*). Recency rendah menunjukkan pelanggan baru saja bertransaksi, artinya masih aktif.
- b. Frequency (F): Menghitung jumlah faktur unik (transaksi) yang dilakukan oleh pelanggan. Semakin tinggi nilainya, semakin sering pelanggan bertransaksi.
- c. Monetary (M): Total nilai uang yang dibelanjakan oleh pelanggan. Dihitung dari  $Quantity \times UnitPrice$  untuk seluruh transaksi pelanggan.

Nilai RFM dihitung untuk masing-masing pelanggan berdasarkan data historis transaksi yang tersedia. Data yang dihasilkan merupakan tabel agregat dengan CustomerID sebagai indeks dan kolom Recency, Frequency, serta Monetary.

#### 2.5.2. Perhitungan Variabel RFM

Sebelum melakukan segmentasi pelanggan dengan metode RFM langkah pertama adalah menghitung tiga metrik utama yang menjadi dasar analisis. Recency mengukur seberapa lama sejak terakhir kali pelanggan melakukan transaksi, dengan cara mengurangi tanggal transaksi terakhir dari tanggal acuan analisis (reference date). Frequency dihitung sebagai jumlah unik transaksi yang dilakukan oleh pelanggan, biasanya berdasarkan nomor invoice. Sementara Monetary mencerminkan total nilai pembelian pelanggan dalam periode pengamatan. Proses ini dilakukan dengan mengelompokkan data berdasarkan ID pelanggan (CustomerID) dan menerapkan agregasi terhadap kolom tanggal transaksi, nomor invoice, dan nilai total pembelian (TotalPrice). Pendekatan ini memungkinkan kita memahami perilaku pelanggan berdasarkan seberapa baru, sering, dan banyak mereka berbelanja.

### 2.6. Algoritma K-Means

#### 2.6.1. Penerapan K-Means

Metode klasterisasi yang digunakan adalah K-Means Clustering, yaitu algoritma pembelajaran tidak terawasi yang bertujuan membagi data ke dalam sejumlah klaster yang telah ditentukan (K). Tujuan dari klasterisasi ini adalah mengelompokkan pelanggan ke dalam segmen-segmen berdasarkan kemiripan nilai RFM mereka.

Untuk menentukan nilai K yang optimal, digunakan Metode Elbow, yang melibatkan:

- a. Menjalankan K-Means untuk berbagai nilai K (dari 1 hingga 10)
- b. Menggambar plot nilai inertia (total jarak kuadrat dalam klaster) terhadap K

Algoritma K-Means adalah metode klastering yang digunakan untuk mengelompokkan data ke dalam sejumlah k kelompok (klaster) berdasarkan kemiripan karakteristik. Algoritma ini bekerja dengan mencari pusat klaster (centroid) secara iteratif, lalu mengelompokkan data ke klaster terdekat hingga pusat klaster stabil.



Setelah jumlah kluster ditentukan, algoritma K-Means dijalankan untuk membagi pelanggan ke dalam kluster-kluster tersebut. Model akan mengelompokkan data pelanggan berdasarkan kemiripan nilai RFM-nya.

### 2.7 Evaluasi Kluster

Setelah proses klusterisasi dilakukan menggunakan algoritma K-Means, langkah evaluasi kluster menjadi penting untuk memastikan bahwa jumlah kluster yang dipilih benar-benar optimal dan hasil segmentasi yang diperoleh bermakna. Dalam penelitian ini, evaluasi kluster dilakukan melalui dua pendekatan utama, yaitu metode Elbow dan interpretasi terhadap hasil RFM di setiap kluster.

Pertama, metode Elbow digunakan untuk menentukan jumlah kluster yang paling sesuai. Metode ini melibatkan penghitungan nilai inerti (jumlah kuadrat jarak dari tiap titik ke pusat klusternya) untuk berbagai jumlah kluster (K). Nilai inerti diplot terhadap nilai K, dan titik di mana penurunan inerti mulai melambat secara signifikan dianggap sebagai jumlah kluster yang optimal.

Kedua, evaluasi dilakukan dengan menganalisis karakteristik masing-masing kluster yang terbentuk. Nilai rata-rata dari metrik RFM untuk setiap kluster dihitung dan dibandingkan. Hasilnya menunjukkan bahwa kluster yang terbentuk memiliki perbedaan yang cukup jelas dalam hal frekuensi transaksi, nilai pembelian, dan jarak waktu dari transaksi terakhir. Hal ini mengindikasikan bahwa segmentasi pelanggan telah berhasil dilakukan secara efektif, karena masing-masing kelompok mencerminkan perilaku pelanggan yang berbeda.

## HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Atribut "CustomerID" digunakan sebagai identifikasi unik untuk masing-masing pelanggan, sedangkan atribut waktu dan nilai transaksi digunakan untuk menghitung indikator perilaku pelanggan seperti seberapa baru mereka bertransaksi (Recency), seberapa sering mereka bertransaksi (Frequency), dan seberapa banyak mereka membelanjakan uang (Monetary). Karena data ini mencerminkan aktivitas pembelian aktual pelanggan, dataset ini sangat cocok untuk digunakan dalam analisis segmentasi berbasis algoritma pembelajaran mesin. Adapun untuk 5 baris pertama dataset dapat dilihat pada Tabel 1 berikut.

**Tabel 1.** Data Pembelian Pelanggan

Data Ke-	Invoice No	StockCode	Description	Quantity	InvoiceDate	UnitPrice	Customer ID	Country
0	536365	85123A	White Hanging Heart T-Light Holder	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	White Metal Lantern	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	Cream Cupid Hearts Coat Hanger	8	12/1/2010 8:26	2.75	17850.0	United Kingdom



Data Ke-	Invoice No	StockCode	Description	Quantity	InvoiceDate	UnitPrice	Customer ID	Country
3	536365	84029G	Knitted Union Flag Hot Water Bottle	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	Red Woolly Hottie White Heart.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
541910	581587	22138	Baking Set 9 Piece Retrosport	3	12/9/2011 12:50	4.95	12680	France

### 3.2. Preprocessing Data

Setelah dilakukan serangkaian proses pembersihan dan pengolahan data terhadap dataset transaksi *e-commerce*, diperoleh total sebanyak 4.339 pelanggan unik yang memiliki data transaksi valid dan lengkap. Dari data tersebut, dilakukan perhitungan metrik RFM untuk masing-masing pelanggan. Nilai-nilai RFM tersebut kemudian dinormalisasi menggunakan metode *StandardScaling* agar setiap variabel berada dalam skala yang sebanding, sehingga dapat digunakan dalam proses klusterisasi menggunakan algoritma *K-Means*.

### 3.3. Perhitungan RFM

Segmentasi pelanggan menggunakan metode RFM (*Recency, Frequency, Monetary*) untuk memahami perilaku pelanggan berdasarkan data transaksi yang tersedia. Nilai *Recency* dihitung sebagai selisih hari antara tanggal referensi (satu hari setelah transaksi terakhir tercatat) dengan tanggal transaksi terakhir setiap pelanggan. *Frequency* dihitung berdasarkan jumlah unik transaksi yang dilakukan oleh masing-masing pelanggan, sedangkan *Monetary* dihitung sebagai total nilai pembelian (*TotalPrice*) yang dihabiskan oleh pelanggan selama periode observasi.

```
reference_date = df["InvoiceDate"].max() + pd.Timedelta(days=1)
rfm = df.groupby("CustomerID").agg(
    {
        "InvoiceDate": lambda x: (reference_date - x.max()).days, #
Recency
        "InvoiceNo": "nunique", # Frequency
        "TotalPrice": "sum", # Monetary
    }
)rfm.columns = ["Recency", "Frequency", "Monetary"]
```

Setelah dilakukan pengolahan data menggunakan metode RFM, diperoleh hasil statistik deskriptif dari masing-masing variabel yang ditampilkan pada Tabel berikut.

**Tabel 2.** Hasil Perhitungan RFM

Deskripsi	Recency	Frequency	Monetary
count	4339.000000	4339.000000	4339.000000



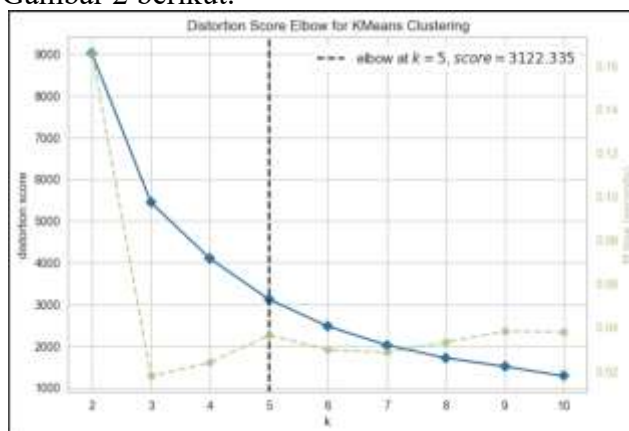
Deskripsi	Recency	Frequency	Monetary
mean	92.518322	4.271952	2048.215911
Std	100.009747	7.705493	8984.248295
Min	1.000000	1.000000	0.000000
25%	18.000000	1.000000	306.455001
50%	51.000000	2.000000	668.560006
75%	142.000000	5.000000	1660.314988
max	374.000000	210.000000	280206.019436

### 3.4. Penerapan K-Means

Untuk menentukan jumlah kluster yang optimal, dilakukan analisis menggunakan metode Elbow. Nilai inerti dihitung untuk berbagai jumlah kluster dari K = 2 hingga K = 10.

```
from yellowbrick.cluster import KElbowVisualizer
elbow = KElbowVisualizer(KMeans(), k=10)
elbow.fit(rfm_scaled)
elbow.show()
```

Dari grafik Elbow yang dihasilkan, tampak bahwa penurunan inerti mulai melambat signifikan pada K = 5, sehingga nilai K = 5 dipilih sebagai jumlah kluster yang paling optimal. Pemilihan ini didasarkan pada titik siku grafik, yang menunjukkan bahwa penambahan kluster setelah titik tersebut tidak memberikan penurunan inerti yang substansial. Visualisasi dari penentuan jumlah kluster ditunjukkan pada Gambar 2 berikut.



**Gambar 2.** Visualisasi pencarian jumlah kluster

### 3.5. Analisis dan Evaluasi Kluster

Setelah algoritma K-Means dijalankan dengan K = 5, setiap pelanggan diklasifikasikan ke dalam salah satu dari lima kluster yang terbentuk. Tabel ringkasan karakteristik tiap kluster pelanggan berdasarkan nilai rata-rata RFM ditampilkan pada Tabel 2.

**Tabel 3.** Nilai rata-rata RFM

Cluster	Recency	Frequency	Monetary
0	15.239	20.819	12349.045
1	248.662	1.551	476.416
2	6.500	120.625	55099.491
3	7.667	43.000	190808.535
4	44.129	3.608	1318.367

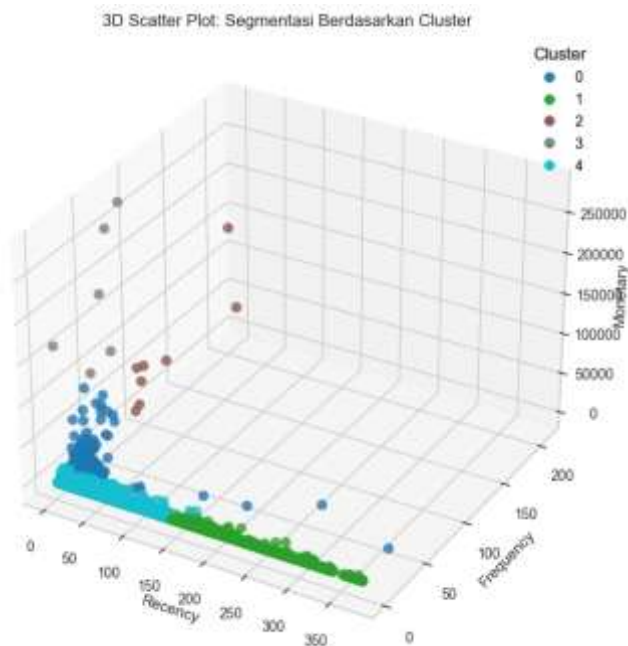




Masing-masing klaster memiliki ciri khas perilaku pelanggan yang berbeda. Adapun analisis dari hasil klaster-klaster tersebut, yaitu :

- a. Klaster 0 terdiri dari pelanggan yang cukup aktif, dengan frekuensi belanja menengah dan nilai pembelian sedang. Walaupun waktu terakhir transaksi mereka cukup lama (recency menengah), mereka masih menunjukkan nilai yang layak untuk dipertahankan melalui strategi aktivasi ulang.
- b. Klaster 1 mencakup pelanggan yang sudah lama tidak melakukan pembelian, frekuensinya rendah, dan kontribusi finansialnya pun kecil. Segmen ini merupakan pelanggan tidak aktif dan bernilai rendah. Strategi pemasaran agresif atau kampanye reaktivasi diperlukan untuk mengembalikan perhatian mereka.
- c. Klaster 2 merupakan klaster dengan pelanggan paling bernilai. Mereka baru saja melakukan transaksi (recency rendah), sangat sering berbelanja (frequency tinggi), dan menghabiskan uang dalam jumlah besar (monetary sangat tinggi). Pelanggan dalam klaster ini adalah target utama untuk strategi retensi, loyalitas, dan personalisasi layanan premium.
- d. Klaster 3 berisi pelanggan yang cukup bernilai. Mereka memiliki loyalitas yang tinggi mirip seperti yang ada pada kluster 2, namun perbedaannya terletak pada frekuensi yang lebih rendah namun pengeluaran yang masih sama tinggi.
- e. Klaster 4 berisi pelanggan yang aktif dengan nilai pembelian dan frekuensi menengah. Mereka masih cukup baru dalam melakukan transaksi terakhir. Pelanggan ini berpotensi dipromosikan menjadi pelanggan premium melalui pendekatan *upselling*, *cross-selling*, atau insentif loyalitas.

Visualisasi dari scatter plot tiga dimensi berdasarkan recency, frequency, dan juga monetary dapat dilihat pada Gambar 3 berikut.



Gambar 3. Scatter Plot 3D

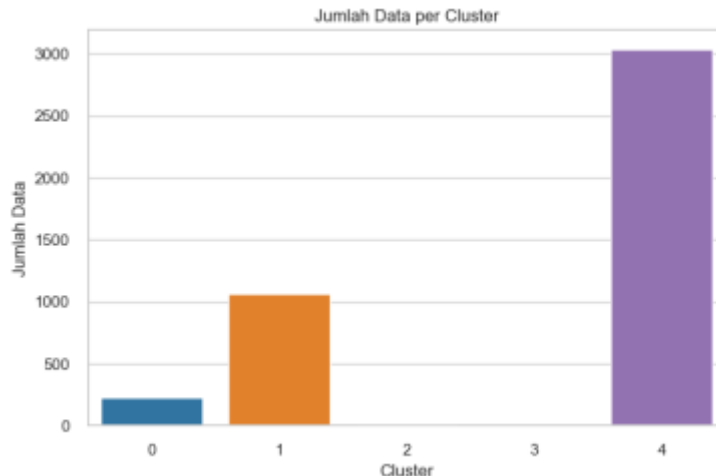
Adapun untuk 5 baris pertama dari dataset RFM hasil pengelompokan menggunakan perhitungan metrik RFM terdapat pada Tabel 3 berikut.

Tabel 3. Tabel Dataset RFM

No	CustomerID	Recency	Frequency	Monetary	Cluster
0	12346.0	326	1	77183.597169	0
1	12347.0	2	7	4309.999988	4
2	12348.0	75	4	1797.239997	4

No	CustomerID	Recency	Frequency	Monetary	Cluster
3	12349.0	19	1	1757.549994	4
4	12350.0	310	1	334.399997	1

Hasil klasterisasi menggunakan metode RFM menghasilkan lima klaster dengan jumlah anggota yang bervariasi. Klaster 4 merupakan klaster dengan jumlah anggota terbanyak, yaitu sebanyak 3.038 pelanggan. Selanjutnya, klaster 1 memiliki 1.061 pelanggan, sedangkan klaster 0 mencakup 226 pelanggan. Klaster 2 dan klaster 3 masing-masing hanya memiliki 8 dan 6 pelanggan, yang menunjukkan bahwa hanya sedikit pelanggan yang memiliki karakteristik RFM ekstrem yang unik, baik sangat tinggi maupun sangat rendah. Ketimpangan jumlah antar klaster ini mencerminkan distribusi pelanggan yang tidak merata dalam hal perilaku belanja dan nilai kontribusinya terhadap perusahaan. Visualisasi jumlah data setiap masing-masing klaster ditunjukkan pada Gambar 4.



**Gambar 4.** Visualisasi Jumlah Data Setiap Klaster

Visualisasi hasil klasterisasi dalam ruang dua dimensi juga menunjukkan bahwa kelima klaster yang terbentuk memiliki distribusi yang cukup terpisah. Hal ini menunjukkan bahwa model K-Means yang digunakan berhasil mengelompokkan pelanggan berdasarkan pola perilaku pembelian yang bermakna. Hasil ini memperkuat bahwa pendekatan segmentasi pelanggan menggunakan kombinasi RFM dan K-Means Clustering efektif dalam mengungkap struktur segmen dalam data pelanggan *e-commerce*. Dengan informasi yang diperoleh dari masing-masing klaster, perusahaan dapat merancang strategi pemasaran yang lebih terarah dan efisien. Alih-alih menerapkan pendekatan seragam untuk semua pelanggan, perusahaan kini dapat menyesuaikan pendekatan mereka berdasarkan karakteristik unik dari setiap segmen. Pendekatan ini tidak hanya dapat meningkatkan kepuasan pelanggan, tetapi juga dapat meningkatkan konversi penjualan dan loyalitas jangka panjang.

### KESIMPULAN

Penelitian ini menunjukkan bahwa metode K-Means Clustering yang dikombinasikan dengan analisis RFM dalam segmentasi pelanggan *e-commerce* yang menghasilkan 5 klaster (Cluster 0–4) dengan karakteristik yang berbeda. Cluster 0 terdiri dari pelanggan yang jarang bertransaksi dan sudah lama tidak membeli, namun memiliki nilai pembelian sangat tinggi (high spender yang tidak aktif). Cluster 1 mencakup pelanggan pasif dengan frekuensi dan nilai pembelian rendah serta kontribusi minimal. Cluster 2 berisi pelanggan dengan perilaku pembelian sedang dalam hal recency, frequency, dan monetary. Cluster 3 merupakan pelanggan terbaik yang sangat loyal, dengan frekuensi tinggi, nilai transaksi besar, dan waktu pembelian yang baru. Cluster 4 menunjukkan pelanggan aktif dan bernilai, dengan recency rendah, frekuensi sedang hingga tinggi, serta nilai transaksi yang cukup besar. Penelitian selanjutnya disarankan untuk mencoba metode lain dalam menentukan jumlah



klaster, seperti Silhouette Score atau Davies-Bouldin Index, guna membandingkan kualitas segmentasi yang dihasilkan, terutama jika data memiliki pola yang lebih kompleks.

### DAFTAR PUSTAKA

- [1] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.
- [2] A. . Fadillah, I. K. A. Mogi, and I. D. M. B. A. Darmawan, "Analisis Data Recency Frequency dan Monetary (RFM) Guna Melakukan Pendekatan Kepada Konsumen Ralali.Com," vol. 2, no. November, pp. 83–88, 2023.
- [3] Nurmilayanti and N. Hartono, "Implementasi Metode Clustering Sebagai Penunjang Strategi dalam Manajemen Pelanggan," *J. Fasilkom*, vol. 13, no. 3, pp. 605–613, 2023, doi: 10.37859/jf.v13i3.5617.
- [4] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejur.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.
- [5] A. A. Alya Putri and S. A. Rahmah, "Implementasi Data Mining Dengan Algoritma K-Means Clustering Untuk Analisis Bisnis Pada Perusahaan Asuransi," *Djtechno J. Teknol. Inf.*, vol. 5, no. 1, pp. 139–152, 2024, doi: 10.46576/djtechno.v5i1.4537.
- [6] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [7] M. Sopiyan, F. Fauziah, and Y. F. Wijaya, "Fraud Detection Using Random Forest Classifier, Logistic Regression, and Gradient Boosting Classifier Algorithms on Credit Cards," *JUITA J. Inform.*, vol. 10, no. 1, p. 77, 2022, doi: 10.30595/juita.v10i1.12050.
- [8] H. Hairani, D. Susilowati, I. Puji Lestari, K. Marzuki, and L. Z. A. Mardedi, "Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 275–282, 2022, doi: 10.30812/matrik.v21i2.1542.
- [9] A. Hermawan, N. R. Jayanti, A. Saputra, and C. Tambunan, "Optimalisasi Strategi Pemasaran Melalui Analisis RFM pada Dataset Transaksi Ritel Menggunakan Python," no. 4, 2024.
- [10] N. Wayan Wardani *et al.*, "Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM," *J. Manaj. Dan Teknol. Inf.*, vol. 12, no. 2, pp. 2087–5312, 2022, [Online]. Available: <https://ojs.mahadewa.ac.id/index.php/jmti>
- [11] F. Khalish, N. M. Piranti, and O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," vol. 8, pp. 5392–5397, 2025.
- [12] K. Z. Wijaya, A. Djunaidi, and F. Mahananto, "Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri," *J. Tek. ITS*, vol. 10, no. 2, 2021, doi: 10.12962/j23373539.v10i2.67707.
- [13] Carudin, "Jurnal Teknologi Terpadu PROBLEM," *J. Teknol. Terpadu Vol.*, vol. 7, no. 2, pp. 77–82, 2021, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/jtt/article/download/318/201>
- [14] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "Implementasi K-Means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model Rfm (Studi Kasus: Lantikya Store Jombang)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4677.
- [15] D. T. Warianti, P. Astagina, R. Julianto, and F. Y. Arini, "Optimasi K-Means Menggunakan Algoritma Firefly Untuk Segmentasi Pelanggan pada E-commerce," vol. 14, no. 3, pp. 775–785, 2024.

