



Penerapan K-Means Clustering pada Data Ritel untuk Segmentasi Pelanggan Berbasis Monetary-Frequency

Application of K-Means Clustering on Retail Data for Customer Monetary-Frequency Segmentation

Rifqi Rahmatullah

Universitas Sebelas April

Email: 220660121018@student.unsap.ac.id

Article Article Info

Article history :

Received : 11-06-2026

Revised : 13-06-2026

Accepted : 15-06-2026

Published : 17-06-2026

Abstract

The increasing competition in the retail industry requires companies to understand customer behavior more deeply to optimize marketing strategies. Customer segmentation is an essential approach to identify consumer characteristics and design more targeted strategies. This study aims to apply the K-Means Clustering data mining algorithm to segment customers in retail sales data using the Superstore Dataset. The methodology involves preprocessing 9,800 transaction records into 793 unique customer records by extracting Monetary and Frequency attributes. The Elbow Method was then applied to determine the optimal number of clusters. The results indicate that the optimal number of clusters is $K=4$. The K-Means algorithm successfully classified customers into four segments, namely superstar customers, loyal customers, potential customers, and at-risk customers. These findings provide data-driven insights that can be utilized by companies to design more effective marketing strategies, enhance customer loyalty, and maximize profitability.

Keywords: Algorithm, Customer Segmentation, Data Mining

Abstrak

Persaingan industri ritel yang semakin ketat menuntut perusahaan untuk memahami perilaku pelanggan secara mendalam guna mendukung optimalisasi strategi pemasaran. Segmentasi pelanggan menjadi pendekatan penting untuk mengidentifikasi karakteristik konsumen dan merancang strategi yang lebih tepat sasaran. Penelitian ini bertujuan untuk menerapkan algoritma data mining K-Means Clustering dalam melakukan segmentasi pelanggan pada data penjualan ritel menggunakan Superstore Dataset. Metodologi penelitian mencakup tahap pra-pemrosesan data, yaitu mengolah 9.800 baris transaksi menjadi 793 data pelanggan unik melalui ekstraksi atribut Monetary dan Frequency. Selanjutnya, metode Elbow digunakan untuk menentukan jumlah cluster yang optimal. Hasil penelitian menunjukkan bahwa jumlah cluster terbaik adalah $K=4$. Algoritma K-Means berhasil mengelompokkan pelanggan ke dalam empat segmen, yaitu segmen pelanggan superstar, pelanggan loyal, pelanggan potensial, serta pelanggan pasif atau berisiko. Temuan ini memberikan wawasan berbasis data yang dapat dimanfaatkan perusahaan untuk menyusun strategi pemasaran yang lebih efektif, meningkatkan loyalitas pelanggan, serta memaksimalkan potensi keuntungan.

Kata Kunci : Algoritma, Segmentasi Pelanggan, Data Mining

PENDAHULUAN

Persaingan lanskap bisnis global saat ini tengah mengalami transformasi fundamental yang didorong oleh dinamika pasar yang semakin kompetitif dan perubahan perilaku konsumen yang cepat, sehingga menuntut perusahaan untuk beradaptasi dengan strategi yang lebih responsif (M. F.



Fadhillah, et al., 2025). Pergeseran paradigma dari pendekatan yang berorientasi pada produk (product-oriented) menuju pendekatan yang berpusat pada pelanggan (customer-oriented) menjadi prasyarat mutlak bagi keberlangsungan bisnis di era modern (M. F. Fadhillah, et al., 2025). Dalam lingkungan industri ritel yang sangat dinamis, pemahaman mendalam mengenai profil dan perilaku pelanggan bukan lagi sekadar keunggulan tambahan, melainkan kebutuhan krusial untuk memenangkan persaingan (I. Yunita, et al., 2025).

Perusahaan ritel kini dihadapkan pada tantangan untuk mengelola volume data transaksi yang sangat besar dan kompleks, yang jika dimanfaatkan dengan benar, dapat menjadi aset strategis untuk pengambilan keputusan (I. Yunita, et al., 2025). Fenomena ini diperkuat oleh fakta bahwa strategi pemasaran yang seragam atau "satu ukuran untuk semua" terbukti tidak lagi efektif dan efisien dalam menjangkau audiens yang semakin terfragmentasi (I. Yunita, et al., 2025). Oleh karena itu, kemampuan untuk melakukan segmentasi pelanggan secara akurat menjadi kunci untuk merancang strategi pemasaran yang lebih personal dan relevan (N. H. Ahsina, et al., 2022).

Untuk mengatasi tantangan pengelolaan data yang masif tersebut, disiplin ilmu Data Mining hadir sebagai solusi metodologis yang sistematis untuk mengekstraksi pengetahuan berharga dari sekumpulan data yang besar. Secara teoretis, Data Mining merupakan proses komputasi yang mengintegrasikan teknik statistik, kecerdasan buatan, dan pembelajaran mesin untuk mengungkap pola tersembunyi yang tidak dapat dideteksi melalui analisis manual (F. SLN, 2023). Salah satu teknik utama dalam Data Mining yang banyak diterapkan untuk tujuan bisnis adalah Clustering, yaitu metode pengelompokan data ke dalam kluster-kluster di mana objek dalam satu kelompok memiliki tingkat kemiripan yang tinggi satu sama lain dibandingkan dengan objek di kelompok lain (N. H. Ahsina, et al., 2022).

Di antara berbagai algoritma clustering, algoritma K-Means diakui sebagai metode yang paling populer dan efisien karena kemampuannya dalam menangani dataset besar dengan waktu komputasi yang relatif cepat (C. H. Ardana, 2024). Keunggulan K-Means terletak pada kesederhanaan implementasinya dan kemudahan interpretasi hasil, yang menjadikannya pilihan ideal untuk aplikasi segmentasi pelanggan di sektor ritel (A. M. Artiarno, et al., 2025). Dalam konteks segmentasi pelanggan, analisis perilaku sering kali didasarkan pada data historis transaksi yang mencerminkan kebiasaan pembelian pelanggan di masa lalu (I. Lewaaelhamd, 2024).

Model Recency, Frequency, Monetary (RFM) adalah kerangka kerja yang telah terbukti kuat dan diakui secara luas untuk mengevaluasi nilai pelanggan berdasarkan tiga dimensi perilaku utama (M. F. Fadhillah, et al., 2025). Dimensi Recency mengukur seberapa baru pelanggan melakukan transaksi, Frequency mengukur seberapa sering pelanggan berbelanja, dan Monetary mengukur total nilai uang yang dikeluarkan pelanggan (M. F. Fadhillah, et al., 2025). Namun, dalam fokus strategi penjualan tertentu, dimensi Monetary dan Frequency sering kali menjadi indikator yang paling krusial untuk menilai profitabilitas dan loyalitas pelanggan secara langsung (E. Febrianty, et al., 2023).

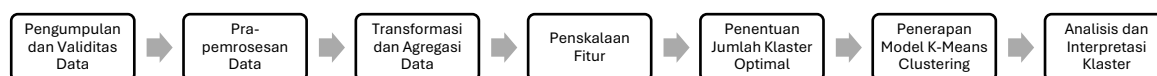
Berbagai penelitian terdahulu telah mendokumentasikan keberhasilan penerapan metode ini dalam berbagai konteks industri. Penelitian yang dilakukan oleh Fadhillah, Suyoso, dan Puspitasari (2024) pada perusahaan distributor B2B menunjukkan bahwa algoritma K-Means mampu menghasilkan segmentasi yang lebih baik dibandingkan metode lain seperti Agglomerative dan DBSCAN. Penelitian ini bermaksud untuk mengisi celah praktis tersebut dengan menerapkan



algoritma K-Means secara khusus pada dimensi Monetary dan Frequency untuk menghasilkan segmentasi yang dapat langsung ditindaklanjuti dalam strategi pemasaran taktis.

METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif yang berlandaskan pada kerangka kerja Knowledge Discovery in Database (KDD) untuk melakukan ekstraksi pengetahuan dari basis data transaksional ritel. Desain penelitian dirancang secara sistematis dan terstruktur untuk mengubah data transaksi mentah menjadi wawasan strategis mengenai segmentasi pelanggan. Tahapan penelitian meliputi: (1) Pengumpulan dan Validitas Data, (2) Pra-pemrosesan Data, (3) Transformasi dan Agregasi Data, (4) Penskalaan Fitur, (5) Penentuan Jumlah Kluster Optimal, (6) Penerapan Model K-Means Clustering, dan (7) Analisis dan Interpretasi Kluster.



Gambar 1. Metode Penelitian

Pengumpulan dan Validitas Data

Data yang digunakan dalam penelitian ini diklasifikasikan sebagai data sekunder, yaitu dataset publik "Sales Forecasting" yang diperoleh dari repositori Kaggle (R. Sahoo, 2026). Dataset ini merekam aktivitas transaksi penjualan dari sebuah "Superstore" global, terdiri dari 9.800 baris data transaksional yang mencakup 18 atribut operasional. Dalam konteks penelitian segmentasi pelanggan berbasis Monetary-Frequency, fokus analisis diberikan pada atribut Customer ID, Order ID, dan Sales. Integritas data diverifikasi untuk memastikan bahwa setiap transaksi memiliki referensi pelanggan yang valid dan nilai penjualan yang logis.

Pra-pemrosesan Data

Pra-pemrosesan data merupakan langkah fundamental dalam KDD untuk memastikan bahwa data yang akan diolah bersih, konsisten, dan siap dimodelkan. Langkah pertama adalah pembersihan data untuk menghilangkan anomali; transaksi tanpa Customer ID dieliminasi karena data tersebut tidak dapat dikaitkan dengan profil pelanggan tertentu (I. R. Jawara and Z. Fatah, 2025). Strategi pra-pemrosesan data yang tepat terbukti dapat meningkatkan kinerja model secara signifikan dan mengurangi tingkat kesalahan (R. Surya, 2025).

Transformasi, Agregasi Data, dan Penskalaan Fitur

Penelitian ini mengekstraksi dua variabel utama: Frequency dihitung berdasarkan jumlah transaksi unik per pelanggan, sedangkan Monetary dihitung berdasarkan total pengeluaran per pelanggan. Pendekatan ini mengubah 9.800 baris data transaksional menjadi 793 profil pelanggan unik yang komprehensif dan siap untuk tahap pemodelan. Selanjutnya, normalisasi data dilakukan karena Monetary bisa mencapai jutaan sementara Frequency hanya puluhan; tanpa penskalaan, fitur dengan nilai nominal besar akan mendominasi perhitungan jarak Euclidean (M. Annas and S. N. Wahab, 2023). Dengan standardisasi, setiap fitur memberikan kontribusi yang seimbang dalam kalkulasi jarak.



Penentuan Jumlah Kluster Optimal

Penelitian ini menggunakan kombinasi Metode Elbow dan Silhouette Score. Metode Elbow digunakan untuk menentukan jumlah kluster terbaik dengan cara memvisualisasikan hasil perbandingan jumlah kluster yang membentuk pola "siku" pada grafik (N. H. Ahsina, et al., 2022). Teknik ini menghitung nilai Sum of Squared Errors (SSE) untuk berbagai variasi nilai K; titik "siku" dianggap sebagai nilai optimal di mana penambahan kluster baru tidak lagi memberikan penurunan varians yang signifikan. Untuk memvalidasi hasil dari Metode Elbow, digunakan metrik validasi internal berupa Silhouette Score yang mengukur kohesi dan separasi antar kluster (I. Yunita, et al., 2025).

Penerapan Model K-Means Clustering

Setelah jumlah kluster optimal ditentukan, algoritma K-Means diterapkan pada dataset yang telah distandarisasi. Mekanisme K-Means bekerja secara iteratif melalui: inisialisasi centroid secara acak, penugasan setiap titik data ke centroid terdekat menggunakan Euclidean Distance, dan pembaruan posisi centroid berdasarkan rata-rata nilai semua anggota kluster. Proses ini diulang hingga kondisi konvergensi tercapai, yaitu ketika posisi centroid tidak lagi berubah (N. H. Ahsina, et al., 2022).

Analisis dan Interpretasi Kluster

Tahap akhir adalah menerjemahkan hasil matematis klustering menjadi wawasan bisnis yang dapat ditindaklanjuti. Profil kluster dianalisis dengan menghitung nilai rata-rata dari atribut asli, yaitu Monetary dan Frequency, untuk setiap kelompok yang terbentuk. Pelanggan kemudian dikategorikan ke dalam persona segmen berdasarkan kombinasi nilai Frequency dan Monetary, seperti pelanggan loyal atau champions, pelanggan potensial, serta pelanggan pasif atau berisiko (I. Yunita, et al., 2025).

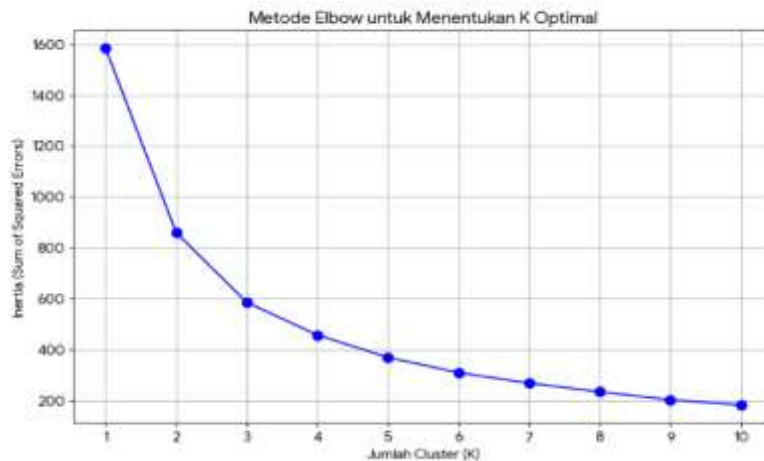
HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil analisis data yang dilakukan menggunakan metodologi clustering algoritma K-Means, serta pembahasan mendalam mengenai implikasi dari pola-pola yang ditemukan. Tujuan utama dari tahapan ini adalah untuk mengekstraksi pengetahuan (knowledge discovery) dari basis data pelanggan yang besar agar menjadi informasi baru yang valid dan potensial untuk pengambilan keputusan strategis (F. Liantoni, 2022).

Penentuan Jumlah Cluster Optimal dengan Metode Elbow

Langkah fundamental dalam penerapan algoritma K-Means adalah penentuan jumlah kluster (K) yang paling representatif terhadap struktur data. Pemilihan K yang tepat krusial untuk menghindari overfitting (terlalu banyak kluster yang tidak bermakna) atau underfitting (terlalu sedikit kluster sehingga menggabungkan kelompok yang heterogen).

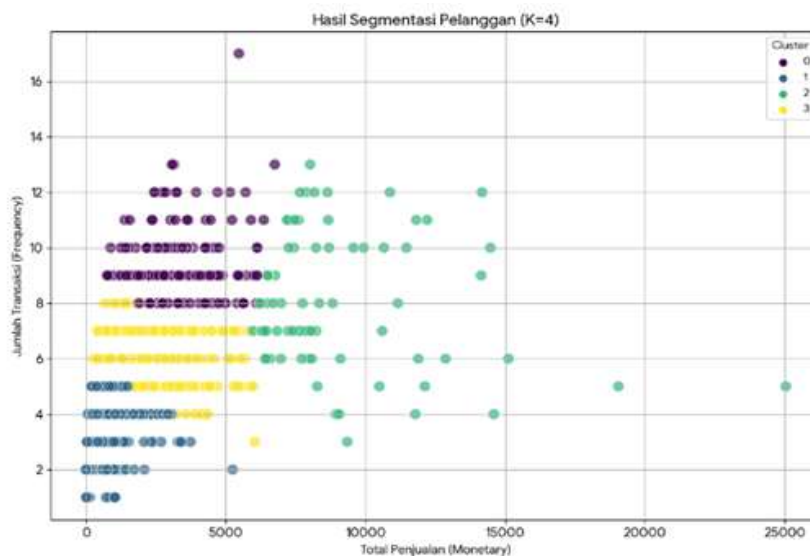
Berdasarkan hasil komputasi yang direpresentasikan pada grafik Metode Elbow, diamati dinamika penurunan nilai inersia sebagai berikut: (1) Fase Penurunan Tajam terjadi saat jumlah kluster ditingkatkan dari K=1 ke K=2, dan berlanjut hingga K=3; (2) Fase Transisi masih terlihat signifikan saat bergerak dari K=3 ke K=4; (3) Fase Konvergensi (Melandai) terjadi setelah melewati titik K=4. Mengacu pada fenomena tersebut, titik K=4 diidentifikasi sebagai "siku" atau titik optimal (N. H. Ahsina, et al., 2022).



Gambar 2. Grafik Evaluasi Jumlah Cluster menggunakan Metode Elbow

Visualisasi Hasil Segmentasi K-Means

Setelah menetapkan parameter $K=4$, algoritma K-Means dijalankan pada dataset yang terdiri dari 793 entitas pelanggan. Data input telah melalui proses pra-pemrosesan dan penskalaan untuk memastikan bahwa setiap atribut, yaitu Monetary (Total Penjualan) dan Frequency (Jumlah Transaksi), memiliki bobot kontribusi yang seimbang dalam perhitungan jarak Euclidean.



Gambar 3. Visualisasi Hasil Segmentasi (K=4) Berdasarkan Atribut Monetary dan Frequency

Dari visualisasi spasial tersebut, terlihat terbentuknya empat partisi data yang distingtif dengan cukup baik, meskipun terdapat beberapa borderline points antar klaster yang berdekatan. Pemisahan visual ini menjadi dasar validasi eksternal bahwa variabel Monetary dan Frequency adalah diskriminator yang efektif dalam membedakan profil pelanggan pada dataset ini.

Profiling dan Analisis Karakteristik Cluster

Untuk menerjemahkan hasil matematis algoritma menjadi wawasan bisnis (actionable insights), dilakukan analisis profil terhadap sentroid dari masing-masing klaster. Hasil rekapitulasi statistik deskriptif untuk setiap klaster disajikan dalam Tabel 1.



Tabel 1. Profil Rata-rata Cluster

Cluster Label	Jumlah Pelanggan	Rata-rata Monetary (Rp)	Rata-rata Jumlah Transaksi (Frequency)	Kategori Segmen
Cluster 2	70	9.184,97	8,01	Superstar Customer
Cluster 0	175	3.413,18	9,39	Loyal Customer
Cluster 3	300	2.572,74	6,11	Potential Customer
Cluster 1	248	1.005,89	3,57	At-Risk Customer

Analisis Cluster 2: The Superstar Customers (Nilai Tinggi, Frekuensi Tinggi)

Klaster ini merupakan segmen yang paling eksklusif, hanya terdiri dari 70 pelanggan atau sekitar 8,8% dari total populasi, namun memiliki dampak finansial yang sangat signifikan. Kelompok ini mencatatkan rata-rata nilai belanja (Monetary) tertinggi sebesar Rp9.184,97 dengan intensitas transaksi rata-rata 8,01 kali. Pelanggan dalam klaster ini dapat dikategorikan sebagai "Paus" (Whales) dalam terminologi bisnis. Dalam konteks Customer Relationship Management (CRM), kelompok ini adalah aset paling berharga; mempertahankan pelanggan tipe ini jauh lebih efisien dibandingkan mengakuisisi pelanggan baru.

Analisis Cluster 0: The Loyal Customers (Nilai Sedang, Frekuensi Tertinggi)

Klaster 0 merepresentasikan kelompok pelanggan yang memiliki tingkat kunjungan paling rutin. Keunikan utama dari klaster ini adalah nilai Frequency rata-rata tertinggi, yaitu 9,39 kali, bahkan sedikit lebih tinggi dibandingkan Superstar Customers. Namun, nilai Monetary mereka berada di tingkat menengah, yaitu Rp3.413,18. Konsistensi belanja mereka menciptakan arus kas (cash flow) yang stabil bagi perusahaan. Berdasarkan konsep Pareto, gabungan Cluster 2 dan Cluster 0 kemungkinan besar menyumbang 80% dari total aktivitas transaksi (Amna et al., 2023).

Analisis Cluster 3: The Potential Customers (Nilai Sedang, Frekuensi Sedang)

Ini adalah segmen dengan populasi terbesar, mencakup 300 pelanggan atau sekitar 37,8% dari total data, dengan rata-rata belanja sebesar Rp2.572,74 dan frekuensi transaksi sebanyak 6,11 kali. Kelompok ini adalah "tulang punggung" volume pelanggan. Potensi pertumbuhan (growth potential) terbesar terletak pada klaster ini karena jumlahnya yang masif; sedikit peningkatan pada frekuensi atau nilai belanja per orang di klaster ini akan memberikan efek multiplikasi yang besar pada total pendapatan.

Analisis Cluster 1: At-Risk / Passive Customers (Nilai Rendah, Frekuensi Rendah)

Klaster ini merupakan kelompok kedua terbesar (248 pelanggan) namun dengan kontribusi ekonomi terendah. Rata-rata Monetary hanya Rp1.005,89 dan frekuensi transaksi sangat rendah di angka 3,57 kali. Pelanggan ini kemungkinan hanya berbelanja saat ada diskon besar, atau pernah mencoba produk sekali lalu tidak kembali lagi. Tingkat keterikatan (engagement) mereka dengan merek sangat lemah dan mereka rentan untuk berpindah ke kompetitor (churn) atau berhenti bertransaksi sama sekali.



Pembahasan Implikasi Manajerial dan Strategi Bisnis

Berdasarkan hasil profilasi di atas, berikut adalah rekomendasi strategi yang diformulasikan untuk masing-masing segmen: (1) Strategi Retensi Eksklusif untuk Cluster 2 (Superstar): membuat program "VIP Membership" yang memberikan privilese khusus seperti layanan prioritas dan akses awal ke produk baru; (2) Strategi Peningkatan Nilai Transaksi untuk Cluster 0 (Loyal): menerapkan Upselling dan Cross-selling, serta mekanisme "bundle pricing" untuk mendorong kenaikan Average Order Value; (3) Strategi Nurturing untuk Cluster 3 (Potential): program gamifikasi dan komunikasi pemasaran yang bersifat persuasif dan rutin; (4) Strategi Re-aktivasi Selektif untuk Cluster 1 (At-Risk): Low-Cost Re-engagement melalui saluran komunikasi murah seperti email otomatis atau notifikasi aplikasi.

KESIMPULAN

Penelitian ini telah berhasil menerapkan algoritma data mining K-Means Clustering untuk melakukan segmentasi pelanggan pada data penjualan ritel dengan menggunakan Superstore Dataset. Berdasarkan hasil analisis menggunakan metode Elbow, jumlah cluster yang paling optimal untuk dataset ini adalah empat. Implementasi K-Means mampu mengidentifikasi segmen pelanggan yang berbeda secara signifikan: (1) pelanggan superstar dengan Monetary dan Frequency sangat tinggi (70 pelanggan, rata-rata Rp9.184,97); (2) pelanggan loyal dengan frekuensi tertinggi namun nilai sedang (175 pelanggan, rata-rata Rp3.413,18); (3) pelanggan potensial dengan nilai dan frekuensi sedang (300 pelanggan, rata-rata Rp2.572,74); dan (4) pelanggan pasif atau berisiko dengan nilai terendah (248 pelanggan, rata-rata Rp1.005,89). Temuan ini menunjukkan bahwa K-Means efektif dalam mengungkap pola perilaku pelanggan yang dapat ditindaklanjuti, sehingga memberikan kontribusi penting bagi perancangan strategi bisnis yang lebih tepat sasaran.

Penelitian selanjutnya disarankan untuk: (1) menambahkan atribut Recency agar analisis model RFM lebih lengkap dan dapat mendeteksi pelanggan yang sudah berhenti berlangganan (churn); (2) menggabungkan data perilaku transaksional dengan data sikap pelanggan seperti tingkat kepuasan, sehingga profil segmen yang dihasilkan lebih holistik; (3) mengeksplorasi algoritma clustering lain seperti DBSCAN atau Hierarchical Clustering untuk membandingkan kinerja; dan (4) menangani outlier ekstrem secara lebih eksplisit sebelum proses klastering untuk memastikan posisi sentroid yang lebih representatif.

DAFTAR PUSTAKA

- A. M. Artiarno, P. Setiaji, and F. Nugraha, "K-Means Clustering untuk Segmentasi Pelanggan: Mengungkap Pola Pembelian Strategi Pemasaran pada Sektor Ritel," *Edumatic J. Pendidikan Informasi*, vol. 9, no. 2, pp. 442–451, 2025.
- Amna et al., *Data Mining*. Padang: PT Global Eksekutif Teknologi, 2023.
- C. H. Ardana, A. A. A. A. Khoyum, and M. Faisal, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024.
- E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, no. September, pp. 122–137, 2023.
- F. Liantoni, *Data Mining Dan Penerapan Metode*, Eureka Media Akasara, 2022.
- F. SLN, *Basic Data Mining from A to Z Dasar Membangun Tindakan Bisnis*. Bandung, 2023.



- I. Lewaaelhamd, "Customer Segmentation Using Machine Learning Model: An Application of RFM Analysis," *Journal of Data Science and Intelligent System*, vol. 2, no. 1, pp. 29–36, 2024.
- I. R. Jawaara and Z. Fatah, "Penggunaan Data Mining Untuk Memprediksi Penjualan Pada Toko Perlengkapan Bangunan Menggunakan Metode Apriori," *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 1, pp. 52–60, 2025.
- I. Yunita, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, "Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran," *Jurnal Media Teknologi dan Sistem Industri*, vol. 9, no. 1, pp. 58-66, 2025.
- M. Annas and S. N. Wahab, "Data Mining Methods: K-Means Clustering Algorithms," *International Journal Cyber and IT Service Management (IJCITSM)*, vol. 3, no. 1, pp. 40–47, 2023.
- M. F. Fadhillah, A. L. A. Suyoso, and I. Puspitasari, "Customer Segmentation with Clustering Algorithm Based on Recency, Frequency, and Monetary (RFM) Attributes," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 48–56, 2025.
- N. H. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 8, no. 3, 2022.
- P. Vania, and B. N. Sari, "Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 547–558, 2023.
- R. Sahoo, "Superstore Sales Dataset." [Online]. Available: <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting> [Accessed: Jan. 1, 2026].
- R. Surya, "Peningkatan Akurasi OCR dalam Pemrosesan Formulir Keuangan melalui Fine-Tuning Transformer dan Strategi Pra-pemrosesan Data," *Jurnal Inovasi Informatika (JII)*, vol. 7, no. 2, pp. 1–12, 2025.
- Y. Yonata, H. Maharani, and C. Viona, "Analisis Clustering Pelanggan Berdasarkan Data Transaksi Penjualan Menggunakan Algoritme K-Means dan Metode Recency, Frequency, Monetary (RFM)," *Jurnal Telematika*, vol. 16, no. 2, pp. 77–84, 2019.