

Segmentasi Pelanggan Menggunakan *K-Means* *Clustering*: Menganalisis Metrik RFM untuk Strategi Pemasaran

Ika Yunita^{1*}, Puti Retno Ali², Maria Angela Kartawidjaja³, Ronald Sukwadi^{4*}

¹ PASSOLVING

Agility Solution Workspace, Jakarta 13820

² Teknik Industri Universitas Muhammadiyah Cirebon

Jl. Watubelah No. 40, Cirebon 45611

^{1,3,4} Program Studi Program Profesi Insinyur Universitas Katolik Indonesia Atma Jaya

Jl. Jenderal Sudirman, Karet Semanggi, Jakarta 12930

^{1*} ikayunita0512@gmail.com

² puti.retnoali@umc.ac.id

³ maria.kw@atmajaya.ac.id

⁴ ronald.sukwadi@atmajaya.ac.id

Customer Segmentation Using *K-Means* Clustering: Analyzing RFM Metrics for Enhanced Marketing Strategies

Dikirimkan : 07, 2024. Diterima : 03, 2025. Dipublikasikan : 03, 2025.

Abstract— *Means clustering algorithm for customer segmentation in the retail sector. The background of the research is based on the need to gain a deeper understanding of customer behavior in order to develop more effective Customer Relationship Management (CRM) strategies tailored to each customer segment. The data used in this study was collected from sales transactions over a period of three years, from April 1, 2021, to August 23, 2023, totaling 62,677 transactions. The primary objective is to enhance CRM strategies by categorizing customers based on their purchasing behavior. The K-Means algorithm was employed to group customers according to their RFM values, while the Elbow Method and Silhouette Analysis were used to determine the optimal number of clusters. As a result, customers were classified into three main segments: General Development, General Maintenance, and Important Maintenance. Each segment is characterized by a unique combination of recency, frequency, and monetary values, providing insights into targeted marketing approaches. Short-term strategies include personalized promotions and targeted campaigns to encourage repeat purchases and increase the average order value. Long-term initiatives focus on developing loyalty programs and VIP services to enhance customer retention and lifetime value. These findings emphasize the effectiveness of data-driven CRM strategies in optimizing customer engagement and profitability in the competitive retail landscape.*

Keywords— CRM; customer segmentation; K-mean clustering; RFM analysis; retail industry

Abstrak— Penelitian ini mengkaji penerapan analisis RFM (*Recency, Frequency, Monetary*) dan algoritma *clustering K-Means* dalam melakukan segmentasi pelanggan di sektor ritel. Latar belakang penelitian ini didasarkan pada pentingnya memahami perilaku pelanggan secara lebih mendalam untuk menyusun strategi *Customer Relationship Management (CRM)* yang lebih efektif dan disesuaikan dengan kebutuhan masing-masing segmen pelanggan. Data yang digunakan berasal dari transaksi penjualan selama periode tiga tahun, yakni dari 1 April 2021 hingga 23 Agustus 2023, dengan total sebanyak 62.677 transaksi. Tujuan utama

penelitian ini adalah untuk meningkatkan strategi CRM melalui pengelompokan pelanggan berdasarkan perilaku pembelian mereka. Algoritma *K-Means* digunakan untuk mengklasifikasikan pelanggan berdasarkan nilai RFM, sementara Metode *Elbow* dan Analisis *Silhouette* diterapkan untuk menentukan jumlah *cluster* yang optimal. Berdasarkan hasil analisis, pelanggan terbagi menjadi tiga segmen utama: *General Development*, *General Maintenance*, dan *Important Maintenance*. Masing-masing segmen ditentukan oleh kombinasi unik dari nilai *recency*, *frequency*, dan *monetary*, yang memberikan wawasan dalam menentukan strategi pemasaran yang lebih tepat sasaran. Strategi jangka pendek mencakup promosi yang dipersonalisasi dan kampanye yang ditargetkan untuk mendorong pembelian berulang dan meningkatkan nilai rata-rata pesanan. Sedangkan inisiatif jangka panjang berfokus pada pengembangan program loyalitas dan layanan VIP untuk meningkatkan retensi pelanggan serta nilai seumur hidup mereka. Hasil penelitian ini menegaskan efektivitas strategi CRM berbasis data dalam mengoptimalkan keterlibatan pelanggan dan profitabilitas perusahaan di lingkungan ritel yang kompetitif.

Kata kunci—Analisis RFM; CRM; industri retail; *K-mean cluster*; segmentasi pelanggan

I. PENDAHULUAN

Pertanyaan mendasar dalam strategi bisnis dan pemasaran adalah mengapa beberapa perusahaan atau merek berkinerja lebih baik dibandingkan pesaing mereka meskipun menawarkan produk serupa. Fenomena ini sering kali membingungkan para peneliti, praktisi, dan investor, terutama ketika perbedaan signifikan dalam volume penjualan muncul di antara merek-merek yang bersaing. Untuk menjawab pertanyaan ini, perusahaan sering menggunakan segmentasi konsumen, di mana pelanggan dikelompokkan berdasarkan kebutuhan, atribut, atau perilaku yang serupa [1]. Segmentasi konsumen, yang merupakan inti dari Teori Segmentasi, Penargetan, dan Pemosisian (STP) oleh Philip Kotler [2], membantu bisnis memahami pelanggan mereka secara mendalam, sehingga memungkinkan penerapan strategi pemasaran yang efektif dan meningkatkan loyalitas serta retensi pelanggan [3], [4]. Salah satu metode yang digunakan adalah melakukan segmentasi [5], [6], [8]. Metode segmentasi seperti analisis RFM dapat digunakan dengan memanfaatkan informasi perilaku untuk mengelompokkan konsumen secara ringkas dan tepat guna [1], [3], [4], [7], [8].

Algoritma *K-Means* biasanya menggunakan dua jenis segmentasi utama: wawasan kualitatif dan kuantitatif. Studi ini menekankan penggunaan wawasan kuantitatif untuk pengelompokan segmentasi. Segmentasi konsumen yang didefinisikan dengan jelas membantu dalam alokasi aset periklanan secara efisien, memungkinkan bisnis untuk fokus pada kelompok konsumen tertentu dan membina hubungan yang langgeng. Industri ritel atau perdagangan adalah sektor kunci di mana segmentasi konsumen dan penambangan data sangat penting karena data yang luas tentang logistik, pendapatan, rasio konsumsi, layanan pengiriman ulang, dan lainnya. Penambangan data ritel atau perdagangan membantu mengenali dan memetakan perilaku dan tren konsumen sepanjang siklus hidup bisnis, meningkatkan layanan pelanggan, strategi penjualan, dan banyak lagi.

Selain itu, pemilihan algoritma *clustering* yang tepat sangat bergantung pada tujuan analisis dan karakteristik data yang digunakan [9]. *Clustering*, sebuah metode pembelajaran tanpa supervisi, dapat diklasifikasikan ke dalam kategori *hard clustering* dan *soft clustering* [10]. *Hard clustering*, seperti *K-Means* dan *hierarchical clustering*, lebih cocok untuk *dataset* yang memiliki batasan *cluster* yang jelas dan tidak tumpang tindih. Sebaliknya, *soft clustering*, seperti Fuzzy C-means, lebih bermanfaat ketika data menunjukkan tumpang tindih antar-*cluster*, karena memungkinkan fleksibilitas dalam keanggotaan *cluster*. Algoritma yang dipilih harus sesuai dengan konteks bisnis dan tujuan spesifik, misalnya untuk mendeteksi pola perilaku pelanggan yang lebih halus atau kompleks. Dengan mempertimbangkan kekuatan dan keterbatasan dari masing-masing algoritma, bisnis dapat mengoptimalkan hasil segmentasi pelanggan. Segmentasi yang efektif, baik melalui *hard* maupun *soft clustering*, memungkinkan perusahaan untuk menyesuaikan strategi pemasaran secara lebih presisi, meningkatkan relevansi penawaran produk, dan memperkuat loyalitas pelanggan.

Model RFM unggul sebagai teknik segmentasi perilaku yang populer [9], [11]. Model RFM mengelompokkan pelanggan berdasarkan *recency* (sejak waktu pembelian terakhir), *frequency* (jumlah total pembelian), dan *monetary* (nilai pembelian total dalam periode waktu tertentu) dengan tujuan untuk mengidentifikasi pelanggan paling berharga bagi upaya pemasaran yang ditargetkan [12], [13].

Segmentasi pelanggan berbasis model RFM telah diilustrasikan melalui berbagai teknik pengelompokan. Penerapan *business intelligence* untuk mengidentifikasi calon pelanggan di industri ritel dengan menyediakan data yang relevan dan tepat waktu melalui Teknik Pengelompokan *K-Means*, yang mengelompokkan pelanggan ke dalam kelompok-kelompok berbeda berdasarkan kesamaan dalam perilaku pembelian mereka, memungkinkan bisnis untuk menargetkan strategi pemasaran yang disesuaikan untuk segmen

pelanggan tertentu [1]. Dengan memanfaatkan *K-Means clustering*, mereka dapat memberikan wawasan yang dapat diimplementasikan, meningkatkan penargetan pelanggan, dan pada akhirnya memperbaiki hasil bisnis di sektor ritel.

Penelitian terdahulu terkait *K-Means* antara lain, pendekatan dua dimensi untuk mempelajari perilaku pelanggan dalam perbankan elektronik dan tradisional di Iran menggunakan *K-Means* dan DBSCAN [11], analisis dinamika keragaman perilaku pelanggan supermarket *multichannel* (baru) menggunakan Analisis Kelas Laten (LCA) [14], dan pemodelan perilaku pelanggan dan klasifikasi dalam ritel melalui analisis *K-Means* [15].

Perusahaan di sektor ritel ini menunjukkan permasalahan penurunan penjualan tanpa alasan yang jelas, kurangnya pemahaman terhadap perilaku pembelian pelanggan, serta rendahnya tingkat retensi dan loyalitas pelanggan. Selain itu, segmentasi pelanggan yang tidak efektif dan strategi pemasaran yang seragam sering kali menghasilkan kinerja program pemasaran yang buruk, terutama di tengah persaingan yang semakin ketat. Investasi pemasaran yang tidak efisien juga menjadi masalah, karena anggaran pemasaran tidak diarahkan kepada segmen pelanggan yang tepat.

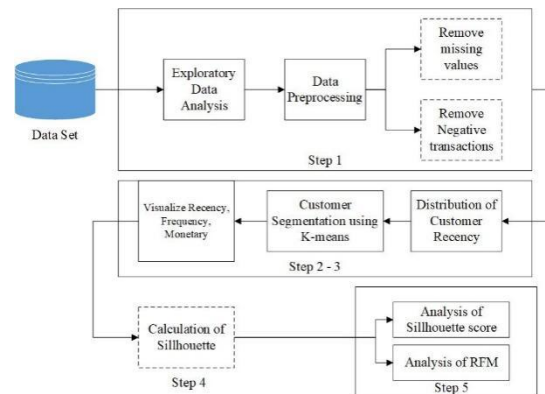
Berdasarkan permasalahan tersebut, studi ini menganalisis perilaku pembelian historis pelanggan untuk memaksimalkan penjualan. Perusahaan di sektor ritel dapat merancang strategi penjualan dan pemasaran, seperti kampanye promosi, diskon musiman, atau penawaran kupon, berdasarkan wawasan statistik untuk meningkatkan penjualan dan meningkatkan retensi pelanggan. Untuk mencapai tujuan ini, segmentasi pelanggan dilakukan melalui algoritma *K-Means*, dengan memanfaatkan nilai RFM dari *database* penjualan. RFM, yang merupakan singkatan dari *Recency* (ketika terakhir kali membeli), *Frequency* (frekuensi pembelian), dan *Monetary* (nilai moneter), memberikan wawasan tentang perilaku pelanggan dengan menganalisis pola pembelian, pembelian terbaru, dan profitabilitas. Analisis *Silhouette* adalah metode yang digunakan untuk memvalidasi *cluster* berdasarkan kohesi (seberapa erat hubungan antar item dalam *cluster*) dan separasi (seberapa jelas *cluster* tersebut dibedakan dari yang lain). Analisis ini membantu menentukan jumlah optimal *cluster* dengan menghitung Koefisien *Silhouette*, yang mempertimbangkan kedua aspek kohesi dan separasi. Algoritma *K-Means* mengelompokkan pelanggan ke dalam *cluster* yang berbeda berdasarkan nilai RFM mereka, memungkinkan perusahaan untuk mengidentifikasi dan menargetkan segmen pelanggan tertentu secara lebih efektif. Dengan memanfaatkan teknik-teknik ini, studi ini bertujuan untuk memberikan kerangka kerja yang kokoh untuk segmentasi pelanggan, membantu perusahaan

ritel menyesuaikan strategi pemasaran mereka untuk segmen pelanggan yang berbeda, dengan tujuan akhir meningkatkan penjualan dan meningkatkan retensi pelanggan.

II. METODOLOGI PENELITIAN

A. Model Resensi Frekuensi Moneter (RFM)

Bagian ini menjelaskan tujuan studi, algoritma yang digunakan, dan kerangka eksperimental yang dirancang untuk mencapai hasil yang diinginkan. *Cluster*, yang didefinisikan sebagai kelompok objek dengan karakteristik yang sama, sangat penting untuk segmentasi pelanggan dan analisis lanjutan. Penggunaan *K-Means* dalam segmentasi pelanggan telah terbukti efektif [16]. *K-Means*, sebuah algoritma *clusterisasi* berbasis prototipe, mengidentifikasi sejumlah *cluster* yang telah ditentukan dengan *centroid*-nya, menawarkan efisiensi komputasi yang superior dan kinerja yang lebih baik pada *dataset* besar dibandingkan metode lainnya [17]. Salah satu keunggulan *K-Means* adalah ketergantungannya pada satu parameter *input*, yaitu 'K,' yang meminimalkan tingkat salah klasifikasi data. Segmentasi pelanggan adalah aplikasi utama dari *K-Means*, dan studi ini menggunakan algoritma tersebut untuk tujuan tersebut. Pendekatan yang diusulkan dibagi menjadi lima langkah, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Kerangka kerja penelitian

Bagian berikut akan membahas secara spesifik setiap langkah:

Langkah 1. Analisis Eksplorasi dan Pra-pemrosesan Data

Metodologi yang diusulkan diimplementasikan pada *dataset* transaksi pelanggan tiga tahun. Data penjualan aktual dari perusahaan terdiri dari 62.882 nilai observasi yang disimpan dalam sebuah *file*, dengan periode transaksi antara 01-04-2021 dan 23-08-2023. Variabel yang diambil meliputi tanggal pembelian, kode pelanggan, nama kota pelanggan, dan total pembelian.

Langkah pertama melibatkan pelaksanaan analisis data eksplorasi (*Exploratory Data Analysis* atau

EDA) untuk mengungkap pola dalam data melalui analisis statistik atau representasi visual. EDA membantu dalam mengidentifikasi pelanggan yang berbeda, menganalisis proporsi pesanan teratas, meneliti integritas data, mengidentifikasi ketidakkonsistenan dalam deskripsi dan kode stok, serta memverifikasi keberadaan nilai *null*. Selain itu, pra-pemrosesan data dilakukan untuk mengidentifikasi dan menangani nomor identifikasi pelanggan yang hilang, transaksi negatif, dan ketidaklengkapan lainnya.

Langkah 2. Pelaksanaan Analisis RFM

Setelah pra-pemrosesan data, Analisis RFM dilakukan dengan memeriksa transaksi terbaru, frekuensi, dan pengeluaran pelanggan. Untuk menentukan variabel *recency*, dipilih tanggal referensi satu hari sebelum transaksi terakhir. Analisis RFM, metode segmentasi pelanggan yang banyak digunakan dalam pemasaran basis data [17], memainkan peran penting, terutama di industri ritel. Setiap pelanggan dievaluasi berdasarkan tiga faktor [17]:

- a) *Recency*: Menunjukkan jumlah hari sebelum tanggal referensi sejak pembelian terakhir. Nilai *recency* yang lebih rendah menunjukkan kunjungan yang lebih sering oleh pelanggan.
- b) *Frequency*: Mewakili selang waktu antara dua pembelian berturut-turut oleh seorang pelanggan. Nilai *frequency* yang lebih tinggi menunjukkan peningkatan kunjungan ke perusahaan.
- c) *Monetary*: Mencerminkan jumlah uang yang dihabiskan oleh pelanggan dalam jangka waktu tertentu. Nilai *monetary* yang lebih tinggi menunjukkan keuntungan yang lebih besar bagi perusahaan.

Langkah 3. Clusterisasi K-Means dan Langkah 4. Perhitungan Silhouette Score

Penggunaan algoritma *K-Means* untuk clusterisasi adalah teknik pembelajaran tanpa pengawasan yang diterapkan dalam analisis data. Metode ini melibatkan penentuan 'K' *centroid* dari *dataset* 'D' dan mengalokasikan titik data yang tidak tumpang tindih ke *cluster* terdekat. Dalam algoritma *K-Means*, jarak *intracluster* dimaksimalkan dibandingkan dengan jarak antar-*cluster*. Melalui proses *iteratif*, titik data dialokasikan kembali ke berbagai *cluster* berdasarkan perhitungan *centroid*. Langkah-langkah algoritma *K-Means* adalah sebagai berikut:

1. Membaca *dataset*, memasukkan jumlah *cluster*
2. Memilih pusat *cluster* awal
3. Menghitung jarak antara data yang tersisa dan pusat *cluster*
4. Mengalokasikan data ke *cluster* di mana pusat *cluster* terdekat berada

5. Menghitung rata-rata data di setiap *cluster* dan menggunakannya sebagai pusat *cluster* baru untuk menghitung nilai SSE untuk semua *cluster*
6. Kembali ke langkah 2 jika posisi *centroid* baru dan *centroid* lama tidak sama.

Formula matematis untuk menghitung secara manual *silhouette* suatu objek diberikan di bawah ini. Identifikasi titik tertentu, seperti (p1, q1), dalam *cluster* 1, di mana objek dalam *cluster* mewakili nilai RFM. Hitung jarak rata-rata dari (p1, q1) ke semua objek dalam *cluster* yang sama (nilai jarak intra-*cluster* a_i). Tentukan jarak rata-rata dari (p1, q1) ke objek di *cluster* lain sebagaimana ditentukan dalam Persamaan (1) [1].

$$\sqrt{\sum_{i=1}^n (p1 - pi) + (q1 - qi)^2} \quad (1)$$

Hitung koefisien *silhouette* $s(i)$ untuk *cluster* 1 menggunakan persamaan yang diberikan dalam Persamaan (2) [1].

$$S(i) = \begin{cases} 1 - \frac{a_i}{b_i}, & a_i < b_i \\ 0, & a_i = b_i \\ \frac{a_i}{b_i} - 1, & a_i > b_i \end{cases} \quad (2)$$

Di mana,

- a_i mewakili jarak rata-rata minimum dari objek (p1, q1) ke semua objek lain dalam *cluster* yang sama.
- b_i mewakili jarak rata-rata minimum dari (p1, q1) ke semua *cluster* lain yang tidak berisi (p1, q1).

Hitung koefisien *silhouette* untuk *cluster* 2, 3, ..., n menggunakan prosedur yang sama seperti yang dijelaskan sebelumnya. *Cluster* yang menunjukkan koefisien *silhouette* tertinggi dianggap paling optimal menurut metode evaluasi ini. Hitung koefisien *silhouette* rata-rata di semua objek untuk mengevaluasi keseluruhan *cluster*.

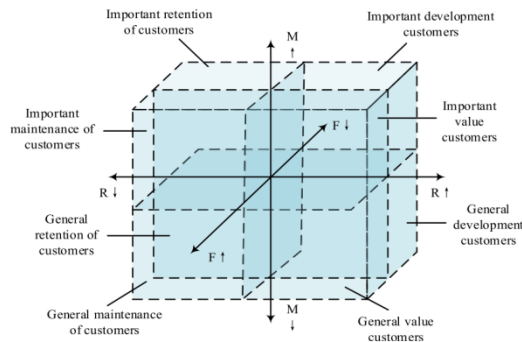
Langkah 5. Evaluasi Cluster:

Setelah mendapatkan nilai untuk setiap atribut RFM, tahap selanjutnya adalah melakukan standarisasi nilai-nilai tersebut. Tujuan standarisasi adalah untuk memudahkan analisis data yang dihasilkan. Standarisasi data dilakukan dengan menggunakan *software* R yang akan menghasilkan *z-score* sebagai nilai akhir yang distandarisasi. Menggunakan *Elbow Effect* atau *Elbow Method* untuk menentukan *cluster* optimal berdasarkan nilainya.

Setelah mengategorikan pelanggan ke dalam *cluster*, langkah selanjutnya adalah menganalisis setiap *cluster* untuk mengidentifikasi karakteristik masing-masing kelompok. Analisis ini akan memandu pengembangan strategi pemasaran yang disesuaikan untuk setiap *cluster*. Pada tahap ini analisis dilakukan dengan menggunakan data *Final Cluster Centers* pada *software* R.

B. Segmentasi Pelanggan

Algoritma *K-Means* menggunakan metrik jarak *Euclidean* untuk membagi pelanggan menjadi beberapa segmen sesuai dengan nilai RFM mereka. Mengacu pada Gambar 2, mengategorikan segmentasi pelanggan mengidentifikasi mereka yang memiliki permintaan lebih tinggi dan mereka yang menunjukkan pertumbuhan yang konsisten. Simbol (\uparrow) mewakili skor kelompok di atas rata-rata keseluruhan. Sedangkan simbol (\downarrow) melambangkan skor kelompok yang berada di bawah rata-rata keseluruhan.



Gambar 2. Diagram klasifikasi pelanggan [8]

Gambar 2 menggambarkan model segmentasi pelanggan berdasarkan 8 *cluster* utama yang dapat diidentifikasi [8]:

1. *Important Retention of Customers*: Pelanggan dengan nilai R rendah, F rendah, dan M tinggi. Pelanggan tidak membeli baru-baru ini (*recency* rendah), tidak membeli secara sering (*frequency* rendah), tetapi menghabiskan jumlah yang signifikan saat mereka membeli (*monetary value* tinggi). Pelanggan ini mungkin pembeli sesekali yang melakukan pembelian besar ketika mereka melakukannya, menunjukkan bahwa mereka menemukan nilai dalam produk tetapi mungkin bukan pembeli reguler.
2. *Important Development of Customers*: Pelanggan dengan nilai R tinggi, F rendah, dan M tinggi. Pelanggan ini telah membeli baru-baru ini (*recency* tinggi), tidak membeli secara sering (*frequency* rendah), tetapi menghabiskan jumlah yang signifikan saat mereka membeli. Pelanggan ini mungkin baru diperoleh atau pelanggan yang kembali yang mulai meningkatkan pengeluarannya, menunjukkan potensi pertumbuhan jika frekuensi pembeliannya dapat ditingkatkan.
3. *Important Value Customers*: Pelanggan dengan nilai R tinggi, F tinggi, dan M tinggi. Pelanggan ini telah membeli baru-baru ini (*recency* tinggi), membeli secara sering (*frequency* tinggi), dan menghabiskan jumlah yang signifikan. Ini adalah pelanggan bernilai tinggi yang loyal (frekuensi pembelian tinggi) dan pengeluaran tinggi. Mereka kemungkinan merupakan

pelanggan inti yang memberikan kontribusi signifikan terhadap pendapatan.

4. *Important Maintenance of Customers*: Pelanggan dengan nilai R rendah, F tinggi, dan M tinggi. Pelanggan ini tidak membeli baru-baru ini (*Recency* rendah), membeli secara sering (*frequency* tinggi), dan menghabiskan jumlah yang signifikan. Pelanggan ini aktif di masa lalu tetapi mungkin memerlukan upaya untuk mengaktifkan kembali mereka karena ketidakaktifan baru-baru ini. Mereka memiliki potensi untuk menjadi pelanggan bernilai tinggi lagi jika upaya retensi berhasil.
5. *General Maintenance of Customers*: Pelanggan dengan nilai R rendah, F rendah, dan M rendah. Pelanggan ini tidak membeli baru-baru ini (*Recency* rendah), tidak membeli secara sering (*frequency* rendah), dan menghabiskan jumlah yang rendah. Pelanggan ini memiliki keterlibatan rendah dan memberikan kontribusi minimal terhadap pendapatan. Mereka mungkin pembeli sesekali yang melakukan pembelian kecil secara jarang.
6. *General Retention of Customers*: Pelanggan dengan nilai R rendah, F tinggi, dan M rendah. Pelanggan ini tidak membeli baru-baru ini (*recency* rendah), membeli secara sering (*frequency* tinggi), tetapi menghabiskan jumlah yang rendah. Pelanggan ini sebelumnya aktif dan membeli secara sering, tetapi jumlah pengeluarannya relatif rendah. Upaya untuk meningkatkan nilai pesanan rata-rata mereka atau mendorong pengeluaran yang lebih tinggi mungkin bermanfaat.
7. *General Development of Customers*: Pelanggan dengan nilai R tinggi, F rendah, dan M rendah. Pelanggan ini telah membeli baru-baru ini (*recency* tinggi), tidak membeli secara sering (*frequency* rendah), dan menghabiskan jumlah yang rendah. Pelanggan ini adalah pembeli baru tetapi belum meningkatkan pengeluaran atau frekuensi mereka. Ada potensi bagi mereka untuk menjadi lebih bernilai jika perilaku pembeliannya berkembang secara positif.
8. *General Value Customers*: Pelanggan dengan nilai R tinggi, F tinggi, dan M rendah. Pelanggan ini telah membeli baru-baru ini (*recency* tinggi), membeli secara sering (*frequency* tinggi), tetapi menghabiskan jumlah yang rendah. Pelanggan ini adalah pembeli yang sering tetapi pembelian individu mereka lebih kecil. Mereka mungkin sensitif terhadap harga atau lebih suka barang dengan biaya rendah meskipun loyal.

III. HASIL PENELITIAN

Perusahaan studi kasus adalah bisnis peternakan terintegrasi dengan sejarah perdagangan lebih dari 10 tahun. Awalnya perusahaan

mengandalkan bisnis peternakan yang berpusat di Yogyakarta dan mulai melakukan ekspansi dan investasi pada bisnis retail makanan beku sejak 5 tahun lalu.

A. Analisis Eksplorasi dan Pemrosesan Awal Data

Untuk menghilangkan variasi RFM periodik, pertama-tama kami melakukan standarisasi data. Data variabel ditampilkan pada Tabel I.

TABEL I
VARIABEL DALAM DATASET TRANSAKSI PELANGGAN

Nama Variabel	Tipe Data	Deskripsi
TanggalPesanan	Date	Tanggal pembelian
IDPelanggan	Numeric	Kode pelanggan
Kota	Nominal	Nama kota pelanggan
Penjualan	Numeric	Total pembelian

Hasil dari pembersihan dan pra-pemrosesan data dalam analisis *clusterisasi* berbasis model RFM menunjukkan pentingnya memastikan kualitas data untuk mencapai segmentasi pelanggan yang akurat. Proses ini mencakup identifikasi pelanggan unik, standarisasi data, dan penanganan ketidakkonsistenan serta nilai yang hilang. Langkah-langkah ini memberikan dasar yang kuat untuk menghitung metrik RFM (*Recency*, *Frequency*, *Monetary*) yang kemudian digunakan untuk mengelompokkan pelanggan. Dengan memastikan setiap pelanggan diwakili oleh skor RFM yang distandarisi, analisis ini memungkinkan perusahaan mendapatkan segmentasi yang informatif dan dapat diimplementasikan. Hasil dari proses ini berkontribusi langsung pada keakuratan segmentasi, yang mendukung strategi pemasaran yang lebih tepat sasaran dan meningkatkan efektivitas dalam memaksimalkan penjualan serta retensi pelanggan.

Langkah-langkah dan tugas penting dalam persiapan data dijabarkan sebagai berikut: Setelah persiapan ini, *dataset* akhir terdiri dari 62.677 transaksi yang dibeli antara 01-04-2021 dan 23-08-2023 (Tabel II).

TABEL II
DATA AWAL RITEL MAKANAN BEKU

No.	ID Pelanggan	Tanggal Pesanan	Kota	Penjualan
1	CUST-15397	1-Apr-21	Jakarta	18,750,000
2	CUST-2111	1-Apr-21	Bogor	25,200,000
3	CUST-7441	1-Apr-21	Bogor	31,500,000

No.	ID Pelanggan	Tanggal Pesanan	Kota	Penjualan
4	CUST-5241	1-Apr-21	Jakarta	44,100,000
5	CUST-14208	1-Apr-21	Tangerang	99,908,000
...
62673	CUST-4245	23-Aug-23	Bogor	44,640,000
62674	CUST-14208	23-Aug-23	Tangerang	8,350,430
62675	CUST-4245	23-Aug-23	Bogor	42,160,000
62676	CUST-14208	23-Aug-23	Tangerang	42,146,575
62677	CUST-4245	23-Aug-23	Bogor	167,400,000

B. RFM Model

Setelah standarisasi nilai RFM, langkah berikutnya adalah menerapkan algoritma *K-Means* untuk melakukan *clusterisasi* pada data yang telah diolah. Dengan 1.150 data pelanggan (*clusterisasi* dari 62.677 data berdasarkan ID Pelanggan) yang telah diproses, algoritma *K-Means* mengelompokkan pelanggan ke dalam beberapa *cluster* berdasarkan nilai *Zrecency*, *Zfrequency*, dan *Zmonetary* (Tabel III). Hasil *clusterisasi* ini memungkinkan perusahaan untuk mengidentifikasi segmen-segmen pelanggan yang berbeda dengan pola pembelian dan pengeluaran yang bervariasi. Visualisasi grafis dari *cluster* membantu dalam memahami pola yang lebih mendalam, memberikan wawasan berharga mengenai pelanggan yang paling berharga dan yang memerlukan upaya retensi. Segmentasi ini memfasilitasi pengembangan strategi pemasaran yang lebih tepat sasaran, seperti penawaran khusus bagi pelanggan dengan frekuensi pembelian tinggi atau program loyalitas untuk segmen dengan *monetary value* yang signifikan, yang pada akhirnya diharapkan dapat meningkatkan retensi dan penjualan secara keseluruhan.

TABEL III
SKOR RFM UNTUK RITEL MAKANAN BEKU

No.	ID Pelanggan	#Transaksi	Zrecency	Zfrequency	Zmonetary
1	CUST-1001	35	0.61209657	-0.0387526	-0.084574
2	CUST-10392	1	-1.1079556	0.1063153	-0.0774931
3	CUST-10393	1	-1.5582697	0.1063153	0.0883478
4	CUST-10394	1	1.3540075	-0.1063153	-0.0886156
5	CUST-10395	4	1.41675618	-0.1003538	-0.0879203

No.	ID Pelanggan	#Transaksi	Zrecency	Zfrequency	Zmonetary
...
1146	CUST-9333	3	0.39063063	0.102341	0.0884855
1147	CUST-9334	12	1.569343	0.0844568	0.0843493
1148	CUST-9335	315	1.4549189	0.51764559	0.01773335
1149	CUST-9336	1	0.62686097	0.1063153	0.0884281
1150	CUST-9337	9	0.0153903	0.0904182	0.0861982

C. Pengelompokan K-Mean dan Skor Silhouette

Dari sintaks R, hitung jumlah total kotak dalam cluster untuk rentang nilai k menggunakan metode Elbow (Tabel IV).

TABEL IV
HASIL METODE ELBOW

k	Within_sum_of_square
2	1522.92527
3	681.076474
4	464.394265
5	228.022017
6	180.470034
7	155.174535
8	95.7863498
9	88.2353280
10	73.7487225
11	66.7499181
12	62.2876139
13	56.5876021
14	54.2657634
15	49.4553727

Metode Elbow menunjukkan bahwa jumlah cluster yang optimal adalah ketika jumlah kuadrat dalam cluster (WCSS) mulai berkurang lebih lambat. Berdasarkan Tabel IV terjadi penurunan yang signifikan dari k = 2 menjadi k=4. Setelah k=4, penurunan menjadi lebih bertahap. Oleh karena itu, k=4 tampaknya merupakan kandidat yang baik untuk jumlah cluster yang optimal menggunakan Metode Elbow.

Langkah selanjutnya dilakukan Metode Silhouette. Metode Silhouette menunjukkan bahwa jumlah cluster yang optimal adalah yang memaksimalkan lebar silhouette rata-rata (Tabel V).

Dari nilai-nilai tersebut, k=2 memiliki skor silhouette tertinggi (0,97049023), namun skor silhouette untuk k=2 terkadang bisa menyesatkan.

Di luar k=2, k=5 memiliki skor silhouette tertinggi berikutnya (0,59173859).

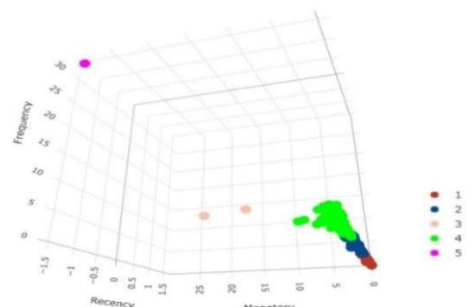
TABEL V
LEBAR SILHOUETTE RATA-RATA

k	Within_sum_of_square
2	0.97049023
3	0.55544118
4	0.58160064
5	0.59173859
6	0.54943122
7	0.53742465
8	0.57385098
9	0.54534302
10	0.53948092
11	0.53070266
12	0.53623871
13	0.50654160
14	0.49631939
15	0.50655256

Mengingat kedua metode, Metode Elbow menyarankan k=4. Metode silhouette menyarankan k=5 sebagai pilihan terbaik berikutnya setelah k=2. Dengan demikian, jumlah cluster yang optimal dapat berupa k=4 atau k=5. Namun, mengingat Metode Silhouette mempertimbangkan seberapa baik keterpisahan cluster, k=5 sedikit lebih optimal.

D. Segmentasi Pelanggan

Visualisasi cluster menggunakan software R dapat dilihat pada Gambar 3. Visualisasi cluster digunakan untuk memahami pola dan struktur yang mendasari data, menginterpretasikan hasil algoritma clustering dengan menunjukkan bagaimana titik-titik data dikelompokkan, menilai kualitas dan efektivitas algoritma clustering, membandingkan dan membedakan cluster yang berbeda untuk menganalisis karakteristiknya, dan secara efektif mengkomunikasikan temuan dan wawasan yang diperoleh dari analisis cluster kepada pemangku kepentingan.



Gambar 3. Diagram klasifikasi pelanggan retail makanan beku

Berdasarkan Gambar 3, cluster yang diperoleh kemudian dikelompokkan berdasarkan cluster pada Gambar 2 sehingga diperoleh segmentasi pelanggan sebagai berikut (Tabel VI).

TABEL VI
DISTRIBUSI SEGMENTASI PELANGGAN RETAIL MAKANAN
BEKU

Kluster	Jumlah Kasus	Segmentasi
1	402	<i>General Development of Customers</i>
2	438	<i>General Maintenance of Customers</i>
3	2	<i>Important Maintenance of Customers</i>
4	307	<i>Important Maintenance of Customers</i>
5	1	<i>Important Maintenance of Customers</i>

IV. PEMBAHASAN

Customer Relationship Management (CRM) sangat penting bagi bisnis yang ingin menumbuhkan loyalitas dan profitabilitas pelanggan jangka panjang, khususnya melalui segmentasi pelanggan menggunakan analisis RFM—Kekinian (R), Frekuensi (F), dan Nilai Moneter (M). Berdasarkan Tabel VI, segmen yang teridentifikasi, *General Development of Customers* (402 kasus) terdiri dari pembeli terkini dengan frekuensi pembelian dan nilai moneter rendah. Dalam jangka pendek, bisnis dapat menerapkan kampanye email yang dipersonalisasi yang menawarkan diskon pada pembelian berikutnya untuk mendorong pembelian berulang. Dalam jangka panjang, memelihara segmen ini melibatkan pengembangan program loyalitas yang memberikan imbalan atas pembelian yang sering dilakukan dengan manfaat eksklusif, sehingga meningkatkan nilai umurnya. Sebaliknya, segmen *General Maintenance of Customers* (438 kasus) terdiri dari pelanggan yang sering membeli namun dengan nilai uang lebih rendah. Strategi jangka pendek mencakup penerapan promosi yang ditargetkan berdasarkan riwayat pembelian untuk meningkatkan nilai pesanan rata-rata [18], [19]. Dalam jangka panjang, mempertahankan keterlibatan rutin melalui buletin yang dipersonalisasi atau rekomendasi produk dapat memperdalam hubungan mereka dengan merek dan meningkatkan retensi [19], [20]. Terakhir, segmen *Important Maintenance of Customers* (310 kasus) mencakup pelanggan bernilai tinggi yang menunjukkan tanda-tanda penurunan aktivitas. Tindakan segera mencakup penjangkauan yang dipersonalisasi untuk mengatasi permasalahan atau memberikan dukungan proaktif. Strategi jangka panjang melibatkan pembuatan program VIP dengan manajemen akun yang dipersonalisasi dan pratinjau eksklusif produk baru untuk memperkuat loyalitas mereka [20]. Dengan menyelaraskan strategi CRM dengan karakteristik dan kebutuhan

berbeda dari setiap segmen RFM, bisnis dapat secara efektif membina hubungan pelanggan, meningkatkan kepuasan, dan mempertahankan pertumbuhan dari waktu ke waktu.

V. KESIMPULAN

Kesimpulan Berdasarkan diskusi rinci dan temuan yang disajikan, penelitian ini menggarisbawahi pentingnya segmentasi pelanggan berbasis RFM dalam meningkatkan strategi manajemen hubungan pelanggan (CRM) di sektor ritel. Melalui pemanfaatan algoritma *clustering K-Means*, penelitian ini secara efektif mengelompokkan pelanggan menjadi tiga kelompok berbeda: *General Development of Customers*, *General Maintenance of Customers*, dan *Important Maintenance of Customers*. Setiap segmen dicirikan oleh perilaku unik terkait keterkinian, frekuensi, dan nilai moneter, yang menawarkan wawasan tentang pendekatan CRM yang disesuaikan. Strategi jangka pendek seperti promosi yang dipersonalisasi dan kampanye yang ditargetkan direkomendasikan untuk memberikan dampak langsung, sementara inisiatif jangka panjang termasuk program loyalitas dan layanan VIP bertujuan untuk menumbuhkan loyalitas pelanggan yang bertahan lama dan meningkatkan nilai seumur hidup. Temuan ini menyoroti peran penting segmentasi berbasis data dalam mengoptimalkan upaya pemasaran dan meningkatkan retensi pelanggan, yang pada akhirnya berkontribusi terhadap pertumbuhan bisnis yang berkelanjutan dan profitabilitas dalam lanskap ritel yang kompetitif.

Ke depan, penelitian di masa depan dapat mengeksplorasi dimensi tambahan segmentasi pelanggan di luar RFM, yang berpotensi mengintegrasikan analisis *psikografis* atau perilaku untuk mendapatkan wawasan yang lebih beragam. Selain itu, menyelidiki efektivitas strategi CRM *omnichannel* dalam beradaptasi dengan perubahan perilaku dan preferensi konsumen akan bermanfaat. Mengatasi permasalahan ini dapat lebih menyempurnakan praktik CRM dan memberikan pemahaman yang lebih mendalam mengenai strategi keterlibatan pelanggan di berbagai segmen pasar.

REFERENSI

- [1] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using *K-Means* algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [2] B. Sharp, J. Dawes, and K. Victory, "The market-based assets theory of brand competition," *J. Retail. Consum. Serv.*, vol. 76, no. April 2023, p. 103566, 2024, doi: 10.1016/j.jretconser.2023.103566.
- [3] S. Gupta, T. Justy, S. Kamboj, A. Kumar, and E.

- Kristoffersen, "Big data and firm marketing performance: Findings from knowledge-based view," *Technol. Forecast. Soc. Change*, vol. 171, no. November 2020, 2021, doi: 10.1016/j.techfore.2021.120986.
- [4] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, "RFM model customer segmentation based on hierarchical approach using FCA[Formula presented]," *Expert Syst. Appl.*, vol. 237, no. PB, p. 121449, 2024, doi: 10.1016/j.eswa.2023.121449.
- [5] B. Ivens, K. Kasper-Brauer, A. Leischnig, and S. C. Thornton, "Implementing customer relationship management successfully: A configurational perspective," *Technol. Forecast. Soc. Change*, vol. 199, no. June 2023, 2024, doi: 10.1016/j.techfore.2023.123083.
- [6] F. Mohammed, R. B. Ahmad, S. B. Hassan, Y. Fazea, and A. I. Alzahrani, "An empirical evidence on the impact of social customer relationship management on the small and medium enterprises performance," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, p. 100248, 2024, doi: 10.1016/j.jjime.2024.100248.
- [7] D. C. Li, W. L. Dai, and W. T. Tseng, "A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7186–7191, 2011, doi: 10.1016/j.eswa.2010.12.041.
- [8] Z. Tang, Y. Jiao, and M. Yuan, "RFM user value tags and XGBoost algorithm for analyzing electricity customer demand data," *Syst. Soft Comput.*, vol. 6, no. April, p. 200098, 2024, doi: 10.1016/j.sasc.2024.200098.
- [9] M. Alves Gomes and T. Meisen, *A review on customer segmentation methods for personalized customer targeting in e-commerce use cases*, vol. 21, no. 3. Springer Berlin Heidelberg, 2023. doi: 10.1007/s10257-023-00640-4.
- [10] S. Singh and S. Srivastava, "Review of Clustering Techniques in Control System," *Procedia Comput. Sci.*, vol. 173, pp. 272–280, 2020, doi: 10.1016/j.procs.2020.06.032.
- [11] M. Hosseini, N. Abdolvand, and S. R. Harandi, "Two-dimensional analysis of customer behavior in traditional and electronic banking," *Digit. Bus.*, vol. 2, no. 2, p. 100030, 2022, doi: 10.1016/j.digbus.2022.100030.
- [12] O. Sokol and V. Holý, "The role of shopping mission in retail customer segmentation," *Int. J. Mark. Res.*, vol. 63, no. 4, pp. 454–470, 2021, doi: 10.1177/1470785320921011.
- [13] Dedi, M. I. Dzulhaq, K. W. Sari, S. Ramdhan, R. Tullah, and Sutarman, "Customer segmentation based on RFM value using K-Means algorithm," in *2019 fourth international conference on informatics and computing*, 2019, pp. 1–7.
- [14] M. Frassetto, M. Ieva, and C. Ziliani, "Online channel adoption in supermarket retailing," *J. Retail. Consum. Serv.*, vol. 59, no. November 2020, p. 102374, 2021, doi: 10.1016/j.jretconser.2020.102374.
- [15] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *J. Retail. Consum. Serv.*, vol. 61, no. June 2020, p. 102566, 2021, doi: 10.1016/j.jretconser.2021.102566.
- [16] W. Qadadeh and S. Abdallah, "Customers Segmentation in the Insurance Company (TIC) Dataset," *Procedia Comput. Sci.*, vol. 144, pp. 277–290, 2018, doi: 10.1016/j.procs.2018.10.529.
- [17] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1251–1257, 2021, doi: 10.1016/j.jksuci.2018.09.004.
- [18] F. Solari, N. Lysova, M. Bocelli, A. Volpi, and R. Montanari, "Perishable Product Inventory Management In The Case Of Discount Policies And Price-Sensitive Demand: Discrete Time Simulation And Sensitivity Analysis," *Procedia Comput. Sci.*, vol. 232, pp. 1233–1241, 2024, doi: 10.1016/j.procs.2024.01.121.
- [19] L. Hu, W. Zhang, S. Du, and X. Sun, "How to achieve targeted advertising with the e-commerce platform's membership system?," *Omega (United Kingdom)*, vol. 130, no. November 2023, 2025, doi: 10.1016/j.omega.2024.103156.
- [20] L. Xu and Z. Meng, "The role of membership fees in online retail market competition," *Res. Int. Bus. Financ.*, vol. 67, no. PA, p. 102089, 2024, doi: 10.1016/j.ribaf.2023.102089.