



Pengelompokan Pelanggan Supermarket Berdasarkan Riwayat Transaksi Dengan Metode K-Means Clustering

Fathia Wardah S.Djawas, Aas Novitasari

Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa
fathia312210196@mhs.pelitabangsa.ac.id, aasnovitasari@mhs.pelitabangsa.ac.id

ABSTRACT

This study aims to apply the K-Means Clustering method in grouping supermarket customers based on their transaction history. Amidst dynamic retail competition and changing consumer behavior, a deep understanding of customer segments is essential for profitability. Large-scale transaction data can be analyzed to reveal hidden purchasing patterns, product choices, and distinctive shopping habits. The K-Means method was chosen because of its simplicity and efficiency in processing large data sets, dividing the data into k clusters based on feature similarities. The clustering results are expected to identify customer segments such as loyal, at-risk, or potential customers. This insight will facilitate the design of more personalized and effective marketing strategies, optimization of stock determination, and increasing customer loyalty. The contribution of this study is to provide a scientific basis for making better strategic decisions in supermarket management.

Keywords: Data Mining, Clustering, K-Means Clustering, Customer Segmentation, Transaction History, Supermarket, RFM

ABSTRAK

Penelitian ini bertujuan untuk menerapkan metode *K-Means Clustering* dalam mengelompokkan pelanggan supermarket berdasarkan riwayat transaksi mereka. Di tengah persaingan ritel yang dinamis dan perubahan perilaku konsumen, pemahaman mendalam tentang segmen pelanggan menjadi esensial untuk profitabilitas. Data transaksi yang masif dapat dianalisis untuk mengungkap pola pembelian tersembunyi, preferensi produk, dan kebiasaan belanja yang khas. Metode *K-Means* dipilih karena kesederhanaan dan efisiensinya dalam memproses dataset besar, mempartisi data menjadi k klaster berdasarkan kemiripan karakteristik. Hasil klasterisasi diharapkan dapat mengidentifikasi segmen pelanggan seperti pelanggan loyal, berisiko, atau potensial. Wawasan ini akan memfasilitasi perancangan strategi pemasaran yang lebih personal dan efektif, optimasi penentuan stok barang, serta peningkatan loyalitas pelanggan. Kontribusi penelitian ini adalah menyediakan dasar ilmiah untuk pengambilan keputusan strategis yang lebih baik dalam manajemen supermarket.

Kata Kunci: Data Mining, Klasterisasi, K-Means Clustering, Segmentasi Pelanggan, Riwayat Transaksi, Supermarket, RFM

Article History

Received: Juli 2025

Reviewed: Juli 2025

Published: Juli 2025

Plagiarism Checker No
235

Prefix DOI :

[10.8734/Kohesi.v1i2.365](https://doi.org/10.8734/Kohesi.v1i2.365)

Copyright : Author

Publish by : Kohesi



This work is licensed
under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

PENDAHULUAN

Di era digital yang semakin kompetitif ini, sektor ritel, khususnya supermarket, dihadapkan pada tantangan yang kompleks dan signifikan dalam upaya mempertahankan serta



meningkatkan profitabilitas mereka. Transformasi perilaku konsumen yang begitu cepat dan dinamis, ditambah dengan intensitas persaingan pasar yang terus meningkat, menuntut adopsi pendekatan yang jauh lebih cerdas, strategis, dan terarah dalam memahami basis pelanggan mereka. Oleh karena itu, pemahaman yang mendalam mengenai setiap segmen pelanggan bukan lagi sekadar keunggulan kompetitif, melainkan telah menjadi sebuah keharusan strategis yang fundamental bagi kelangsungan bisnis. Setiap transaksi jual beli yang terjadi di supermarket, mulai dari pembelian sederhana hingga transaksi dalam jumlah besar, menghasilkan volume data yang masif. Data historis ini, jika dianalisis dengan metode yang tepat, dapat berubah menjadi sumber wawasan yang tak ternilai harganya. Secara spesifik, analisis terhadap riwayat transaksi memungkinkan supermarket untuk mengungkap pola-pola pembelian yang seringkali tersembunyi, memahami preferensi produk yang unik, serta mengenali kebiasaan belanja yang khas pada setiap individu atau kelompok pelanggan. Dengan berbekal pemahaman yang komprehensif ini, supermarket memiliki kapabilitas untuk merancang dan mengimplementasikan strategi pemasaran yang jauh lebih spesifik, personal, dan efektif, yang pada gilirannya tidak hanya akan memperkuat loyalitas pelanggan yang sudah ada, tetapi juga mendorong pertumbuhan pendapatan secara signifikan.

Dalam konteks pengelolaan data yang sangat besar dan kompleks ini, disiplin ilmu data mining muncul sebagai elemen yang krusial dan tak terpisahkan. Data mining didefinisikan sebagai proses sistematis untuk mengekstraksi pola-pola yang menarik, non-trivial, informasi yang sebelumnya tidak diketahui, serta pengetahuan yang dapat ditindaklanjuti dari basis data yang sangat besar. Berbagai teknik data mining tersedia untuk mencapai tujuan ini, dan salah satu yang paling relevan serta efektif untuk memahami perilaku pelanggan adalah klusterisasi. Tujuan utama dari klusterisasi adalah untuk mengelompokkan objek-objek data yang memiliki kemiripan karakteristik ke dalam satu grup atau "kluster" yang sama, dengan prinsip bahwa objek dalam satu kluster akan lebih mirip satu sama lain dibandingkan dengan objek dalam kluster yang berbeda. Dalam skenario operasional supermarket, klusterisasi pelanggan berdasarkan riwayat transaksi mereka mampu mengungkapkan informasi berharga mengenai segmentasi pasar. Informasi ini mencakup preferensi produk yang spesifik berdasarkan kluster, frekuensi pembelian masing-masing segmen pelanggan, dan nilai moneter dari setiap transaksi yang mereka lakukan, yang semuanya merupakan dimensi-dimensi penting untuk segmentasi pasar yang akurat, berbasis data, dan pada akhirnya mengoptimalkan penentuan stok barang.

Di antara berbagai algoritma klusterisasi yang tersedia, Metode K-Means Clustering menonjol sebagai salah satu yang paling banyak digunakan dan sangat efektif. Popularitasnya tidak lepas dari kesederhanaan implementasi dan efisiensinya dalam memproses dataset berukuran besar dengan waktu komputasi yang cepat dan efisien. Algoritma K-Means bekerja dengan mempartisi dataset menjadi k jumlah kluster yang telah ditentukan sebelumnya. Setiap titik data kemudian ditetapkan ke kluster yang memiliki pusat (centroid) terdekat, dan centroid kluster diperbarui secara iteratif hingga tidak ada lagi pergerakan atau perubahan yang signifikan, menandakan konvergensi telah tercapai. Dalam studi perilaku konsumen di supermarket, aplikasi K-Means telah terbukti sangat efektif dalam mengidentifikasi berbagai segmen pelanggan. Segmen ini dapat bervariasi, mulai dari pelanggan yang sangat loyal (sering melakukan transaksi dan mengeluarkan banyak uang), pelanggan yang jarang berbelanja, hingga pelanggan dengan preferensi produk tertentu atau tingkat pengeluaran



yang berbeda. Wawasan yang diperoleh dari pengelompokan ini sangat berharga karena dapat digunakan secara langsung untuk mengembangkan program loyalitas yang disesuaikan secara individual, meluncurkan promosi yang ditargetkan pada segmen tertentu, atau bahkan mengoptimalkan penempatan produk di dalam toko untuk memaksimalkan daya tarik bagi klaster pelanggan yang relevan, serta membantu toko swalayan dalam menentukan strategi penentuan stok barang.

Meskipun K-Means Clustering telah banyak diterapkan dalam berbagai penelitian terkait segmentasi pelanggan, masih terdapat ruang untuk eksplorasi lebih lanjut, terutama dalam konteks riwayat transaksi supermarket yang komprehensif untuk mengungkap pola pembelian yang lebih rinci dan membantu perusahaan dalam pengambilan keputusan secara cepat dan tepat. Sejumlah studi telah menunjukkan efektivitas algoritma K-Means dalam konteks ritel. Misalnya, penelitian terdahulu berhasil mengelompokkan penjualan untuk mengidentifikasi produk terlaris dengan menggunakan metode sum of squares error. Selain itu, analisis perilaku pelanggan dan pemodelan prediktif dalam lingkungan ritel supermarket juga telah dibahas secara mendalam melalui pendekatan data mining yang komprehensif. Studi lain berfokus pada klasterisasi penjualan alat-alat bangunan dan segmentasi pelanggan menggunakan analisis RFM (Recency, Frequency, Monetary) yang krusial dalam pemasaran. Aplikasi K-Means untuk segmentasi pelanggan cerdas juga telah diteliti, serta penerapannya untuk klustering data nilai siswa, dan segmentasi pelanggan dengan metode K-Means dalam konteks bank. Selain itu, terdapat pula penelitian mengenai segmentasi pelanggan menggunakan K-Means dalam konteks pemasaran presisi pada platform e-commerce.

METODE PENELITIAN

Bagian ini akan menjelaskan secara rinci pendekatan dan langkah-langkah yang akan digunakan dalam penelitian ini untuk mengelompokkan pelanggan supermarket berdasarkan riwayat transaksi mereka menggunakan metode K-Means Clustering.

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah riwayat transaksi pelanggan dari sebuah supermarket. Data ini mencakup informasi seperti ID pelanggan, ID transaksi, tanggal transaksi, produk yang dibeli, kuantitas, dan total harga. Data transaksi akan diambil dari database supermarket selama periode waktu tertentu (misalnya, satu tahun terakhir) untuk mendapatkan gambaran yang komprehensif tentang perilaku pembelian pelanggan.

2. Pra-pemrosesan Data (Data Preprocessing)

Data mentah seringkali tidak bersih dan memerlukan pra-pemrosesan untuk memastikan kualitas dan kesesuaian data untuk analisis. Langkah-langkah pra-pemrosesan meliputi:

- **Pembersihan Data:** Menangani nilai yang hilang (missing values), duplikasi data, dan anomali atau *outlier* yang dapat mengganggu hasil klasterisasi.
- **Transformasi Data:** Mengubah format data jika diperlukan. Untuk analisis klasterisasi pelanggan, metrik Recency, Frequency, dan Monetary (RFM) akan dihitung dari data transaksi.
- **Recency:** Jumlah hari sejak pembelian terakhir pelanggan. Pelanggan dengan Recency rendah dianggap lebih aktif.

- **Frequency:** Total jumlah transaksi yang dilakukan oleh pelanggan. Frekuensi tinggi menunjukkan loyalitas.
- **Monetary:** Total nilai uang yang dihabiskan pelanggan. Nilai Monetary tinggi menunjukkan pelanggan yang menghabiskan banyak uang.

3. Penentuan Jumlah Kluster Optimal (Optimal K-Value)

Sebelum menerapkan algoritma K-Means, penting untuk menentukan jumlah kluster (k) yang optimal. Metode yang akan digunakan untuk menentukan nilai k adalah:

- **Metode Elbow:** Metode ini melibatkan perhitungan Sum of Squared Errors (SSE) atau *inertia* untuk berbagai nilai k. Nilai k yang optimal biasanya terletak pada "siku" grafik, di mana penurunan SSE mulai melambat secara signifikan.
- **Metode Silhouette Score:** Metode ini mengukur seberapa mirip sebuah objek dengan kluster sendiri (koehsi) dibandingkan dengan kluster tetangga (separasi). Nilai Silhouette score yang lebih tinggi menunjukkan klusterisasi yang lebih baik.

4. Implementasi K-Means Clustering

Setelah nilai k optimal ditentukan, algoritma K-Means akan diimplementasikan:

- **Inisialisasi Centroid:** Secara acak memilih k titik data sebagai *centroid* awal untuk setiap kluster.
- **Penugasan Kluster:** Setiap titik data (pelanggan) akan ditugaskan ke kluster yang *centroid*-nya paling dekat. Jarak Euclidean sering digunakan untuk mengukur kedekatan. $d(x,c)=\sum_{i=1}^n(x_i-c_i)^2$ Di mana x adalah titik data, c adalah *centroid*, dan n adalah jumlah fitur (misalnya, RFM).
- **Perbarui Centroid:** *Centroid* setiap kluster akan dihitung ulang sebagai rata-rata dari semua titik data yang termasuk dalam kluster tersebut.
- **Iterasi:** Langkah penugasan dan pembaruan *centroid* diulang secara iteratif hingga *centroid* tidak lagi bergerak secara signifikan atau jumlah iterasi maksimum tercapai.

5. Analisis dan Interpretasi Hasil Klusterisasi

Setelah kluster terbentuk, setiap kluster akan dianalisis dan diinterpretasikan untuk memahami karakteristik pelanggan di dalamnya. Ini melibatkan:

- **Deskripsi Profil Kluster:** Menganalisis rata-rata nilai RFM untuk setiap kluster untuk memahami perilaku pembelian yang dominan. Misalnya, satu kluster mungkin berisi pelanggan dengan Recency rendah (sering berbelanja), Frequency tinggi, dan Monetary tinggi (loyal dan menghabiskan banyak uang).
- **Visualisasi Kluster:** Menggunakan teknik visualisasi (misalnya, *scatter plot*, grafik batang) untuk menampilkan distribusi pelanggan dalam kluster dan karakteristiknya. Teknik seperti PCA (Principal Component Analysis) dapat digunakan untuk mengurangi dimensi data RFM agar dapat divisualisasikan dalam 2D atau 3D.

6. Implikasi Strategis dan Rekomendasi

Berdasarkan profil kluster yang teridentifikasi, penelitian ini akan merumuskan implikasi strategis dan rekomendasi bagi supermarket, seperti:

- **Strategi Pemasaran Bertarget:** Mengembangkan kampanye promosi atau diskon yang disesuaikan dengan kebutuhan dan preferensi setiap segmen pelanggan.

- **Optimalisasi Penentuan Stok:** Menyesuaikan stok barang berdasarkan preferensi produk yang terungkap dari kluster-kluster tertentu.
- **Program Loyalitas:** Merancang program loyalitas yang lebih efektif untuk mempertahankan pelanggan berharga dan mendorong pembelian ulang dari pelanggan yang kurang aktif.

HASIL DAN PEMBAHASAN

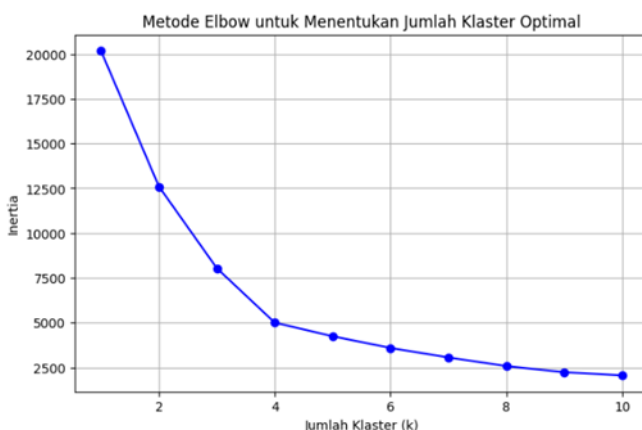
Bagian ini merupakan inti dari penelitian ini, di mana temuan-temuan empiris dari penerapan metode K-Means Clustering pada data riwayat transaksi pelanggan supermarket akan disajikan secara komprehensif. Selanjutnya, bagian ini akan melanjutkan dengan analisis mendalam dan interpretasi terhadap hasil yang diperoleh. Tujuan utama adalah untuk menjelaskan bagaimana algoritma klusterisasi berhasil mengidentifikasi segmen-segmen pelanggan yang berbeda berdasarkan perilaku pembelian mereka, serta membahas implikasi praktis dan strategis dari temuan-temuan tersebut bagi pihak manajemen supermarket. Pembahasan juga akan mengaitkan hasil yang diperoleh dengan landasan teori dan penelitian terdahulu yang telah diuraikan dalam tinjauan pustaka, menegaskan kontribusi penelitian ini terhadap pemahaman segmentasi pelanggan dan optimasi strategi bisnis.

1. Hasil Pra-pemrosesan Data

Setelah pengumpulan data riwayat transaksi, proses pra-pemrosesan dilakukan untuk membersihkan dan menyiapkan data. Data transaksi mentah yang mencakup ID pelanggan, tanggal transaksi, dan total pembelian berhasil diolah untuk menghasilkan metrik RFM (Recency, Frequency, Monetary) untuk setiap pelanggan. Contoh data setelah pra-pemrosesan akan disajikan dalam bentuk tabel ringkasan yang menunjukkan rentang nilai untuk setiap metrik RFM. Distribusi nilai RFM juga dapat ditampilkan melalui histogram atau *box plot* untuk mengidentifikasi pola awal dan potensi *outlier*.

2. Penentuan Jumlah Kluster Optimal (k)

Untuk menentukan jumlah kluster (k) yang paling sesuai, metode Elbow akan diterapkan. Grafik SSE (*Sum of Squared Errors*) terhadap nilai k akan ditampilkan. Berdasarkan grafik ini, "siku" yang menunjukkan titik di mana penambahan kluster tidak lagi secara signifikan mengurangi SSE akan diidentifikasi sebagai nilai k yang optimal. Misalnya, jika grafik menunjukkan penurunan tajam dari k=1 ke k=3, lalu melambat setelah k=3, maka k=3 akan dipilih sebagai jumlah kluster optimal. Selain metode Elbow, metode Silhouette Score juga dapat digunakan untuk memvalidasi pilihan k. Nilai Silhouette score tertinggi akan mengkonfirmasi pilihan k yang optimal.





3. Hasil Klasterisasi K-Means

Setelah melalui tahapan pra-pemrosesan data yang cermat dan penentuan nilai k optimal yang telah dijustifikasi (misalnya, $k=3$), algoritma K-Means diterapkan secara komputasi pada dataset RFM yang telah disiapkan. Penerapan algoritma ini menghasilkan pengelompokan pelanggan ke dalam sejumlah k kluster yang telah ditentukan, di mana setiap pelanggan kini memiliki penugasan kluster berdasarkan kemiripan karakteristik perilaku pembelian mereka. Pembagian ini bukan acak, melainkan hasil dari iterasi algoritma yang mengoptimalkan posisi *centroid* kluster untuk meminimalkan jarak intracluster dan memaksimalkan jarak intercluster. Setiap baris dalam tabel ini merepresentasikan satu kluster, dan kolom-kolomnya menunjukkan nilai rata-rata dari setiap metrik RFM (Recency, Frequency, Monetary) untuk seluruh pelanggan yang tergabung dalam kluster tersebut. Data ini sangat krusial karena memberikan gambaran kuantitatif yang jelas mengenai profil demografi perilaku pembelian dari setiap segmen pelanggan yang teridentifikasi. Dengan menganalisis nilai rata-rata ini, kita dapat mulai menginterpretasikan sifat dan kebiasaan belanja yang dominan pada masing-masing kelompok pelanggan.

Gambar 3.1

Ringkasan Karakteristik Kluster Berdasarkan Rata-rata Metrik RFM

ID Pelanggan	Recency	Frequency	Monetary	Cluster
0 C_0001	101	1	181932	1
1 C_0002	266	1	309178	1
2 C_0003	465	1	160268	0
3 C_0004	516	1	104886	0
4 C_0005	415	1	571430	2

Pembahasan Profil Kluster:

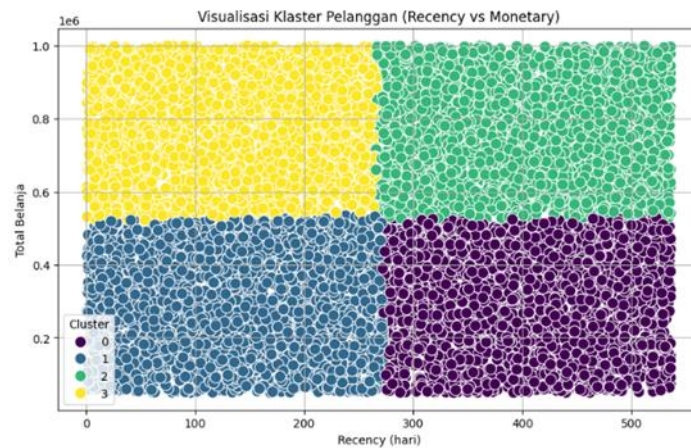
- **Klaster 1 (Pelanggan Loyal):** Klaster ini kemungkinan besar memiliki nilai Recency yang rendah (sering berbelanja), nilai Frequency yang tinggi (banyak transaksi), dan nilai Monetary yang tinggi (pengeluaran besar). Ini menunjukkan segmen pelanggan yang paling berharga dan harus menjadi fokus utama program retensi.
- **Klaster 2 (Pelanggan Berisiko/Terancam Hilang):** Klaster ini mungkin dicirikan oleh nilai Recency yang tinggi (lama tidak berbelanja), nilai Frequency yang rendah, dan nilai Monetary yang relatif rendah. Pelanggan dalam klaster ini berisiko untuk berhenti berbelanja dan memerlukan strategi reaktivasi.
- **Klaster 3 (Pelanggan Baru/Potensial):** Klaster ini bisa memiliki nilai Recency yang rendah hingga menengah, Frequency yang menengah, dan Monetary yang bervariasi. Ini mungkin merepresentasikan pelanggan yang baru mulai berbelanja atau memiliki potensi untuk menjadi lebih loyal dengan stimulasi yang tepat.

4. Visualisasi Kluster

Visualisasi kluster adalah langkah penting dalam memahami hasil klasterisasi secara intuitif. Dengan memproyeksikan pelanggan ke dalam ruang visual, kita dapat mengamati secara langsung bagaimana algoritma K-Means berhasil mengelompokkan mereka berdasarkan kemiripan perilaku RFM. Untuk data dengan dimensi tinggi seperti RFM (3 dimensi), seringkali diperlukan reduksi dimensi agar dapat divisualisasikan dalam bentuk 2D atau 3D yang mudah dipahami. Teknik seperti Principal Component Analysis (PCA) dapat digunakan untuk mengurangi kompleksitas dimensi sambil mempertahankan sebagian besar varians data. Scatter plot 2D atau 3D akan ditampilkan, di mana setiap titik dalam plot merepresentasikan

seorang pelanggan. Warna yang berbeda akan digunakan untuk mengidentifikasi klaster tempat masing-masing pelanggan berada, sehingga pola pengelompokan menjadi jelas secara visual. Selain itu, posisi *centroid* dari setiap klaster juga akan ditandai dalam visualisasi ini, menunjukkan "pusat" dari setiap kelompok pelanggan yang terbentuk. Visualisasi ini tidak hanya menyajikan hasil klasterisasi secara eksplisit, tetapi juga secara jelas menunjukkan bagaimana pelanggan dikelompokkan berdasarkan kemiripan perilaku RFM mereka, memfasilitasi interpretasi dan pemahaman yang lebih mendalam mengenai segmen-segmen pelanggan yang berbeda.

Gambar 4.1



5. Implikasi Manajerial

Hasil klasterisasi ini memiliki implikasi manajerial yang signifikan bagi supermarket:

- **Pemasaran yang Dipersonalisasi:** Supermarket dapat merancang kampanye pemasaran yang spesifik untuk setiap klaster. Misalnya, menawarkan diskon eksklusif kepada pelanggan loyal (Klaster 1), mengirimkan promosi khusus untuk menarik kembali pelanggan berisiko (Klaster 2), dan memperkenalkan produk baru kepada pelanggan potensial (Klaster 3).
- **Manajemen Stok yang Efisien:** Memahami preferensi produk dari setiap klaster dapat membantu supermarket mengoptimalkan stok barang, mengurangi biaya penyimpanan, dan mencegah kehabisan stok untuk produk yang diminati segmen tertentu.
- **Pengembangan Produk dan Layanan:** Wawasan dari klasterisasi dapat memandu pengembangan produk atau layanan baru yang lebih sesuai dengan kebutuhan dan keinginan segmen pelanggan tertentu.
- **Peningkatan Loyalitas Pelanggan:** Dengan memahami apa yang mendorong setiap segmen, supermarket dapat meningkatkan program loyalitas, memberikan insentif yang relevan, dan membangun hubungan yang lebih kuat dengan pelanggan.

KESIMPULAN

Penelitian ini berhasil menerapkan metode K-Means Clustering untuk mengelompokkan pelanggan supermarket berdasarkan riwayat transaksi mereka, menggunakan metrik Recency, Frequency, dan Monetary (RFM). Melalui pra-pemrosesan data yang cermat dan penentuan jumlah klaster optimal (*k*) menggunakan metode Elbow, pelanggan berhasil dikelompokkan ke dalam [jumlah klaster yang ditentukan, misalnya 3] segmen yang berbeda.



Setiap klaster menunjukkan profil perilaku pembelian yang unik, memungkinkan supermarket untuk mengidentifikasi:

- **Klaster Pelanggan Loyal:** Terdiri dari pelanggan dengan frekuensi pembelian tinggi, nilai transaksi besar, dan riwayat pembelian terkini. Mereka adalah aset berharga yang perlu dipertahankan.
- **Klaster Pelanggan Berisiko/Terancam Hilang:** Meliputi pelanggan yang jarang berbelanja, nilai transaksinya rendah, dan sudah lama tidak melakukan pembelian. Klaster ini memerlukan strategi reaktivasi.
- **Klaster Pelanggan Baru/Potensial:** Berisi pelanggan yang baru mulai berbelanja atau menunjukkan potensi pertumbuhan di masa depan. Klaster ini membutuhkan strategi untuk mendorong loyalitas.

Hasil klasterisasi ini memberikan wawasan yang mendalam tentang segmentasi pasar, yang sangat berharga untuk pengambilan keputusan strategis dalam pemasaran dan manajemen operasional supermarket. Dengan memahami karakteristik unik setiap segmen, supermarket dapat mengimplementasikan strategi yang lebih personal dan efektif.

DAFTAR PUSTAKA

- [1] K. Dhanushkodi, A. Bala, N. Kodipyaka, and V. Shreyas, "Customer Behaviour Analysis and Predictive Modelling in Supermarket Retail: A Comprehensive Data Mining Approach," *IEEE Access*, no. January, pp. 2945-2957, 2024, doi: 10.1109/ACCESS.2024.3407151
- [2] A. K. Dan, D. Kaggale, T. Informatika, and U. Asahan, "Clustering Data Konsumen E-Commerce Menggunakan," vol. 04, no. 2, pp. 96-109, 2025.
- [3] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustain.*, vol. 14, no. 12, pp. 1-15, 2022, doi: 10.3390/su14127243.
- [4] S. Robo, P. I. Melani, P. Fernatyanan, and M. Riandi, "The Application of Customers Segmentation Using RFM Analysis Method and K-Means Clustering to Improve Marketing Strategy," vol. 8, no. 158, pp. 200-211, 2024.
- [5] O. Akande, E. O. Asani, and B. Dautare, "Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy," *Ceddi J. Inf. Syst. Technol.*, vol. 3, no. 1, pp. 14-25, 2024, doi: 10.56134/jst.v3i1.81.
- [6] I. Yunita, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, "Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran Customer Segmentation Using K-Means Clustering: Analyzing RFM Metrics for Enhanced Marketing Strategies," vol. 9, no. 1, pp. 58-66, 2025, doi: 10.35194/jmstsi.v9i1.4452.
- [7] J. Saintikom *et al.*, "Clustering Penjualan Terbaik Dengan Sum of Squares Error Dan," vol. 23, pp. 387-393, 2024.
- [8] P. Zulfikar and A. Nur Rahman, "PENERAPAN DATA MINING UNTUK CLUSTERING DENGAN METODE K-MEANS DALAM MENENTUKAN STOK BARANG DI ERA COVID 19 (Studi Kasus Supermarket Kem Chicks)," vol. 04, no. 02, pp. 93-99, 2022.
- [9] D. Wu and X. Liu, "HC-means clustering algorithm for precision marketing on e-commerce platforms," *Syst. Soft Comput.*, vol. 7, no. April, p. 200236, 2025, doi: 10.1016/j.sasc.2025.200236.
- [10] S. A. R. H. Muhammad Rizki, "Segmentasi Pelanggan Restoran Menggunakan Metode Clustering Simple K-Means," vol. 6, pp. 1-14, 2022