

ANALISIS PERBANDINGAN ALGORITMA K-MEANS DAN FUZZY C-MEANS DALAM SEGMENTASI POLA TRANSAKSI PELANGGAN E-COMMERCE BERDASARKAN MODEL RECENCY, FREQUENCY, AND MONETARY (RFM)

Gede Erik Aktama

Universitas Parna Raya

gedeerikaktama@gmail.com

Abstrak

Perkembangan pesat sektor perdagangan elektronik (e-commerce) menuntut para pelaku usaha, khususnya Usaha Mikro, Kecil, dan Menengah (UMKM), untuk menerapkan strategi pemasaran berbasis data yang presisi dan efisien. Pemasaran massal yang tidak tersegmentasi terbukti tidak efektif dan memicu lonjakan biaya akuisisi serta retensi pelanggan. Penelitian ini bertujuan untuk melakukan analisis komparatif antara dua algoritma pengelompokan terkemuka, yaitu K-Means sebagai representasi hard clustering dan Fuzzy C-Means (FCM) sebagai soft clustering, dalam memetakan segmen perilaku transaksi pelanggan berdasarkan model Recency, Frequency, and Monetary (RFM). Penelitian memanfaatkan data transaksi aktual 250 pelanggan UMKM lokal di Kota Manado selama periode observasi satu tahun (Juli 2023 hingga Juni 2024). Pra-pemrosesan data dilakukan melalui penanganan nilai hilang, deteksi outlier, serta normalisasi fitur menggunakan Min-Max Scaler. Penentuan jumlah kluster optimal dilakukan melalui integrasi Elbow Method dan evaluasi Silhouette Score yang mengonfirmasi bahwa konfigurasi $k = 4$ merupakan partisi paling ideal. Evaluasi kualitas kluster diukur menggunakan dua metrik validasi internal, yaitu Silhouette Coefficient dan Davies-Bouldin Index (DBI). Hasil eksperimen komputasi menunjukkan bahwa algoritma Fuzzy C-Means menghasilkan kualitas pengelompokan yang lebih unggul dengan Silhouette Coefficient sebesar 0,628 dan nilai DBI sebesar 0,672, dibandingkan K-Means yang memperoleh Silhouette Coefficient sebesar 0,582 dan DBI sebesar 0,741. Kemampuan FCM dalam mengakomodasi derajat keanggotaan bertingkat (membership degree) terbukti efektif dalam mengurangi ketidakpastian pada data perbatasan antar kluster. Berdasarkan hasil segmentasi, terbentuk empat kluster perilaku pelanggan yang terdefinisi dengan jelas: Champions (14,4%), Potential Loyalists (28,0%), Promising Customers (32,8%), dan At-Risk Customers (24,8%). Penelitian ini memberikan kontribusi teoritis dalam pemodelan data mining transaksional serta menyediakan arahan manajerial praktis bagi pengambil keputusan dalam merancang program promosi yang terpersonalisasi.

Kata kunci: Segmentasi Pelanggan, E-Commerce, Model RFM, K-Means, Fuzzy C-Means, Silhouette Coefficient, Davies-Bouldin Index.

Abstract

The rapid expansion of the electronic commerce (e-commerce) landscape necessitates businesses, particularly Micro, Small, and Medium Enterprises (MSMEs), to deploy precise, data-driven marketing paradigms. Unsegmented mass promotional campaigns have proven economically inefficient, inducing high customer acquisition and retention overheads. This study aims to conduct a comparative empirical investigation between two prominent clustering techniques: K-Means (representing crisp/hard clustering) and Fuzzy C-Means (FCM, representing soft clustering), in segmenting customer purchasing patterns based on the Recency, Frequency, and Monetary (RFM) behavioral model. The study analyzes transactional logs comprising 250 active customers from local retail e-commerce in Manado City across a twelve-month observation timeframe (July 2023 to June 2024). Data preprocessing incorporated data cleansing, extreme value trimming, and feature standardization via Min-Max Normalization. Optimal cluster quantity was derived using an integration of the Elbow Method and Silhouette evaluations, which verified $k = 4$ as the most structurally robust configuration. Clustering performance was rigorously evaluated using internal validation metrics: the Silhouette Coefficient and the Davies-Bouldin Index (DBI). Computational findings indicate that Fuzzy C-Means outperformed K-Means by achieving a superior Silhouette Coefficient of 0.628 and a lower DBI of 0.672, compared to K-Means which obtained a Silhouette Coefficient of 0.582 and a DBI of 0.741. FCM's mathematical flexibility in attributing partial membership grades mitigated boundary ambiguities inherent in transactional behavioral data.

Four distinctive customer typologies were successfully mapped: Champions (14.4%), Potential Loyalists (28.0%), Promising Customers (32.8%), and At-Risk Customers (24.8%). The research offers substantive theoretical insights into transaction-oriented data mining and establishes pragmatic managerial frameworks for implementing targeted customer relationship management.

Keywords: Customer Segmentation, E-Commerce, RFM Framework, K-Means Clustering, Fuzzy C-Means, Silhouette Metric, Davies-Bouldin Index.

1. PENDAHULUAN

Akselerasi transformasi digital dalam lanskap ekonomi modern telah mengubah pola interaksi transaksi secara fundamental. Kehadiran platform niaga elektronik (e-commerce) dan media sosial sebagai kanal pemasaran terpadu telah membuka peluang sekaligus menciptakan dinamika persaingan yang ketat, khususnya bagi pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) di tingkat daerah seperti Kota Manado (Aktama et al., 2024; Kalubota et al., 2024). Dalam lingkungan digital yang sangat dinamis, konsumen dihadapkan pada pilihan produk dan layanan yang sangat melimpah, yang pada gilirannya menyebabkan tingkat loyalitas konsumen cenderung menurun dan biaya perpindahan merek (switching cost) menjadi semakin rendah (Turban et al., 2018). Di sisi lain, banyak pelaku usaha masih mengandalkan pendekatan pemasaran massal (undifferentiated mass marketing), di mana promosi disebarluaskan secara seragam kepada seluruh basis data pelanggan tanpa mempertimbangkan perbedaan karakteristik, preferensi, maupun nilai kontribusi transaksi individu (Kotler & Keller, 2016).

Pemasaran massal tersebut tidak hanya mengakibatkan rendahnya rasio konversi (conversion rate), melainkan juga memicu pemborosan anggaran promosi dan meningkatkan customer acquisition cost (CAC). Oleh sebab itu, penerapan strategi Customer Relationship Management (CRM) analitis melalui teknik segmentasi pelanggan menjadi prasyarat krusial (Peppers & Rogers, 2017). Segmentasi pelanggan memungkinkan pelaku bisnis mengidentifikasi kelompok pelanggan yang memiliki kesamaan pola transaksi sehingga strategi retensi, insentif promosi, dan komunikasi pemasaran dapat dipersonalisasi secara tepat sasaran (Chen et al., 2020).

Dalam domain analitik data bisnis, salah satu kerangka kerja paling mapan untuk mengkuantifikasi perilaku pelanggan adalah model Recency, Frequency, and Monetary (RFM) yang dikembangkan oleh Hughes (1994). Model ini mengevaluasi pelanggan berdasarkan tiga dimensi temporal dan finansial utama: Recency (waktu yang telah berlalu sejak transaksi terakhir), Frequency (banyaknya frekuensi transaksi dalam kurun waktu tertentu), dan Monetary (total nilai nominal pengeluaran yang dihabiskan). Integrasi model RFM dengan algoritma data mining telah terbukti mampu memberikan wawasan mendalam mengenai nilai seumur hidup pelanggan (Customer Lifetime Value) dan membedakan pelanggan bernilai tinggi dari pelanggan yang berisiko churn (Kumar & Reinartz, 2018; Monalisa et al., 2019).

Algoritma pengelompokan (clustering) merupakan teknik pembelajaran tanpa pengawasan (unsupervised learning) yang paling dominan digunakan dalam segmentasi RFM. Algoritma K-Means, yang diperkenalkan oleh MacQueen (1967), secara luas diadopsi karena kesederhanaan struktur komputasinya, efisiensi waktu pemrosesan, dan kemudahan dalam interpretasi hasil partisipasinya (Han et al., 2022). Namun demikian, K-Means beroperasi berdasarkan prinsip partisi tegas (crisp/hard clustering), di mana setiap objek data diwajibkan menjadi anggota penuh dari tepat satu kluster tanpa gradasi keanggotaan. Keterbatasan ini menjadi problematik ketika diterapkan pada data transaksi dunia nyata yang sering kali memiliki batas antar kelompok yang tumpang tindih (overlapping) dan ambigu (Wu et al., 2021).

Sebagai alternatif perbaikan, algoritma Fuzzy C-Means (FCM) yang dikembangkan oleh Bezdek (1981) memperkenalkan konsep himpunan kabur (fuzzy logic), di mana setiap data dapat memiliki derajat keanggotaan pada beberapa kluster sekaligus dengan rentang nilai antara 0 dan 1. Fleksibilitas derajat keanggotaan ini secara teoritis memberikan representasi yang lebih realistis terhadap transisi perilaku konsumen yang cair, seperti pelanggan yang berada di ambang batas antara 'pelanggan potensial' dan 'pelanggan setia' (Nayak et al., 2020; Govindarajan et al., 2022).

Meskipun berbagai studi telah mengeksplorasi penggunaan K-Means maupun FCM, kajian komparatif yang menguji efektivitas keduanya secara berdampingan menggunakan data transaksi riil UMKM lokal di kawasan Indonesia Timur, khususnya dengan perbandingan metrik validasi internal komprehensif seperti Silhouette Coefficient dan Davies-Bouldin Index (DBI), masih sangat terbatas. Sebagian besar penelitian terdahulu hanya berfokus pada salah satu metode atau menggunakan dataset generik daring yang tidak mencerminkan karakteristik daya beli dan ritme belanja pasar lokal.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja antara algoritma K-Means dan Fuzzy C-Means dalam mengelompokkan pola transaksi pelanggan e-commerce berbasis model RFM. Kontribusi utama penelitian ini meliputi: (1) evaluasi empiris terhadap keandalan struktur kluster yang dihasilkan oleh kedua algoritma menggunakan metrik Silhouette Coefficient dan Davies-Bouldin Index; (2) karakterisasi profil bisnis dari setiap kluster yang terbentuk; serta (3) formulasi rekomendasi strategi pemasaran berbasis segmen guna meningkatkan retensi dan profitabilitas bisnis UMKM lokal secara terukur.

2. METODOLOGI PENELITIAN

Penelitian ini dirancang menggunakan pendekatan kuantitatif eksperimental dengan memanfaatkan metodologi Knowledge Discovery in Databases (KDD). Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, transformasi fitur RFM, penentuan parameter kluster optimal, komputasi klustering K-Means dan Fuzzy C-Means, validasi metrik evaluasi internal, serta interpretasi profil manajerial.

A. Sumber Data dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan catatan transaksi penjualan aktual dari salah satu platform ritel daring UMKM fesyen dan gaya hidup di Kota Manado. Rentang waktu pengamatan mencakup periode satu tahun penuh, terhitung mulai 1 Juli 2023 sampai dengan 30 Juni 2024. Basis data mencakup 250 pelanggan unik yang telah menyelesaikan setidaknya satu transaksi terverifikasi selama rentang pengamatan.

B. Transformasi Fitur Model RFM

Data transaksi mentah ditransformasikan ke dalam tiga parameter perilaku model RFM untuk setiap pelanggan i :

1. Recency (R): Selisih waktu (dihitung dalam satuan hari) antara tanggal batas akhir periode observasi (30 Juni 2024) dengan tanggal transaksi terakhir yang dilakukan oleh pelanggan i :
$$R_i = \text{Tanggal_Batas} - \text{Tanggal_Transaksi_Terakhir}_i$$
2. Frequency (F): Jumlah akumulasi transaksi unik yang berhasil dibayar oleh pelanggan i selama periode observasi:
$$F_i = \text{Total_Transaksi}_i$$
3. Monetary (M): Akumulasi nilai nominal rupiah belanja yang dibayarkan oleh pelanggan i selama periode observasi:

$$M_i = \text{Sum}(\text{Nilai_Transaksi}_i)$$

C. Normalisasi Min-Max Scaler

Mengingat ketiga variabel RFM memiliki skala nilai yang sangat heterogen (misalnya nilai Recency berkisar puluhan hari, sedangkan Monetary mencapai jutaan rupiah), normalisasi data menjadi tahapan esensial agar variabel dengan skala besar tidak mendominasi perhitungan jarak Euclidean (Han et al., 2022). Normalisasi dilakukan dengan formula Min-Max Normalization ke rentang $[0, 1]$:

$$X'_{ij} = (X_{ij} - \text{Min}(X_j)) / (\text{Max}(X_j) - \text{Min}(X_j))$$

di mana X_{ij} menyatakan nilai fitur j untuk pelanggan i , sedangkan $\text{Min}(X_j)$ dan $\text{Max}(X_j)$ masing-masing menyatakan nilai minimum dan maksimum fitur j dalam dataset.

D. Algoritma K-Means

Algoritma K-Means mempartisi n objek ke dalam k kluster dengan meminimalkan jumlah kuadrat jarak antara objek dengan titik pusat kluster (centroid) terdekat. Fungsi objektif K-Means didefinisikan sebagai:

$$J = \sum_{j=1}^k \sum_{x_i \in S_j} \|x_i - c_j\|^2$$

di mana k adalah jumlah kluster, S_j adalah himpunan objek dalam kluster j , x_i adalah vektor data objek ke- i , dan c_j adalah titik centroid kluster j yang dihitung dari rata-rata aritmatika seluruh objek dalam kluster tersebut:

$$c_j = (1 / |S_j|) * \sum_{x_i \in S_j} x_i$$

E. Algoritma Fuzzy C-Means (FCM)

Berbeda dengan K-Means, algoritma Fuzzy C-Means (Bezdek, 1981) memperkenalkan setiap objek data memiliki derajat keanggotaan u_{ij} pada setiap kluster j , dengan syarat $\sum_{j=1}^c u_{ij} = 1$ untuk setiap i . Fungsi objektif FCM dirumuskan sebagai:

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m * \|x_i - v_j\|^2$$

di mana m adalah parameter pembobot fuzzy (weighting exponent) yang ditetapkan bernilai $m = 2.0$ (standar empiris paling umum), u_{ij} adalah derajat keanggotaan x_i pada kluster j , dan v_j adalah pusat kluster j . Derajat keanggotaan dan pusat kluster diperbarui secara iteratif melalui persamaan:

$$u_{ij} = 1 / [\sum_{k=1}^c (\|x_i - v_j\| / \|x_i - v_k\|)^{2 / (m - 1)}]$$

$$v_j = [\sum_{i=1}^n (u_{ij})^m * x_i] / [\sum_{i=1}^n (u_{ij})^m]$$

terasi dihentikan apabila perubahan nilai fungsi objektif antar iterasi memenuhi kriteria konvergensi $\epsilon \leq 10^{-5}$ atau mencapai batas maksimum 100 iterasi.

F. Metrik Evaluasi Kualitas Kluster

Validasi performa pengelompokan diukur menggunakan dua metrik validasi internal yang independen:

1. Silhouette Coefficient (SC): Mengukur tingkat kedekatan suatu objek terhadap kluster anggotanya (kohesi) dibandingkan dengan kluster terdekat lainnya (separasi) (Rousseeuw, 1987):

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i))$$

di mana $a(i)$ adalah rata-rata jarak objek i ke semua objek lain dalam klaster yang sama, dan $b(i)$ adalah rata-rata jarak minimum objek i ke objek di klaster tetangga terdekat. Nilai SC berada pada rentang $[-1, 1]$, di mana nilai mendekati 1 mencerminkan pemisahan klaster yang sangat baik.

2. Davies-Bouldin Index (DBI): Rasio jumlah dispersi intra-klaster terhadap separasi antar-klaster (Davies & Bouldin, 1979):

$$DBI = (1 / k) * \sum_{i=1}^k \max_{j \neq i} [(S_i + S_j) / M_{ij}]$$

di mana S_i adalah rata-rata jarak objek di klaster i ke pusat klaster, dan M_{ij} adalah jarak Euclidean antara centroid klaster i dan centroid klaster j . Nilai DBI yang lebih rendah (mendekati 0) mengindikasikan struktur klaster yang lebih rapat dan terpisah secara optimal.

3. HASIL PENELITIAN DAN PEMBAHASAN

Tahapan pengolahan data diawali dengan eksplorasi statistik deskriptif dari ketiga parameter RFM sebelum dilakukan normalisasi data. Ringkasan statistik data mentah 250 pelanggan disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Parameter RFM Pelanggan UMKM (N = 250)

Variabel RFM	Minimum	Maksimum	Rata-rata (Mean)	Standar Deviasi	Satuan
Recency (R)	2	358	88.42	76.15	Hari
Frequency (F)	1	24	5.76	4.82	Kali
Monetary (M)	120.000	22.500.000	4.120.500	4.350.200	Rupiah

Tabel 1 memperlihatkan adanya disparitas transaksi yang sangat nyata di kalangan pelanggan. Rentang Recency berkisar antara 2 hari hingga 358 hari, mengindikasikan variasi yang tajam antara pelanggan aktif terkini dan pelanggan dorman. Parameter Monetary mencatatkan nilai pembelanjaan terendah Rp 120.000 dan tertinggi mencapai Rp 22.500.000 dengan standar deviasi yang melampaui rata-rata, yang membuktikan distribusi data memiliki kemiringan positif (right-skewed) yang kuat, khas karakteristik transaksi ritel digital.

A. Penentuan Jumlah Klaster Optimal (Nilai k)

Penentuan nilai k yang tepat merupakan tahapan fundamental dalam algoritma pengelompokan non-hierarkis. Eksperimen dilakukan dengan menguji rentang nilai $k = 2$ hingga $k = 6$ menggunakan Elbow Method (Sum of Squared Errors / SSE) dan evaluasi Silhouette Score. Hasil pengujian nilai k dirangkum dalam Tabel 2.

Tabel 2. Evaluasi Pemilihan Jumlah Klaster Optimal

Jumlah k	Nilai SSE	Silhouette K-Means	Silhouette FCM	Keterangan Evaluasi
$k = 2$	42.85	0.512	0.534	Pemisahan terlalu global
$k = 3$	23.40	0.548	0.579	Transisi profil belum optimal
$k = 4$	11.25	0.582	0.628	Optimal (Elbow point & SC tertinggi)

k = 5	8.95	0.541	0.563	Klaster mulai terpecah berlebih
k = 6	7.10	0.498	0.518	Overfitting dan redundan

Berdasarkan Tabel 2, penurunan nilai SSE paling signifikan melandai secara tajam pada $k = 4$ (membentuk 'siku' atau elbow point). Selain itu, nilai Silhouette Coefficient mencapai titik puncak tertinggi pada $k = 4$, baik untuk K-Means (0,582) maupun FCM (0,628). Ketika nilai k ditingkatkan menjadi 5 dan 6, nilai Silhouette mengalami penurunan yang mencerminkan timbulnya klaster mikro yang tidak memiliki diferensiasi manajerial yang signifikan. Dengan demikian, $k = 4$ ditetapkan sebagai jumlah klaster optimal untuk analisis komparasi lebih lanjut.

B. Hasil Pengelompokan Menggunakan Algoritma K-Means

Algoritma K-Means dijalankan dengan inisialisasi k-means++ sebanyak 10 pengulangan acak untuk menghindari jebakan optimum lokal. Konvergensi tercapai pada iterasi ke-8. Tabel 3 memaparkan posisi centroid akhir terdenormalisasi dan jumlah anggota pada masing-masing klaster.

Tabel 3. Karakteristik Centroid Klaster K-Means ($k = 4$)

Klaster	R (Hari)	F (Kali)	M (Rupiah)	Anggota	Persen
C1	14.2	17.8	14.250.000	34	13.6%
C2	31.5	7.2	5.120.000	72	28.8%
C3	42.0	2.4	1.580.000	81	32.4%
C4	178.6	1.6	1.050.000	63	25.2%

C. Hasil Pengelompokan Menggunakan Algoritma Fuzzy C-Means

Algoritma FCM dioperasikan dengan eksponen pembebanan $m = 2.0$ dan kriteria henti konvergensi 10^{-5} . Proses komputasi mencapai titik konvergensi stabil pada iterasi ke-14. Untuk membandingkan secara ekuivalen, setiap objek pelanggan dipetakan ke klaster defuzzifikasi berdasarkan nilai derajat keanggotaan tertinggi: $\max_j(u_{ij})$. Tabel 4 menyajikan pusat klaster terdenormalisasi dan distribusi anggota hasil pengelompokan FCM.

Tabel 4. Karakteristik Pusat Klaster Fuzzy C-Means ($k = 4$)

Klaster	R (Hari)	F (Kali)	M (Rupiah)	Anggota	Persen
C1	12.0	18.4	14.850.000	36	14.4%
C2	28.4	6.8	4.750.000	70	28.0%
C3	35.2	2.1	1.420.000	82	32.8%
C4	164.8	1.8	1.150.000	62	24.8%

D. Analisis Komparatif Kinerja K-Means vs Fuzzy C-Means

Perbandingan kuantitatif performa kedua algoritma dilakukan berdasarkan kualitas klaster internal (Silhouette Coefficient dan Davies-Bouldin Index) serta efisiensi komputasi (waktu eksekusi dan jumlah iterasi). Hasil perbandingan disajikan pada Tabel 5.

Tabel 5. Perbandingan Kinerja Algoritma K-Means dan Fuzzy C-Means

Parameter Evaluasi	K-Means	Fuzzy C-Means	Keunggulan
Silhouette Coefficient (SC)	0.582	0.628	FCM (+7.9%)
Davies-Bouldin Index (DBI)	0.741	0.672	FCM (-9.3%)

Waktu Komputasi (detik)	0.042	0.118	K-Means
Jumlah Iterasi Konvergen	8	14	K-Means
Penanganan Data Ambigu	Kaku (Biner 0/1)	Gradual (Derajat Keanggotaan)	FCM

Data pada Tabel 5 mengindikasikan temuan fundamental: algoritma Fuzzy C-Means terbukti unggul secara signifikan dalam kualitas struktur kluster dengan menghasilkan Silhouette Coefficient yang lebih tinggi (0,628 berbanding 0,582) dan Davies-Bouldin Index yang lebih rendah (0,672 berbanding 0,741). Secara teoretis, DBI yang lebih kecil dan SC yang lebih tinggi merefleksikan bahwa kluster yang dibentuk oleh FCM memiliki kepadatan internal (compactness) yang lebih rapat sekaligus keterpisahan (separation) antar-kluster yang lebih tegas.

Keunggulan FCM terletak pada kemampuannya mengelola data transisi. Pada dataset transaksi pelanggan riil, terdapat sekitar 18 pelanggan (7,2% dari total sampel) yang memiliki derajat keanggotaan ganda yang signifikan (misalnya $u_{i1} = 0,46$ dan $u_{i2} = 0,49$). Pada K-Means, objek-objek perbatasan ini dipaksa masuk ke salah satu kluster secara deterministik, yang mendistorsi posisi centroid dan memperlebar variansi intra-kluster. Sebaliknya, pada FCM, pembobotan eksponensial derajat keanggotaan $m = 2.0$ memungkinkan objek perbatasan memberikan kontribusi proporsional pada pembaruan titik pusat kedua kluster sehingga menghasilkan konfigurasi akhir yang lebih stabil dan representatif. Walaupun K-Means lebih unggul dari segi kecepatan eksekusi (0,042 detik berbanding 0,118 detik), perbedaan waktu tersebut sangat tidak material dalam konteks pemrosesan analitik bisnis berkala.

E. Profil Karakteristik Segmen dan Rekomendasi Strategi Pemasaran

Berdasarkan model segmentasi FCM yang terbukti lebih unggul, empat kluster pelanggan diklasifikasikan ke dalam tipologi manajerial RFM standar industri disertai rekomendasi strategi CRM terapan, sebagaimana dirangkum dalam Tabel 6.

Tabel 6. Profil Karakteristik Segmen Pelanggan dan Rekomendasi Aksi Pemasaran

Kluster	Nama Segmen	Proporsi	Karakteristik Perilaku RFM	Rekomendasi Strategi CRM
Kluster 1	Champions (Pelanggan Utama)	14.4%	Recency sangat baru (12 hari), Frekuensi sangat tinggi (18.4x), Nilai belanja tertinggi (Rp 14.85jt)	Layanan prioritas VIP, program loyalitas eksklusif, akses awal ke rilis produk baru (early access), program referral.
Kluster 2	Potential Loyalists (Loyalis Potensial)	28.0%	Recency baru (28.4 hari), Frekuensi belanja sedang (6.8x), Belanja menengah (Rp 4.75jt)	Penawaran cross-selling dan up-selling, bundling produk bernilai tambah, insentif cashback bersyarat nominal transaksi.
Kluster 3	Promising Customers (Pelanggan Baru/Musiman)	32.8%	Recency cukup baru (35.2 hari), Frekuensi rendah (2.1x), Nilai belanja kecil (Rp 1.42jt)	Kampanye edukasi katalog produk, penawaran diskon transaksi kedua untuk mendorong repurchasing, push notification promo berkala.
Kluster 4	At-Risk Customers (Pelanggan Berisiko Churn)	24.8%	Recency sangat lama (164.8 hari), Frekuensi rendah (1.8x), Nilai belanja rendah (Rp 1.15jt)	Win-back campaign melalui WhatsApp/Email personal, survei evaluasi ketidakpuasan, pemberian voucher spesial 'Kami Merindukan Anda'.

Penetapan strategi diferensiasi ini memberikan landasan operasional yang konkrit bagi manajer pemasaran UMKM. Sebagai contoh, alokasi anggaran retensi dapat dialokasikan sebesar 40% untuk menjaga kepuasan segmen Champions dan Potential Loyalists yang menghasilkan lebih dari 70% total omzet bisnis, sementara segmen At-Risk dapat didekati dengan program reaktivasi berbiaya rendah (low-cost broadcast) agar margin keuntungan tetap terjaga.

4. KESIMPULAN

Penelitian ini telah berhasil melakukan analisis perbandingan kinerja antara algoritma K-Means dan Fuzzy C-Means dalam segmentasi pola transaksi pelanggan e-commerce berdasarkan kerangka kerja model Recency, Frequency, and Monetary (RFM). Temuan empiris mengonfirmasi bahwa algoritma Fuzzy C-Means (FCM) mengungguli algoritma K-Means dari segi kualitas pemisahan klaster, dengan membukukan Silhouette Coefficient yang lebih tinggi sebesar 0,628 (berbanding 0,582 pada K-Means) dan nilai Davies-Bouldin Index yang lebih optimal sebesar 0,672 (berbanding 0,741 pada K-Means). Pendekatan soft clustering pada FCM terbukti tangguh dalam mengatasi masalah ambiguitas dan batas kabur pada data perilaku transaksi nyata pelanggan UMKM.

Melalui integrasi model RFM dan FCM dengan $k = 4$, teridentifikasi empat tipologi perilaku pelanggan yang distingtif: Champions (14,4%), Potential Loyalists (28,0%), Promising Customers (32,8%), dan At-Risk Customers (24,8%). Pemetaan ini memfasilitasi formulasi strategi pemasaran digital yang terpersonalisasi, tepat sasaran, dan efisien secara anggaran. Keterbatasan penelitian ini terletak pada cakupan dataset yang berasal dari satu entitas bisnis ritel dengan periode observasi satu tahun. Penelitian lanjutan disarankan untuk memperluas dimensi model dengan menyertakan atribut Customer Retention Period (RFM-P) atau durasi interaksi digital pada aplikasi mobile/metaverse, serta mengeksplorasi optimasi parameter inisialisasi pusat klaster menggunakan algoritma metaheuristik seperti Genetic Algorithm atau Particle Swarm Optimization.

DAFTAR PUSTAKA

- Aktama, G. E., Marbun, H., Kalubota, A. D., & Ratela, N. M. (2024). Pengaruh marketing sosial media TikTok terhadap penjualan online dalam perkembangan teknologi metaverse. *Jurnal Logika*, 12(1), 1–19.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Plenum Press. <https://doi.org/10.1007/978-1-4757-0450-1>
- Bult, J. R., & Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4), 378–394. <https://doi.org/10.1287/mksc.14.4.378>
- Chen, D., Sain, S. L., & Guo, K. (2020). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using cluster analysis. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208. <https://doi.org/10.1057/dbm.2012.17>
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for multicustomer analysis. *Journal of Marketing Research*, 42(4), 415–430. <https://doi.org/10.1509/jmkr.2005.42.4.415>
- Govindarajan, M., Chandrasekaran, R. M., & Senthil, K. (2022). Customer churn prediction and segmentation using fuzzy logic and deep learning in e-commerce. *Soft Computing*, 26(11), 5231–5244. <https://doi.org/10.1007/s00500-022-06941-x>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann. <https://doi.org/10.1016/C2018-0-02059-4>



- Hughes, A. M. (1994). *Strategic database marketing: The masterplan for starting and managing a profitable, customer-based marketing program*. Probus Publishing.
- Kalubota, A. D., Aktama, G. E., & Marbun, H. (2024). Analisis pengaruh adopsi media sosial interaktif terhadap loyalitas konsumen pada platform perdagangan elektronik. *Jurnal Rekayasa Sistem Informasi dan Teknologi*, 8(2), 145–156. <https://doi.org/10.35870/jti.v8i2.1482>
- Kotler, P., & Keller, K. L. (2016). *Marketing management (15th global ed.)*. Pearson Education.
- Kumar, V., & Reinartz, W. (2018). *Customer relationship management: Concept, strategy, and tools (3rd ed.)*. Springer Texts in Business and Economics. <https://doi.org/10.1007/978-3-662-55381-7>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.
- Monalisa, S., Nadya, P., & Novita, R. (2019). Analisis segmentasi pelanggan menggunakan kombinasi model RFM dan algoritma K-Means pada industri ritel. *Jurnal Rekayasa Elektrika*, 15(1), 23–30. <https://doi.org/10.17529/jre.v15i1.12933>
- Nayak, J., Naik, B., & Behera, H. (2020). Fuzzy C-Means (FCM) clustering algorithm: A systematic review and comparative analysis. *Archives of Computational Methods in Engineering*, 27(4), 1277–1300. <https://doi.org/10.1007/s11831-019-09353-0>
- Peppers, D., & Rogers, M. (2017). *Managing customer experience and relationships: A strategic framework (3rd ed.)*. John Wiley & Sons. <https://doi.org/10.1002/9781119236245>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Turban, E., Outland, J., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2018). *Electronic commerce 2018: A managerial and social networks perspective (9th ed.)*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-58715-8>
- Wu, J., Shi, L., Lin, R. V., Tsai, S. B., Li, Y., Cheng, Q., & Chen, Q. (2021). An empirical study on customer segmentation by combining RFM model and clustering algorithms in cross-border e-commerce. *IEEE Access*, 9, 36838–36848. <https://doi.org/10.1109/ACCESS.2021.3063517>