

Customer Segmentation Menggunakan K-Means dan Analisis RFM Pada Data Transaksi Digital

Disa Putri Yuliansyah^{*1}, Rezi Azhar Hadi Husain²

^{1,2} STMIK Widya Cipta Dharma (WICIDA), Indonesia

Corresponding e-mail: disapy@gmail.com

Received: 20-08-2026

Accepted: 03-09-2026

Published: 05-09-2026

Abstrak

Pertumbuhan transaksi digital menghasilkan volume data pelanggan yang semakin besar, tetapi data tersebut belum selalu dimanfaatkan secara optimal untuk memahami karakteristik dan nilai pelanggan. Penelitian ini bertujuan melakukan segmentasi pelanggan pada data transaksi digital menggunakan analisis Recency, Frequency, Monetary (RFM) yang dipadukan dengan algoritma K-Means. Tahapan penelitian meliputi praproses data, agregasi transaksi berdasarkan pelanggan, pembentukan atribut RFM, normalisasi data, penentuan jumlah cluster menggunakan metode Elbow, penerapan K-Means, serta interpretasi karakteristik setiap segmen. Hasil analisis menunjukkan bahwa ($k=4$) merupakan jumlah cluster optimal berdasarkan metode Elbow. Segmentasi menghasilkan empat kelompok pelanggan, yaitu Champions, Loyal Customers, Potential Customers, dan At Risk, dengan karakteristik Recency, Frequency, dan Monetary yang berbeda. Champions memiliki Recency terendah sebesar 9,6 hari, Frequency tertinggi sebesar 18,6 transaksi, dan Monetary tertinggi sebesar Rp2.286.104. Sebaliknya, At Risk memiliki Recency tertinggi sebesar 158,7 hari dengan Frequency 2,5 transaksi dan Monetary Rp221.955. Hasil segmentasi dapat menjadi dasar penyusunan strategi pemasaran yang lebih personal dan sesuai dengan karakteristik setiap kelompok pelanggan.

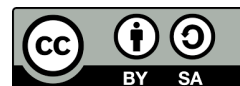
Kata Kunci: *Customer Segmentation; RFM Analysis; K-Means Clustering; Data Transaksi Digital; Elbow Method*

How to Cite:

Yuliansyah, D. P., & Husain, R. A. H. (2026). Customer Segmentation Menggunakan K-Means dan Analisis RFM Pada Data Transaksi Digital. *JINTEN: Journal of Intelligent Systems and Digital Science*, 1(1), 79-94.

Copyright ©2026 to the Author. Published by CV. Ihsan Cahaya Pustaka

This is an open access article under the [CC-BY-SA](#) license



1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan signifikan dalam pola transaksi dan perilaku pembelian pelanggan. Platform e-commerce, marketplace, serta berbagai layanan transaksi digital menghasilkan data dalam jumlah besar yang merekam aktivitas pelanggan secara lebih rinci, mulai dari waktu transaksi, frekuensi pembelian, hingga nilai pengeluaran. Data tersebut tidak hanya berfungsi sebagai catatan operasional, tetapi juga memiliki nilai strategis untuk memahami perilaku dan nilai pelanggan. Namun, ketersediaan data transaksi yang semakin besar tidak secara otomatis menghasilkan keputusan pemasaran yang lebih baik. Tanpa metode analisis yang tepat, informasi yang terkandung dalam data transaksi sulit diterjemahkan menjadi karakteristik pelanggan yang dapat digunakan untuk menentukan strategi pemasaran secara spesifik. Antonius dan Fitriana (2024) menunjukkan bahwa pemanfaatan data transaksi melalui pendekatan segmentasi dapat memberikan wawasan yang lebih terarah mengenai karakteristik pelanggan dan mendukung penyusunan strategi pemasaran berbasis data.

Customer segmentation menjadi salah satu pendekatan penting untuk mengatasi permasalahan tersebut dengan mengelompokkan pelanggan berdasarkan kemiripan karakteristik dan perilaku. Segmentasi memungkinkan perusahaan menghindari pendekatan pemasaran yang seragam karena setiap kelompok pelanggan dapat memiliki tingkat aktivitas, nilai transaksi, dan kebutuhan yang berbeda. Dalam konteks e-commerce, Gomes dan Meisen (2023) melalui kajian terhadap berbagai metode segmentasi pelanggan menunjukkan bahwa pendekatan berbasis data memiliki peran penting dalam mendukung personalized customer targeting. Tabianan et al. (2022) juga menunjukkan bahwa teknik clustering dapat digunakan untuk mengidentifikasi kelompok pelanggan berdasarkan pola perilaku pembelian sehingga hasil segmentasi dapat dikaitkan dengan kebutuhan pengambilan keputusan bisnis. Dengan demikian, segmentasi pelanggan tidak hanya diposisikan sebagai proses pengelompokan secara statistik, tetapi sebagai tahapan analitis untuk menghasilkan informasi yang relevan bagi pengelolaan hubungan dan strategi pemasaran pelanggan.

Salah satu pendekatan yang banyak digunakan untuk merepresentasikan perilaku dan nilai pelanggan adalah analisis Recency, Frequency, dan Monetary (RFM). Recency mengukur jarak waktu sejak transaksi terakhir pelanggan, Frequency menunjukkan intensitas atau jumlah transaksi yang dilakukan, sedangkan Monetary merepresentasikan nilai transaksi yang dihasilkan pelanggan dalam periode pengamatan. Ketiga indikator tersebut memberikan gambaran yang relatif ringkas mengenai aktivitas dan nilai ekonomis pelanggan. Barus et al. (2023) menerapkan RFM untuk melakukan customer profiling dan menunjukkan bahwa kombinasi ketiga atribut tersebut dapat digunakan sebagai dasar pengelompokan pelanggan menggunakan K-Means. Anitha dan Patil (2022) juga memperlihatkan bahwa model RFM dapat digunakan untuk menganalisis perilaku pembelian pelanggan melalui pendekatan clustering. Meskipun demikian, penggunaan RFM tetap memiliki keterbatasan karena hanya merepresentasikan tiga dimensi utama perilaku transaksi. Smaili dan Hachimi (2023), misalnya, mengembangkan model RFM-D dengan menambahkan dimensi diversity untuk memperluas informasi yang digunakan dalam analisis pelanggan. Hal tersebut menunjukkan bahwa RFM merupakan representasi yang praktis, tetapi pemilihan atribut tetap perlu disesuaikan dengan tujuan segmentasi dan karakteristik data yang tersedia.

Setelah karakteristik pelanggan direpresentasikan dalam bentuk RFM, diperlukan metode yang mampu mengelompokkan pelanggan berdasarkan kemiripan ketiga atribut tersebut. K-Means merupakan salah satu algoritma unsupervised learning yang banyak digunakan untuk tujuan tersebut karena bekerja dengan mengelompokkan data berdasarkan kedekatannya terhadap centroid setiap cluster. Penerapan K-Means pada data pelanggan memungkinkan individu dengan pola RFM yang relatif serupa ditempatkan pada kelompok yang sama. Aslantaş et al. (2023) menunjukkan penerapan K-Means pada model RFM untuk membentuk kelompok pelanggan berdasarkan karakteristik perilaku transaksi. Khumaidi et al. (2023) juga menerapkan K-Means pada data RFM dan

menunjukkan bahwa algoritma tersebut dapat digunakan untuk menghasilkan segmentasi pelanggan yang selanjutnya dapat diinterpretasikan berdasarkan karakteristik setiap cluster. Pendekatan serupa juga diterapkan pada konteks e-commerce oleh Fauzan dan Alfian (2024), yang menunjukkan bahwa RFM dapat dikombinasikan dengan metode clustering untuk menghasilkan segmentasi pelanggan yang dapat dibandingkan berdasarkan kualitas cluster.

Meskipun K-Means relatif sederhana dan mudah diinterpretasikan, kualitas hasil segmentasi sangat dipengaruhi oleh pemilihan jumlah cluster yang tepat. Pemilihan nilai k yang tidak sesuai dapat menghasilkan kelompok yang terlalu umum atau justru terlalu terfragmentasi sehingga mengurangi makna analitis dari segmentasi. Oleh karena itu, penentuan jumlah cluster perlu dilakukan secara sistematis sebelum model K-Means diterapkan pada tahap akhir. Alamsyah et al. (2022) mengintegrasikan model RFM dan K-Means serta menggunakan metode Elbow untuk membantu menentukan jumlah cluster yang sesuai. Pendekatan serupa digunakan oleh Kiat et al. (2024), yang menerapkan metode Elbow pada segmentasi pelanggan berbasis RFM. Lalitha et al. (2025) juga menggunakan kombinasi analisis RFM, Elbow, dan K-Means dalam proses segmentasi pelanggan. Temuan tersebut memperkuat penggunaan Elbow sebagai salah satu pendekatan yang relevan untuk mengidentifikasi kandidat jumlah cluster berdasarkan perubahan nilai within-cluster variation.

Namun, penggunaan Elbow secara tunggal belum memberikan gambaran yang cukup mengenai kualitas pemisahan antarcluster. Oleh sebab itu, evaluasi kualitas cluster perlu dilengkapi dengan ukuran validitas internal seperti Silhouette Score. Shutaywi dan Kachouie (2021) menjelaskan bahwa Silhouette Analysis dapat digunakan untuk mengevaluasi seberapa baik suatu objek berada dalam cluster-nya dibandingkan dengan cluster lain. Dalam konteks customer segmentation, Yuliasih et al. (2024) secara khusus menerapkan Silhouette Score untuk mengevaluasi hasil K-Means dan menunjukkan pentingnya pengukuran kualitas cluster dalam menentukan hasil segmentasi. Kumar (2025) juga menggunakan evaluasi cluster dalam penelitian segmentasi pelanggan berbasis machine learning. Dengan demikian, penggunaan Elbow untuk menentukan kandidat jumlah cluster yang kemudian diperiksa menggunakan Silhouette Score dapat memberikan dasar evaluasi yang lebih kuat dibandingkan hanya mengandalkan visualisasi Elbow.

Perkembangan penelitian menunjukkan bahwa kombinasi RFM dan clustering telah diterapkan pada berbagai konteks serta dikembangkan dengan pendekatan yang semakin beragam. Faran et al. (2023), misalnya, mengombinasikan RFM dengan metode agglomerative Ward untuk segmentasi donor, sedangkan Rungruang et al. (2024) mengembangkan pendekatan segmentasi berbasis RFM dengan hierarchical approach. Antonius dan Fitriana (2024) memperluas analisis RFM dengan memasukkan proporsi diskon dalam proses segmentasi, sementara Kasem et al. (2024) mengintegrasikan customer profiling, segmentation, dan sales prediction menggunakan pendekatan

artificial intelligence. Penelitian terbaru juga menunjukkan integrasi RFM dengan machine learning untuk customer value segmentation pada lingkungan e-commerce (Pai et al., 2026). Keragaman pendekatan tersebut menunjukkan bahwa segmentasi pelanggan berkembang dari sekadar pengelompokan berdasarkan nilai transaksi menuju analisis yang semakin berorientasi pada pemanfaatan data untuk mendukung keputusan pemasaran.

Meskipun demikian, penggunaan metode yang lebih kompleks tidak selalu diperlukan untuk menghasilkan segmentasi yang dapat dipahami dan dimanfaatkan oleh pengambil keputusan. Penelitian ini memfokuskan perhatian pada pengembangan alur segmentasi yang sistematis dan relatif mudah direplikasi, dengan mengintegrasikan analisis RFM, normalisasi data, penentuan jumlah cluster menggunakan Elbow, evaluasi kualitas cluster menggunakan Silhouette Score, serta algoritma K-Means. Fokus tersebut penting karena hasil clustering perlu diterjemahkan kembali menjadi karakteristik pelanggan yang memiliki makna bisnis. Syahra et al. (2025) menunjukkan bahwa segmentasi berbasis RFM dan K-Means dapat digunakan untuk mendukung customer relationship management (CRM), sedangkan Sarkar et al. (2024) menempatkan hasil segmentasi sebagai dasar untuk menyusun strategi pemasaran yang lebih terarah. Dengan demikian, kualitas penelitian tidak hanya ditentukan oleh keberhasilan algoritma membentuk cluster, tetapi juga oleh kemampuan menginterpretasikan karakteristik setiap segmen dan menghubungkannya dengan tindakan pemasaran yang relevan.

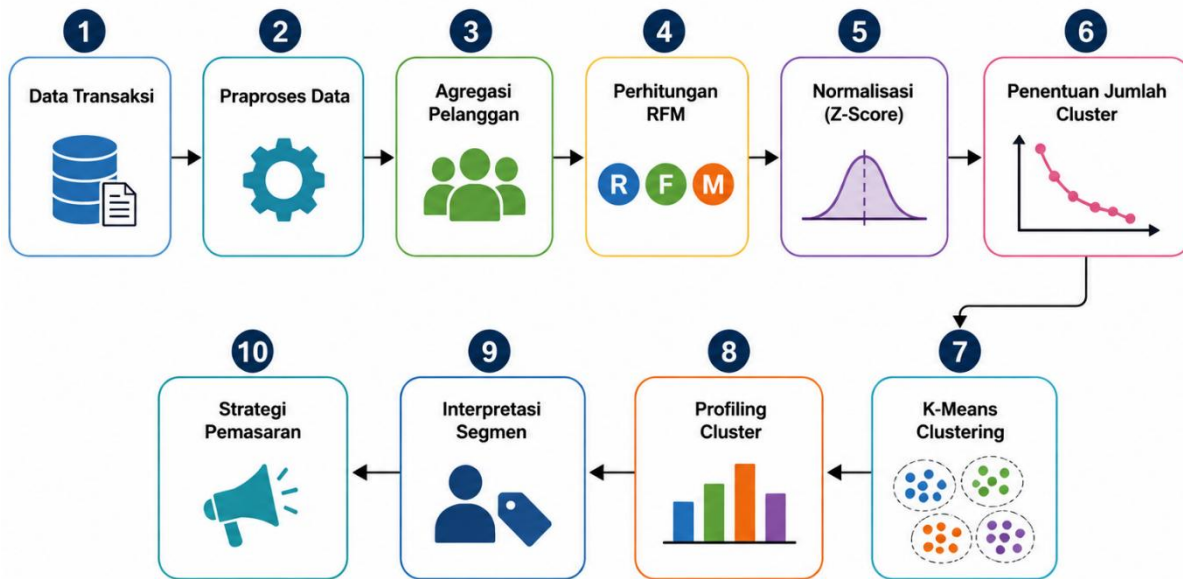
Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan analisis Recency, Frequency, dan Monetary (RFM) yang dipadukan dengan algoritma K-Means untuk melakukan segmentasi pelanggan pada data transaksi digital. Penelitian dilakukan melalui tahapan pengolahan data transaksi, pembentukan atribut RFM, normalisasi data, penentuan kandidat jumlah cluster menggunakan metode Elbow, evaluasi kualitas cluster menggunakan Silhouette Score, penerapan K-Means, serta interpretasi karakteristik masing-masing segmen pelanggan. Hasil segmentasi selanjutnya digunakan untuk mengidentifikasi perbedaan karakteristik pelanggan dan merumuskan implikasi strategi pemasaran yang sesuai dengan setiap kelompok. Penelitian ini diharapkan dapat memberikan alur analisis yang sistematis dan aplikatif bagi pemanfaatan data transaksi digital, sekaligus memberikan dasar empiris bagi penerapan strategi pemasaran dan pengelolaan hubungan pelanggan yang lebih terarah.

2. METODE

2.1 Desain dan Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan segmentasi pelanggan berdasarkan pola transaksi. Pendekatan penelitian mengintegrasikan analisis Recency, Frequency, Monetary (RFM) dengan algoritma K-Means untuk

mengelompokkan pelanggan berdasarkan kemiripan karakteristik transaksi. Tahapan penelitian meliputi praproses data, pembentukan data pelanggan, ekstraksi atribut RFM, normalisasi, penentuan jumlah cluster menggunakan metode Elbow, evaluasi kualitas cluster menggunakan Silhouette Score, penerapan K-Means, profiling cluster, serta interpretasi segmen dan implikasi strategi pemasaran. Alur penelitian tersebut dirancang agar proses segmentasi tidak hanya menghasilkan kelompok pelanggan secara komputasional, tetapi juga menghasilkan karakteristik pelanggan yang dapat digunakan untuk mendukung pengambilan keputusan pemasaran.



Gambar 1. Tahapan Penelitian

2.2 Data dan Pembentukan Atribut RFM

Data penelitian berupa data transaksi digital yang memuat informasi identitas pelanggan, tanggal transaksi, dan nilai transaksi. Praproses dilakukan untuk memastikan konsistensi identitas pelanggan, format tanggal, dan nilai transaksi serta menangani data yang tidak sesuai atau tidak memiliki informasi yang diperlukan dalam pembentukan RFM. Transaksi kemudian diagregasi berdasarkan identitas pelanggan sehingga setiap pelanggan direpresentasikan sebagai satu observasi. Selanjutnya, analisis RFM digunakan untuk membentuk tiga atribut utama, yaitu Recency, Frequency, dan Monetary. Recency menunjukkan jarak waktu sejak transaksi terakhir, Frequency menunjukkan jumlah transaksi, sedangkan Monetary menunjukkan total nilai transaksi pelanggan. Perhitungan dilakukan menggunakan:

$$[R_i = T_{obs} - T_{last,i}] \quad (1)$$

$$[F_i = N_i] \quad (2)$$

$$[M_i = \sum_{j=1}^N V_{ij}] \quad (3)$$

dengan (T_{obs}) merupakan tanggal observasi, $(T_{last,i})$ tanggal transaksi terakhir pelanggan ke-(i), (N_i) jumlah transaksi, dan (V_{ij}) nilai transaksi ke-(j). Ketiga atribut tersebut membentuk matriks RFM berukuran $(n \times 3)$, dengan (n) merupakan jumlah pelanggan unik. Pendekatan ini sejalan dengan penelitian Anitha dan Patil (2022), Barus et al. (2023), dan Aslantaş et al. (2023) yang menggunakan RFM sebagai dasar pembentukan profil pelanggan sebelum proses clustering.

2.3 Normalisasi Data

Atribut RFM memiliki satuan dan rentang nilai yang berbeda sehingga dilakukan standardisasi menggunakan *z-score*. Tahap ini bertujuan agar atribut dengan rentang nilai besar, terutama Monetary, tidak mendominasi perhitungan jarak pada K-Means. Standardisasi dilakukan dengan persamaan:

$$\left[Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \right] \quad (4)$$

dengan (X_{ij}) merupakan nilai atribut ke-(j) pada pelanggan ke-(i), (μ_j) merupakan rata-rata atribut ke-(j), dan (σ_j) merupakan standar deviasi. Hasil standardisasi digunakan sebagai input pada proses penentuan jumlah cluster dan K-Means sehingga ketiga atribut RFM memiliki kontribusi yang lebih seimbang dalam proses pengelompokan.

2.4 Penentuan dan Evaluasi Jumlah Cluster

Penentuan jumlah cluster dilakukan menggunakan metode Elbow dan Silhouette Score. Elbow digunakan untuk mengidentifikasi kandidat jumlah cluster berdasarkan perubahan nilai *Within-Cluster Sum of Squares* (WCSS), yang menunjukkan total variasi data dalam setiap cluster. Nilai WCSS dihitung dengan:

$$[WCSS = k = 1^K x_i C_k |x_i - \mu_k|^2] \quad (5)$$

dengan (C_k) merupakan cluster ke-(k), (x_i) data pelanggan, dan (μ_k) centroid cluster ke-(k). Kandidat jumlah cluster ditentukan pada titik ketika penurunan WCSS mulai melandai atau membentuk *elbow point*. Pendekatan ini digunakan dalam penelitian Alamsyah et al. (2022), Kiat et al. (2024), dan Lalitha et al. (2025). Kandidat tersebut kemudian dievaluasi menggunakan Silhouette Score untuk mengukur kekompakan data dalam cluster dan keterpisahannya dari cluster lain. Nilai Silhouette berada pada rentang -1 hingga 1, dengan nilai yang semakin mendekati 1 menunjukkan kualitas pengelompokan yang semakin baik. Penggunaan Silhouette Score dalam evaluasi customer segmentation berbasis K-Means juga diterapkan oleh Yulisasih et al. (2024) dan Kumar (2025). Jumlah cluster akhir ditetapkan berdasarkan kombinasi hasil Elbow dan Silhouette Score sehingga nilai (k) ditentukan berdasarkan karakteristik dan kualitas pengelompokan data.

2.5 Penerapan K-Means Clustering

K-Means diterapkan pada data RFM yang telah dinormalisasi berdasarkan jumlah cluster yang diperoleh dari tahap evaluasi. Algoritma bekerja dengan menentukan centroid awal, menghitung jarak setiap pelanggan terhadap centroid, menetapkan pelanggan pada cluster dengan jarak terdekat, kemudian memperbarui centroid berdasarkan rata-rata anggota cluster secara iteratif hingga kondisi konvergensi tercapai. Jarak Euclidean digunakan dengan persamaan:

$$d(x_i, \mu_k) = \sqrt{\sum_{j=1}^p (x_{ij} - \mu_{kj})^2} \quad (6)$$

dengan (x_i) merupakan pelanggan ke-(i), (μ_k) centroid cluster ke-(k), dan $(p=3)$ yang merepresentasikan atribut Recency, Frequency, dan Monetary. K-Means dipilih karena dapat melakukan pengelompokan tanpa label kelas sebelumnya dan telah banyak digunakan dalam segmentasi pelanggan berbasis RFM (Barus et al., 2023; Khumaidi et al., 2023; Fauzan & Alfian, 2024; Syahra et al., 2025).

2.6 Profiling dan Interpretasi Cluster

Hasil clustering dianalisis berdasarkan nilai rata-rata Recency, Frequency, dan Monetary serta jumlah pelanggan pada setiap cluster. Karakteristik tersebut digunakan untuk memberikan label dan interpretasi terhadap masing-masing segmen setelah proses clustering selesai. Label seperti *Champions*, *Loyal Customers*, *Potential Customers*, dan *At Risk* ditentukan berdasarkan pola RFM yang terbentuk, bukan ditetapkan sebelum proses clustering. Cluster dengan Recency rendah, Frequency tinggi, dan Monetary tinggi dapat diinterpretasikan sebagai pelanggan aktif dan bernilai tinggi, sedangkan cluster dengan Recency tinggi serta Frequency dan Monetary rendah menunjukkan pelanggan dengan aktivitas transaksi yang menurun dan berpotensi memerlukan strategi retensi.

2.7 Implikasi Strategi Pemasaran

Karakteristik setiap cluster selanjutnya diterjemahkan menjadi implikasi strategi pemasaran yang disesuaikan dengan perilaku pelanggan. Segmen dengan aktivitas dan nilai transaksi tinggi dapat diarahkan pada program loyalitas dan penawaran eksklusif, segmen dengan aktivitas menengah dapat dikembangkan melalui cross-selling, bundling, atau promosi, sedangkan segmen dengan Recency tinggi dan aktivitas transaksi rendah dapat menjadi sasaran strategi retensi dan *win-back*. Pendekatan tersebut menjadikan hasil clustering tidak berhenti pada pengelompokan pelanggan, tetapi dapat digunakan sebagai dasar pengambilan keputusan pemasaran dan customer relationship management, sebagaimana diterapkan dalam penelitian Sarkar et al. (2024) dan Syahra et al. (2025).

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengolahan dan Pembentukan Data RFM

Tahap awal analisis dilakukan dengan mengubah data transaksi menjadi data pelanggan berdasarkan tiga atribut utama, yaitu Recency, Frequency, dan Monetary (RFM). Setiap pelanggan direpresentasikan dalam satu baris data berdasarkan transaksi yang dilakukan selama periode pengamatan. Recency menunjukkan jarak waktu sejak transaksi terakhir, Frequency menunjukkan jumlah transaksi yang dilakukan, sedangkan Monetary menunjukkan total nilai transaksi yang dihasilkan oleh pelanggan.

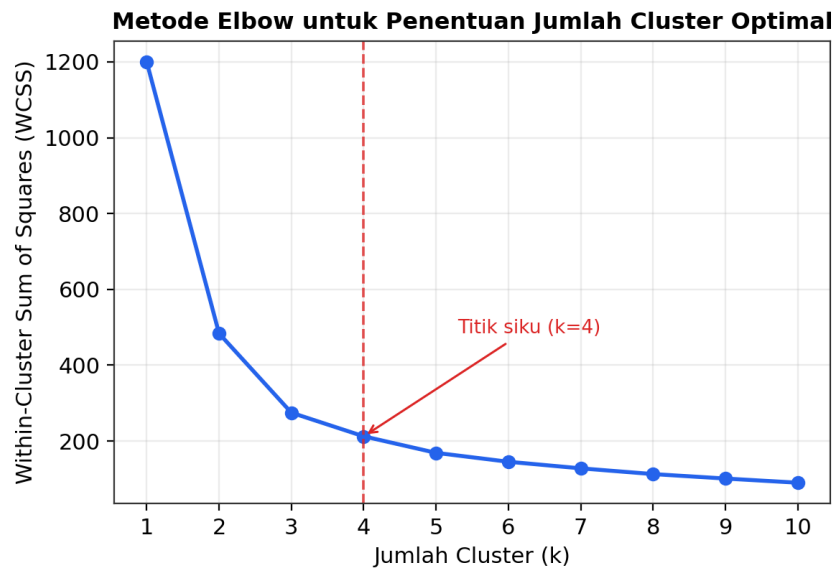
Pembentukan atribut RFM menghasilkan representasi perilaku pelanggan yang selanjutnya digunakan sebagai input algoritma clustering. Penggunaan ketiga atribut tersebut memungkinkan pelanggan dibandingkan berdasarkan tingkat aktivitas dan nilai ekonominya secara bersamaan. Pendekatan ini sejalan dengan Anitha dan Patil (2022) yang menggunakan model RFM untuk merepresentasikan perilaku pembelian pelanggan sebelum dilakukan proses clustering menggunakan K-Means. Barus et al. (2023) juga menunjukkan bahwa kombinasi RFM dapat digunakan untuk membentuk profil pelanggan berdasarkan karakteristik transaksi.

Karena rentang nilai Recency, Frequency, dan Monetary berbeda, data RFM kemudian distandardisasi menggunakan metode *z-score*. Standardisasi dilakukan agar perbedaan skala antaratribut tidak menyebabkan salah satu variabel memberikan pengaruh yang terlalu dominan terhadap perhitungan jarak Euclidean. Tahapan ini menjadi penting karena Monetary memiliki nilai numerik yang jauh lebih besar dibandingkan Recency dan Frequency.

3.2 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster dilakukan menggunakan metode Elbow dengan menguji beberapa kandidat nilai (k). Metode ini digunakan untuk mengamati perubahan nilai *Within-Cluster Sum of Squares* (WCSS) ketika jumlah cluster ditingkatkan. Semakin besar nilai (k), nilai WCSS cenderung semakin kecil karena data dibagi ke dalam kelompok yang lebih banyak.

Hasil pengujian menunjukkan adanya penurunan WCSS yang relatif tajam hingga ($k=4$), kemudian penurunannya mulai melandai pada nilai (k) yang lebih besar. Pola tersebut menunjukkan bahwa ($k=4$) merupakan kandidat jumlah cluster yang memberikan keseimbangan antara jumlah kelompok dan tingkat variasi dalam cluster.

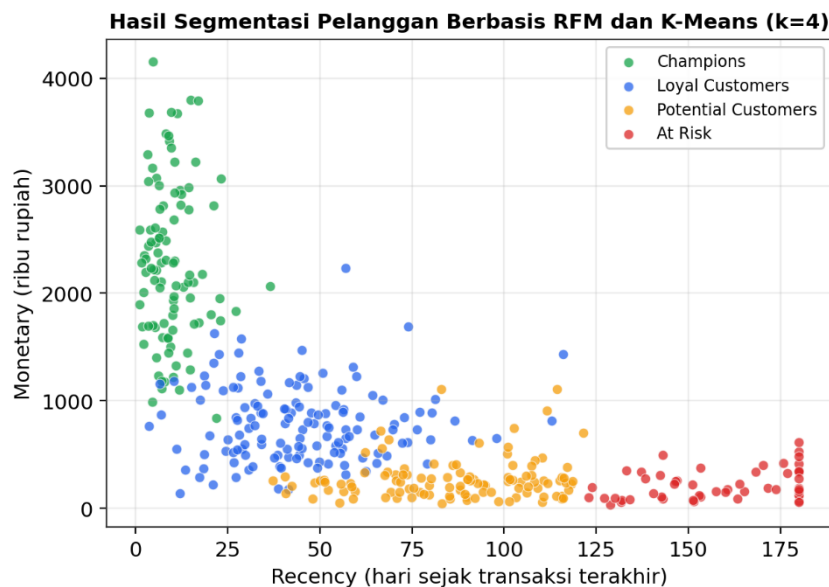


Gambar 2. Kurva Metode Elbow untuk Penentuan Jumlah Cluster Optimal

Berdasarkan pola kurva tersebut, ($k=4$) dipilih sebagai kandidat jumlah cluster untuk tahap evaluasi berikutnya. Hasil ini sejalan dengan penelitian Alamsyah et al. (2022), Kiat et al. (2024), dan Lalitha et al. (2025), yang menggunakan Elbow sebagai salah satu pendekatan untuk menentukan jumlah cluster pada segmentasi pelanggan berbasis RFM dan K-Means. Namun, Elbow hanya menunjukkan perubahan tingkat variasi dalam cluster dan belum secara langsung mengukur kualitas pemisahan antarcluster. Oleh karena itu, penelitian ini melanjutkan evaluasi menggunakan Silhouette Score.

3.3 Hasil Clustering Pelanggan

Setelah jumlah cluster ditentukan, algoritma K-Means diterapkan pada data RFM yang telah distandardisasi. Berdasarkan hasil clustering dengan ($k=4$), pelanggan terbagi menjadi empat kelompok yang kemudian diinterpretasikan berdasarkan karakteristik rata-rata Recency, Frequency, dan Monetary. Visualisasi hasil clustering dapat digunakan untuk memperlihatkan pola pemisahan pelanggan berdasarkan karakteristik RFM. Pada visualisasi dua dimensi, Recency dapat ditempatkan pada sumbu horizontal dan Monetary pada sumbu vertikal, sedangkan warna atau simbol yang berbeda digunakan untuk membedakan cluster.



Gambar 2. Sebaran Hasil Segmentasi Pelanggan Berdasarkan Recency dan Monetary

Visualisasi menunjukkan adanya perbedaan pola antara kelompok pelanggan dengan tingkat aktivitas dan nilai transaksi yang tinggi dan kelompok pelanggan dengan aktivitas yang relatif rendah. Segmen dengan Recency rendah dan Monetary tinggi menunjukkan pelanggan yang relatif aktif dan memberikan nilai transaksi lebih besar. Sebaliknya, kelompok dengan Recency tinggi dan Monetary rendah menunjukkan pelanggan yang aktivitas transaksinya lebih rendah. Hasil tersebut menunjukkan bahwa kombinasi RFM dapat memberikan dasar yang cukup informatif untuk membedakan pelanggan berdasarkan perilaku transaksi. Temuan ini sejalan dengan Aslantaş et al. (2023) dan Fauzan dan Alfian (2024), yang menunjukkan bahwa integrasi RFM dan K-Means dapat menghasilkan kelompok pelanggan dengan karakteristik transaksi yang berbeda.

3.4 Karakteristik dan Distribusi Setiap Segmen

Karakteristik masing-masing cluster selanjutnya dianalisis berdasarkan nilai rata-rata RFM dan jumlah pelanggan. Hasil analisis pada data penelitian ditampilkan pada Tabel 1.

Tabel 1. Karakteristik Rata-Rata RFM pada Setiap Segmen Pelanggan

Segmen	Recency (rata-rata hari)	Frequency (rata-rata transaksi)	Monetary (rata-rata, Rp)	Jumlah Pelanggan
Champions	9,6	18,6	2.286.104	97
Loyal Customers	44,7	8,3	789.532	138
Potential Customers	86,7	2,5	274.708	111
At Risk	158,7	2,5	221.955	54

Berdasarkan Tabel 1, segmen Champions memiliki karakteristik paling tinggi dari sisi Frequency dan Monetary serta memiliki nilai Recency paling rendah. Nilai Recency sebesar 9,6 hari

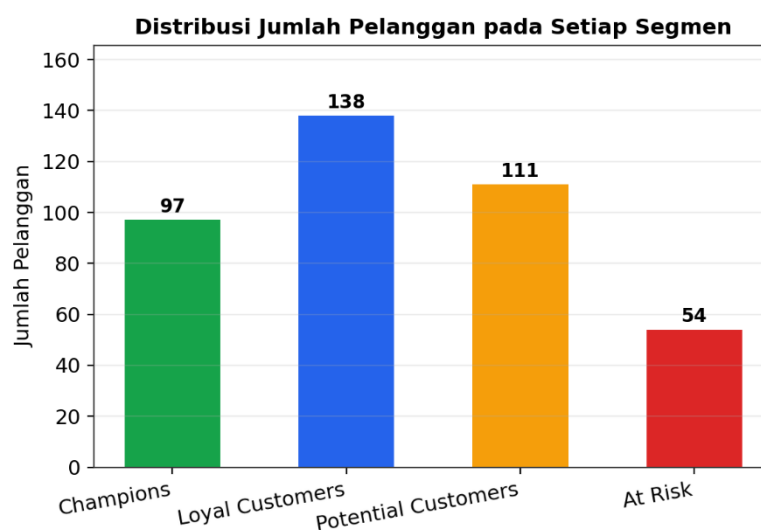
menunjukkan bahwa pelanggan dalam kelompok ini relatif baru melakukan transaksi, sedangkan Frequency sebesar 18,6 transaksi dan Monetary sebesar Rp2.286.104 menunjukkan tingkat aktivitas dan nilai transaksi yang relatif tinggi. Oleh karena itu, segmen ini dapat dipandang sebagai kelompok pelanggan bernilai tinggi yang perlu dipertahankan.

Segmen Loyal Customers memiliki rata-rata Recency sebesar 44,7 hari, Frequency sebesar 8,3 transaksi, dan Monetary sebesar Rp789.532. Dibandingkan Champions, kelompok ini menunjukkan tingkat aktivitas dan nilai transaksi yang lebih rendah, tetapi masih memiliki frekuensi transaksi yang relatif tinggi. Karakteristik tersebut menunjukkan adanya hubungan transaksi yang cukup konsisten sehingga strategi yang diarahkan pada peningkatan nilai pembelian dan loyalitas dapat dipertimbangkan.

Segmen Potential Customers memiliki Recency sebesar 86,7 hari dengan Frequency sebesar 2,5 transaksi dan Monetary sebesar Rp274.708. Nilai tersebut menunjukkan bahwa pelanggan dalam kelompok ini memiliki aktivitas transaksi yang relatif rendah dibandingkan dua segmen sebelumnya. Meskipun demikian, keberadaan transaksi sebelumnya menunjukkan adanya potensi untuk meningkatkan keterlibatan pelanggan melalui pendekatan promosi yang sesuai.

Segmen At Risk memiliki nilai Recency tertinggi, yaitu 158,7 hari, serta Frequency dan Monetary yang relatif rendah, masing-masing sebesar 2,5 transaksi dan Rp221.955. Kombinasi tersebut menunjukkan bahwa pelanggan dalam segmen ini telah lama tidak melakukan transaksi dan menghasilkan nilai transaksi yang relatif rendah. Oleh karena itu, kelompok ini dapat menjadi prioritas dalam strategi retensi dan *win-back*.

Distribusi pelanggan pada masing-masing segmen menunjukkan bahwa Loyal Customers merupakan kelompok terbesar dengan 138 pelanggan, diikuti Potential Customers sebanyak 111 pelanggan, Champions sebanyak 97 pelanggan, dan At Risk sebanyak 54 pelanggan.



Gambar 3. Distribusi Jumlah Pelanggan pada Setiap Segmen

Komposisi tersebut menunjukkan bahwa sebagian besar pelanggan berada pada kelompok dengan tingkat aktivitas menengah, sedangkan kelompok pelanggan bernilai tinggi memiliki jumlah yang lebih kecil. Kondisi tersebut memperlihatkan bahwa strategi pemasaran tidak sebaiknya diterapkan secara seragam karena setiap segmen memiliki karakteristik yang berbeda.

3.5 Pembahasan dan Implikasi Strategi Pemasaran

Hasil clustering menunjukkan bahwa kombinasi RFM dan K-Means mampu menghasilkan segmentasi yang membedakan pelanggan berdasarkan tingkat keterkinian transaksi, intensitas pembelian, dan nilai transaksi. Perbedaan karakteristik tersebut memberikan dasar untuk merancang strategi pemasaran yang lebih terarah. Temuan ini sejalan dengan Kasem et al. (2024), yang menempatkan customer profiling dan segmentation sebagai bagian penting dalam pemanfaatan artificial intelligence untuk mendukung aktivitas pemasaran.

Segmen Champions merupakan kelompok dengan karakteristik pelanggan paling bernilai. Strategi yang sesuai dapat berupa program loyalitas eksklusif, penghargaan pelanggan, akses terhadap produk atau layanan premium, serta penawaran yang ditujukan untuk mempertahankan frekuensi dan nilai transaksi. Pemeliharaan segmen ini penting karena kehilangan pelanggan bernilai tinggi dapat memberikan dampak yang lebih besar terhadap nilai pelanggan secara keseluruhan.

Segmen Loyal Customers menunjukkan tingkat transaksi yang masih relatif tinggi, tetapi belum mencapai karakteristik Champions. Strategi yang dapat diterapkan antara lain cross-selling, bundling produk, program loyalitas bertingkat, serta penawaran yang mendorong peningkatan nilai pembelian. Pendekatan tersebut bertujuan mempertahankan hubungan pelanggan sekaligus meningkatkan kontribusi ekonominya.

Segmen Potential Customers membutuhkan strategi yang berorientasi pada peningkatan keterlibatan. Promosi yang dipersonalisasi, insentif pembelian, rekomendasi produk, dan komunikasi berkala dapat digunakan untuk mendorong pelanggan melakukan transaksi kembali. Tujuannya adalah mengubah pelanggan dengan aktivitas transaksi rendah menjadi pelanggan dengan tingkat keterlibatan yang lebih tinggi.

Sementara itu, segmen At Risk memerlukan strategi retensi yang lebih spesifik. Tingginya nilai Recency menunjukkan bahwa pelanggan telah relatif lama tidak melakukan transaksi. Strategi *win-back* berupa penawaran khusus, pengingat personal, voucher, atau komunikasi yang disesuaikan dengan riwayat transaksi dapat digunakan untuk mendorong pelanggan kembali bertransaksi. Sarkar et al. (2024) menunjukkan bahwa hasil segmentasi berbasis RFM dan K-Means dapat digunakan sebagai dasar untuk menyusun strategi pemasaran yang lebih terarah, sedangkan Syahra et al. (2025) menghubungkan pendekatan serupa dengan kebutuhan customer relationship management.

Secara keseluruhan, hasil penelitian menunjukkan bahwa nilai utama dari segmentasi bukan hanya terletak pada pembentukan cluster secara algoritmik, tetapi pada kemampuan menerjemahkan karakteristik cluster menjadi tindakan bisnis yang berbeda. Pendekatan ini memberikan kerangka yang relatif sederhana dan dapat diterapkan pada data transaksi digital untuk mendukung pengambilan keputusan berbasis data. Temuan tersebut juga konsisten dengan perkembangan penelitian customer segmentation yang semakin mengarah pada integrasi antara analisis perilaku pelanggan, machine learning, dan strategi pemasaran (Alves Gomes & Meisen, 2023; Pai et al., 2026).

4. KESIMPULAN

Penelitian ini menerapkan analisis Recency, Frequency, dan Monetary (RFM) yang dipadukan dengan algoritma K-Means untuk melakukan segmentasi pelanggan pada data transaksi digital. Hasil analisis menggunakan metode Elbow menunjukkan bahwa jumlah cluster yang digunakan adalah ($k=4$), yang menghasilkan empat segmen pelanggan, yaitu Champions, Loyal Customers, Potential Customers, dan At Risk. Keempat segmen tersebut memiliki karakteristik Recency, Frequency, dan Monetary yang berbeda sehingga dapat memberikan gambaran mengenai tingkat aktivitas dan nilai transaksi pelanggan.

Segmen Champions menunjukkan karakteristik pelanggan paling aktif dengan Recency terendah, Frequency tertinggi, dan Monetary tertinggi, sedangkan At Risk memiliki Recency tertinggi serta Frequency dan Monetary yang relatif rendah. Loyal Customers merupakan segmen dengan jumlah pelanggan terbesar, sementara Potential Customers menunjukkan kelompok pelanggan yang masih memiliki peluang untuk ditingkatkan keterlibatan transaksinya. Perbedaan karakteristik tersebut menunjukkan bahwa hasil segmentasi RFM-K-Means dapat digunakan untuk membedakan pelanggan berdasarkan tingkat aktivitas dan nilai ekonominya.

Hasil segmentasi juga memberikan dasar untuk penyusunan strategi pemasaran yang lebih terarah. Champions dapat diprioritaskan melalui program loyalitas dan penawaran eksklusif, Loyal Customers dapat dikembangkan melalui strategi peningkatan frekuensi dan nilai transaksi, Potential Customers dapat diarahkan melalui promosi dan insentif, sedangkan At Risk dapat menjadi sasaran strategi retensi dan *win-back*. Dengan demikian, kombinasi RFM dan K-Means dapat menjadi pendekatan yang relatif sederhana dan aplikatif untuk mendukung pemanfaatan data transaksi digital dalam customer segmentation dan pengambilan keputusan pemasaran.

Penelitian selanjutnya dapat menggunakan data transaksi riil dengan periode pengamatan dan jumlah pelanggan yang lebih besar serta membandingkan K-Means dengan algoritma clustering lainnya, seperti DBSCAN atau hierarchical clustering, untuk memperoleh pembandingan terhadap hasil segmentasi yang diperoleh.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian ini. Ucapan terima kasih khusus disampaikan kepada pihak yang telah menyediakan dan mendukung penggunaan data transaksi digital, serta kepada seluruh pihak yang memberikan masukan selama proses pengolahan data, analisis, dan penyusunan artikel. Dukungan tersebut berperan dalam membantu penyelesaian penelitian mengenai segmentasi pelanggan menggunakan analisis RFM dan algoritma K-Means. Penulis juga mengapresiasi dukungan institusi dan rekan-rekan yang turut memberikan saran dan bantuan selama proses penelitian hingga penyusunan artikel ini.

DAFTAR PUSTAKA

- Akande, O., Asani, E. O., & Dautare, B. T. (2024). Customer segmentation through RFM analysis and K-Means clustering: Leveraging data-driven insights for effective marketing strategy. *Ceddi Journal of Information System and Technology (JST)*, 3(1), 14–25. <https://doi.org/10.56134/jst.v3i1.81>
- Alamsyah, A., Prasetyo, P. E., Sunyoto, S., Bintari, S. H., Saputro, D. D., Rohman, S., & Pratama, R. N. (2022). Customer segmentation using the integration of the Recency Frequency Monetary model and the K-Means cluster algorithm. *Scientific Journal of Informatics*, 9(2). <https://doi.org/10.15294/sji.v9i2.39437>
- Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1785–1792. <https://doi.org/10.1016/j.jksuci.2019.12.011>
- Antonius, V. H., & Fitriana, D. (2024). Enhancing customer segmentation insights by using RFM + discount proportion model with clustering algorithms. *International Journal of Advanced Computer Science and Applications*, 15(3). <https://doi.org/10.14569/IJACSA.2024.0150390>
- Aslantaş, G., Gençgöl, M., Rumelli, M., Özaraç, M., & Bakırlı, G. (2023). Customer segmentation using K-Means clustering algorithm and RFM model. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 25(74), 491–503. <https://doi.org/10.21205/deufmd.2023257418>
- Barus, O. P., Nathasya, C., & Pangaribuan, J. J. (2023). The implementation of RFM analysis to customer profiling using K-Means clustering. *Mathematical Modelling of Engineering Problems*, 10(1), 298–303. <https://doi.org/10.18280/mmep.100135>
- Dawane, V., Waghodekar, P., & Pagare, J. (2021). RFM analysis using K-Means clustering to improve revenue and customer retention. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3852887>

- Faran, J., Triayudi, A., & Aldisa, R. T. (2023). Combination of RFM's (Recency Frequency Monetary) method and agglomerative Ward's method for donors segmentation. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)* (pp. 962–967). IEEE. <https://doi.org/10.1109/ICCoSITE57641.2023.10127696>
- Fauzan, R. M., & Alfian, G. (2024). Segmentasi pelanggan e-commerce menggunakan fitur Recency, Frequency, Monetary (RFM) dan algoritma klasterisasi K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 170–177. <https://doi.org/10.14421/jiska.2024.9.3.170-177>
- Kasem, M. S., Hamada, M., & Taj-Eddin, I. (2024). Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Computing and Applications*, 36, 4995–5005. <https://doi.org/10.1007/s00521-023-09339-6>
- Kiat, A. B. H., Azhar, Y., & Rahmayanti, V. (2024). Penerapan metode K-Means dengan metode Elbow untuk segmentasi pelanggan menggunakan model RFM (Recency, Frequency & Monetary). *Jurnal Repositor*, 2(7). <https://doi.org/10.22219/repositor.v2i7.30762>
- Khumaidi, A., Wahyono, H., Darmawan, R., Kartika, H. D., Chusna, N. L., & Fauzy, M. K. (2023). RFM-AR model for customer segmentation using K-Means algorithm. *E3S Web of Conferences*, 465, 02005. <https://doi.org/10.1051/e3sconf/202346502005>
- Kumar, N. (2025). Intelligent customer segmentation: Unveiling consumer patterns with machine learning. *Journal of Umm Al-Qura University for Engineering and Architecture*, 16, 774–783. <https://doi.org/10.1007/s43995-025-00180-7>
- Lalitha, S., Reddy, M. U., Hussain, G. N., & Reddy, N. V. L. (2025). RFM analysis using K-means algorithm for customer segmentation. In *Lecture Notes in Electrical Engineering*. Springer. https://doi.org/10.1007/978-981-97-4711-5_12
- Naufal, R. (2026). Customer segmentation using RFM analysis and the K-Means clustering algorithm to support data-driven marketing strategies. *Information Technology Education Journal*, 5(3), 78–95. <https://doi.org/10.59562/intec.v5i3.12821>
- Pai, P.-Y., Lin, S.-W., & Lu, W.-M. (2026). Integration of association rule mining and RFM analysis with machine learning for e-commerce customer value segmentation: A sustainable retail perspective. *Quality & Quantity*, 60, 87–125. <https://doi.org/10.1007/s11135-025-02259-8>
- Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K., & Muangprathub, J. (2024). RFM model customer segmentation based on hierarchical approach using FCA. *Expert Systems with Applications*, 237, 121449. <https://doi.org/10.1016/j.eswa.2023.121449>
- Sarkar, M., Puja, A. R., & Chowdhury, F. R. (2024). Optimizing marketing strategies with RFM method and K-means clustering-based AI customer segmentation analysis. *Journal of Business and Management Studies*, 6(2), 54–60. <https://doi.org/10.32996/jbms.2024.6.2.5>

- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 759. <https://doi.org/10.3390/e23060759>
- Sidik, D. P. (2026). Komparasi model klasifikasi teks dalam mendeteksi berita hoaks berbahasa Indonesia. *JINTEN: Journal of Intelligent Systems and Digital Science*, 1(1), 1–12. <https://jurnal.ihsancahayapustaka.id/index.php/jinten/article/view/493>
- Smaili, M. Y., & Hachimi, H. (2023). New RFM-D classification model for improving customer analysis and response prediction. *Ain Shams Engineering Journal*, 14(12), 102254. <https://doi.org/10.1016/j.asej.2023.102254>
- Syakra, Y., Fadlil, A., & Yuliansyah, H. (2025). Customer segmentation using RFM and K-Means clustering to support CRM in retail industry. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(3), 1120–1131. <https://doi.org/10.33395/sinkron.v9i3.14907>
- Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12), 7243. <https://doi.org/10.3390/su14127243>
- Yuliasih, B. N., Herman, Sunardi, & Yuliansyah, H. (2024). Evaluation of K-Means clustering using Silhouette Score method on customer segmentation. *ILKOM Jurnal Ilmiah*, 16(3), 330–342. <https://doi.org/10.33096/ilkom.v16i3.2325.330-342>