

Segmentasi Pelanggan dan Target Pemasaran (*Targeting*) Menggunakan Algoritma *K-Means* *Clustering* untuk Kampanye Iklan yang Efektif

Customer Segmentation and Target Market Determination Using K-Means Clustering Algorithm for Effective Advertising Campaigns

Roswanda Nuraini^{*1}, Zaenur Rozikin², Suprpto³

^{1,2,3} Program Studi Teknik Informatika
Universitas Pelita Bangsa, Indonesia

E-mail: ^{*1}roswanda2004@gmail.com

Article Info:	Received 14 Juni 2026	Revised 15 Juni 2026	Accepted 28 Juni 2026	Published: 29 Juni 2026
----------------------	--------------------------	-------------------------	--------------------------	----------------------------

Abstrak

Perkembangan teknologi informasi menyebabkan meningkatnya volume data transaksi pelanggan yang dimiliki perusahaan, namun data tersebut sering kali belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan pemasaran, khususnya dalam penyusunan rekomendasi strategi *targeting* yang tepat sasaran. Oleh karena itu, penelitian ini bertujuan melakukan segmentasi pelanggan berdasarkan perilaku transaksi menggunakan metode *Recency, Frequency, Monetary (RFM)* dan algoritma *K-Means Clustering*. Data yang digunakan terdiri atas 5.000 pelanggan dengan 32.295 transaksi. Setelah melalui tahap *preprocessing* yang meliputi pemeriksaan struktur data, pengecekan *missing value*, penghapusan duplikasi, *filtering dataset*, dan transformasi tipe data, diperoleh 3.864 pelanggan dengan 28.009 transaksi. Selanjutnya, pembentukan variabel *RFM* menghasilkan 3.522 pelanggan yang memenuhi kualifikasi agregasi untuk proses segmentasi. Data kemudian dinormalisasi menggunakan metode *Min-Max Scaling*. Proses *clustering* dilakukan menggunakan algoritma *K-Means* dengan penentuan jumlah cluster menggunakan *Elbow Method* dan evaluasi menggunakan *Silhouette Score*. Hasil penelitian menunjukkan bahwa empat cluster dipilih dengan nilai *Silhouette Score* sebesar 0,3787 karena menghasilkan segmentasi yang lebih representatif untuk kebutuhan interpretasi bisnis. Berdasarkan karakteristik nilai *centroid RFM*, empat segmen yang terbentuk terdiri atas *VIP Customer, Loyal Customer, Regular Customer, dan At-Risk Customer*. Hasil segmentasi menunjukkan bahwa setiap segmen memiliki karakteristik perilaku transaksi yang berbeda sehingga menghasilkan rekomendasi strategi *targeting* yang berbeda untuk masing-masing segmen pelanggan. Dengan demikian, hasil penelitian ini dapat menjadi dasar analitis dalam penyusunan rekomendasi strategi *targeting* yang lebih terarah dan berbasis data serta berpotensi mendukung pengambilan keputusan pemasaran dan upaya mempertahankan pelanggan bernilai tinggi.

Kata Kunci: *Customer Relationship Management (CRM); K-Means Clustering; Perilaku Pelanggan; Rekomendasi Strategi Targeting; RFM; Segmentasi Pelanggan.*

Abstract

Advances in information technology have led to an increase in the volume of customer transaction data held by companies; however, this data is often not utilized optimally to support marketing decision-making, particularly in formulating targeted strategic recommendations. Therefore, this study aims to segment customers based on their transaction behavior using the Recency, Frequency, Monetary (RFM) method and the K-Means Clustering algorithm. The dataset consists of 5,000 customers with 32,295 transactions. After undergoing a preprocessing stage that included examining the data structure, checking for missing values, removing duplicates, filtering the dataset, and transforming data types, the final dataset comprised 3,864 customers with 28,009 transactions. Subsequently, the creation of RFM variables resulted in 3,522 customers who met the aggregation criteria for the segmentation process. The data was then normalized using the Min-Max Scaling method. The clustering process was performed using the K-Means algorithm, with the number of clusters determined using the Elbow Method and evaluated using the Silhouette Score. The research results show that four clusters were selected with a Silhouette Score of 0.3787 because they produced a more representative segmentation for business interpretation needs. Based on the characteristics of the RFM centroid values, the four segments formed consist of VIP Customers, Loyal Customers, Regular Customers, and At-Risk Customers. The segmentation results show that each segment has distinct transactional behavior characteristics, leading to different targeting strategy recommendations for each customer segment. Thus, the results of this study can serve as an analytical foundation for developing more targeted, data-driven targeting strategy recommendations and have the potential to support marketing decision-making and efforts to retain high-value customers.

Keywords: Customer Behavior; Customer Relationship Management (CRM); Customer Segmentation; K-Means Clustering; RFM; Targeting Strategy Recommendation.

This is an open access article under the CC BY-SA license.



1. PENDAHULUAN

Perkembangan ekonomi digital dan penetrasi *e-commerce* telah meningkatkan volume data transaksi pelanggan secara signifikan. Kondisi ini menciptakan peluang sekaligus tantangan bagi perusahaan dalam mengolah data menjadi informasi yang bernilai untuk mendukung pengambilan keputusan pemasaran [1]. Pertumbuhan pengguna *e-commerce* di tingkat global maupun di Indonesia menunjukkan bahwa aktivitas transaksi digital terus meningkat seiring meningkatnya penetrasi internet dan adopsi platform digital [2], [3], [4]. Di tengah kondisi tersebut, perusahaan mulai beralih dari pemasaran massal menuju pemasaran berbasis data (*data-driven marketing*) untuk meningkatkan efektivitas pengambilan keputusan pemasaran [5].

Pemasaran berbasis data memanfaatkan analisis perilaku pelanggan untuk mendukung proses segmentasi pelanggan sehingga perusahaan dapat menyusun rekomendasi strategi *targeting* yang lebih tepat sasaran [6]. Personalisasi pemasaran berbasis data terbukti meningkatkan relevansi komunikasi pemasaran, keterlibatan pelanggan, serta efektivitas pemasaran [7], [8], [9]. Namun, pada praktiknya masih banyak perusahaan yang belum memanfaatkan data historis transaksi secara optimal sehingga keputusan pemasaran masih dilakukan secara umum dan berpotensi menyebabkan pemborosan biaya promosi (*marketing waste*) [5], [10].

Salah satu pendekatan yang banyak digunakan untuk segmentasi pelanggan adalah model *Recency, Frequency, Monetary* (RFM), yang merepresentasikan perilaku transaksi pelanggan berdasarkan waktu transaksi terakhir, frekuensi transaksi, dan nilai pembelian. Model RFM banyak diintegrasikan dengan algoritma *K-Means Clustering* karena mampu menghasilkan segmentasi pelanggan yang mudah diinterpretasikan dan mendukung pengambilan keputusan

pemasaran [11], [12]. Penentuan jumlah *cluster* umumnya dilakukan menggunakan *Elbow Method* dan dievaluasi menggunakan *Silhouette Score* untuk memperoleh hasil segmentasi yang representatif [13], [14].

Berbagai penelitian terdahulu telah menerapkan integrasi RFM dan *K-Means Clustering* pada berbagai sektor dan menunjukkan bahwa pendekatan tersebut mampu menghasilkan segmentasi pelanggan yang efektif [12], [14], [15], [16], [17]. Meskipun demikian, sebagian besar penelitian masih berfokus pada kualitas hasil *clustering* dan karakteristik setiap segmen pelanggan, sedangkan pemanfaatan hasil segmentasi sebagai dasar penyusunan rekomendasi strategi *targeting* untuk mendukung pengambilan keputusan pemasaran masih relatif terbatas. Selain itu, justifikasi pemilihan jumlah *cluster* umumnya hanya didasarkan pada indikator statistik tanpa mempertimbangkan kebutuhan interpretasi bisnis.

Berdasarkan *research gap* tersebut, penelitian ini bertujuan menerapkan metode RFM dan algoritma *K-Means Clustering* untuk melakukan segmentasi pelanggan berdasarkan perilaku transaksi. Penelitian ini memberikan tiga kontribusi utama, yaitu menerapkan proses segmentasi melalui tahapan *preprocessing* dan pembentukan variabel RFM, memberikan justifikasi pemilihan jumlah *cluster* berdasarkan pertimbangan statistik dan kebutuhan interpretasi bisnis, serta menyusun rekomendasi strategi *targeting* berdasarkan karakteristik setiap segmen pelanggan. Dengan demikian, hasil penelitian diharapkan dapat menjadi dasar analitis dalam penyusunan rekomendasi strategi *targeting* untuk mendukung pengambilan keputusan pemasaran berbasis data secara lebih terarah dan efisien.

2. METODE

Penelitian ini menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri atas enam tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [23]. Tahapan penelitian meliputi *preprocessing* data, pembentukan variabel *Recency*, *Frequency*, *Monetary* (RFM), normalisasi menggunakan *Min-Max Scaling*, proses *clustering* menggunakan algoritma *K-Means Clustering*, serta evaluasi hasil menggunakan *Elbow Method* dan *Silhouette Score*. Alur penelitian ditunjukkan pada Gambar 1.

Dataset yang digunakan merupakan dataset sekunder *e-commerce* yang diperoleh dari platform Kaggle dan terdiri atas dataset pelanggan (*customers*) serta dataset transaksi (*transactions*) [28]. Dataset awal mencakup 5.000 pelanggan dengan 12 atribut dan 32.295 transaksi dengan 10 atribut. Tahap *preprocessing* meliputi pemeriksaan struktur data, identifikasi *missing value*, penghapusan data tidak lengkap, penghapusan data duplikat, penyaringan transaksi yang tidak valid, serta integrasi dataset pelanggan dan transaksi. Setelah proses tersebut diperoleh 3.864 pelanggan dan 28.009 transaksi yang digunakan sebagai dasar pembentukan variabel RFM. Selanjutnya dilakukan agregasi data sehingga hanya pelanggan yang memiliki riwayat transaksi valid yang digunakan pada proses segmentasi.

Tahap *Business Understanding* dilakukan untuk mengidentifikasi permasalahan bisnis serta menetapkan tujuan penelitian, yaitu melakukan segmentasi pelanggan berdasarkan perilaku transaksi sebagai dasar penyusunan rekomendasi strategi *targeting*. Tahap *Data Understanding* digunakan untuk memahami karakteristik dataset, sedangkan *Data Preparation* meliputi proses pembersihan data, integrasi dataset, pembentukan variabel RFM, dan normalisasi menggunakan metode *Min-Max Scaling*.

Variabel RFM terdiri atas *Recency*, *Frequency*, dan *Monetary* yang dihitung menggunakan Persamaan (1), (2), (3). Ketiga variabel tersebut digunakan sebagai fitur utama pada proses segmentasi pelanggan menggunakan algoritma *K-Means Clustering*.

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

dengan R_i adalah nilai *Recency* pelanggan ke- i , T_{ref} adalah tanggal referensi analisis, dan T_{last} adalah tanggal transaksi terakhir pelanggan.

$$F_i = \sum_{j=1}^n \text{Transaksi}_{ij} \quad (2)$$

dengan F_i adalah jumlah transaksi yang dilakukan pelanggan ke- i selama periode pengamatan.

$$M_i = \sum_{j=1}^n (\text{Quantity}_{ij} \times \text{Price}_{ij}) \quad (3)$$

dengan M_i adalah total nilai pembelian pelanggan ke- i , Quantity_{ij} adalah jumlah produk yang dibeli, dan Price_{ij} adalah harga produk. Ketiga variabel tersebut selanjutnya digunakan sebagai fitur utama dalam proses segmentasi pelanggan menggunakan algoritma *K-Means Clustering*.

Proses *clustering* dilakukan menggunakan algoritma *K-Means Clustering* dengan parameter *random_state* = 42, inialisasi centroid menggunakan *k-means++*, dan *max_iter* = 300. Jumlah *cluster* ditentukan menggunakan *Elbow Method* dan dievaluasi menggunakan *Silhouette Score*. Persamaan *Euclidean Distance* dan *Silhouette Score* masing-masing ditunjukkan pada Persamaan (4) dan Persamaan (5).

Persamaan *Euclidean Distance* (K-Means)

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

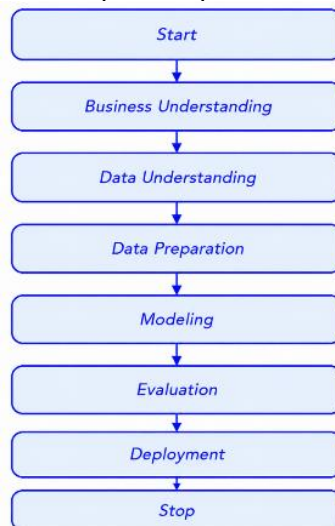
dengan $d(x, y)$ adalah jarak *Euclidean* antara dua titik data, x_i adalah nilai atribut ke- i pada data x , y_i adalah nilai atribut ke- i pada data y , dan n adalah jumlah atribut yang digunakan.

Persamaan *Silhouette Score*

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

dengan $s(i)$ adalah nilai *Silhouette* untuk data ke- i , $a(i)$ adalah rata-rata jarak data ke- i terhadap seluruh anggota dalam *cluster* yang sama, dan $b(i)$ adalah rata-rata jarak minimum data ke- i terhadap anggota pada *cluster* terdekat lainnya.

Tahap *Evaluation* dilakukan dengan menganalisis karakteristik setiap *cluster* berdasarkan nilai centroid RFM. Selanjutnya setiap *cluster* diberi label sesuai karakteristik perilaku pelanggan. Tahap *Deployment* tidak dilakukan dalam bentuk implementasi sistem, tetapi berupa penyusunan rekomendasi strategi *targeting* berdasarkan karakteristik masing-masing segmen pelanggan sebagai dasar pendukung pengambilan keputusan pemasaran.



Gambar 1. Tahapan Penelitian Menggunakan Metodologi CRISP-DM.

3. HASIL

3.1 Hasil *Data Preparation* dan Pembentukan Variabel RFM

Tahap *data preparation* dilakukan untuk menghasilkan dataset yang siap digunakan dalam proses segmentasi pelanggan. Dataset pelanggan dan transaksi diintegrasikan

menggunakan atribut *customer_id*, kemudian dilakukan pemeriksaan *missing value*, penghapusan data duplikat, penyaringan transaksi yang tidak valid, serta transformasi tipe data pada atribut tanggal transaksi.

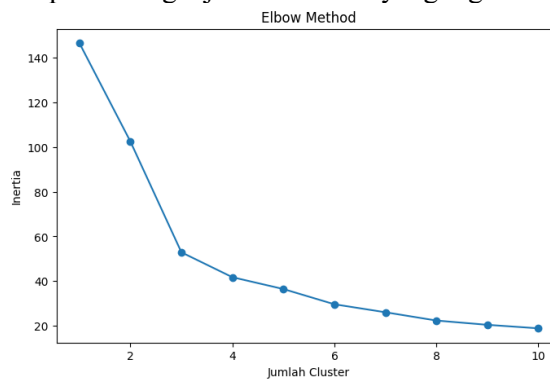
Selanjutnya dibentuk variabel *Recency*, *Frequency*, dan *Monetary* (RFM). Variabel *Recency* dihitung berdasarkan selisih antara tanggal referensi dan tanggal transaksi terakhir pelanggan, *Frequency* berdasarkan jumlah transaksi selama periode pengamatan, sedangkan *Monetary* berdasarkan total nilai transaksi setiap pelanggan.

Hasil pembentukan variabel RFM menghasilkan 3.522 pelanggan yang siap digunakan pada proses *clustering*. Jumlah tersebut lebih sedikit dibandingkan 3.864 pelanggan hasil *preprocessing* awal karena pada proses *inner join* dan agregasi hanya pelanggan yang memiliki riwayat transaksi valid yang dapat membentuk variabel RFM. Pelanggan yang tidak memiliki transaksi aktif atau seluruh transaksinya tereliminasi pada tahap pembersihan data tidak disertakan dalam proses analisis. Setelah itu, data RFM dinormalisasi menggunakan *Min-Max Scaling* sehingga seluruh atribut berada pada rentang 0–1 sebelum dilakukan proses *clustering*.

3.2 Hasil Penentuan Jumlah Cluster Menggunakan *Elbow Method*

Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method* dengan menganalisis perubahan nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa nilai *k*. Pengujian dilakukan pada rentang jumlah *cluster* yang berbeda untuk memperoleh jumlah *cluster* yang paling optimal.

Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 2, terlihat bahwa penurunan nilai WCSS mulai melandai pada titik $k = 4$. Titik tersebut menunjukkan terbentuknya elbow point sehingga empat *cluster* dipilih sebagai jumlah *cluster* yang digunakan dalam penelitian ini.



Gambar 2. Hasil *Elbow Method*

Pada penelitian ini, pengujian dilakukan dengan variasi jumlah *cluster* mulai dari $k = 1$ hingga $k=10$. Hasil perhitungan nilai WCSS ditunjukkan pada Tabel 1.

Tabel 1. Hasil Perhitungan WCSS

Jumlah Cluster (k)	WCSS
1	146,461
2	102,434
3	52,883
4	41,752
5	36,465
6	29,651
7	26,079
8	22,374
9	20,460
10	18,895

Berdasarkan Tabel 1, nilai *Within Cluster Sum of Squares* (WCSS) mengalami penurunan yang signifikan dari $k = 1$ sebesar 146,461 menjadi 52,883 pada $k = 3$. Penurunan ini

menunjukkan bahwa penambahan jumlah cluster masih memberikan peningkatan kualitas clustering yang cukup besar.

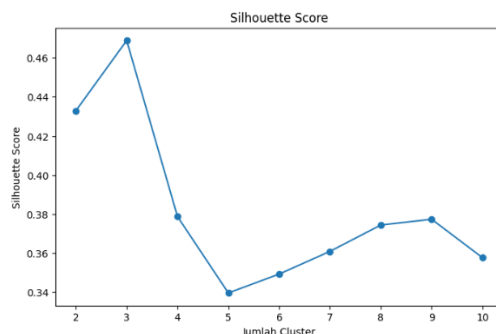
Namun, mulai $k = 4$, penurunan WCSS mulai melambat menjadi 41,752 dan terus menurun secara gradual hingga $k = 10$. Pola ini menunjukkan adanya *diminishing returns*, yaitu penambahan jumlah *cluster* tidak lagi memberikan peningkatan yang signifikan terhadap homogenitas *cluster*. Oleh karena itu, titik elbow diidentifikasi pada $k = 4$ sebagai jumlah *cluster* optimal.

3.3 Hasil Evaluasi Menggunakan *Silhouette Score*

Setelah jumlah *cluster* ditentukan, dilakukan evaluasi menggunakan *Silhouette Score* untuk mengukur kualitas *cluster* yang terbentuk. Nilai *Silhouette Score* digunakan untuk mengukur tingkat kohesi dalam *cluster* dan separasi antar *cluster*. Tabel 2 menunjukkan hasil evaluasi *Silhouette Score* pada beberapa nilai k .

Tabel 2. Hasil Evaluasi *Silhouette Score*

Jumlah Cluster (k)	<i>Silhouette Score</i>
2	0,4329
3	0,4689
4	0,3787
5	0,3396
6	0,3492
7	0,3609
8	0,3744
9	0,3773
10	0,3576



Gambar 3. Grafik *Silhouette Score*

Nilai *Silhouette Score* tertinggi diperoleh pada $k = 3$ sebesar 0,4689, sedangkan pada $k = 4$ sebesar 0,3787. Hasil evaluasi ini menjadi salah satu dasar dalam penentuan jumlah *cluster*.

3.4 Hasil Clustering Pelanggan

Setelah jumlah *cluster* optimal ditentukan sebanyak empat *cluster* berdasarkan hasil *Elbow Method* dan *Silhouette Score*, proses *clustering* dilakukan menggunakan algoritma K-Means terhadap data RFM yang telah dinormalisasi. Hasil *clustering* menghasilkan empat kelompok pelanggan dengan jumlah anggota yang berbeda pada setiap *cluster*. Distribusi pelanggan pada masing-masing *cluster* ditunjukkan pada Tabel 3.

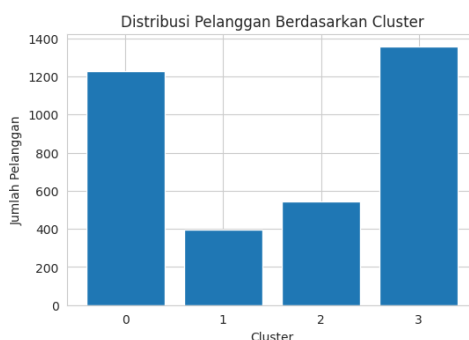
Tabel 3. Distribusi Pelanggan Berdasarkan Cluster

Cluster	Jumlah Pelanggan	Persentase (%)
0	1.229	34,90%
1	394	11,19%
2	544	15,45%
3	1.355	38,46%
Total	3.522	100%

Berdasarkan Tabel 3, *Cluster 3* memiliki jumlah pelanggan terbesar yaitu 1.355 pelanggan (38,46%), sedangkan *Cluster 1* memiliki jumlah pelanggan terkecil yaitu 394 pelanggan

(11,19%). *Cluster 0* dan *Cluster 2* masing-masing terdiri atas 1.229 pelanggan (34,90%) dan 544 pelanggan (15,45%). Perbedaan jumlah anggota pada setiap *cluster* menunjukkan bahwa karakteristik pelanggan tidak tersebar secara merata sehingga diperlukan rekomendasi strategi *targeting* yang berbeda untuk setiap segmen pelanggan.

Untuk memberikan gambaran visual mengenai distribusi pelanggan pada setiap *cluster*, hasil *clustering* ditampilkan dalam bentuk diagram batang sebagaimana ditunjukkan pada Gambar 4.



Gambar 4. Distribusi Pelanggan Berdasarkan *Cluster*

3.5 Interpretasi Centroid dan Penentuan Nama *Cluster*

Karakteristik masing-masing *cluster* diperoleh berdasarkan nilai rata-rata *Recency*, *Frequency*, dan *Monetary* sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Karakteristik *Cluster* Berdasarkan Nilai RFM

<i>Cluster</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
0	95,33	7,33	6.555,93
1	516,77	2,77	2.375,49
2	101,36	13,45	12.203,57
3	177,84	2,69	2.337,48

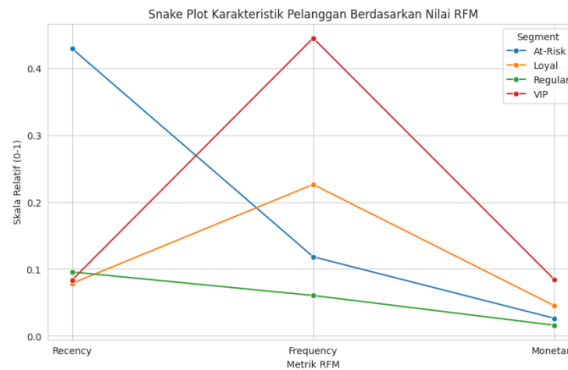
Berdasarkan karakteristik yang diperoleh dari hasil analisis *centroid*, setiap *cluster* kemudian diberikan nama segmen pelanggan yang merepresentasikan perilaku transaksi masing-masing kelompok. Penamaan segmen dilakukan dengan mempertimbangkan kombinasi nilai *Recency*, *Frequency*, dan *Monetary* pada setiap *cluster*. Hasil penentuan nama segmen pelanggan ditunjukkan pada Tabel 5.

Tabel 5. Penentuan Nama Segmen Pelanggan

<i>Cluster</i>	Karakteristik Utama	Nama Segmen
0	<i>Frequency</i> dan <i>Monetary</i> tinggi	<i>Loyal Customer</i>
1	<i>Recency</i> sangat tinggi	<i>At-Risk Customer</i>
2	<i>Frequency</i> dan <i>Monetary</i> tertinggi	<i>VIP Customer</i>
3	<i>Frequency</i> dan <i>Monetary</i> rendah	<i>Regular Customer</i>

Berdasarkan Tabel 5, Berdasarkan Tabel 5, empat *cluster* yang terbentuk diinterpretasikan sebagai *VIP Customer*, *Loyal Customer*, *Regular Customer*, dan *At-Risk Customer* sesuai karakteristik nilai RFM masing-masing.

Untuk mempermudah proses visualisasi karakteristik setiap segmen pelanggan, nilai RFM yang telah dinormalisasi kemudian divisualisasikan menggunakan *Snake Plot* sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Snake Plot Karakteristik Pelanggan Berdasarkan Nilai RFM

Gambar 5 menunjukkan perbedaan pola nilai RFM antarcluster yang menjadi dasar interpretasi karakteristik segmen pelanggan.

4. PEMBAHASAN

4.1 Analisis Karakteristik Segmen Pelanggan

Hasil segmentasi menunjukkan bahwa kombinasi metode RFM dan algoritma *K-Means Clustering* mampu membedakan pelanggan berdasarkan perilaku transaksi sehingga menghasilkan empat segmen dengan karakteristik yang berbeda. Perbedaan karakteristik tersebut menunjukkan adanya variasi tingkat loyalitas dan nilai ekonomi pelanggan, sehingga setiap segmen memerlukan pendekatan pengelolaan pelanggan yang berbeda. Pelanggan dengan nilai transaksi dan frekuensi pembelian yang tinggi merupakan aset penting yang perlu dipertahankan melalui strategi retensi, sedangkan pelanggan dengan aktivitas transaksi yang menurun memerlukan upaya reaktivasi untuk mengurangi risiko kehilangan pelanggan. Temuan ini menunjukkan bahwa segmentasi berbasis RFM tidak hanya mampu mengidentifikasi pola perilaku pelanggan, tetapi juga menyediakan dasar analitis dalam penyusunan rekomendasi strategi *targeting* yang lebih tepat sasaran [11], [12].

4.2 Analisis Distribusi Pelanggan

Distribusi pelanggan menunjukkan bahwa segmen *Regular Customer* memiliki jumlah pelanggan terbanyak (1.355 pelanggan), diikuti *Loyal Customer* (1.229 pelanggan), *VIP Customer* (544 pelanggan), dan *At-Risk Customer* (394 pelanggan). Distribusi ini menunjukkan bahwa sebagian besar pelanggan masih berada pada kategori menengah yang berpotensi ditingkatkan menjadi pelanggan loyal maupun pelanggan bernilai tinggi, sementara jumlah pelanggan VIP relatif lebih sedikit tetapi memberikan kontribusi transaksi yang lebih besar.

Tabel 6. Perbandingan *Centroid* $k = 3$ dan $k = 4$

Konfigurasi	Cluster	Recency	Frequency	Monetary	Interpretasi
$k = 3$	C0	121.10	10.84	11,522.34	High Activity
$k = 3$	C1	160.94	4.58	3,113.68	Low Value
$k = 3$	C2	202.56	9.78	63,872.27	High Spender
$k = 4$	C0	104.41	4.02	3,057.51	Regular
$k = 4$	C1	100.85	11.30	10,353.83	Loyal
$k = 4$	C2	202.56	9.78	63,872.27	VIP
$k = 4$	C3	503.91	4.32	3,539.40	At-Risk

Berdasarkan Tabel 6, konfigurasi $k = 3$ masih menghasilkan segmentasi yang bersifat agregatif sehingga pelanggan bernilai tinggi (VIP) dan pelanggan loyal belum dapat dipisahkan secara jelas. Sebaliknya, konfigurasi $k = 4$ menghasilkan segmentasi yang lebih representatif terhadap variasi perilaku pelanggan, terutama pada dimensi *Recency* dan *Monetary*, sehingga lebih mudah diinterpretasikan untuk kebutuhan bisnis.

Meskipun nilai *Silhouette Score* tertinggi diperoleh pada konfigurasi $k = 3$ sebesar 0,4689, penelitian ini menggunakan konfigurasi $k = 4$ karena menghasilkan segmentasi yang

lebih representatif untuk kebutuhan interpretasi bisnis. Oleh karena itu, pemilihan jumlah *cluster* tidak hanya mempertimbangkan kualitas statistik, tetapi juga kemampuan model dalam menghasilkan segmen pelanggan yang dapat dimanfaatkan sebagai dasar penyusunan rekomendasi strategi targeting. Perbandingan kedua konfigurasi disajikan pada Tabel 7.

Tabel 7. Komparasi Parameter Metrik Komputasi dan Dampak Manajerial $k = 3$ vs $k = 4$

Parameter Evaluasi	Konfigurasi Statistik ($k = 3$)	Struktur	Konfigurasi Kebutuhan Strategi Bisnis ($k = 4$)	Dampak Informasi	Akurasi terhadap Kampanye Iklan
<i>Silhouette Score</i>	0,4689 (Tertinggi secara matematis)	secara	0,3787 (Valid secara statistik)		Konfigurasi $k = 3$ unggul secara matematis, namun mengorbankan ketajaman pengelompokan operasional pasar.
Nilai WCSS (Inertia)	52,883		41,752 (Lebih Rendah/Kompak)		Penurunan nilai WCSS pada $k = 4$ membuktikan <i>cluster</i> secara geometris jauh lebih rapat dan homogen.
Struktur Segmen Pelanggan Aktif	<i>Melebur/Bercampur</i> menjadi satu segmen makro (<i>High-Value Customer</i>)		<i>Terpisah secara Tegas</i> menjadi segmen VIP Customer (544) dan Loyal Customer (1.229).		Konfigurasi $k = 3$ menghilangkan batas pengeluaran ekstrem kelompok VIP, memicu bias pada <i>centroid</i> nilai <i>Monetary</i> .
Alokasi Biaya Promosi (<i>Marketing Waste</i>)	Risiko Tinggi: Pelanggan VIP dan Loyal menerima bentuk iklan dan insentif diskon yang sama rata.		Risiko Kampanye dispesifikasikan untuk Premium untuk <i>Cross-selling</i> Loyal).	Rendah: iklan (Retensi VIP vs untuk	Konfigurasi $k = 4$ meminimalkan pemborosan biaya dengan menargetkan insentif hanya pada segmen yang sensitif promosi.

Meskipun nilai *Silhouette Score* tertinggi diperoleh pada $k = 3$ sebesar 0,4689, konfigurasi $k = 4$ dipilih karena memberikan manfaat yang lebih besar dalam interpretasi bisnis. Pada konfigurasi $k = 3$, pelanggan *VIP Customer* dan *Loyal Customer* masih tergabung dalam satu kelompok sehingga karakteristik pelanggan bernilai tinggi belum dapat dibedakan secara jelas. Sebaliknya, konfigurasi $k = 4$ mampu memisahkan pelanggan menjadi empat segmen, yaitu *VIP Customer*, *Loyal Customer*, *Regular Customer*, dan *At-Risk Customer*, sehingga menghasilkan dasar yang lebih baik dalam penyusunan rekomendasi strategi targeting. Selain itu, penurunan nilai WCSS dari 52,883 menjadi 41,752 menunjukkan bahwa penambahan satu cluster masih memberikan peningkatan kualitas pengelompokan yang layak tanpa mengurangi validitas model [13].

Hasil penelitian ini sejalan dengan penelitian Hermawan dkk. (2024) [22] dan Febrianty dkk. (2023) [25] yang menunjukkan bahwa integrasi metode RFM dan algoritma K-Means mampu mengidentifikasi kelompok pelanggan berdasarkan perilaku transaksi. Perbedaannya terletak pada distribusi pelanggan, di mana penelitian ini didominasi oleh segmen *Regular Customer*, sedangkan penelitian sebelumnya didominasi oleh pelanggan dengan loyalitas tinggi. Perbedaan tersebut dipengaruhi oleh karakteristik dataset yang digunakan, sehingga menunjukkan bahwa struktur *cluster* tidak hanya dipengaruhi oleh algoritma, tetapi juga oleh karakteristik data dan konteks bisnis yang dianalisis.

4.3 Analisis Rekomendasi Strategi Targeting Berdasarkan Segmen Pelanggan

Hasil segmentasi pelanggan selanjutnya digunakan sebagai dasar dalam penyusunan rekomendasi strategi *targeting* yang lebih tepat sasaran. Setiap segmen pelanggan memiliki karakteristik yang berbeda sehingga memerlukan pendekatan pemasaran yang berbeda pula.

Tabel 8. Rekomendasi Strategi *Targeting* berdasarkan Segmen Pelanggan

Segmen	Karakteristik Utama	Rekomendasi Strategi <i>Targeting</i> per Segmen
VIP Customer	Frequency dan Monetary tertinggi, Recency rendah.	Email Marketing Premium & Akses Eksklusif: Pengiriman katalog digital khusus produk premium via email, pemberian <i>early access</i> (akses awal) 3 hari lebih cepat untuk produk baru sebelum rilis massal, serta undangan program loyalitas berbasis poin prioritas.
Loyal Customer	Frequency cukup tinggi, Monetary menengah.	Program Cross-selling & Up-selling Terpersonalisasi: Penawaran paket <i>bundling</i> produk yang relevan dengan riwayat transaksi mereka, serta pemberian voucher potongan harga otomatis jika nilai transaksi mereka dinaikkan (<i>up-selling</i>).
Regular Customer	Frequency dan Monetary rendah, tetapi rutin.	Insentif Peningkatan Transaksi: Pemberian diskon berkala pada hari-hari tertentu (misal <i>payday promo</i>) dan voucher gratis ongkos kirim dengan syarat minimum pembelian untuk mendorong peningkatan frekuensi belanja.
At-Risk Customer	Recency sangat tinggi (sudah lama tidak bertransaksi).	Kampanye Reaktivasi (Win-Back Campaign): Iklan berbasis <i>retargeting</i> di media sosial serta pengiriman pesan personalisasi/SMS penawaran khusus berisi "voucher eksklusif rindu belanja" dengan masa kedaluwarsa pendek (7 hari) untuk memicu transaksi kembali.

Berdasarkan karakteristik masing-masing segmen, rekomendasi strategi *targeting* disusun sesuai nilai *Recency*, *Frequency*, dan *Monetary* (RFM) pelanggan. *VIP Customer* diprioritaskan untuk program loyalitas dan layanan eksklusif, *Loyal Customer* difokuskan pada strategi *cross-selling* dan *up-selling*, *Regular Customer* diarahkan melalui promosi berkala untuk meningkatkan aktivitas transaksi, sedangkan *At-Risk Customer* memerlukan program reaktivasi guna mendorong pelanggan kembali bertransaksi. Perbedaan rekomendasi tersebut menunjukkan bahwa hasil segmentasi dapat menjadi dasar penyusunan rekomendasi strategi *targeting* yang lebih tepat sasaran.

4.4 Implikasi Hasil Penelitian terhadap Penyusunan Rekomendasi Strategi *Targeting*

Hasil segmentasi pelanggan dan rekomendasi strategi *targeting* yang dihasilkan dalam penelitian ini bersifat sebagai *decision support system* dalam penyusunan rekomendasi strategi *targeting*. Dengan kata lain, rekomendasi *targeting* yang disusun pada Tabel 8 masih berada pada tahap konseptual berdasarkan karakteristik masing-masing segmen pelanggan, dan belum diuji melalui implementasi kampanye iklan secara langsung di lingkungan bisnis nyata. Oleh karena itu, penelitian ini tidak dapat menyimpulkan secara empiris peningkatan efektivitas kampanye iklan, melainkan hanya memberikan dasar analitis untuk membantu pengambilan keputusan pemasaran yang lebih terarah dan berbasis data. Penerapan metode RFM dan algoritma *K-Means Clustering* dalam penelitian ini memberikan implikasi rekomendasi strategis berupa penyusunan rekomendasi strategi *targeting* berbasis data (*data-driven*), di mana hasil segmentasi digunakan sebagai dasar analitis dalam merancang rekomendasi strategi *targeting* yang lebih terarah, efisien, serta berpotensi mendukung optimalisasi alokasi anggaran dan pengurangan pemborosan biaya promosi di masa mendatang.

5. KESIMPULAN

Penelitian ini berhasil menerapkan metode *Recency*, *Frequency*, *Monetary* (RFM) dan algoritma *K-Means Clustering* untuk melakukan segmentasi pelanggan berdasarkan perilaku transaksi. Dataset awal terdiri atas 5.000 pelanggan dan 32.295 transaksi, yang setelah tahap *preprocessing* menghasilkan 3.864 pelanggan dan 28.009 transaksi, dengan 3.522 pelanggan

memenuhi kualifikasi agregasi untuk pemodelan RFM. Hasil evaluasi menunjukkan bahwa nilai *Silhouette Score* tertinggi diperoleh pada $k = 3$ sebesar 0,4689, namun penelitian ini memilih $k = 4$ dengan nilai *Silhouette Score* 0,3787 karena menghasilkan segmentasi yang lebih sesuai untuk kebutuhan interpretasi bisnis. Empat segmen pelanggan yang terbentuk yaitu *VIP Customer* (544 pelanggan), *Loyal Customer* (1.229 pelanggan), *Regular Customer* (1.355 pelanggan), dan *At-Risk Customer* (394 pelanggan). Hasil segmentasi ini berfungsi sebagai *decision support system* dalam membantu penyusunan rekomendasi strategi *targeting* yang lebih terarah berdasarkan karakteristik masing-masing segmen pelanggan. Penelitian selanjutnya disarankan untuk membandingkan metode ini dengan algoritma lain seperti *K-Medoids*, *Hierarchical Clustering*, atau *DBSCAN* untuk memperoleh hasil segmentasi yang lebih komprehensif.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak terdapat konflik kepentingan dalam penelitian ini, baik yang bersifat finansial maupun non-finansial. Penulis juga menyatakan bahwa penelitian ini dilakukan secara independen tanpa adanya pengaruh dari pihak mana pun terhadap proses penelitian, analisis data, maupun penyusunan artikel.

DAFTAR PUSTAKA

- [1] S. Rahayu, H. W. Halawa, A. F. Abdillah, and A. Mujayanah, "Pemanfaatan Big Data Analytics untuk Analisis Pola Perilaku Konsumen E-commerce Strategis," *JUTECH: Journal Education and Technology*, vol. 6, no. 1, pp. 64–73, 2025, doi: 10.31932/jutech.v6i1.4834.
- [2] Y. Syahra, A. Fadlil, and H. Yuliansyah, "Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry," *SINKRON: Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1120–1131, 2025, doi: 10.33395/sinkron.v9i3.14907.
- [3] R. M. Fauzan and G. Alfian, "Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma *Clusterisasi* K-Means," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 170–177, 2024, doi: 10.14421/jiska.2024.9.3.170-177.
- [4] O. Akande, E. O. Asani, and B. T. Dautare, "Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy," *Ceddi Journal of Information System and Technology*, vol. 3, no. 1, pp. 14–25, 2024, doi: 10.56134/jst.v3i1.81.
- [5] V. Soni, "A Meta-Analytic Examination of the Effects of Personalized Digital Marketing on Consumer Purchasing," *International Journal of Computer Trends and Technology*, vol. 72, no. 9, pp. 32–43, 2024, doi: 10.14445/22312803/IJCTT-V72I9P106.
- [6] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *Journal of Marketing Analytics*, vol. 11, no. 4, pp. 677–692, 2023, doi: 10.1057/s41270-023-00235-5.
- [7] S. Septiani, Musthofa, and P. Seviawani, "Penggunaan Big Data untuk Personalisasi Layanan dalam Bisnis E-commerce," *ADI Bisnis Digital Interdisiplin Jurnal*, vol. 5, no. 1, pp. 51–57, 2024, doi: 10.34306/abdi.v5i1.1098.
- [8] I. Ramadhani and N. H. Salisah, "Peran AI dalam Personalisasi Periklanan Digital: Studi Kasus Spotify," *Communicator Sphere*, vol. 4, no. 2, pp. 122–131, 2024, doi: 10.55397/cps.v4i2.120.
- [9] Z. Zhang, "Market Segmentation and Personalized Marketing Strategy Optimization Driven by Big Data Analysis," in *Advances in Transdisciplinary Engineering*, IOS Press, 2024, pp. 240–249, doi: 10.3233/ATDE240434.
- [10] S. P. L. Geni, H. Safitri, F. Merry, M. Wati, and Haviluddin, "Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian," *JUKI: Jurnal Komputer dan Informatika*, vol. 7, no. 1, pp. 89–99, 2025, doi: 10.53842/juki.v7i1.1049.

- [11] I. Yunita, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, "Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran," *Jurnal Media Teknik dan Sistem Industri*, vol. 9, no. 1, pp. 58–66, 2025, doi: 10.35194/jmtsi.v9i1.4452.
- [12] M. F. Fadhillah, A. L. A. Suyoso, and I. Puspitasari, "Segmentasi Pelanggan dengan RFM dan K-Means Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 48–56, 2024, doi: 10.57152/malcom.v5i1.1491.
- [13] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering Optimization in RFM Analysis Based on K-Means," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 470–477, 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.
- [14] Y. Putri, D. Aldo, and W. Ilham, "Retail Marketing Strategy Optimization Using K-Means Clustering," *SINKRON: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 4, pp. 2155–2163, 2024, doi: 10.33395/sinkron.v8i4.14000.
- [15] J. Banjarnahor et al., "Analyzing Consumer Shopping Interest via Social Media Ads Using K-Means and C4.5," *Jurnal Sisfokom*, vol. 13, no. 3, pp. 416–421, 2024, doi: 10.32736/sisfokom.v13i3.2228.
- [16] S. Kumar, R. Rani, S. K. Pippal, and R. Agrawal, "Customer Segmentation in E-Commerce: K-Means vs Hierarchical Clustering," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 23, no. 1, pp. 119–128, 2025, doi: 10.12928/telkomnika.v23i1.26384.
- [17] R. Martiansah, S. Monalisa, F. Muttakin, and M. Fronita, "Customer Segmentation Analysis Through RFM-D Model and K-Means Algorithm," *Jurnal Sistem Cerdas*, vol. 8, no. 1, pp. 1–11, 2025, doi: 10.37396/jsc.v8i1.504.
- [18] K. Gopalakrishnan, "Customer Segmentation Using K-Means Clustering for Targeted Marketing in Banking," *International Journal of Artificial Intelligence & Machine Learning*, vol. 3, no. 2, pp. 8–17, 2024, doi: 10.5281/zenodo.13627403.
- [19] 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 Teknologi dan Sistem Informasi*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.
- [20] 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.
- [21] P. A. Wicaksana, I. B. A. Swamardika, and R. S. Hartati, "Literature Review Analisis Perilaku Pelanggan Menggunakan RFM Model," *Majalah Ilmiah Teknologi Elektro*, vol. 21, no. 1, pp. 21–30, 2022, doi: 10.24843/mite.2022.v21i01.p04.
- [22] A. Hermawan et al., "Optimalisasi Strategi Pemasaran Melalui Analisis RFM pada Dataset Ritel," *Jurnal Manajemen Riset Inovasi*, vol. 2, no. 4, pp. 254–267, 2024, doi: 10.55606/mri.v2i4.3246.
- [23] B. Al Hamid and T. Wiradinata, "RFM-Based Customer Segmentation Using K-Means Clustering for Marketing Strategy Optimization at Queen Audio," *Journal of Applied Informatics and Computing*, vol. 10, no. 2, pp. 1828–1833, 2026, doi: 10.30871/jaic.v10i2.12302.
- [24] N. Zahro, N. A. Maori, and G. W. N. Wibowo, "Integration of RFM Method and K-Means Clustering for Customer Segmentation Effectiveness," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 5, no. 1, pp. 12–21, 2025, doi: 10.20895/dinda.v5i1.1649.
- [25] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.

- [26] R. Siagian, P. Sirait, and A. Halim, "Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce," *SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 2, pp. 260–270, 2022, doi: 10.32520/stmsi.v11i2.1337.
- [27] R. D. Nugraha, D. D. Adelia, and D. Rivaldi, "Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma K-Means Clustering," *Digital Transformation Technology (Digitech)*, vol. 5, no. 2, pp. 141–148, 2025, doi: 10.47709/digitech.v5i2.6340.
- [28] D. Chen, L. S. Sain, and K. Guo, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," *Journal of Database Marketing & Customer Strategy Management*, vol. 19, no. 3, pp. 197–208, 2012, doi: 10.1057/dbm.2012.17.