

Analisis Segmentasi Pelanggan Menggunakan RFM dan K-Means Clustering sebagai Dasar Penyusunan Aturan Pendukung Keputusan

Meisya Dwi Andini*, Rafa Nadira Catra, Weli Ratri Homausyah, Haaniyah Aurelia, Allsela Meiriza, Ken Ditha Tania, Zaqqi Yamani

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: ^{1,*}meisyadwiandini16@gmail.com, ²rafanadirac@gmail.com, ³weliratrihomausyah@gmail.com,

⁴haaniyaharelia@gmail.com, ⁵allsela_meiriza@yahoo.co.id, ⁶kenya.tania@gmail.com, ⁷zaqqi_yamani@unsri.ac.id

Submitted: 09/03/2026; Accepted: 31/03/2026; Published: 31/03/2026

Abstrak—Salah satu metode penting dalam mendukung inisiatif Customer Relationship Management (CRM) berbasis data adalah segmentasi pelanggan. Namun, dalam praktiknya, hasil segmentasi sering kali hanya digunakan untuk analisis deskriptif dan belum dimanfaatkan secara lebih lanjut dalam proses pendukung keputusan. Penelitian ini bertujuan untuk memanfaatkan hasil segmentasi pelanggan menggunakan pendekatan Recency, Frequency, Monetary (RFM) dan algoritma K-Means sebagai dasar penyusunan rekomendasi keputusan. Tahapan penelitian meliputi pra-pemrosesan data, perhitungan nilai RFM, normalisasi menggunakan metode Min-Max Scaling, serta penentuan jumlah kluster menggunakan Elbow Method dan Silhouette Score. Hasil evaluasi menunjukkan bahwa jumlah kluster terbaik adalah empat dengan nilai Silhouette Score sebesar 0,61, yang mengindikasikan kualitas pemisahan kluster yang cukup baik. Segmentasi pelanggan menghasilkan empat kategori, yaitu High Value/VIP Customers, Loyal Customers, Potential Customers, dan Low Value/Dormant Customers, yang masing-masing memiliki karakteristik perilaku transaksi yang berbeda. Karakteristik tersebut kemudian diinterpretasikan dalam bentuk aturan keputusan berbasis logika IF–THEN, misalnya pelanggan dengan nilai Recency rendah, Frequency tinggi, dan Monetary tinggi direkomendasikan strategi loyalty reward dan upselling. Hasil penelitian ini menunjukkan bahwa segmentasi pelanggan tidak hanya dapat digunakan untuk analisis deskriptif, tetapi juga dapat dimanfaatkan sebagai dasar penyusunan rekomendasi pemasaran secara praktis, meskipun masih bersifat sederhana dan berbasis heuristik. Kontribusi penelitian ini adalah mengintegrasikan hasil segmentasi RFM-K-Means dengan aturan keputusan IF–THEN untuk menghasilkan rekomendasi strategi pemasaran yang lebih aplikatif dalam mendukung pengambilan keputusan berbasis data.

Kata Kunci: CRM; Knowledge-Based Decision Support; K-Means Clustering; RFM

Abstract—One of the important methods in supporting data-driven Customer Relationship Management (CRM) initiatives is customer segmentation. However, in practice, segmentation results are often limited to descriptive analysis and are not further utilized in decision-support processes. This study aims to utilize customer segmentation results based on the Recency, Frequency, Monetary (RFM) approach and the K-Means algorithm as a basis for developing decision-support recommendations. The research stages include data preprocessing, RFM value calculation, normalization using the Min-Max Scaling method, and determining the optimal number of clusters using the Elbow Method and Silhouette Score. The evaluation results indicate that the optimal number of clusters is four, with a Silhouette Score of 0.61, which reflects a moderately good level of cluster separation. The segmentation results classify customers into four categories: High Value/VIP Customers, Loyal Customers, Potential Customers, and Low Value/Dormant Customers, each exhibiting distinct transactional behavior characteristics. These characteristics are then interpreted into decision rules using IF–THEN logic; for example, customers with low Recency, high Frequency, and high Monetary values are recommended strategies such as loyalty rewards and upselling. The findings suggest that customer segmentation can be extended beyond descriptive analysis and utilized as a practical basis for marketing decision-making, although the approach remains relatively simple and heuristic-based. The contribution of this study is to integrate RFM-KMeans segmentation results with IF–THEN decision rules to generate more applicable marketing strategy recommendations in supporting data-driven decision making.

Keywords: CRM; Knowledge-Based Decision Support; K-Means Clustering; RFM

1. PENDAHULUAN

Transformasi digital berperan signifikan dalam memperkuat strategi Customer Relationship Management (CRM) dengan memungkinkan perusahaan mengelola interaksi pelanggan secara lebih sistematis, responsif, dan terintegrasi [1]. Perkembangan teknologi digital, seperti big data, cloud computing, dan artificial intelligence, telah mendorong perubahan paradigma dalam pengelolaan hubungan pelanggan dari yang bersifat konvensional menjadi berbasis data dan analitik. Meskipun demikian, dalam praktiknya pengambilan keputusan bisnis masih kerap didasarkan pada intuisi manajerial. Pendekatan intuitif ini sering kali dipengaruhi oleh pengalaman subjektif dan bias kognitif, sehingga berpotensi menghasilkan keputusan yang kurang optimal. Kondisi ini berpotensi mengurangi efektivitas implementasi CRM karena keputusan yang diambil belum sepenuhnya bertumpu pada analisis data yang terukur [2]. Pemanfaatan analitik data dalam kerangka business intelligence terbukti mampu meningkatkan kualitas pengambilan keputusan berbasis informasi yang lebih objektif dan terstruktur [3]. Dengan adanya sistem business intelligence, data yang sebelumnya tersebar dapat diintegrasikan, dianalisis, dan disajikan dalam bentuk informasi yang mudah dipahami oleh pengambil keputusan. Integrasi teknologi digital dengan pendekatan analitik menjadi penting untuk meningkatkan akurasi, kecepatan, serta objektivitas dalam perumusan strategi pemasaran berbasis data [4]. Selain itu, integrasi ini juga memungkinkan perusahaan untuk merespons perubahan perilaku pelanggan secara real-time, sehingga strategi yang dihasilkan menjadi lebih adaptif dan kompetitif dalam menghadapi dinamika pasar yang terus berkembang.

Dalam konteks CRM, segmentasi pelanggan merupakan tahapan fundamental untuk memahami heterogenitas perilaku konsumen. Segmentasi memungkinkan perusahaan mengidentifikasi kelompok pelanggan dengan karakteristik yang serupa sehingga strategi pemasaran dapat disesuaikan secara lebih spesifik dan efektif. Salah satu pendekatan yang banyak digunakan untuk proses segmentasi adalah teknik clustering dalam data mining yang bertujuan mengelompokkan data berdasarkan kemiripan karakteristik. Pendekatan ini menjadi semakin relevan seiring meningkatnya volume dan kompleksitas data pelanggan yang tidak lagi dapat dianalisis secara manual. Dalam pemanfaatan teknologi kecerdasan buatan dan machine learning, pendekatan unsupervised learning seperti K-Means memungkinkan identifikasi pola tersembunyi dalam data tanpa memerlukan label sebelumnya [2]. Hal ini memberikan fleksibilitas dalam eksplorasi data yang belum terstruktur. Berbagai penelitian menunjukkan bahwa metode clustering seperti K-Means, K-Medoids, maupun DBSCAN mampu mengidentifikasi pola dalam dataset yang kompleks [5]. Masing-masing metode memiliki keunggulan dan kelemahan tersendiri, misalnya K-Means unggul dalam efisiensi komputasi, sedangkan DBSCAN lebih mampu menangani data dengan bentuk cluster yang tidak beraturan. Selain itu, penerapan clustering pada berbagai domain data juga sering dievaluasi menggunakan indikator kualitas cluster seperti Davies Bouldin Index (DBI) untuk memastikan hasil pengelompokan yang optimal [6]. Evaluasi ini penting untuk memastikan bahwa cluster yang dihasilkan memiliki tingkat homogenitas internal yang tinggi dan heterogenitas antar cluster yang jelas. Hal ini membuka peluang penerapan segmentasi yang lebih adaptif terhadap perubahan pasar, sehingga strategi pemasaran dapat diperbarui sesuai dinamika tren konsumen. Namun, proses segmentasi sering kali berhenti pada tahap pengelompokan (clustering) tanpa dilanjutkan pada pemanfaatan hasil analisis secara lebih lanjut [4]. Adapun keterkaitan antara hasil clustering dan sistem pendukung keputusan menyebabkan potensi informasi strategis dari data pelanggan belum dimanfaatkan secara optimal [7]. Dengan kata lain, hasil analisis yang dihasilkan belum sepenuhnya diterjemahkan menjadi tindakan nyata dalam strategi bisnis. Oleh karena itu, diperlukan pendekatan yang tidak hanya menghasilkan segmentasi, tetapi juga mampu menerjemahkannya menjadi rekomendasi yang aplikatif dalam pengambilan keputusan strategis [8]. Pendekatan ini diharapkan dapat menjembatani kesenjangan antara analisis data dan implementasi strategi di lapangan.

Sejumlah penelitian sebelumnya menunjukkan adanya keterbatasan dalam pemanfaatan hasil segmentasi pelanggan. Penelitian oleh Ridwan (2025) [1] menekankan transformasi segmentasi berbasis data perilaku, namun belum mengintegrasikan hasilnya ke dalam sistem pendukung keputusan yang operasional sehingga implementasinya dalam konteks bisnis masih terbatas pada analisis eksploratif. Selanjutnya, Wang (2025) [2] menggunakan pendekatan algoritma berbasis optimasi untuk segmentasi, tetapi berfokus pada peningkatan akurasi model tanpa mengaitkan hasilnya dengan strategi bisnis yang implementatif, sehingga nilai praktisnya dalam pengambilan keputusan masih belum optimal. Penelitian Titus (2025) [7] melakukan komparasi berbagai teknik clustering dalam kerangka RFM, namun hanya berhenti pada evaluasi performa algoritma tanpa menghasilkan rekomendasi strategis yang dapat langsung diterapkan oleh praktisi bisnis. Sementara itu, penelitian yang dilakukan oleh Akande (2024) [8] telah mengombinasikan RFM dan K-Means untuk menghasilkan insight pemasaran, tetapi belum menyusun mekanisme formal seperti aturan keputusan (decision rules) yang sistematis sehingga interpretasi hasil masih bergantung pada subjektivitas analis. Dengan demikian, terdapat kesenjangan penelitian (research gap) yang jelas, yaitu belum adanya integrasi yang komprehensif antara proses segmentasi pelanggan, evaluasi kualitas cluster, serta penerjemahan hasilnya ke dalam rekomendasi keputusan berbasis aturan yang aplikatif. Kesenjangan ini menunjukkan perlunya pendekatan yang tidak hanya kuat secara metodologis, tetapi juga relevan secara praktis dalam mendukung pengambilan keputusan berbasis data.

Model segmentasi berbasis RFM (Recency, Frequency, Monetary) yang dikombinasikan dengan algoritma K-Means telah banyak digunakan untuk mengelompokkan pelanggan berdasarkan pola perilaku pembelian [10]. Pendekatan RFM memberikan ukuran kuantitatif terkait waktu transaksi terakhir, intensitas pembelian, serta kontribusi nilai moneter pelanggan. Ketiga dimensi ini mampu merepresentasikan nilai pelanggan secara komprehensif dari sisi aktivitas dan kontribusi ekonomi. Sementara itu, K-Means berfungsi mengidentifikasi kelompok pelanggan yang memiliki karakteristik homogen [10]. Proses iteratif dalam K-Means memungkinkan pembentukan cluster yang optimal berdasarkan jarak antar data. Kombinasi keduanya membentuk landasan analitis yang cukup kuat untuk mendukung penyusunan rekomendasi dalam konteks strategi CRM [11]. Selain itu, pendekatan ini juga relatif sederhana dan mudah diimplementasikan, sehingga banyak digunakan dalam berbagai studi dan aplikasi industri.

Integrasi analisis RFM, proses clustering, dan mekanisme interpretasi hasil dalam satu kerangka kerja menghasilkan pendekatan yang lebih komprehensif [4]. Pendekatan ini tidak hanya berfokus pada pengelompokan data, tetapi juga pada bagaimana hasil tersebut dapat dimanfaatkan secara strategis. Hasil segmentasi tidak hanya berfungsi sebagai deskripsi kelompok pelanggan, tetapi juga dapat dimanfaatkan sebagai input bagi proses pendukung keputusan [8]. Melalui kerangka ini, perusahaan dapat merancang strategi pemasaran yang lebih tepat sasaran, seperti penentuan promosi yang berbeda untuk setiap segmen pelanggan. Selain itu, perusahaan juga dapat meningkatkan efisiensi alokasi sumber daya dengan memprioritaskan segmen pelanggan yang memiliki nilai tinggi. Dengan demikian, strategi yang dihasilkan menjadi lebih responsif terhadap perubahan perilaku pelanggan serta berorientasi pada peningkatan loyalitas dan profitabilitas [12].

Pemanfaatan hasil segmentasi RFM-K-Means dalam konteks pendukung keputusan memberikan nilai tambah praktis dalam manajemen pemasaran. Pendekatan ini memungkinkan penyusunan rekomendasi berdasarkan karakteristik masing-masing segmen sehingga proses pengambilan keputusan menjadi lebih terarah dan terukur [13].

Selain itu, pendekatan ini juga mendukung personalisasi layanan dan perancangan kampanye pemasaran yang sesuai dengan profil pelanggan, sehingga efektivitas komunikasi dan kepuasan pelanggan dapat ditingkatkan secara berkelanjutan [12]. Dalam jangka panjang, penerapan pendekatan ini berpotensi meningkatkan retensi pelanggan dan nilai seumur hidup pelanggan (customer lifetime value).

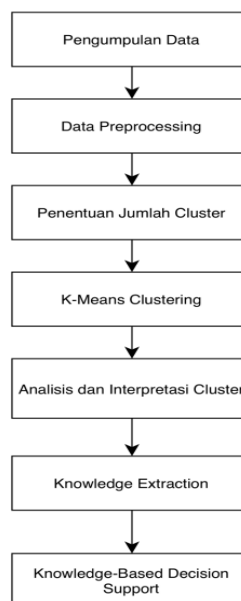
Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis dan memanfaatkan hasil segmentasi pelanggan menggunakan pendekatan RFM dan algoritma K-Means sebagai dasar penyusunan rekomendasi pendukung keputusan. Berbeda dengan penelitian yang berfokus pada pembentukan cluster, penelitian ini menekankan pada pemanfaatan hasil segmentasi untuk mendukung proses pengambilan keputusan. Penelitian ini tidak hanya melakukan segmentasi pelanggan, tetapi juga mengaitkan karakteristik cluster dengan rekomendasi strategi pemasaran. Melalui representasi aturan keputusan berbasis logika IF–THEN, hasil analisis data diterjemahkan menjadi rekomendasi strategi yang lebih sistematis dalam mendukung pengambilan keputusan pemasaran berbasis data. Pendekatan ini diharapkan dapat menjembatani kesenjangan antara analisis data dan implementasi strategi bisnis secara nyata.

Kontribusi penelitian ini terletak pada pengembangan kerangka terintegrasi yang menghubungkan proses segmentasi pelanggan berbasis RFM dan K-Means dengan sistem pendukung keputusan melalui representasi aturan IF–THEN. Penelitian ini tidak hanya menghasilkan cluster pelanggan, tetapi juga menyediakan mekanisme interpretasi yang sistematis untuk menerjemahkan karakteristik setiap segmen menjadi rekomendasi strategi pemasaran yang aplikatif. Selain itu, penelitian ini memperkaya literatur dengan menggabungkan evaluasi kualitas cluster (DBI) dan pendekatan decision rules dalam satu alur analisis yang komprehensif, sehingga dapat meningkatkan akurasi, objektivitas, dan implementabilitas hasil analisis dalam konteks CRM berbasis data [6] [8]. Lebih lanjut, penelitian ini juga memberikan kontribusi praktis bagi organisasi dalam mengoptimalkan pemanfaatan data pelanggan sebagai dasar pengambilan keputusan yang lebih efektif dan berbasis bukti (evidence-based decision making).

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan tujuan menganalisis dan memanfaatkan hasil segmentasi pelanggan dalam kerangka pendukung keputusan dengan menggunakan metode pengelompokan K-Means dan analisis Recency, Frequency, Monetary (RFM). Kumpulan data, yang mencakup data transaksi pelanggan, diambil dari Kumpulan Data Segmentasi Pelanggan yang tersedia di platform Kaggle [14], yang meskipun praktis namun memiliki keterbatasan dalam merepresentasikan kompleksitas data dunia nyata. Proses penelitian ini meliputi pengumpulan data, prapemrosesan data, penyusunan variabel RFM, normalisasi data, penentuan jumlah kluster, pengelompokan menggunakan teknik K-Means, serta interpretasi hasil segmentasi untuk menyusun pengetahuan sebagai dasar rekomendasi keputusan. Gambar 1 menggambarkan alur kerja penelitian.



Gambar 1. Flowchart Metodologi

Berdasarkan Gambar 1, flowchart metodologi penelitian ini menggambarkan tahapan sistematis yang dimulai dari pengumpulan data pelanggan sebagai dasar analisis, kemudian dilanjutkan dengan proses data preprocessing untuk membersihkan, mentransformasi, dan menyiapkan data agar siap diolah. Selanjutnya dilakukan penentuan jumlah cluster yang optimal menggunakan metode evaluasi tertentu sebelum menerapkan algoritma K-Means

Clustering untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik, khususnya nilai RFM. Hasil clustering kemudian dianalisis dan diinterpretasikan untuk memahami profil masing-masing segmen pelanggan. Tahap berikutnya adalah knowledge extraction, yaitu proses mengubah hasil analisis menjadi pengetahuan yang lebih terstruktur, seperti pola perilaku atau aturan sederhana. Pada tahap akhir, pengetahuan tersebut dimanfaatkan dalam knowledge-based decision support untuk menghasilkan rekomendasi yang mendukung pengambilan keputusan strategis, sehingga keseluruhan proses mengalir dari data mentah hingga menjadi dasar keputusan yang lebih terarah dan berbasis data.

2.2 Dataset Penelitian

Dataset yang digunakan adalah data “Customer Segmentation Dataset” diperoleh dari situs web Kaggle. Kumpulan data ini berisi informasi transaksi penjualan dari sebuah bisnis ritel yang menjual berbagai macam barang secara daring. Data transaksi tersebut terdiri dari 541.909 catatan transaksi pelanggan yang telah terstruktur dalam kurun waktu 2010-2011, sehingga memudahkan proses analisis namun berpotensi tidak sepenuhnya mencerminkan permasalahan data nyata seperti noise, inkonsistensi, dan ketidakseimbangan data. Untuk mengidentifikasi karakteristik pelanggan berdasarkan aktivitas pembelian mereka, variabel RFM disimpulkan dari data transaksi tersebut [15].

2.3 Data Preprocessing

Tujuan dari tahap persiapan data adalah untuk membersihkan dan menyiapkan data sebelum dianalisis. Data duplikat dihilangkan, nilai yang hilang ditangani, serta dilakukan pengelompokan berdasarkan CustomerID. Selain itu, dilakukan pemeriksaan terhadap outlier yang berpotensi mempengaruhi hasil clustering, meskipun tidak semua data ekstrem dihilangkan agar tetap mempertahankan karakteristik asli data. Atribut CustomerID kemudian digunakan untuk memproses dan menggabungkan data transaksi pada tingkat pelanggan. Tujuan dari metode penggabungan ini adalah untuk mengumpulkan data yang diperlukan guna menghitung variabel nilai moneter, frekuensi, dan kedekatan waktu [16].

2.4 Perhitungan RFM

Recency adalah lamanya waktu yang telah berlalu sejak transaksi terakhir pelanggan; Frequency adalah seberapa sering pelanggan melakukan pembelian dalam rentang waktu tertentu; dan Monetary adalah nilai total transaksi pelanggan. Ketiga metrik ini memberikan gambaran numerik mengenai tingkat keterlibatan pelanggan dalam aktivitas pembelian [17]. Pada tahap normalisasi, digunakan metode Min-Max Scaling untuk mengubah rentang nilai setiap variabel ke dalam skala yang seragam, sehingga tidak ada atribut yang mendominasi proses clustering.

2.5 Normalisasi Data

Untuk mengurangi bias yang disebabkan oleh perbedaan skala antar variabel, nilai-nilai variabel RFM dinormalisasi sebelum tahap pengelompokan. Metode normalisasi yang digunakan adalah pendekatan Min-Max Scaling untuk mengubah rentang data ke skala yang seragam. Prosedur ini bertujuan untuk memastikan bahwa setiap atribut memiliki kontribusi yang seimbang dan tidak mendominasi proses pengelompokan data [18].

2.6 Penentuan Jumlah Cluster

Dengan menghitung *Within Cluster Sum of Squares (WCSS)*, pendekatan *Elbow* digunakan untuk menghitung jumlah kluster yang ideal. Teknik ini digunakan untuk menentukan titik ideal di mana penambahan kluster tidak lagi menyebabkan nilai WCSS menurun secara signifikan. Untuk mengevaluasi kualitas pemisahan antar kluster, *Silhouette Score* juga digunakan untuk memvalidasi jumlah kluster. Untuk memastikan bahwa jumlah kluster akhir stabil dan representatif, kombinasi kedua metode ini diterapkan [19].

2.7 Proses Clustering K-Means

Pelanggan dikelompokkan menggunakan algoritma *K-Means* berdasarkan tingkat kesamaan atribut data mereka. Titik pusat awal ditetapkan sebagai pusat kluster setelah jumlah kluster yang diinginkan ditentukan. Jarak Euklides kemudian digunakan untuk menentukan jarak antara setiap titik data dan titik pusat tersebut. Kluster data dengan sifat yang relatif seragam dihasilkan dengan mengulangi prosedur tersebut hingga posisi titik pusat menjadi stabil [20].

2.8 Perancangan Knowledge-Based Decision Support

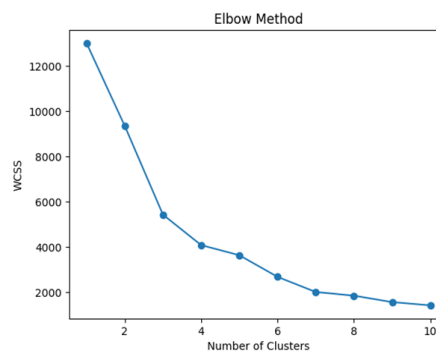
Perancangan pendekatan pendukung keputusan berdasarkan hasil segmentasi pelanggan merupakan tahap terakhir dari penelitian ini. Atribut setiap kluster, yang dihasilkan oleh analisis RFM, dipahami sebagai informasi eksplisit yang menggambarkan pola perilaku pelanggan. Logika IF-THEN kemudian digunakan untuk menyampaikan pengetahuan ini dalam bentuk aturan keputusan sederhana yang bersifat heuristik. Pendekatan ini tidak melibatkan pengembangan arsitektur sistem berbasis pengetahuan secara formal, seperti mesin inferensi atau basis pengetahuan terstruktur, melainkan berfokus pada interpretasi hasil clustering menjadi rekomendasi praktis. Pedoman ini menghubungkan taktik yang disarankan untuk setiap segmen pasar dengan ciri-ciri perilaku konsumen. Akibatnya,

hasil segmentasi dapat berfungsi sebagai landasan pendukung pengambilan keputusan dalam manajemen hubungan pelanggan selain menghasilkan pengelompokan statistik pelanggan [21].

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Jumlah Cluster

Untuk menghasilkan pengelompokan data yang stabil, informatif, dan menunjukkan kualitas pemisahan yang baik antar cluster, *Elbow Method* dan *Silhouette Score* digunakan dalam penelitian ini untuk menghitung jumlah cluster. Memilih nilai K yang tepat merupakan langkah krusial dalam pendekatan *K-Means* untuk segmentasi konsumen. Maka dari itu, pendekatan *Elbow Method* berguna untuk menunjukkan variasi keseluruhan di dalam setiap cluster guna menentukan jumlah cluster yang ideal berdasarkan nilai *Within Cluster Sum of Squares (WCSS)*.



Gambar 2. Hasil *Elbow Method*

Terlihat dari Gambar 2 bahwa anggota dalam suatu cluster lebih homogen ketika nilai WCSS lebih rendah. Nilai WCSS pada gambar ini menunjukkan penurunan yang signifikan hingga K = 4, setelah itu nilainya mulai stabil. Titik siku terlihat pada K = 4, yang mengindikasikan bahwa kualitas pengelompokan tidak meningkat secara signifikan dengan penambahan cluster lebih lanjut setelah K = 4.

Skor *Silhouette* dihitung sebagai teknik penilaian tambahan guna mengonfirmasi temuan pendekatan *Elbow* dalam penelitian ini. Teknik ini menilai sejauh mana objek dalam satu kelompok dapat dibedakan dari objek dalam kelompok lain. Pemisahan kelompok yang lebih baik ditandai oleh nilai yang mendekati 1 pada skala -1 hingga 1. Hasil berikut ini diperoleh dari perhitungan Skor *Silhouette* untuk berbagai nilai K.

Untuk menentukan jumlah cluster yang optimal, dilakukan evaluasi menggunakan metode *Silhouette Score* pada beberapa variasi nilai K. Pengujian ini bertujuan untuk mengukur kualitas pemisahan antar cluster serta tingkat kedekatan data dalam satu cluster. Semakin tinggi nilai *Silhouette Score* (mendekati 1), maka semakin baik kualitas pengelompokan yang dihasilkan. Hasil perhitungan *Silhouette Score* untuk masing-masing nilai K disajikan pada Tabel 1.

Tabel 1. Hasil *Silhouette Score*

K	Silhouette Score
2	0.56
3	0.59
4	0.61
5	0.43
6	0.49
7	0.51

Dibandingkan dengan perhitungan jumlah cluster alternatif, perhitungan *Silhouette Score* menunjukkan bahwa nilai K = 4 menghasilkan skor maksimum sebesar 0,61, yang mengindikasikan kualitas pemisahan cluster yang cukup baik, namun belum menunjukkan separasi yang sangat kuat. Oleh karena itu, hasil ini masih bersifat relatif dan dapat diperkuat dengan metrik validasi tambahan seperti *Davies-Bouldin Index* atau *Calinski-Harabasz Index* untuk memastikan konsistensi kualitas cluster. Studi ini menyimpulkan bahwa terdapat empat cluster berdasarkan hasil evaluasi menggunakan metode *Elbow* dan *Silhouette Score*, dengan mempertimbangkan keseimbangan antara interpretabilitas dan kualitas segmentasi.

3.2 Hasil Clustering RFM

Untuk memahami karakteristik setiap cluster yang terbentuk, dilakukan perhitungan nilai rata-rata *Recency*, *Frequency*, dan *Monetary (RFM)* pada masing-masing cluster. Nilai rata-rata ini digunakan untuk mengidentifikasi pola perilaku pelanggan dalam setiap kelompok hasil clustering. Hasil perhitungan tersebut disajikan pada Tabel 2.

Tabel 2. Rata-rata RFM tiap *Cluster*

Cluster	Recency	Frequency	Monetary
0	14,672986	22,047393	12,453.225924
1	247,564030	1,551789	478,107581
2	6,384615	82,692308	127.338,313846
3	42,910580	3,655748	1.349,698298

Fitur-fitur yang ditampilkan dalam Tabel 2, yaitu skor rata-rata Recency, Frequency, dan Monetary (RFM), diperoleh dari hasil segmentasi K-means, dengan jumlah kluster sebanyak empat. Nilai-nilai ini memberikan indikasi awal mengenai perbedaan perilaku pelanggan antar cluster, namun interpretasi lebih lanjut diperlukan untuk memahami makna bisnis dari setiap kelompok. Berdasarkan skor Recency, seorang pelanggan dapat dianggap lebih aktif jika waktu yang berlalu sejak transaksi terakhirnya relatif singkat. Sementara itu, skor Monetary menunjukkan nilai total pembelian pelanggan, dan skor Frequency menunjukkan jumlah transaksi.

3.3 Interpretasi Segmentasi Pelanggan

Selain karakteristik RFM, distribusi jumlah pelanggan pada setiap cluster juga dianalisis untuk mengetahui proporsi segmentasi yang terbentuk. Informasi ini penting untuk memahami dominasi segmen tertentu dalam dataset serta sebagai dasar dalam penyusunan strategi pemasaran. Jumlah customer pada masing-masing cluster disajikan pada Tabel 3.

Tabel 1. Jumlah *Customer* dari Tiap *Cluster*

Cluster	Jumlah Customer
3	3.053
1	1.062
0	211
2	13

Berdasarkan hasil dari clustering dengan mengimplementasikan metode K-Means dengan pendekatan RFM diperoleh empat cluster customer dengan karakteristik yang berbeda. Analisis dilakukan berdasarkan nilai rata-rata RFM pada setiap clusternya, jumlah anggota dalam masing-masing cluster, serta visualisasi distribusi pelanggan dalam dimensi Frequency dan Monetary. Perbedaan nilai ini menunjukkan adanya variasi perilaku pembelian yang signifikan, meskipun interpretasi yang dihasilkan masih bersifat deskriptif dan bergantung pada pendekatan analisis yang digunakan.

Cluster 2 - High Value / VIP Customers

Pada Cluster ini sebanyak 13 customer memiliki nilai Recency yang paling rendah yaitu 6.38, nilai Frequency tertinggi sebesar 82.69, dan nilai Monetary tertinggi sebesar 127,338.313846. Hal ini menunjukkan bahwa customer dalam cluster ini sangat aktif melakukan transaksi, frekuensi pembelian yang tinggi, serta memberikan kontribusi yang paling besar. Meskipun jumlah customer pada cluster ini hanya 13 orang, tetapi kontribusi Monetarynya sangat besar dibandingkan dengan cluster yang lainnya. Oleh karena itu, cluster ini dapat dikategorikan sebagai customer high value atau VIP customer, meskipun ukuran cluster yang sangat kecil perlu diperhatikan dalam pengambilan keputusan strategis.

Cluster 0 - Loyal Customers

Cluster ini menunjukkan nilai Recency yang relatif rendah yaitu sebesar 14.67, nilai Frequency yang cukup tinggi sebesar 22.05, dan nilai Monetary sebesar 12,453.225924. Hal ini menunjukkan bahwa customer cukup sering melakukan transaksi dengan nilai pembelian yang stabil. Oleh karena itu, cluster ini dapat dikategorikan sebagai loyal customer, meskipun klasifikasi ini bersifat interpretatif berdasarkan pola RFM dan belum melalui validasi perilaku pelanggan secara longitudinal.

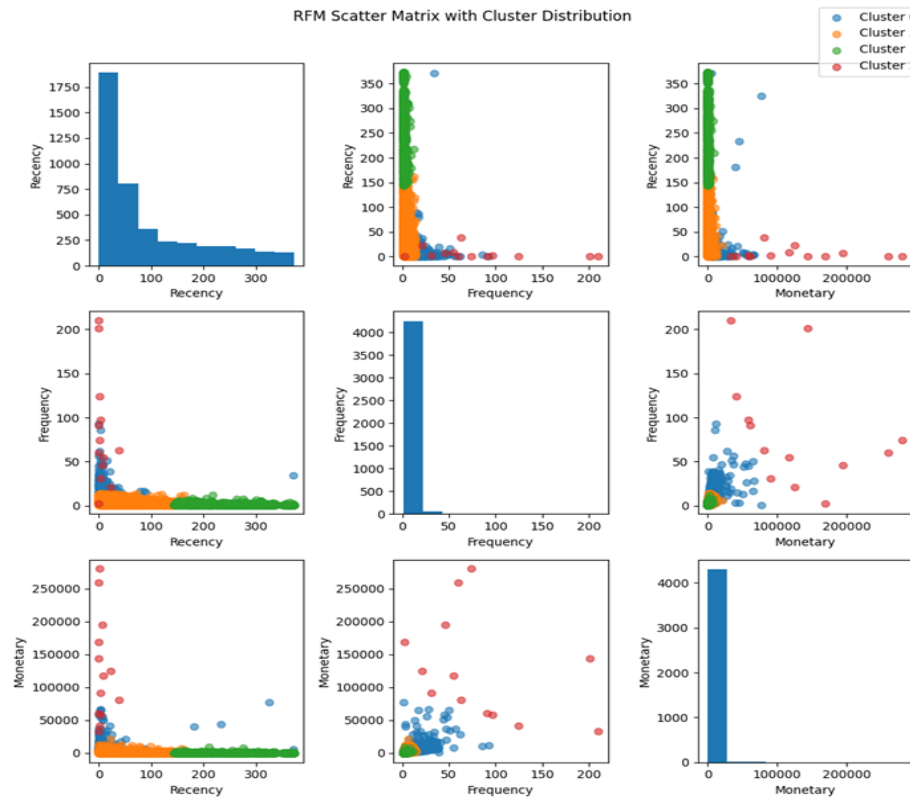
Cluster 3 - Potential Customers

Customer pada cluster ini memiliki nilai Recency sedang sebesar 42,91, nilai Frequency rendah-sedang sebesar 3.66, serta nilai Monetary sebesar 1,349.70. Hal ini menunjukkan bahwa pada cluster ini memiliki jumlah anggota terbesar namun kontribusi terhadap pendapatannya relatif moderat. Oleh karena itu, cluster ini dapat dikategorikan sebagai potential customer, dengan catatan bahwa penentuan potensi masih bersifat asumsi berbasis pola data saat ini.

Cluster 1 - Low-Value / Dormant Customers

Cluster 1 ini menunjukkan nilai Recency yang tinggi yaitu sebesar 247,56, nilai Frequency rendah sebesar 1.55, serta nilai Monetary rendah sebesar 478.11. Hal ini menunjukkan bahwa customer pada cluster ini sudah lama tidak melakukan transaksi dan memiliki kontribusi pendapatan yang sangat rendah. Oleh karena itu, cluster ini dapat dikategorikan sebagai dormant customer, yang mengindikasikan potensi churn berdasarkan indikator RFM.

Dilakukan visualisasi menggunakan RFM Scatter Matrix untuk menampilkan distribusi masing-masing cluster berdasarkan Recency, Frequency, dan Monetary. Visualisasi ini bertujuan untuk mendukung interpretasi hasil clustering, meskipun pendekatan ini merupakan teknik visualisasi standar yang tidak secara langsung menghasilkan wawasan baru tanpa analisis lanjutan.



Gambar 1 Visualisasi Distribusi Cluster Pelanggan Berdasarkan Dimensi RFM

Berdasarkan visualisasi diatas, memperlihatkan bahwa distribusi cluster menunjukkan pemisahan yang cukup jelas pada dimensi Monetary dan Frequency, namun masih terdapat potensi overlap antar cluster. Cluster 2 yang berwarna merah memiliki kecenderungan nilai Monetary dan Frequency yang lebih tinggi dibandingkan dengan cluster yang lain. Sementara itu, Cluster 1 yang berwarna hijau didominasi dengan nilai Recency yang tinggi dengan Frequency dan Monetary yang relatif rendah. Dan pada Cluster 0 dan Cluster 3 terlihat menunjukkan pola distribusi yang lebih moderat dengan variasi nilai yang lebih tersebar.

3.4 Pengembangan Model Knowledge-Based Decision Support

Berdasarkan hasil segmentasi pelanggan yang telah diperoleh pada tahap clustering sebelumnya, setiap cluster memiliki karakteristik perilaku transaksi yang berbeda. Perbedaan karakteristik tersebut kemudian digunakan sebagai dasar dalam penyusunan pendekatan pendukung keputusan berbasis aturan sederhana.

Proses implementasi dilakukan dengan merepresentasikan karakteristik nilai Recency, Frequency, dan Monetary (RFM) ke dalam bentuk aturan keputusan berbasis logika IF-THEN. Representasi ini bersifat konseptual dan digunakan untuk menghubungkan kondisi perilaku pelanggan dengan tindakan strategis yang direkomendasikan, tanpa melibatkan pengembangan sistem berbasis pengetahuan secara formal seperti mesin inferensi atau arsitektur DSS yang kompleks.

Aturan keputusan yang dihasilkan dalam penelitian ini dirumuskan sebagai berikut :

- IF *Recency* sangat rendah AND *Frequency* sangat tinggi AND *Monetary* sangat tinggi, THEN pelanggan dikategorikan sebagai *High Value/VIP*, dengan strategi *priority service* dan *exclusive promotion*.
- IF *Recency* rendah AND *Frequency* tinggi AND *Monetary* tinggi, THEN pelanggan dikategorikan sebagai *Loyal Customer*, dengan strategi *loyalty reward* dan *upselling*.
- IF *Recency* sedang AND *Frequency* rendah-sedang AND *Monetary* sedang, THEN pelanggan dikategorikan sebagai *Potential Customer*, dengan strategi *promotional campaign* dan *engagement program*.
- IF *Recency* tinggi AND *Frequency* rendah AND *Monetary* rendah, THEN pelanggan dikategorikan sebagai *Dormant Customer*, dengan strategi *re-activation* dan *special discount*.

Rangkuman hubungan antara karakteristik RFM, kategori pelanggan, dan strategi yang direkomendasikan disajikan pada Tabel 4.

Tabel 2 Basis Pengetahuan Model *Knowledge-Based Decision Support*

Cluster	Karakteristik RFM	Kategori Pelanggan	Rekomendasi Strategi
Cluster 2	Recency sangat rendah, Frequency sangat tinggi, Monetary sangat tinggi	High Value / VIP	Priority service, exclusive promotions

Cluster	Karakteristik RFM	Kategori Pelanggan	Rekomendasi Strategi
Cluster 0	Recency rendah, Frequency tinggi, Monetary tinggi	Loyal Customer	Loyalty reward, upselling
Cluster 3	Recency sedang, Frequency rendah-sedang, Monetary sedang	Potential Customer	Promotion campaign, engagement program
Cluster 1	Recency tinggi, Frequency rendah, Monetary rendah	Dormant Customer	Re-activation campaign, special discount

Implementasi pendekatan ini menunjukkan bahwa hasil segmentasi pelanggan dapat dimanfaatkan untuk menghasilkan rekomendasi strategi yang bersifat praktis. Namun, rekomendasi yang dihasilkan masih bersifat umum dan belum disesuaikan secara spesifik dengan konteks bisnis tertentu. Dengan demikian, integrasi antara data mining dan pendekatan berbasis aturan ini lebih tepat dipandang sebagai dukungan analitis untuk pengambilan keputusan, bukan sebagai pengembangan model DSS yang kompleks atau adaptif.

4. KESIMPULAN

Penelitian ini menerapkan segmentasi pelanggan menggunakan model RFM dan algoritma K-Means untuk mengelompokkan pelanggan berdasarkan perilaku transaksi. Hasil evaluasi menggunakan metode Elbow dan Silhouette Score menunjukkan bahwa jumlah cluster optimal adalah empat, dengan nilai Silhouette Score sebesar 0,61 yang mengindikasikan kualitas pemisahan yang cukup baik. Setiap cluster memiliki karakteristik berbeda yang kemudian diinterpretasikan menjadi kategori pelanggan seperti High Value, Loyal, Potential, dan Dormant. Hasil segmentasi ini selanjutnya dimanfaatkan untuk menyusun aturan keputusan berbasis logika IF–THEN sebagai bentuk dukungan analitis dalam pengambilan keputusan pemasaran. Dengan demikian, penelitian ini menunjukkan bahwa hasil clustering tidak hanya dapat digunakan untuk analisis deskriptif, tetapi juga dapat diinterpretasikan menjadi rekomendasi praktis, meskipun masih bersifat sederhana dan berbasis heuristik. Keterbatasan penelitian ini terletak pada penggunaan dataset publik yang relatif terstruktur serta pendekatan rekomendasi yang belum didukung oleh arsitektur sistem pendukung keputusan yang formal. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan data yang lebih kompleks dari lingkungan nyata, menambahkan metrik evaluasi clustering yang lebih beragam, serta mengembangkan sistem pendukung keputusan yang lebih komprehensif dan adaptif. Kontribusi penelitian ini adalah menyediakan pendekatan terintegrasi yang menghubungkan hasil segmentasi RFM-K-Means dengan aturan keputusan berbasis IF–THEN, sehingga menghasilkan rekomendasi strategi pemasaran yang lebih sistematis, mudah diinterpretasikan, dan dapat langsung digunakan sebagai dasar pengambilan keputusan berbasis data. Saran untuk penelitian selanjutnya adalah mengintegrasikan metode clustering dengan teknik machine learning lain seperti supervised learning untuk validasi hasil segmentasi, mengembangkan sistem pendukung keputusan berbasis dashboard interaktif, serta menguji pendekatan ini pada data real-time agar mampu memberikan rekomendasi yang lebih dinamis dan adaptif terhadap perubahan perilaku pelanggan.

REFERENCES

- [1] I. B. Ridwan, "Transforming Customer Segmentation with Unsupervised Learning Models and Behavioral Data in Digital Commerce," *Int. J. Res. Publ. Rev.*, vol. 6, no. 5, pp. 2232–2249, May 2025, doi: 10.55248/gengpi.6.0525.1652.
- [2] G. Wang, "Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering," *PLoS One*, vol. 20, no. 2 February, Feb. 2025, doi: 10.1371/journal.pone.0318519.
- [3] M. M. Triputra, A. Rifai, and K. D. Tania, "Evaluasi Kualitas Pendidikan Dasar di Sumatera Selatan Menggunakan Business Intelligence Model," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 601–613, Mar. 2025, doi: 10.52436/1.jpti.618.
- [4] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, "RFM model customer segmentation based on hierarchical approach using FCA," *Expert Syst. Appl.*, vol. 237, p. 121449, 2024, doi: <https://doi.org/10.1016/j.eswa.2023.121449>.
- [5] A. Clustering, E. T. Lakeisyah, K. D. Tania, and M. Afrina, "Komparasi Klasterisasi Data Historis Gempa Bumi Menggunakan DBSCAN, K-Means, dan Agglomerative Clustering," *Build. Informatics, Technol. Sci.*, vol. 7, no. 3, pp. 1674–1683, 2025, doi: <https://doi.org/10.47065/bits.v7i3.8426>.
- [6] M. D. Akhda and K. D. Tania, "Comparison of K-Means and K-Medoids Algorithms for Clustering Poverty Data in South Sumatra Using DBI Evaluation," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 15, no. 2, pp. 233–245, 2024, doi: 10.31849/digitalzone.v15i2.23624.
- [7] T. Z. F. Titus Zira Fate, D. S. E. Dogo Siyani Ezra, and I. I. S. Jandir Isaac Samuel, "Comparative Analysis of Clustering Techniques for Customer Segmentation: Evaluating K-Means, Hierarchical, and DBSCAN Models alongside RFM Frameworks to Enhance Marketing Strategies through Behavioral, Demographic, and Transactional Insights," *Int. J. Adv. Eng. Manag.*, vol. 7, no. 4, pp. 34–43, Apr. 2025, doi: 10.35629/5252-07043443.
- [8] O. N. Akande, H. B. 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," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, IEEE, 2024, pp. 1–8.
- [9] I. B. Ridwan, "Transforming Customer Segmentation with Unsupervised Learning Models and Behavioral Data in Digital Commerce," *Int. J. Res. Publ. Rev.*, vol. 6, no. 5, pp. 2232–2249, May 2025, doi: 10.55248/gengpi.6.0525.1652.
- [10] B. G. Vo, H. D. S. Van, N. D. Van, and H. D. Huynh, "Customer Segmentation: Automatic K-Optimization and RFM-Based K-Means Clustering," *Association for Computing Machinery (ACM)*, Feb. 2025, pp. 173–178. doi:



- 10.1145/3731763.3731805.
- [11] M. S. E. Kasem, M. Hamada, and I. Taj-Eddin, “Customer profiling, segmentation, and sales prediction using AI in direct marketing,” *Neural Comput. Appl.*, vol. 36, no. 9, pp. 4995–5005, Mar. 2024, doi: 10.1007/s00521-023-09339-6.
- [12] M. Sarkar, A. R. Puja, and F. R. Chowdhury, “Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis,” *J. Bus. Manag. Stud.*, vol. 6, no. 2, pp. 54–60, Mar. 2024, doi: 10.32996/jbms.2024.6.2.5.
- [13] J. Chitra and J. Heikal, “Customer segmentation using the K-Means Clustering algorithm in Foreign Banks in Indonesia,” *Indones. Account. Res. J.*, vol. 11, no. 4, pp. 230–241, 2024.
- [14] R. M. Fauzan and G. Alfian, “Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma Klasterisasi K-Means,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 170–177, 2024, doi: 10.14421/jiska.2024.9.3.170-177.
- [15] D. A. Imanuel and G. Alfian, “Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 283–292, 2025, doi: 10.25126/jtiik.2025128619.
- [16] A. T. Widiyanto and A. Witanti, “Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global),” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.
- [17] A. A. Rahma, A. Faqih, and A. R. Rinaldi, “Optimalisasi Strategi Pemasaran melalui Segmentasi Pelanggan dengan Analisis RFM dan Algoritma K-Means untuk Bisnis Ritel,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 2, p. 338, 2025, doi: 10.26798/jiko.v9i2.1737.
- [18] S. E. Saqila, I. P. Ferina, and A. Iskandar, “Analisis Perbandingan Kinerja Clustering Data Mining Untuk Normalisasi Dataset,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 356, 2023, doi: 10.30865/json.v5i2.6919.
- [19] M. Guntara and N. Lutfi, “Optimasi Cacah Klaster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coefficient dan Elbow Method,” *JuTI “Jurnal Teknol. Informasi,”* vol. 2, no. 1, p. 43, 2023, doi: 10.26798/juti.v2i1.944.
- [20] S. A. Perdana, S. F. Florentin, and A. Santoso, “Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift,” *Sebatik*, vol. 26, no. 2, pp. 446–457, Dec. 2022, doi: 10.46984/sebatik.v26i2.1991.
- [21] I. H. Witten and E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations,” *ACM SIGMOD Rec.*, vol. 31, no. 1, pp. 76–77, Mar. 2002, doi: 10.1145/507338.507355.