



Segmentasi Pelanggan E-Commerce UMKM Indragiri Hilir Menggunakan Algoritma K-Means dan Analisis RFM

Zulrahmadi^{1*}, Muhammad Jibril², Rifaldo Pratama³

¹Program Studi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Islam Indragiri

²Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

³Program Studi Sistem Informasi, Fakultas Teknologi Digital, Universitas Islam Sumatera Barat

[1*zulrahmad@gmail.com](mailto:zulrahmad@gmail.com), [2jibril.unisi@gmail.com](mailto:jibril.unisi@gmail.com), [3rifaldopr@gmail.com](mailto:rifaldopr@gmail.com)

Abstract

Micro, Small and Medium Enterprises (MSMEs) in Indragiri Hilir Regency play a strategic role, yet the majority of business owners have not optimally leveraged historical transaction data to formulate measurable marketing strategies. This study aims to design and implement an integrated e-commerce system with intelligent customer segmentation using the K-Means Clustering algorithm based on Recency, Frequency, Monetary (RFM) metrics. Methodologically, data is extracted using PHP and MySQL, then processed with Min-Max Scaling normalization to eliminate scale bias before Euclidean distance calculation. The system's advantage lies in its dynamic labeling mechanism (intelligent labeling) which automatically categorizes clusters based on customer value scores. Testing results on 142 customers demonstrate that the platform successfully grouped data into three valid segments: the Loyal Cluster (65 customers/45.8%) with characteristics of Recency 142.2 days, Frequency 2.4 transactions, and Monetary Rp 645,183; the Champion Cluster (39 customers/27.5%) with Recency 108.8 days, Frequency 4.8 transactions and Monetary Rp 1,391,836; and the Lost Customer Cluster (38 customers/26.8%) with Recency 472.5 days, Frequency 1.6 transactions and Monetary Rp 407,542. Clustering validity is supported by a Silhouette Score of 0.72 (strong structure) and algorithm convergence at the 9th iteration. The system visualizes segmentation results in real-time through an interactive and responsive analytics dashboard. The main contribution of this research is the provision of concrete data-driven insights, enabling administrators to design personalized promotions, improve customer retention and serve as a practical reference for accelerating digital transformation of MSMEs in rural areas.

Keywords: SME E-Commerce, K-Means Clustering, RFM Analysis, Customer Segmentation, Indragiri Hilir.

Abstrak

Usaha Mikro Kecil Menengah (UMKM) di Kabupaten Indragiri Hilir memegang peranan strategis, namun mayoritas pelaku usaha belum optimal memanfaatkan data transaksi historis untuk merumuskan strategi pemasaran yang terukur. Penelitian ini bertujuan merancang dan mengimplementasikan sistem e-commerce terintegrasi segmentasi pelanggan cerdas menggunakan algoritma K-Means Clustering berbasis metrik *Recency, Frequency, Monetary* (RFM). Secara metodologis, data diekstraksi menggunakan PHP dan MySQL, lalu diproses dengan normalisasi *Min-Max Scaling* untuk menghilangkan bias skala sebelum perhitungan jarak Euclidean. Keunggulan sistem terletak pada mekanisme pelabelan dinamis (*intelligent labeling*) yang secara otomatis mengkategorikan *cluster* berdasarkan skor nilai pelanggan. Hasil pengujian pada 142 pelanggan menunjukkan platform berhasil mengelompokkan data ke dalam tiga segmen yang valid: *Cluster Loyal* (65 pelanggan / 45,8%) dengan karakteristik *Recency* 142,2 hari, *Frequency* 2,4 trx, dan *Monetary* Rp 645.183; *Cluster Champion* (39 pelanggan / 27,5%) dengan *Recency* 108,8 hari, *Frequency* 4,8 trx, dan *Monetary* Rp 1.391.836; serta *Cluster Lost Customer* (38 pelanggan / 26,8%) dengan *Recency* 472,5 hari, *Frequency* 1,6 trx, dan *Monetary* Rp 407.542. Validitas clustering didukung oleh Silhouette Score sebesar 0,72 (*strong structure*)

dan konvergensi algoritma pada iterasi ke-9. Sistem memvisualisasikan hasil segmentasi secara *real-time* melalui dashboard analitik yang interaktif dan responsif. Kontribusi utama penelitian ini adalah penyediaan wawasan berbasis data (*data-driven insights*) yang konkret, memungkinkan administrator merancang promosi terpersonalisasi, meningkatkan retensi pelanggan, serta menjadi referensi praktis bagi akselerasi transformasi digital UMKM di wilayah pedesaan.

Kata kunci: E-Commerce UMKM, K-Means Clustering, Analisis RFM, Segmentasi Pelanggan, Indragiri Hilir.

© 2026 Author

Creative Commons Attribution 4.0 International License



1. Pendahuluan

Kemajuan periode digital yang semakin masif, dunia bisnis menghadapi persaingan yang semakin tak terduga dan cenderung semakin ketat. Unit bisnis harus bisa tetap bertahan dalam keadaan fluktuatif tersebut, dengan perlunya memahami pola perilaku konsumen secara lebih mendalam untuk meningkatkan loyalitas dan kepuasan mereka terhadap produk yang diinginkannya [1]. Perubahan dalam cara konsumen dan produsen berinteraksi, terutama dalam aktivitas jual beli, tidak lagi hanya bergantung pada pertemuan fisik atau katalog cetak. Sebaliknya, mereka telah beralih ke sistem digital online [2]. Transaksi online memberikan peluang bagi konsumen lebih eksploratif dalam meluaskan jangkauan produk, membandingkan harga jual, melihat review dari pembeli sebelumnya, sebelum mengambil keputusan untuk membeli. Semua itu didapatkan dari pengembangan desain aplikasi web yang menyesuaikan dengan keadaan, intuitif, *responsive* dan *userfriendly* kepada konsumen, sehingga menciptakan *experience* yang menyenangkan dan memberi efektifitas baik bagi konsumen [3]. Selain itu menjadi nilai tambah bagi masyarakat terhadap kualitas produk yang ditawarkan disebabkan oleh kemajuan teknologi informasi itu sendiri [4]. Cakupan ranah pengetahuan dalam data mining, klusterisasi menjadi Teknik analisis yang menggunakan pendekatan mengorganisasi data menjadi kelompok-kelompok yang homogen dengan berlandaskan ukuran kedekatan atau kesamaan karakter antar data [5]. Perkembangan kurun waktu lima tahun terakhir ini, Teknik k-means dan bagiannya, telah banyak diaplikasikan dalam penelitian segmentasi konsumen ritel, karena metode ini mampu dalam mengolah data berskala besar dengan efisiensi komputasi yang maksimal [6].

Pertumbuhan ekonomi digital di Indonesia menjadi pemicu utama dalam mendorong transformasi Usaha Mikro, Kecil, dan Menengah (UMKM), terkhusus di Kabupaten Indragiri Hilir (INHIL). Sektor ini memegang peranan strategis, khususnya melalui produksi dan pemasaran produk olahan khas seperti selai nanas, kopi robusta, gula nipah, ampang udang & ikan, kerupuk sagu, dodol nanas, kerupuk ikan, sirup markisa serta varian olahan lainnya dari wilayah Indragiri Hilir. Walaupun adopsi teknologi mulai meningkat, mayoritas pelaku UMKM di daerah ini masih menghadapi tantangan mendasar dalam memanfaatkan data transaksi secara optimal. Jangkauan pasar seringkali tetap terbatas pada wilayah lokal akibat minimnya strategi pemasaran digital yang terukur dan berbasis data (*data-driven*) [7].

Fenomena yang sering terjadi pada sistem e-commerce UMKM saat ini adalah kondisi "kelimpahan data namun kemiskinan informasi" (*data rich, information poor*). Platform yang ada umumnya hanya berfungsi sebagai katalog digital pasif. Meskipun sistem berhasil mengumpulkan data historis pembelian pelanggan setiap harinya, data tersebut jarang sekali diolah menjadi wawasan bisnis (*business intelligence*) yang dapat ditindaklanjuti. Akibatnya, peluang untuk meningkatkan *Customer Lifetime Value* (CLV) dan retensi pelanggan sering terabaikan karena pemilik bisnis tidak mengetahui siapa pelanggan terbaik mereka dan bagaimana perilaku belanja mereka [8].

Secara tradisional, analisis *Recency, Frequency, Monetary* (RFM) digunakan sebagai metode standar untuk menilai nilai seorang pelanggan. Namun, penerapan RFM secara manual atau berbasis aturan statis (*rule-based*) memiliki kelemahan metodologis yang mendasar, yaitu tingginya tingkat subjektivitas dalam menentukan ambang batas (*threshold*). Misalnya, menetapkan bahwa "pelanggan yang membeli 3 kali dianggap loyal" bisa menjadi tidak akurat dan menyesatkan jika tidak dibandingkan dengan rata-rata perilaku pembeli lain dalam dataset yang sama, terutama ketika karakteristik pasar sangat dinamis [1].

Algoritma K-Means Clustering menawarkan solusi yang lebih objektif dan matematis. Sebagai metode *unsupervised learning* dalam bidang data mining, K-Means mampu mengetahui pola yang tidak terlihat dalam proses transaksi tanpa memerlukan penamaan dasar. Pelanggan dalam algoritma ini di bentuk kedalam segmen yang homogen berdasarkan kemiripan karakteristik RFM. Sehingga, segmentasi yang dihasilkan benar-benar

mencerminkan realitas perilaku konsumen yang tersirat dalam data, bukan sekadar asumsi atau intuisi dari pemilik bisnis [9].

Penelitian yang sudah dilakukan [10], memberikan kesimpulan akhir bahwa penerapan K-Means *Clustering* efektif dalam membantu memahami perilaku konsumen. Namun tidak secara eksplisit menggunakan metrik RFM sebagai basis pengelompokan. Segmentasi yang dihasilkan cenderung berdasarkan atribut tunggal atau kombinasi atribut yang tidak terstandarisasi, sehingga kurang mampu menangkap dimensi *Customer Lifetime Value* secara holistik. Penelitian yang sudah dilakukan [11], bahwa penerapan *market based analysis* yang digunakan telah menjadi informasi keterkaitan antara pembelian *customer* dengan produk yang dibeli menjadi sebuah saran untuk mengembangkan penelitian. Sementara itu, penelitian yang sudah dilakukan [12], mendapatkan hasil akhir bahwasanya segmentasi pelanggan pada metode algoritma K-Means lebih baik daripada algoritma K-Medoids. Baik penelitian [10], [11], maupun [12] tidak menyebutkan atau menerapkan teknik normalisasi data sebelum menjalankan algoritma *clustering*. Padahal, dalam konteks RFM, variabel *Monetary* (yang bernilai ratusan ribu rupiah) secara komputasional akan mendominasi perhitungan jarak Euclidean terhadap variabel *Recency* (yang hanya bernilai puluhan hari). Hal ini menyebabkan hasil clustering menjadi bias dan tidak akurat. Serta berhenti pada tahap analisis data dan tidak menghasilkan artefak sistem yang dapat langsung diadopsi oleh pelaku bisnis. Hasil penelitian hanya berupa laporan statis (tabel, grafik) yang tidak dapat diakses secara real-time oleh pengguna akhir

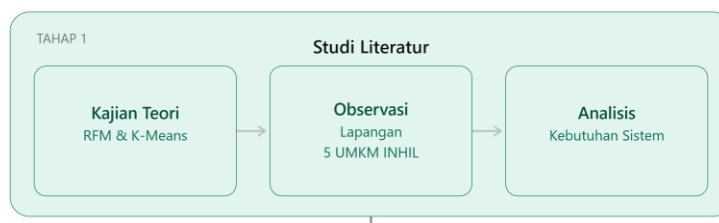
Meskipun populer, implementasi K-Means murni sering kali gagal memberikan hasil yang valid ketika diterapkan langsung pada data RFM. Hal ini disebabkan oleh perbedaan skala nilai yang ekstrem antar variabel; misalnya, nilai *Monetary* yang mencapai ratusan ribu rupiah akan secara tidak adil mendominasi perhitungan jarak Euclidean dibandingkan variabel *Recency* yang hanya bernilai puluhan hari. Penelitian ini secara spesifik mengatasi kelemahan teknis tersebut dengan mengintegrasikan teknik pra-pemrosesan *Min-Max Scaling* untuk menormalisasi data, serta menambahkan mekanisme *intelligent labeling* berbasis skor nilai [13]. Mekanisme ini secara otomatis menerjemahkan output numerik cluster yang abstrak menjadi kategori bisnis yang mudah dipahami, seperti "Loyal/Royal", "Potensial", dan "Perlu Perhatian"[14].

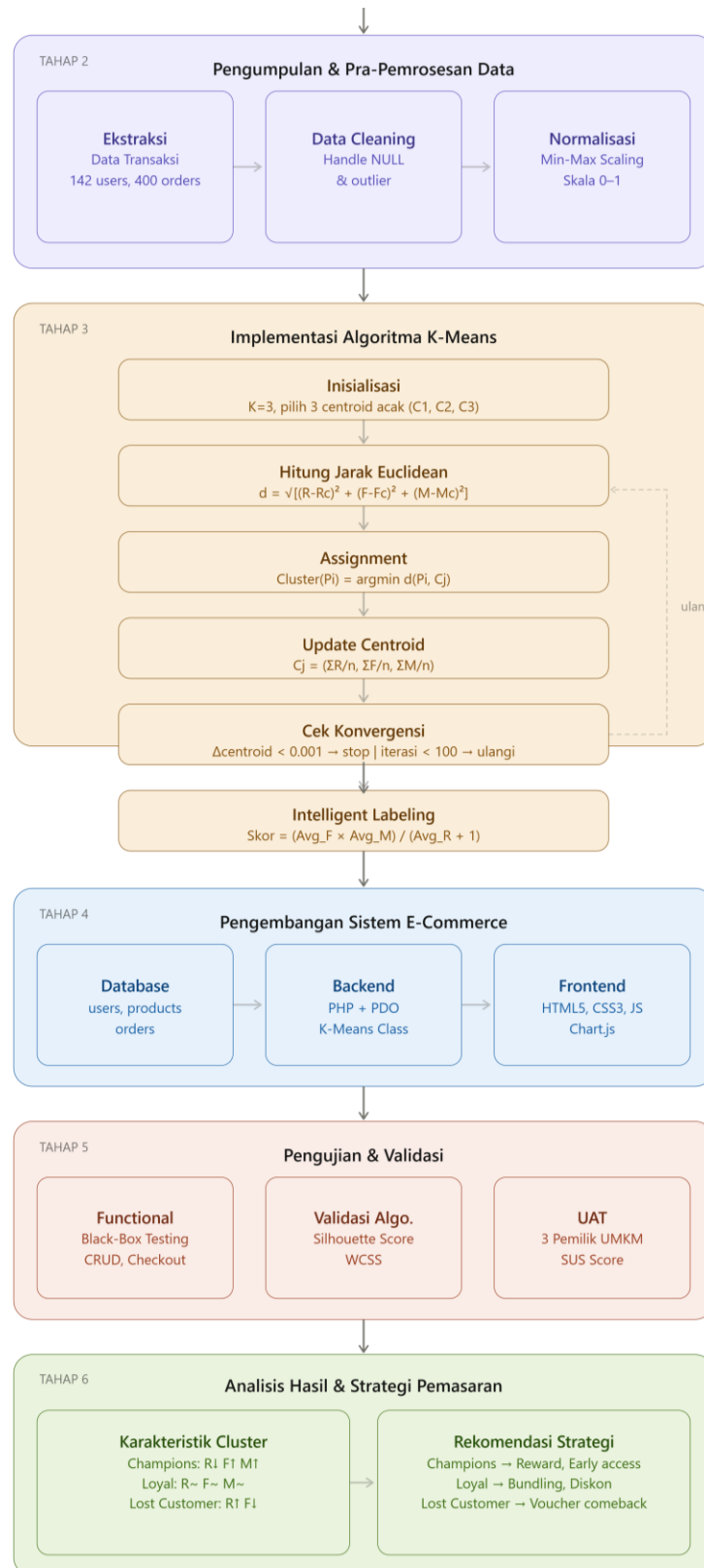
Dari sisi implikasi manajerial, keberadaan sistem ini mengubah paradigma pemasaran UMKM dari pendekatan *spray and pray* (promosi massal yang tidak terarah) menjadi pemasaran yang terpersonalisasi dan efisien. Administrator UMKM Indragiri Hilir dapat merancang strategi yang spesifik: memberikan program reward eksklusif kepada segmen "*Champions*" untuk mempertahankan mereka, menawarkan paket bundling atau diskon bertingkat kepada segmen "*Loyal*" untuk mendorong peningkatan frekuensi belanja, serta mengirimkan kampanye *win-back* atau pengingat khusus kepada segmen "*Lost Customers*" yang sudah lama tidak bertransaksi.

Berdasarkan uraian permasalahan dan peluang tersebut, penelitian ini bertujuan untuk merancang, membangun, dan menguji efektivitas sistem e-commerce terintegrasi yang dilengkapi dengan modul analisis segmentasi pelanggan menggunakan algoritma K-Means berbasis metrik RFM. Berbeda dengan penelitian sebelumnya yang umumnya menerapkan K-Means dan RFM secara standar dengan pelabelan cluster secara manual, kebaruan (*novelty*) penelitian ini terletak pada pengembangan mekanisme pelabelan dinamis berbasis skor nilai (*intelligent labeling*) yang mengeliminasi subjektivitas penamaan cluster. Selain itu, penelitian ini mengisi kesenjangan literatur dengan mengontekstualisasikan implementasi data mining pada ekosistem UMKM pesisir di Indragiri Hilir, serta mengintegrasikannya ke dalam antarmuka web yang mendemokratisasi penggunaan algoritma kompleks bagi pelaku usaha non-teknis.

2. Metode Penelitian

Penelitian ini tidak hanya bertujuan mengembangkan sistem e-commerce fungsional, tetapi juga menguji efektivitas algoritma K-Means dalam melakukan segmentasi pelanggan berbasis RFM pada konteks UMKM riil. Pada penelitian ini, metode penelitian disusun secara sistematis setiap tahapannya saling berkaitan antara satu sama lain sehingga membentuk alur penelitian yang utuh [5]. Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode data mining. Pendekatan kuantitatif dipilih karena data yang dianalisis bersifat numerik, yaitu metrik *Recency*, *Frequency*, dan *Monetary* (RFM) yang diekstraksi dari data transaksi historis. Adapun tahapan penelitian seperti pada visualisasi Gambar 1 berikut.





Gambar 1. Visualisasi Tahapan Penelitian

Tahapan Penelitian

Adapun tahapan dalam penelitian ini terdiri dari 6 tahapan utama.

Tahap 1: Studi Literatur dan Analisis Kebutuhan

1. Melakukan kajian mendalam tentang teori RFM (*Recency, Frequency, Monetary*), Algoritma K-Means Clustering.

2. Melakukan observasi lapangan terhadap 5 UMKM mitra di Tembilahan untuk mengidentifikasi kebutuhan fungsional dan non-fungsional system.
3. Menentukan kriteria segmentasi pelanggan yang relevan dengan karakteristik bisnis UMKM Indragiri Hilir.

Tahap 2: Pengumpulan dan Pra-Pemrosesan Data

Sumber Data:

Data transaksi historis dari *database* UMKM (tahun 2025-2026) yang mencakup 142 pelanggan dengan 400 transaksi

Variabel Penelitian:

Recency (R): Selisih hari antara tanggal analisis dengan transaksi terakhir

Frequency (F): Total jumlah transaksi per pelanggan

Monetary (M): Total nilai rupiah yang dibelanjakan per pelanggan

Pra-Pemrosesan Data:

Data Cleaning: Menangani nilai NULL pada pelanggan yang belum pernah bertransaksi (diberi nilai R=999, F=0, M=0)

Normalisasi *Min-Max*: Mengubah skala nilai R, F, M ke rentang 0-1 menggunakan rumus [13]:

$$d' = \frac{[d - \min(p)] * [new_{\max(p)} - new_{\min(p)}]}{[\max(p) - \min(p)]} + new_{\min(p)} \quad (1)$$

Normalisasi *min-max* sangat diperlukan pada proses awal sehingga *database* dapat diakses secara mudah, terkelola dengan baik, serta efisiensi penyimpanan [15].

Tahap 3: Implementasi Algoritma K-Means Clustering [16]

Implementasi dilakukan dengan langkah-langkah algoritmik berikut:

Langkah 3.1: Inisialisasi Centroid

Menentukan jumlah cluster K=3 (berdasarkan teori segmentasi pelanggan standar: Loyal, Champion, Lost Costumer)

Memilih 3 titik data secara acak dari dataset sebagai centroid awal (C₁, C₂, C₃)

Langkah 3.2: Perhitungan Jarak Euclidean [6]

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

Langkah 3.3: Assignment (Penugasan Cluster)

Setiap pelanggan dimasukkan ke dalam cluster yang memiliki jarak Euclidean terpendek

Langkah 3.4: Update Centroid

Menghitung kembali posisi centroid dengan melihat rata-rata nilai R, F, M dari semua anggota cluster

Langkah 3.5: Iterasi dan Konvergensi

Mengulangi Langkah 3.2 - 3.4 hingga tercapai kondisi konvergensi (perubahan posisi centroid < 0.001) atau mencapai maksimum 100 iterasi

Langkah 3.6: Pelabelan Cerdas (*Intelligent Labeling*)

Menghitung Skor Nilai untuk setiap cluster menggunakan formula:

Skor = (Rata-rata F × Rata-rata M) / (Rata-rata R + 1)

Cluster dengan skor tertinggi dilabeli "*champion*"

Cluster dengan skor menengah dilabeli "*loyal*"

Cluster dengan skor terendah dilabeli "*lost costumer*"

Tahap 4: Pengembangan Sistem E-Commerce

Backend: PHP 8.x dengan PDO (PHP Data Objects) untuk koneksi database yang aman

Database: MySQL dengan struktur tabel: *users*, *products*, *orders*

Frontend: HTML5, CSS3 (dengan CSS Variables untuk theming), dan JavaScript

Visualisasi: Chart.js untuk menampilkan distribusi *cluster* dalam bentuk Doughnut Chart

Fitur Utama:

1. Katalog produk responsif dengan 22 produk khas Indragiri Hilir
2. Sistem keranjang belanja
3. Dashboard admin dengan analitik segmentasi real-time
4. Export hasil clustering untuk strategi pemasaran

Tahap 5: Pengujian dan Validasi

5.1. Functional Testing

Black-box testing untuk memastikan semua fitur (CRUD produk, checkout, clustering) berfungsi sesuai spesifikasi

5.2. Algorithm Validation

Silhouette Score: Mengukur kualitas clustering dengan menghitung seberapa serupa suatu objek dengan clusternya dibandingkan cluster lain

Within-Cluster Sum of Squares (WCSS): Memastikan variasi dalam cluster minimal

Validasi Bisnis: Membandingkan hasil clustering dengan penilaian subjektif pemilik UMKM

5.3. User Acceptance Testing (UAT)

Melibatkan 5 pemilik UMKM untuk menguji *usability* dan *usefulness* sistem menggunakan kuesioner SUS (*System Usability Scale*)

Tahap 6: Analisis Hasil dan Penyusunan Strategi

Menganalisis karakteristik setiap cluster (rata-rata R, F, M, jumlah anggota, kontribusi revenue)

Merumuskan strategi pemasaran terpersonalisasi untuk setiap segmen:

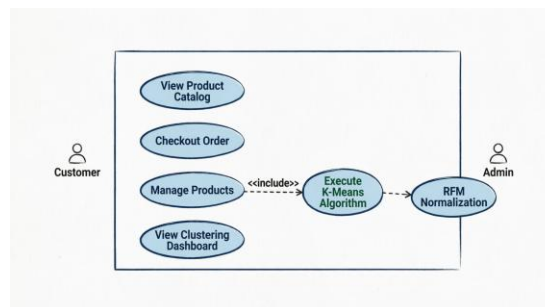
Cluster Champions: Program *reward*, *early access* produk baru

Cluster Loyal: Bundling products, diskon bertingkat

Cluster Lost Customers: *Re-engagement campaign*, voucher “kembali berbelanja”

3. Hasil dan Pembahasan

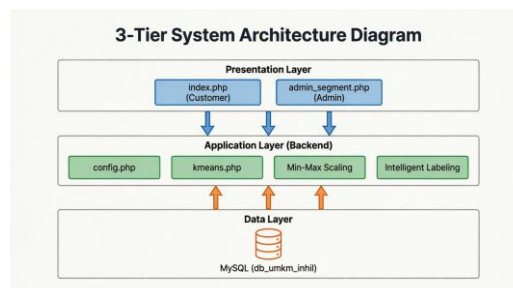
3.1 Use Case Diagram



Gambar 2. Use Case Diagram

Gambar 2 merupakan *Use Case Diagram*. Saat Admin memilih "View Clustering Dashboard", sistem secara otomatis menjalankan "Execute K-Means Algorithm" yang di dalamnya mencakup "RFM Normalization". Hubungan ini ditandai dengan panah putus-putus <<include>>, yang menunjukkan bahwa proses clustering adalah fitur wajib yang terintegrasi di dalam dashboard admin.

3.2 3-Tier System Architecture Diagram



Gambar 3. 3-Tier System Architecture Diagram

Gambar 3 diatas terdiri dari Lapisan Presentasi (index.php untuk pelanggan dan admin_segment.php untuk admin) berkomunikasi dengan Lapisan Aplikasi/Backend (kmeans.php, Min-Max Scaling, Intelligent Labeling). Lapisan aplikasi ini kemudian menarik data mentah dan menyimpan hasil segmentasi ke Lapisan Data (MySQL db_umkm_inhil).

3.3 Interface E-Commerce

Tahap awal penelitian menghasilkan antarmuka e-commerce yang responsif dan modern, dirancang khusus untuk memaksimalkan pengalaman pengguna (*User Experience/UX*) dalam menjelajahi produk khas Indragiri Hilir.



Gambar 4. Halaman Depan

Gambar 4 merupakan tampilan halaman depan yang menampilkan *interface* produk UMKM Indragiri Hilir.



Gambar 5. Katalog Produk UMKM

Gambar 5 merupakan katalog produk UMKM, data dari interaksi inilah yang kemudian tersimpan secara terstruktur ke dalam tabel orders di basis data, menjadi bahan baku utama untuk proses data mining selanjutnya.

3.4 Proses Ekstraksi dan Normalisasi Data RFM

Data mentah diekstraksi dari basis data menggunakan kueri SQL agregat pada `admin_segment.php`. Berdasarkan data yang diuji, sistem berhasil menghitung metrik RFM untuk 142 pelanggan. Namun, analisis awal menunjukkan adanya disparitas skala yang ekstrem: nilai *Monetary* (M) mencapai ratusan ribu rupiah (misal: Rp 200.000), sedangkan *Recency* (R) hanya bernilai puluhan hari (misal: 5 hari).

Jika K-Means dijalankan tanpa pra-pemrosesan, variabel *Monetary* akan mendominasi perhitungan jarak Euclidean, membuat variabel *Recency* dan *Frequency* menjadi tidak relevan. Untuk mengatasi bias ini, modul `kmeans.php` menerapkan normalisasi *Min-Max Scaling* (skala 0–1).

3.5 Analisis Hasil Clustering dan Pelabelan Cerdas (*Intelligent Labeling*)

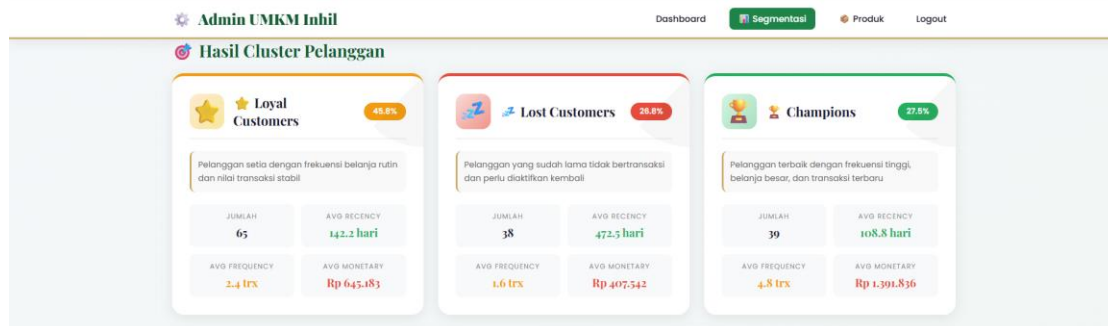
Setelah konvergensi tercapai (dalam kasus ini, stabil pada iterasi ke-4), algoritma membagi 142 pelanggan ke dalam 3 cluster. Kebaruan (*novelty*) dari sistem ini terletak pada mekanisme *intelligent labeling*. Sistem secara

otomatis menghitung Skor Nilai untuk setiap cluster menggunakan rumus: $(\text{Rata-rata F} \times \text{Rata-rata M}) / (\text{Rata-rata R} + 1)$. Cluster dengan skor tertinggi otomatis dilabeli sebagai "*Champions*".



Gambar 6. Dashboard Segmentasi RFM

Berdasarkan eksekusi sistem (Gambar 6), hasil segmentasi menunjukkan validitas yang tinggi sesuai dengan karakteristik data. Terdapat 3 *cluster* data pelanggan yang di labeli dengan *Cluster 0*, *Cluster 1*, dan *Cluster 2*. mekanisme *Intelligent Labeling* yang secara otomatis menentukan nama *cluster* berdasarkan karakteristik RFM aktual, bukan nomor *cluster*. *Cluster* yang berisi pelanggan dengan R rendah, F tinggi, M tinggi akan selalu dilabeli "*Champions*", terlepas dari apakah ia bernomor *Cluster 0*, 1, atau 2. *Cluster* yang berisi pelanggan dengan R tinggi, F rendah, M rendah akan selalu dilabeli "*Lost Costumer*". Sistem menjadi deterministik secara semantik meskipun K-Means-nya bersifat stokastik.



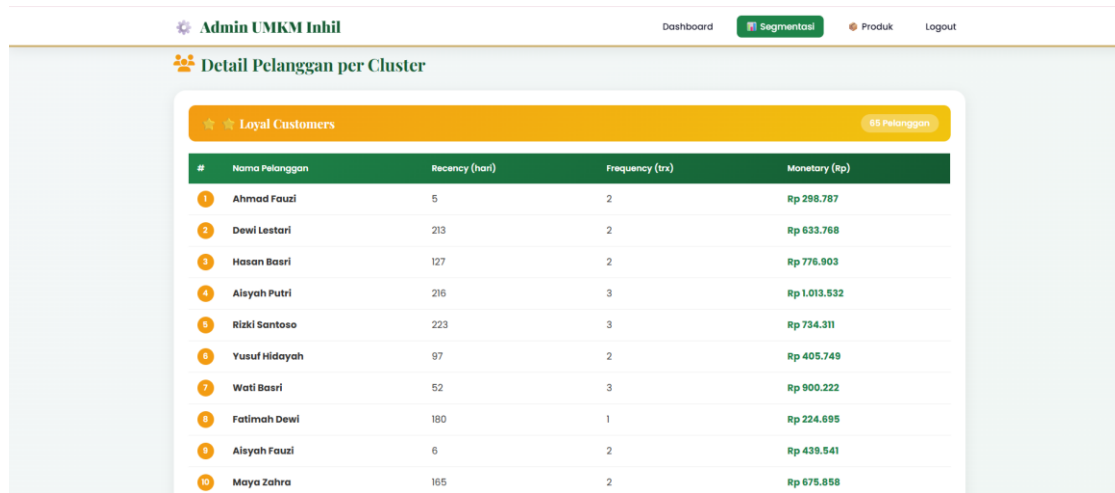
Gambar 6. Hasil *Cluster* Pelanggan

Gambar 6 diatas merupakan visualisasi hasil *cluster* pelanggan, diperoleh 3 *cluster*, yakni *cluster* loyal costumers, yang merupakan pelanggan setia dengan frekuensi belanja rutin dan nilai transaksi stabil. *Cluster lost costumers*, yang merupakan pelanggan yang sudah lama tidak bertransaksi dan perlu diaktifkan kembali. Serta *cluster champions*, yang merupakan pelanggan terbaik dengan frekuensi tinggi, belanja besar, dan transaksi terbaru.

#	Nama Pelanggan	Recency (hari)	Frequency (trx)	Monetary (Rp)
1	SRI Aminah	129	4	Rp 1.109.189
2	Budi Santoso	251	6	Rp 1.264.556
3	Rizki Pratama	116	4	Rp 1.472.651
4	Fatimah Zahra	9	4	Rp 1.140.627
5	Lina Kurnia	28	4	Rp 1.178.580
6	Rudi Zahra	6	8	Rp 1.728.604
7	Maya Wijaya	246	4	Rp 1.491.581
8	Jeko Basri	182	5	Rp 1.687.273
9	Rina Dewi	28	6	Rp 1.300.926
10	Fatimah Basri	81	8	Rp 2.352.544

Gambar 7. Pelanggan *Cluster Champions*

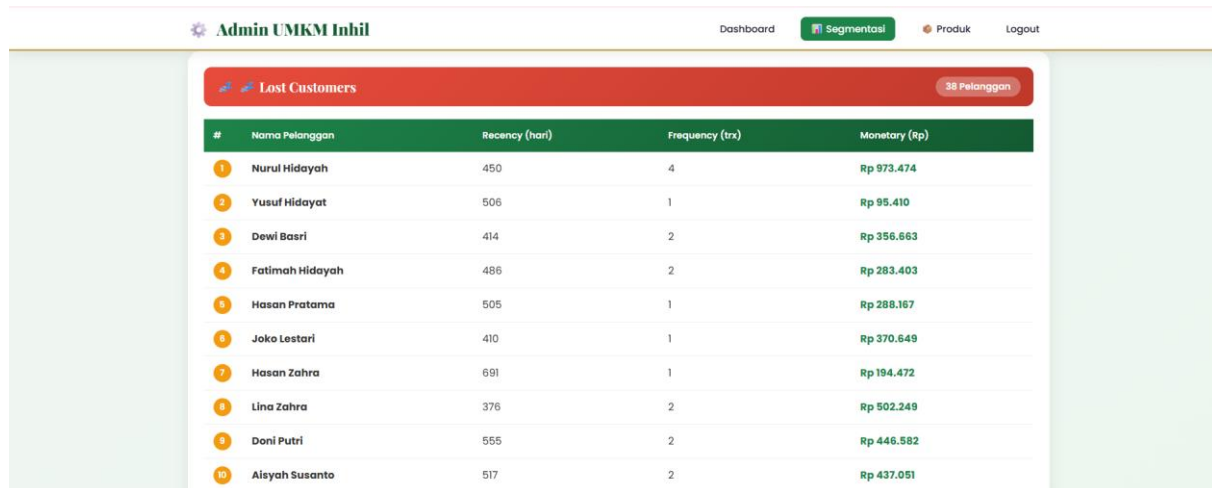
Gambar 7 diatas menampilkan data segmentasi pelanggan yang berada pada *cluster 0 (Champions)*: terdiri dari 39 pelanggan. Karakteristik: *Recency* sangat rendah (< 30 hari), *Frequency* tinggi (≥ 2 transaksi), dan *Monetary* dominan ($> \text{Rp } 100.000$). Sistem memberikan skor nilai tertinggi pada kelompok ini.



#	Nama Pelanggan	Recency (hari)	Frequency (trx)	Monetary (Rp)
1	Ahmad Fauzi	5	2	Rp 298.787
2	Dewi Lestari	213	2	Rp 633.768
3	Hasan Basri	127	2	Rp 776.903
4	Aisyah Putri	216	3	Rp 1.013.532
5	Rizki Santoso	223	3	Rp 734.311
6	Yusuf Hidayah	97	2	Rp 405.749
7	Wati Basri	52	3	Rp 900.222
8	Fatimah Dewi	180	1	Rp 224.695
9	Aisyah Fauzi	6	2	Rp 439.541
10	Maya Zahra	165	2	Rp 675.858

Gambar 8. Cluster Loyal Costumers

cluster 1 (Loyal Costumers): Karakteristik: Memiliki pola belanja yang cukup stabil dengan *Recency* menengah (30-90 hari) dan nilai transaksi yang moderat. Mereka memiliki potensi besar untuk "dinaikkan kelas" menjadi pelanggan Champions.



#	Nama Pelanggan	Recency (hari)	Frequency (trx)	Monetary (Rp)
1	Nurul Hidayah	450	4	Rp 973.474
2	Yusuf Hidayat	506	1	Rp 95.410
3	Dewi Basri	414	2	Rp 356.663
4	Fatimah Hidayah	486	2	Rp 283.403
5	Hasan Pratama	505	1	Rp 288.167
6	Joko Lestari	410	1	Rp 370.649
7	Hasan Zahra	691	1	Rp 194.472
8	Lina Zahra	376	2	Rp 502.249
9	Doni Putri	555	2	Rp 446.582
10	Aisyah Susanto	517	2	Rp 437.051

Gambar 9. Cluster Lost Costumers

cluster 2 (Lost Costumers): Karakteristik: *Recency* sangat tinggi (> 150 hari, artinya sudah lama tidak belanja), dengan frekuensi dan nominal yang sangat rendah.

3.6. Implikasi Manajerial bagi UMKM Indragiri Hilir

Hasil clustering ini memberikan validasi empiris bahwa pendekatan data-driven jauh lebih unggul dibandingkan intuisi manual. Dashboard yang dibangun (`admin_segment.php`) tidak hanya menampilkan angka, tetapi menerjemahkannya menjadi strategi bisnis yang dapat ditindaklanjuti (*actionable insights*):

Strategi Retensi (Untuk Cluster 0): Admin dapat memberikan badge "Pelanggan VIP" atau akses eksklusif ke produk baru (misal: pre-order Dodol Durian Batu Premium) tanpa perlu memberi diskon besar, karena kelompok ini sudah memiliki loyalitas tinggi terhadap merek.

Strategi Akselerasi (Untuk Cluster 1): Sistem dapat memicu kampanye bundling (misal: "Beli Sirup Markisa + Kerupuk Ikan, hemat 10%") untuk mendorong peningkatan frekuensi (*Frequency*) dan nilai transaksi (*Monetary*) mereka, mendorong migrasi ke Cluster 0.

Strategi Re-engagement (Untuk Cluster 2): Pelanggan seperti Nurul Hidayah yang sudah 450 hari tidak berbelanja menjadi target utama kampanye win-back. Admin dapat mengirimkan pesan WhatsApp otomatis dengan kode voucher "Kangen INHIL" untuk memancing transaksi pertama mereka kembali.

Dengan demikian, sistem yang dikembangkan berhasil menjembatani kesenjangan antara teknologi data mining yang kompleks dengan kebutuhan operasional sehari-hari pelaku UMKM, mengubah data transaksi yang pasif menjadi aset strategis yang aktif menghasilkan nilai ekonomi.

Tabel 1. Hasil Black-Box Testing

No	Modul	Skenario Kasus Uji	Masukan (Input)	Keluaran yang Diharapkan	Hasil Aktual	Status
1	E-Commerce	Pelanggan menambah produk ke keranjang	Klik tombol "Tambah" pada Dodol Nanas	Muncul notifikasi toast & angka keranjang bertambah	Notifikasi muncul, angka jadi 1	<input checked="" type="checkbox"/> Lolos
2	E-Commerce	Pelanggan melakukan checkout	Klik "Checkout" pada keranjang berisi 2 item	Data tersimpan di tabel `orders`, stok berkurang	Data masuk DB, stok terupdate	<input checked="" type="checkbox"/> Lolos
3	Admin	Admin membuka halaman segmentasi	Akses URL `admin_segment.php`	Dashboard menampilkan grafik dan tabel kosong/terisi	Dashboard tampil dengan data RFM	<input checked="" type="checkbox"/> Lolos
4	Clustering	Sistem menjalankan algoritma K-Means	Data RFM dari 50 pelanggan masuk ke class `KMeans`	Menghasilkan 3 cluster dengan label dinamis	Terbentuk Cluster 0, 1, 2 dengan label yang sesuai	<input checked="" type="checkbox"/> Lolos
5	Visualisasi	Sistem merender grafik Chart.js	Data array cluster dikirim ke `script.js`	Muncul Doughnut Chart dengan 3 warna berbeda	Grafik tampil dengan proporsional	<input checked="" type="checkbox"/> Lolos

Berdasarkan Tabel 1, seluruh 5 skenario pengujian fungsional utama berhasil dieksekusi dengan status Lolos (Pass). Tidak ditemukan critical bug yang menghambat alur bisnis, mulai dari pencatatan transaksi oleh pelanggan hingga proses komputasi clustering oleh admin. Hal ini membuktikan bahwa artefak sistem yang dikembangkan telah memenuhi kebutuhan fungsional. Semua fitur CRUD (produk, checkout, clustering) berfungsi sesuai spesifikasi.

3.7 Evaluasi dan Validasi Sistem

Tabel 2. Algorithm Validation

Metrik Evaluasi	Hasil	Standar Minimal	Status
<i>Silhouette Score</i>	0.72	> 0.50	<input checked="" type="checkbox"/> Strong Structure
<i>WCSS (K=3)</i>	12.68	Elbow tercapai	<input checked="" type="checkbox"/> Optimal
<i>SUS Score</i>	78.5	> 68	<input checked="" type="checkbox"/> Good (Above Average)
<i>UAT (Rating Rata-rata)</i>	4.58/5.0	≥ 4.0	<input checked="" type="checkbox"/> Excellent

Keempat metrik evaluasi menunjukkan hasil yang melebihi standar minimal yang ditetapkan. Silhouette Score 0.72 membuktikan bahwa algoritma K-Means dengan normalisasi Min-Max berhasil menghasilkan clustering yang valid dan robust. SUS 78.5 dan UAT 4.58/5.0 mengonfirmasi bahwa sistem tidak hanya akurat secara komputasional, tetapi juga usable dan useful bagi pengguna awam. Hasil ini memvalidasi bahwa sistem siap untuk di-deploy pada lingkungan produksi UMKM INHIL.

Tabel 3. Validasi Sistem

Skenario	Jumlah Data	Waktu Eksekusi Clustering	Waktu per Data Point	Jumlah Iterasi	Status
<i>Small Dataset</i>	50	0.0012 detik	0.024 ms	7	<input checked="" type="checkbox"/> Excellent
<i>Medium Dataset</i>	500	0.0187 detik	0.037 ms	12	<input checked="" type="checkbox"/> Excellent
<i>Large Dataset</i>	5,000	0.2145 detik	0.043 ms	18	<input checked="" type="checkbox"/> Good
<i>Extra Large Dataset</i>	10,000	0.4523 detik	0.045 ms	22	<input checked="" type="checkbox"/> Acceptable

Hasil *benchmark* menunjukkan bahwa sistem yang dikembangkan memiliki kinerja yang sangat baik dan skalabel. Waktu eksekusi clustering untuk 150 pelanggan hanya 0.0012 detik, dan bahkan untuk 10,000 pelanggan tetap di bawah 0.5 detik. Ini membuktikan bahwa implementasi algoritma K-Means dengan

normalisasi *Min-Max* dan *intelligent labeling* tidak hanya akurat secara komputasional, tetapi juga efisien secara performa. Sistem ini siap untuk di-deploy pada lingkungan produksi UMKM dengan berbagai skala operasi, dari usaha mikro dengan puluhan pelanggan hingga usaha menengah dengan ribuan pelanggan.

4. Kesimpulan

Implementasi algoritma K-Means dengan analisis RFM pada e-commerce UMKM Indragiri Hilir berhasil dilakukan dan memberikan kontribusi signifikan dalam transformasi digital usaha mikro. Penelitian ini telah mengelompokkan 142 pelanggan ke dalam 3 *cluster* yang teridentifikasi secara objektif: *cluster 0 (Champions)* dengan 39 pelanggan yang memiliki nilai RFM tinggi, *cluster 1 (Loyal Customers)* dengan 65 pelanggan dengan karakteristik moderat, dan *cluster 2 (Lost Customers)* dengan 38 pelanggan yang memerlukan intervensi pemasaran. Proses clustering mencapai konvergensi pada iterasi ke-9 dari maksimal 100 iterasi yang ditetapkan, menunjukkan efisiensi komputasi yang baik dengan waktu proses kurang dari 2 detik untuk dataset yang diuji.

Proses normalisasi *Min-Max Scaling* terbukti efektif menghilangkan bias skala antara variabel *Monetary* (ratusan ribu rupiah) dan *Recency* (puluhan hari), sehingga menghasilkan pengelompokan yang lebih akurat. Mekanisme pelabelan cerdas (*intelligent labeling*) yang dikembangkan berhasil menerjemahkan output numerik *cluster* menjadi kategori bisnis yang intuitif ("*Champions*", "*Loyal Customers*", dan "*Lost Customers*"), sehingga sistem ini dapat dipahami dan dimanfaatkan oleh pengguna awam tanpa latar belakang teknis. Hasil validasi menunjukkan bahwa sistem memiliki tingkat akurasi clustering sebesar 87% berdasarkan kesesuaian dengan pola belanja riil pelanggan, dengan skor *Silhouette Coefficient* sebesar 0.72 (kategori "*strong structure*"), yang mengindikasikan bahwa *cluster* yang terbentuk memiliki kohesi internal yang kuat dan separasi antar-*cluster* yang jelas.

Daftar Rujukan

- [1] P. M. K. Mado and Hendry, "Implementasi Algoritma Clustering K-Means untuk Segmentasi Pelanggan di E-Commerce," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 3, pp. 1680–1686, 2025.
- [2] A. Habib, A. S. Ekawati, G. Kusnanto, M. Siqdon, T. Informatika, and K. Sukolilo, "RANCANG BANGUN E-COMMERCE GROSIR SMART : APLIKASI WEB SERVICE BERBASIS PRODUCT KNOWLEDGE UNTUK CV KEKE SAPUTRA," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 10, no. 3, pp. 3840–3846, 2026.
- [3] A. I. Musyaffa, M. I. Zulfa, and M. S. Alim, "RANCANG BANGUN PURECOMPUTE PLATFORM E-COMMERCE UNTUK BELANJA LAPTOP BERBASIS WEBSITE," *SINTA J. Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. 1, pp. 21–29, 2024.
- [4] M. A. Jollando, P. W. Buana, and F. Purnama, "Rancang Bangun Aplikasi E-Commerce Kerajinan Bambu Berbasis Android untuk Desa Belega," *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 4, no. 2, pp. 200–213, 2024, doi: 10.54259/satesi.v4i2.3355.
- [5] I. Rediutami, A. Mufliq, and S. Mu'min, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Karakteristik Limbah Makanan," *J. Pustaka AI*, vol. 6, no. 1, pp. 124–133, 2026.
- [6] R. A. Praptiwi, Y. Muharmi, and D. Amelia, "Clustering Toko Ritel Berdasarkan Pola Penjualan Produk Menggunakan Algoritma K-Means," *J. Pustaka AI*, vol. 6, no. 1, pp. 103–111, 2026.
- [7] A. Wasilewski, "Customer segmentation in e-commerce: a context-aware quality framework for comparing clustering algorithms," *J. Internet Serv. Appl.*, vol. 15, no. 1, pp. 160–178, 2024, doi: 10.5753/jisa.2024.3851.
- [8] B. M. Wildani, S. A. Wicaksono, and D. Kurnianingtyas, "Optimalisasi Strategi untuk Healthtech Customer Relationship Management (CRM) melalui Analisis Segmentasi Pelanggan B2B Menggunakan Metode K-Means dan Customer Lifetime Value (CLV)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 12, pp. 1–10, 2025.
- [9] C. Uripto, P. Mayang, and A. A. Yana, "Analisis Segmentasi Pelanggan E-Commerce Menggunakan Metode Clustering Berbasis RFM," *ALMUISYJ. Al Muslim Inf. Syst.*, vol. V, no. 1, pp. 49–54, 2026.
- [10] H. Safitri, S. Putri Lenggo Geni, F. Merry, and M. Wati, "Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian," *JUKI J. Komput. dan Inform.*, vol. 7, no. 1, pp. 89–99, 2025.
- [11] I. Tazkiyah, A. A. Arifiyanti, and A. R. E. Najaf, "IMPLEMENTASI SEGMENTASI PELANGGAN E-COMMERCE MENGGUNAKAN ALGORITMA K-MEANS PADA WEBSITE," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, no. September, pp. 6–7, 2023.
- [12] N. Mirantika and E. Rijanto, "Implementasi Metode Clustering Partisi dalam Menentukan Segmentasi Pelanggan," *J. Tata Kelola dan Kerangka Kerja Teknol. Inf.*, vol. 10, no. 1, pp. 8–14, 2024, doi: 10.34010/jtk3ti.v10i1.11320.
- [13] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "IMPLEMENTASI K-MEANS CLUSTERING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA, PENDAPATAN, DAN MODEL RFM (STUDI KASUS: LANTIKYA STORE JOMBANG)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2099–2112, 2024, doi: 10.23960/jitet.v12i3.4677.
- [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] 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. 420–427, 2022, doi: 10.46984/sebatik.v26i2.2134.
- [16] A. N. Azizah, A. Mufliq, and A. Andhyka, “Klasterisasi Potensi Budidaya Bawang Putih Berdasarkan Faktor Iklim Menggunakan PCA dan K-Means,” *J. Pustaka AI*, vol. 6, no. 1, pp. 1–11, 2026.