

Segmentasi Customer Pada Industri Ritel Menggunakan Teknik *Clustering K-Means*

Evy Nurmiati^{1*}, Arrisa Aprilani Nurindah²

^{1,2}Universitas Islam Negeri Jakarta Syarif Hidayatullah, Indonesia

¹evy.nurmiati@uinjkt.ac.id, ²arrisa.aprilani21@mhs.uinjkt.ac.id



Histori Artikel:

Diajukan: 3 Juli 2025

Disetujui: 20 Juli 2025

Dipublikasi: 22 Juli 2025

Kata Kunci:

*Segmentasi
Pelanggan, RFM, K-
Means Clustering,
Data Mining,
Transmart Carrefour*

*Digital Transformation
Technology (Digitech) is an
Creative Commons License This
work is licensed under a
Creative Commons Attribution-
NonCommercial 4.0
International (CC BY-NC 4.0).*

Abstrak

Persaingan yang semakin ketat dalam industri ritel menuntut perusahaan untuk memahami perilaku pelanggan secara lebih mendalam guna menyusun strategi pemasaran yang tepat sasaran. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan pada industri ritel dengan studi kasus Transmart Carrefour menggunakan pendekatan analisis RFM (Recency, Frequency, Monetary) dan teknik clustering K-Means. Data yang digunakan diperoleh dari dataset "Retail Sales Dataset" melalui platform Gigasheet, yang mencakup lebih dari 1.000 data transaksi pelanggan. Analisis RFM digunakan untuk mengukur perilaku pembelian pelanggan berdasarkan keterkinian transaksi, frekuensi pembelian, dan nilai total transaksi. Nilai RFM kemudian dijadikan input untuk algoritma K-Means Clustering guna mengelompokkan pelanggan ke dalam beberapa segmen homogen. Berdasarkan metode Elbow, jumlah kluster optimal yang diperoleh adalah 4. Hasil klusterisasi menunjukkan bahwa Kluster 2 merupakan segmen pelanggan paling potensial dengan nilai Frequency dan Monetary tertinggi, sedangkan Kluster 3 merupakan segmen pelanggan kurang aktif. Analisis centroid menunjukkan perbedaan karakteristik signifikan antar kluster, yang dapat dimanfaatkan untuk perancangan strategi pemasaran yang lebih efektif dan personal. Dengan memahami karakteristik tiap segmen, perusahaan dapat meningkatkan efisiensi kampanye pemasaran, retensi pelanggan, serta loyalitas pelanggan secara keseluruhan.

PENDAHULUAN

Industri ritel merupakan salah satu sektor penting dalam perekonomian Indonesia, dengan kontribusi signifikan terhadap Produk Domestik Bruto (PDB) dan lapangan pekerjaan. Di tengah persaingan yang ketat, memahami perilaku dan kebutuhan pelanggan menjadi kunci utama bagi para pelaku bisnis ritel untuk meningkatkan penjualan, profitabilitas, dan loyalitas pelanggan (Adriansyah dan Hermanto, 2021).

Transmart Carrefour merupakan salah satu ritel terkemuka di Indonesia yang menyediakan berbagai produk mulai dari kebutuhan sehari-hari, elektronik, hingga *fashion*. Sebagai bagian dari Trans Retail Indonesia, Transmart Carrefour terus berupaya untuk memenuhi kebutuhan konsumennya yang beragam dan terus berkembang. Dalam persaingan yang ketat di industri ritel, memahami perilaku dan preferensi konsumen menjadi faktor kunci guna memperkuat kepuasan dan loyalitas konsumen.

Salah satu hambatan utama yang menjadi tantangan bagi Transmart Carrefour adalah bagaimana mengidentifikasi dan mengategorikan pelanggan berdasarkan karakteristik serta pola belanja mereka merupakan langkah krusial dalam menyusun strategi pemasaran yang lebih tepat sasaran dan bersifat personal. sehingga dapat mengoptimalkan pengalaman berbelanja dan menarik lebih banyak pelanggan.

Algoritma *K-Means Clustering* merupakan salah satu teknik yang paling umum digunakan dalam proses segmentasi pelanggan. Teknik ini memungkinkan data pelanggan dikelompokkan ke dalam kluster-kluster berdasarkan karakteristik yang sama, seperti frekuensi pembelian, nilai transaksi, dan waktu pembelian terakhir. Dengan demikian, perusahaan dapat menyusun strategi pemasaran yang lebih terarah dan personalisasi layanan yang sesuai dengan karakteristik masing-masing segmen pelanggan (Ardiansyah dan Wibowo, 2022).

Beberapa penelitian sebelumnya telah mengaplikasikan metode *K-Means Clustering* dalam konteks industri ritel dan menunjukkan hasil yang signifikan, seperti penelitian oleh (Maresti *et al*, 2025) menerapkan algoritma *K-Means* untuk segmentasi produk berdasarkan data transaksi ritel. Penelitian ini berhasil mengidentifikasi tiga kluster utama, dengan variabel harga sebagai faktor dominan dalam karakteristik kluster. Penelitian selanjutnya (Haryadi *et al*, 2024) melakukan segmentasi pelanggan berdasarkan pendapatan tahunan dan skor pengeluaran menggunakan metode *K-Means*. Hasilnya, lima kluster pelanggan berhasil diidentifikasi, memberikan wawasan mendalam tentang perilaku belanja konsumen. Penelitian oleh (Awalina dan Rahayu, 2023) mengaplikasikan algoritma *K-Means Clustering* pada data transaksi e-retail untuk mengoptimalkan strategi pemasaran, penelitian ini menunjukkan bahwa segmentasi pelanggan dapat meningkatkan efektivitas komunikasi

pemasaran dan loyalitas pelanggan. Penelitian selanjutnya oleh (Anam *et al*, 2022) Menggunakan algoritma K-Means Clustering untuk melakukan segmentasi pelanggan berdasarkan data supermarket. Hasilnya, dua kluster utama berhasil diidentifikasi, yaitu konsumen prioritas dan konsumen biasa, yang dapat digunakan untuk merancang strategi pemasaran yang lebih tepat sasaran.

Dari uraian diatas penelitian ini bertujuan untuk menerapkan Oleh karena itu, penelitian ini bertujuan untuk menerapkan teknik *K-Means Clustering* dalam segmentasi pelanggan Transmart Carrefour. Dengan menganalisis data transaksi pelanggan, diharapkan dapat diidentifikasi kluster-kluster pelanggan yang memiliki karakteristik serupa, Kondisi tersebut memberikan dasar yang kuat bagi perusahaan untuk merumuskan strategi pemasaran yang lebih efektif, yang pada gilirannya berkontribusi terhadap peningkatan kepuasan dan loyalitas pelanggan.

STUDI LITERATUR

Customer Relationship Management (CRM)

Customer Relationship Management (CRM) merupakan bagian dari operasional usaha bisnis yang bertujuan untuk menarik dan meningkatkan loyalitas pelanggan. Fokus utama dari Customer Relationship Management (CRM) terletak pada upaya untuk memaksimalkan kontribusi pelanggan, mendorong loyalitas jangka panjang, dan mempertahankan pelanggan secara berkelanjutan, melalui penerapan strategi yang berorientasi pada pemenuhan kebutuhan individual pelanggan. CRM dikembangkan dengan tujuan untuk memperkuat keterikatan antara perusahaan dan pelanggan, melalui pendekatan komprehensif yang memungkinkan perusahaan menggabungkan nilai pelanggan, preferensi, ekspektasi, serta pola perilaku yang diperoleh melalui analisis data historis transaksi konsumen (Peppard, 2000). Dengan menerapkan CRM secara efektif, perusahaan dapat mempercepat siklus penjualan, meningkatkan kesetiaan pelanggan, serta meningkatkan pendapatan.

Model RFM (*Recency, Frequency, Monetary*)

Pendekatan RFM merupakan metode konvensional yang telah lama diterapkan dan secara luas digunakan dalam mengukur keterlibatan serta nilai hubungan antara perusahaan dan pelanggannya. (Schijns *et al*, 1996). Model RFM sering kali digunakan untuk segmentasi pasar. Model ini menyimpan informasi mengenai kapan terakhir kali pembelian dilakukan (*recency*), seberapa sering pelanggan melakukan pembelian (*frequency*), dan rata-rata jumlah uang yang dibelanjakan (*monetary*) (Lee, 2014). Keunggulan model RFM terletak pada relevansinya karena beroperasi menggunakan beberapa variabel yang dapat diamati serta memiliki objektivitas dalam menentukan segmentasi pelanggan (Adiana *et al*, 2018).

Nilai *recency*, *frequency*, dan *monetary* dikelompokkan menjadi lima kategori, yaitu dengan nilai 1, 2, 3, 4, dan 5. Nilai *recency* ditentukan berdasarkan tanggal transaksi terakhir atau jarak waktu antara transaksi terakhir dengan waktu saat ini. Pelanggan yang melakukan transaksi paling baru akan mendapatkan nilai 5, sedangkan yang melakukan transaksi paling lama akan mendapatkan nilai 1. Untuk *frequency*, pelanggan yang sering melakukan transaksi diberikan nilai 5, sementara yang jarang bertransaksi diberikan nilai 1. Pelanggan yang menghabiskan banyak uang dalam transaksi memiliki nilai *monetary* tinggi, yakni 5, sementara yang mengeluarkan sedikit uang memiliki nilai *monetary* rendah, yaitu 1 (Cheng *et al*, 2009).

Segmentasi Pelanggan

Segmentasi pelanggan merupakan metode yang digunakan perusahaan untuk membagi basis pelanggan mereka ke dalam beberapa kelompok berdasarkan kesamaan karakteristik dan perilaku. Tujuan utamanya adalah untuk memperoleh pemahaman yang lebih baik tentang perilaku dan preferensi pelanggan, serta mengidentifikasi kelompok dengan kebutuhan yang serupa. Proses ini menjadi langkah awal dalam menyusun strategi yang lebih terarah dan personal. Setelah kelompok pelanggan berhasil diidentifikasi, perusahaan dapat merancang tindakan yang sesuai untuk masing-masing segmen, seperti menerapkan strategi pemasaran yang berbeda-beda. Seiring perkembangan teknologi dan kemudahan akses terhadap data, segmentasi pelanggan kini dapat digabungkan dengan metode analitik lanjutan seperti analitik prediktif dan machine learning. Penggunaan pendekatan ini memungkinkan perusahaan mendapatkan wawasan yang lebih mendalam mengenai perilaku pelanggan serta memperkirakan kebutuhan dan preferensi mereka di masa depan (Angelie, 2017).

Data Mining

Data mining telah menarik perhatian yang signifikan dalam komunitas sistem informasi dan masyarakat pada umumnya dalam beberapa tahun terakhir. Hal ini disebabkan oleh ketersediaan yang melimpah dari jumlah data yang besar serta kebutuhan mendesak untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan yang dapat diterapkan. Informasi dan pengetahuan yang dihasilkan dari proses ini dapat dimanfaatkan untuk berbagai aplikasi seperti analisis pasar, deteksi penipuan, retensi pelanggan, pengendalian produksi, dan eksplorasi ilmiah (Han & Kamber, 2012). Data mining merupakan proses mencari hubungan yang signifikan, pola, dan tren baru dengan menyelidiki sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika (Larose, 2014).

Rapid Miner

RapidMiner merupakan perangkat lunak berbasis open source yang dirancang untuk mendukung proses

analisis data, termasuk di dalamnya data mining, text mining, serta analisis prediktif. Dengan menggunakan berbagai teknik deskriptif dan prediktif, Rapid Miner memberikan wawasan yang mendalam kepada pengguna untuk mendukung pengambilan keputusan yang optimal. Terdapat sekitar 500 operator data mining dalam Rapid Miner, termasuk operator untuk input, output, preprocessing data, dan visualisasi. Software ini berdiri sendiri sebagai alat analisis data dan mesin data mining yang dapat diintegrasikan ke dalam produk-produknya sendiri. Rapid Miner ditulis dalam bahasa Java, memungkinkannya berjalan di berbagai sistem operasi (Ardiansyah, 2019).

Microsoft Excel

Microsoft Excel merupakan perangkat lunak lembar kerja (spreadsheet) yang dikembangkan oleh Microsoft dan berfungsi untuk mengelola serta menganalisis data numerik dalam format tabel. Aplikasi ini digunakan secara luas untuk berbagai kebutuhan pengolahan data, baik dalam konteks bisnis, pendidikan, maupun penelitian.

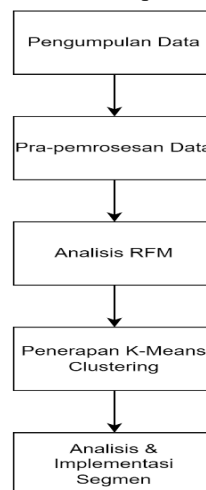
Jupyter Lab

Jupyter Lab adalah pilihan pertama pada tahap awal analisis, terutama saat membuat model RFM. Dalam praktiknya, Jupyter digunakan untuk membersihkan data, menghitung nilai RFM, dan membuat visualisasi awal untuk membantu memahami karakteristik umum pelanggan.

METODE

Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan menerapkan metode analisis RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means Clustering* sebagai teknik segmentasi pelanggan. Pendekatan ini dipilih karena mampu mengolah data transaksi pelanggan secara objektif dan menghasilkan pengelompokan berdasarkan pola perilaku belanja yang terukur. Proses segmentasi ini dilakukan dengan memanfaatkan data transaksi ritel dari *platform Gigasheet*, yang kemudian diolah menggunakan perangkat lunak RapidMiner dan *Microsoft Excel*. Tahapan-tahapan tersebut dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Tahapan penelitian ini dimulai dengan melakukan analisis RFM untuk mengukur nilai pelanggan berdasarkan tiga dimensi utama: keterkinian transaksi (*recency*), frekuensi transaksi (*frequency*), dan nilai moneter dari transaksi (*monetary*). Analisis RFM bertujuan untuk mengidentifikasi pelanggan yang paling berharga dan memahami perilaku pembelian mereka. Setelah nilai RFM untuk setiap pelanggan dihitung, data tersebut akan digunakan sebagai input untuk teknik *Clustering K-Means*. Teknik *K-Means* akan mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan kesamaan nilai RFM mereka. Segmen-segmen ini akan dianalisis lebih lanjut untuk memberikan wawasan mendalam tentang karakteristik dan perilaku masing-masing kelompok pelanggan, yang kemudian dapat digunakan untuk merancang strategi pemasaran yang lebih efektif dan terarah.

Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari *Platform Gigasheet*, sebuah penyedia data yang menawarkan berbagai dataset untuk tujuan analisis dan penelitian. Dataset yang digunakan adalah "*Retail Sales Dataset*" yang mencakup informasi transaksi penjualan ritel yang relevan untuk analisis RFM dan penerapan teknik *Clustering K-Means*. Dataset ini merupakan data sekunder yang tersedia untuk penggunaan publik dan disertai dengan lisensi yang memungkinkan penggunaan untuk keperluan penelitian dan analisis. Sebelum menggunakan data, pastikan untuk memeriksa dan mematuhi ketentuan lisensi yang disediakan oleh *Gigasheet*.

untuk memastikan penggunaan data yang sesuai dan legal. Dataset ini terdiri dari beberapa atribut penting, antara lain:

- Customer ID*: Identifikasi unik untuk setiap pelanggan
- Date Order*: Waktu pemesanan transaksi
- Total Amount*: Total nilai transaksi

Meskipun kumpulan data ini berisi banyak informasi, peneliti akan fokus pada tiga atribut utama ini untuk analisis RFM dan tujuan segmentasi pelanggan. Pemilihan ini didasarkan pada relevansi atribut-atribut tersebut dalam membentuk model RFM yang menjadi inti analisis segmentasi pelanggan dalam penelitian ini. Namun, harap diperhatikan bahwa peneliti melakukan proses pembersihan dan pra-pemrosesan data untuk memastikan kualitas dan relevansi data sebelum menggunakannya untuk analisis. Ini termasuk menangani nilai yang hilang, menghilangkan duplikat, dan normalisasi data. Hal ini dijelaskan lebih rinci pada bagian teknik pengolahan data.

Pemilihan sumber data ini konsisten dengan tujuan CRM dalam menggunakan data pelanggan secara efektif untuk meningkatkan pemahaman pelanggan dan mengoptimalkan strategi bisnis. Dengan menganalisis pola transaksi masa lalu, kita dapat mengidentifikasi segmen pelanggan yang berbeda dan mengembangkan pendekatan yang lebih personal untuk melayani pelanggan tersebut.

Teknik Pengolahan Data

Teknik pengolahan data dilakukan untuk memastikan kualitas dan relevansi data serta mempersiapkannya untuk analisis RFM dan pengelompokan dengan *K-Means Clustering*. Teknik pengolahan data dilakukan dalam empat tahap utama yaitu pemilihan data, pembersihan data, dan pembuatan tabel pivot.

Metode Analisis

Metode analisis yang digunakan dalam penelitian ini melibatkan dua tahap utama: analisis RFM dan teknik *Clustering K-Means*. Pertama, analisis RFM dilakukan untuk menghitung nilai keterkinian, frekuensi, dan nilai moneter dari transaksi pelanggan. Nilai-nilai ini membantu mengidentifikasi pelanggan yang paling berharga berdasarkan perilaku pembelian mereka. Setelah nilai RFM dihitung, teknik *Clustering K-Means* diterapkan untuk mengelompokkan pelanggan ke dalam segmen-segmen yang homogen. Algoritma *K-Means* akan mengelompokkan pelanggan berdasarkan kesamaan nilai RFM mereka, menghasilkan segmen-segmen yang dapat dianalisis lebih lanjut. Proses ini memungkinkan pemahaman yang lebih mendalam tentang karakteristik dan perilaku masing-masing segmen, yang selanjutnya digunakan untuk merancang strategi pemasaran yang lebih efektif dan terarah.

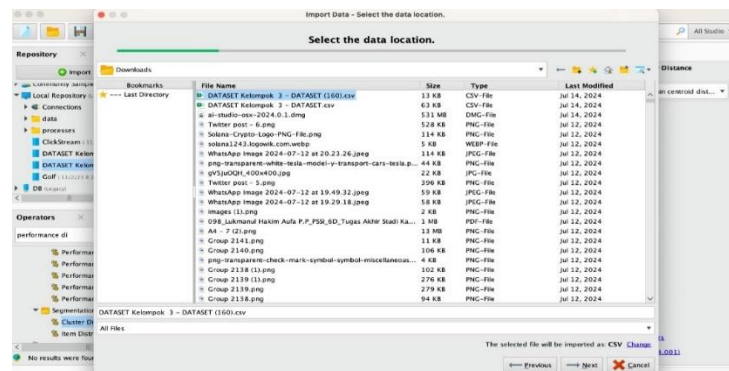
HASIL

Pengolahan Data

Pada tahap pengolahan data, kami mengambil dataset "Retail Sales Dataset" dari *Gigasheet* yang mencakup informasi transaksi penjualan ritel yang relevan. Data ini meliputi tanggal transaksi, jumlah transaksi, dan nilai moneter transaksi dari pelanggan. Dataset Transaksi dapat dilihat pada Tabel 1 dan Input Data dapat dilihat pada Gambar 2.

Tabel 1 Dataset Transaksi

<i>Customer ID</i>	<i>Gender</i>	<i>Age</i>	<i>Product Category</i>	<i>Quantity</i>	<i>Price per Unit</i>	<i>Total Amount</i>
CUST001	Male	34	Beauty	3	50	150
CUST002	Female	26	Clothing	2	500	1000
CUST003	Male	50	Electronics	1	30	30
...	...	...	...	...	...	...
CUST1000	Male	47	Electronics	4	30	120



Gambar 2. Input Data

Pra-pemrosesan Data

Setelah mengumpulkan data dari dataset "Retail Sales Dataset" yang disediakan oleh *Gigasheet*, langkah pertama yang kami ambil adalah melakukan pra-pemrosesan data secara menyeluruh. Pra-pemrosesan ini bertujuan untuk membersihkan data dari kesalahan, mengidentifikasi serta menghilangkan duplikasi, serta mengisi nilai yang hilang agar dataset yang dihasilkan menjadi lebih berkualitas dan siap untuk analisis lebih lanjut. Proses ini krusial untuk memastikan integritas data sebelum kami melangkah ke tahap analisis lebih lanjut seperti analisis RFM dan penerapan teknik *Clustering K-Means*. Tabel pivot dapat dilihat pada Tabel 2.

Tabel 2 Pivot

No.	Customer ID	Date	SUM of Order	Total Amount
1	CUST001	2023-11-24	3	150
2	CUST002	2023-02-27	2	1000
3	CUST003	2023-01-13	1	30
...	...	...	...	...
1000	CUST1000	2023-04-12	4	120

Analisis RFM

Setelah proses pra-pemrosesan data selesai, langkah selanjutnya dalam analisis adalah melakukan analisis RFM (*Recency, Frequency, Monetary*). Analisis RFM dilakukan untuk setiap pelanggan dengan tujuan untuk mengukur dan mengevaluasi nilai dan pola perilaku pembelian mereka. Pertama, kami menghitung nilai *Recency*, yang mengukur seberapa baru pelanggan melakukan transaksi terakhir kali. Frekuensi (*Frequency*) dihitung untuk menentukan seberapa sering pelanggan melakukan transaksi dalam periode waktu tertentu. Sedangkan nilai *Monetary* mengukur total nilai uang yang dihabiskan oleh pelanggan dalam periode yang sama. Tabel RFM dapat dilihat pada Tabel 3.

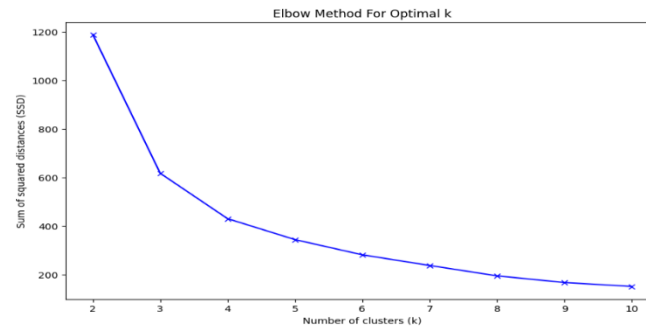
Tabel 3 RFM

Customer ID	R Score	F Score	M Score	RFM Score
CUST001	3	3	1	7
CUST002	1	2	2	5
CUST003	1	1	1	3
..	...	...	...	...
CUST1000	1	3	1	5

Analisis RFM ini memberikan gambaran yang jelas tentang segmentasi pelanggan berdasarkan tingkat keaktifan dan nilai pembelian mereka. Hasil dari analisis RFM akan digunakan sebagai basis untuk melakukan *Clustering* menggunakan teknik *K-Means*.

Penerapan K-Means Clustering**Elbow Method**

Pertama, kami menghitung nilai WCSS (*Within-Cluster Sum of Squares*) untuk berbagai nilai k, yang mengukur variabilitas data dalam setiap kluster. Kami kemudian membuat plot dari nilai WCSS terhadap jumlah kluster k, mencari titik di mana penurunan WCSS mulai melambat secara signifikan. Titik ini, yang terlihat sebagai lengkungan "siku" dalam grafik, menandai titik "*elbow*" di mana penambahan kluster lebih lanjut tidak memberikan penurunan yang signifikan dalam variabilitas data. Dengan demikian, nilai k yang dipilih adalah nilai di sekitar titik *elbow* ini, yang memberikan keseimbangan optimal antara kompleksitas model dan kemampuan model untuk menggambarkan variasi dalam data. Grafik *elbow method* dapat dilihat pada Gambar 3.



Gambar 3. Elbow Method

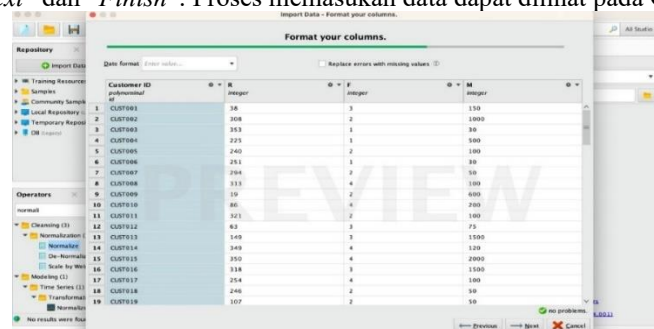
Pada grafik tersebut, sumbu horizontal mewakili jumlah kluster k , sedangkan sumbu vertikal mewakili nilai WCSS. Saat nilai k meningkat dari 1 ke 4, WCSS umumnya akan menurun secara signifikan. Ini karena dengan penambahan kluster, model *K-Means* memiliki lebih banyak fleksibilitas untuk menyesuaikan data ke dalam kelompok-kelompok yang lebih kecil dan lebih homogen.

Namun, setelah mencapai $k=4$, penurunan dalam WCSS mulai melambat secara tajam, membentuk lengkungan yang menyerupai "siku" dalam grafik. Titik di mana lengkungan ini terbentuk adalah titik "*elbow*". Di titik ini, penambahan kluster lebih lanjut tidak memberikan penurunan signifikan dalam nilai WCSS. Oleh karena itu, $k=4$ adalah nilai yang dipilih karena memberikan keseimbangan terbaik antara kompleksitas model (jumlah kluster) dan kemampuan model untuk menggambarkan variasi dalam data dengan baik.

Konvergensi

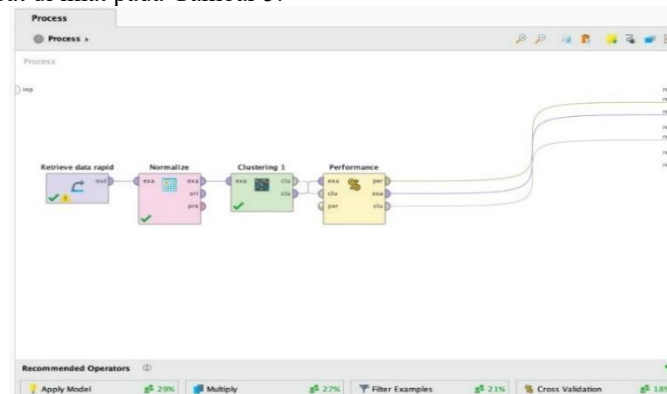
Setelah melakukan *Elbow Method* untuk menentukan jumlah kluster optimal, langkah selanjutnya dengan menggunakan Rapid Miner yaitu menerapkan algoritma *K-Means Clustering* untuk dataset yang ada.

Pertama, masukkan data .CSV dengan cara menarik operator "*Read CSV*" ke dalam margin "*Process*". Lalu pada tab "*Parameters*", pilih *Import Configuration Wizard* dan cari data yang ingin dimasukkan. Lalu klik "*Next*", pada bagian *Customer ID*, ubah *role* menjadi "*id*" dan pastikan tipe daya atribut lainnya adalah "*Int (Integer)*". Lalu klik "*Next*" dan "*Finish*". Proses memasukan data dapat dilihat pada Gambar 4.



Gambar 4. Memasukan Data

Setelah itu tarik operators lainnya seperti "*Normalize*", "*Clustering K-Means*", dan "*Performances*". Pada operator "*Clustering K-Means*" sesuaikan jumlah "*K*" atau kluster pada tab "*Parameters*". Dalam kasus ini, $K=4$ sesuai dengan hasil *Elbow Method* yang sebelumnya sudah dilakukan. Lalu di *Run*. Tampilan *design process Clustering K-Means* dapat di lihat pada Gambar 5.

Gambar 5. Tampilan Design Process *Clustering K-Means*

Pada operator “*Normalize*”, kita bisa melihat hasil data yang sudah di normalisasi setelah di “*Run*”. Data hasil normalisasi dapat dilihat pada Gambar 6.

Row No.	Customer ID	R	F	M
1	CUST001	-1.368	0.429	-0.546
2	CUST002	1.197	-0.454	0.971
3	CUST003	1.624	-1.337	-0.761
4	CUST004	0.409	-1.337	0.079
5	CUST005	0.551	-0.454	-0.636
6	CUST006	0.656	-1.337	-0.761
7	CUST007	1.064	-0.454	-0.725
8	CUST008	1.245	1.312	-0.636
9	CUST009	-1.548	-0.454	0.257
10	CUST010	-0.912	1.312	-0.457
11	CUST011	1.321	-0.454	-0.636
12	CUST012	-1.130	0.429	-0.680
13	CUST013	-0.313	0.429	1.864
14	CUST014	1.586	1.312	-0.600
15	CUST015	1.596	1.312	2.757
16	CUST016	1.292	0.429	1.864

ExampleSet (1,000 examples, 1 special attribute, 3 regular attributes)

Gambar 6. Data Hasil Normalisasi

Setelah di “*Run*”, akan keluar hasil klasterisasi dari jumlah kluster yang sudah ditentukan. Hasil model *Clustering* dapat dilihat pada Gambar 7.

Cluster Model

```
Cluster 0: 265 items
Cluster 1: 231 items
Cluster 2: 269 items
Cluster 3: 235 items
Total number of items: 1000
```

Gambar 7. Model *Clustering*

Rapid Miner akan secara otomatis menetapkan centroid awal, menghitung jarak antara titik data dan centroid, dan mengelompokkan titik data ke dalam kluster yang sesuai. Proses ini berulang hingga konvergensi tercapai, di mana tidak ada lagi perubahan signifikan dalam *assignment* kluster atau perubahan centroid melebihi ambang tertentu. Hasil *Clustering* yang dihasilkan dapat dievaluasi lebih lanjut menggunakan metrik evaluasi seperti *Cluster Distance Performance* di Rapid Miner. Metrik ini membantu dalam memahami seberapa baik titik data dikelompokkan dalam kluster yang homogen berdasarkan jarak antara titik data dan centroid kluster, sehingga kita akan mendapatkan model *Cluster* dengan nilai terbaik.

Analisis & Implementasi Segmen

Dari hasil *Clustering* kita menggunakan algoritma *K-Means*, kita berhasil mengidentifikasi tiga kluster utama dari data pelanggan. Berikut adalah ringkasan dari nilai RFM untuk setiap kluster. Tabel cluster dapat dilihat Tabel 4.

Tabel 4 *Cluster*

Cluster 1			
Customer_id	R	F	M
CUST002	1	2	2
CUST003	1	1	1
CUST004	2	1	2
...	...	...	...
CUST996	2	1	1

Cluster 2			
Customer_id	R	F	M

CUST013	2	3	3
CUST015	1	3	3
CUST016	1	3	3
...	...	...	...
CUST990	2	2	2

<i>Cluster 3</i>			
<i>Customer_id</i>	R	F	M
CUST001	3	3	1
CUST009	3	2	2
CUST012	3	3	1
...	...	...	...
CUST999	3	3	1

<i>Cluster 4</i>			
<i>Customer_id</i>	R	F	M
CUST008	3	3	1
CUST010	3	2	2
CUST014	3	3	1
...	...	...	...
CUST1000	3	3	1

Dari analisis tersebut, terlihat bahwa *Cluster 2* memiliki nilai RFM yang paling tinggi dibandingkan dengan *Cluster 1* dan *Cluster 3*. *Cluster 1* menunjukkan nilai RFM yang berada di bawah *Cluster 2*, menunjukkan bahwa pelanggan dalam kluster ini cenderung memiliki frekuensi transaksi dan nilai pembelian yang lebih rendah dibandingkan dengan *Cluster 2*. Sementara itu, *Cluster 3* memiliki rata-rata RFM score terendah di antara ketiganya, menandakan bahwa pelanggan dalam kluster ini adalah yang paling jarang bertransaksi dan memberikan nilai pembelian yang lebih rendah dibandingkan dengan kluster lainnya. Informasi ini memberikan pemahaman yang lebih mendalam tentang perilaku dan karakteristik setiap segmen pelanggan, yang sangat berharga untuk mengarahkan strategi pemasaran yang lebih terfokus dan efektif serta untuk meningkatkan pengalaman pelanggan secara keseluruhan. Berikut adalah Tabel 5 untuk data centroid dari hasil *Clustering* menggunakan algoritma *K-Means*:

Tabel 5 Tabel Centroid

<i>Cluster</i>	R	F	M
<i>Cluster 1</i>	0.918	-0.817	-0.400
<i>Cluster 2</i>	-0.059	0.754	1.581
<i>Cluster 3</i>	-0.952	-0.743	-0.457
<i>Cluster 4</i>	0.112	1.030	-0.579

Dari tabel tersebut, kita dapat melihat bahwa *Cluster 1* memiliki nilai rata-rata *Recency* sebesar 0.918, *Frequency* sebesar -0.817, dan *Monetary* sebesar -0.400. Ini menunjukkan bahwa pelanggan dalam *Cluster 1* cenderung memiliki transaksi yang cukup sering, namun dengan nilai pembelian yang lebih rendah dibandingkan dengan kluster lainnya.

Cluster 2 memiliki nilai rata-rata *Recency* sebesar -0.059, *Frequency* sebesar 0.754, dan *Monetary* sebesar 1.581. Pelanggan dalam kluster ini adalah yang paling aktif dan berharga, dengan frekuensi transaksi tertinggi dan nilai pembelian yang paling besar. Hal ini menandakan bahwa *Cluster 2* terdiri dari pelanggan yang sangat bernilai bagi perusahaan.

Cluster 3 menunjukkan nilai rata-rata *Recency* sebesar -0.952, *Frequency* sebesar -0.743, dan *Monetary* sebesar -0.457. Pelanggan dalam kluster ini memiliki frekuensi transaksi yang terendah dibandingkan dengan kluster lainnya, menunjukkan bahwa mereka cenderung tidak sering berbelanja. *Cluster 4* menunjukkan nilai rata-

rata *Recency* sebesar 0.112, *Frequency* sebesar 1.030, *Monetary* sebesar -0.579. Pelanggan dalam kluster ini memiliki frekuensi transaksi yang cukup positif, namun dengan nilai pembelian yang lebih rendah dibandingkan dengan kluster lainnya. Untuk melihat tabel perankingan *cluster* dapat dilihat pada Tabel 6.

Tabel 6 Perankingan *Cluster*

No	Clusters	RFM SCORE
1	Cluster 2	2,276
2	Cluster 4	0,563
3	Cluster 1	-0,299
4	Cluster 3	-2,152

Dengan mempertimbangkan karakteristik masing-masing kluster berdasarkan RFM *score*, kami akan memberikan label yang tepat untuk setiap segmen pelanggan:

1. **Cluster 1 - Occasional Shoppers:** Kluster ini memiliki rata-rata RFM *score* yang cukup rendah, ini menunjukkan bahwa pelanggan dalam kluster ini memiliki tingkat transaksi yang cukup rendah dan memberikan nilai pembelian yang lebih rendah dibandingkan dengan kluster lainnya, tetapi masih menunjukkan kecenderungan cerdas dalam pengeluaran mereka.
2. **Cluster 2 - Prime Buyers:** Kluster ini memiliki nilai RFM yang paling tinggi dibandingkan dengan Cluster 1 dan Cluster 3, menandakan bahwa pelanggan dalam kluster ini adalah yang paling bernilai dan aktif secara finansial.
3. **Cluster 3 - Sporadic Spenders:** Kluster ini memiliki nilai RFM sedang dengan frekuensi transaksi yang tidak teratur dan potensi pembelian yang meningkat. Ini menunjukkan bahwa pelanggan dalam kluster ini jarang bertransaksi dan memiliki nilai pembelian yang sangat rendah, dengan sedikit interaksi dan kontribusi terhadap total pendapatan.
4. **Cluster 4 - Smart Spenders:** Kluster ini memiliki nilai RFM di bawah Cluster 2, menunjukkan bahwa pelanggan dalam kluster ini cenderung memiliki frekuensi transaksi dan nilai pembelian yang lebih rendah dibandingkan dengan Cluster 2.

Tabel 7 Segmentasi *Customer* & Rencana Strategi

Segmentasi	Karakteristik	Strategi
<i>Prime Buyers</i>	- Nilai RFM tinggi - Nilai pembelian tinggi	- Pertahankan loyalitas dengan program VIP - <i>Personalize</i> penawaran dan komunikasi
<i>Smart Spenders</i>	- Nilai RFM moderat - Nilai pembelian moderat	- Tawarkan promosi dan diskon yang menarik - Edukasi pelanggan tentang nilai produk dan layanan
<i>Occasional Shoppers</i>	- Nilai RFM cukup rendah - Nilai pembelian cukup rendah	- Tingkatkan kunjungan dengan promo khusus - Dorong pembelian berulang dengan insentif
<i>Sporadic Spenders</i>	- Nilai RFM terendah di antara semua kluster - Nilai pembelian terendah	- Tingkatkan engagement dengan konten edukatif - Tawarkan insentif untuk meningkatkan frekuensi pembelian

KESIMPULAN

Berdasarkan hasil dari penelitian ini menunjukkan bahwa penerapan analisis RFM dan algoritma *K-Means Clustering* berhasil mengidentifikasi empat segmen utama pelanggan Transmart Carrefour, yaitu *Prime Buyers*, *Smart Spenders*, *Occasional Shoppers*, dan *Sporadic Spenders*. Hasil analisis menunjukkan bahwa Cluster 2 (*Prime Buyers*) memiliki skor RFM tertinggi sebesar 2,276, menunjukkan pelanggan dengan aktivitas belanja paling sering dan nilai transaksi tertinggi, sehingga menjadi segmen paling potensial. Sebaliknya, Cluster 3 (*Sporadic Spenders*) memiliki skor RFM terendah sebesar -2,152, menggambarkan pelanggan yang jarang bertransaksi dengan nilai belanja rendah. Segmentasi ini memungkinkan perusahaan menyusun strategi

pemasaran yang lebih tepat, seperti program loyalitas untuk *Prime Buyers* dan kampanye promosi atau edukasi untuk klaster dengan nilai rendah. Dengan demikian, metode ini terbukti efektif dalam menjawab rumusan masalah, yaitu bagaimana mengelompokkan pelanggan berdasarkan perilaku belanja guna mendukung strategi pemasaran yang lebih personal dan efisien.

REFERENSI

- Adiana, D. A., Suwendra, I. W., & Sudarma, K. (2018). Analisis Loyalitas Pelanggan Menggunakan RFM Model dan Clustering. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(2), 210-217.
- Adriansyah, R., & Hermanto, B. (2021). Strategi Pemasaran Ritel di Era Digital. *Jurnal Bisnis dan Manajemen*, 9(1), 33-42.
- Angelie, Y. (2017). Segmentasi Pelanggan Berbasis Data Mining pada Industri Ritel. *Jurnal Sistem Informasi*, 13(2), 97-105.
- Anam, M. K., Putri, Y. A., & Saputro, D. R. (2022). Segmentasi Pelanggan Menggunakan K-Means Clustering pada Data Supermarket. *Jurnal Teknologi dan Sistem Informasi*, 10(1), 13-20.
- Ardiansyah, M. I. (2019). Implementasi Data Mining Menggunakan RapidMiner. *Jurnal Informatika*, 6(2), 88-93.
- Ardiansyah, R., & Wibowo, S. (2022). Segmentasi Pelanggan Ritel Menggunakan Metode Clustering. *Jurnal Teknologi Informasi dan Komunikasi*, 11(1), 45-52.
- Awalina, N., & Rahayu, F. (2023). Optimalisasi Strategi Pemasaran Melalui Segmentasi Pelanggan Online Retail dengan K-Means Clustering. *Jurnal Ilmiah Teknologi Informasi*, 14(1), 29-37.
- Cheng, C. S., Chen, T. H., & Tu, C. T. (2009). An RFM Analysis of Customer Behavior. *International Journal of Business and Management*, 4(2), 56-66.
- Han, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Haryadi, B., Nugroho, R., & Lestari, T. (2024). Customer Segmentation dengan Metode K-Means Clustering: Studi Kasus Data Ritel. *Jurnal Sistem dan Teknologi Informasi*, 8(2), 101-110.
- Khairina, N., & Harahap, M. K. (2018). Menjaga Kerahasiaan Data dengan Steganografi Kombinasi LSB-2 dengan LSB-3. *SinkrOn - Jurnal & Penelitian Teknik Informatika*, 3(1), 286-288.
- Khairina, N., Harahap, M. K., & Lubis, J. H. (2018). The Authenticity of Image using Hash MD5 and Steganography Least Significant Bit. *International Journal of Information System & Technology*, 2(1), 1-6.
- Larose, D. T. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2nd ed.). Wiley.
- Lee, J. (2014). Enhancing Customer Value through RFM Model Analysis. *Journal of Marketing Development and Competitiveness*, 8(1), 94-102.
- Maresti, A., Prasetyo, R., & Wahyuni, E. (2025). Segmentasi Produk Ritel Menggunakan K-Means Clustering. *Jurnal Teknologi dan Informasi*, 15(1), 1-9.
- Peppard, J. (2000). Customer Relationship Management (CRM) in Financial Services. *European Management Journal*, 18(3), 312-327.