

PENERAPAN ALGORITMA FUZZY C-MEANS UNTUK SEGMENTASI PELANGGAN BERBASIS MODEL RFM PADA DATA TRANSAKSI E-COMMERCE

Dimas Pramudya¹, Nana Suarna², Agus Bahtiar³, Indra Wiguna Marthanu⁴, Kaslani⁵

Program Studi Teknik Informatika¹²
Program Studi Sistem Informasi³
Program Studi Sistem Informasi⁴⁵
STMIK IKMI Cirebon.
email: pramudyad08@gmail.com¹

(*) Corresponding Author : pramudyad08@gmail.com;
Published : 30 Januari 2026

Abstract— This study aims to design and implement an e-commerce customer segmentation model using a combination of the Recency, Frequency, Monetary (RFM) method and the Fuzzy C-Means (FCM) algorithm. The RFM method is used to evaluate customer behavior based on the time of the last purchase, transaction frequency, and total purchase value, while FCM is applied to perform segmentation based on fuzzy membership that reflects the degree of customer involvement in each cluster. The research procedure includes loading and cleaning transaction data, calculating RFM scores, normalizing data using Min-Max scaling, evaluating the number of clusters using the Davies-Bouldin and Xie-Beni indices, and implementing the FCM algorithm. The results show that five optimal clusters can group customers into representative segments, such as loyal, potential, and dormant customers. The visualization of the results shows a clear distribution between segments and helps in identifying more effective customer retention strategies. This model provides a data-driven adaptive approach that is relevant to the dynamics of today's digital customer behavior and supports more targeted business decision-making.

Keywords: Customer segmentation, Fuzzy C-Means, Recency, Frequency, Monetary, e-commerce, data mining.

Abstrak— Penelitian ini bertujuan untuk merancang dan mengimplementasikan model segmentasi pelanggan e-commerce menggunakan integrasi metode Recency, Frequency, Monetary (RFM) dan algoritma Fuzzy C-Means (FCM). Metode RFM digunakan untuk mengevaluasi perilaku pelanggan berdasarkan waktu pembelian terakhir, frekuensi transaksi, dan total nilai pembelian, sedangkan FCM diaplikasikan untuk melakukan segmentasi berdasarkan keanggotaan fuzzy yang mencerminkan derajat keterlibatan pelanggan dalam setiap kluster. Prosedur penelitian meliputi pemuatan dan pembersihan data transaksi, perhitungan skor RFM, normalisasi data menggunakan Min-Max scaling, evaluasi jumlah kluster menggunakan indeks Davies-Bouldin dan Xie-Beni, serta implementasi algoritma FCM. Hasil penelitian menunjukkan bahwa lima kluster optimal dapat mengelompokkan pelanggan ke dalam segmen yang representatif, seperti pelanggan loyal, potensial, dan dormant. Visualisasi hasil menunjukkan distribusi yang jelas antar segmen dan membantu dalam mengidentifikasi strategi retensi pelanggan yang lebih efektif. Model ini memberikan pendekatan adaptif berbasis data yang relevan dengan dinamika perilaku pelanggan digital masa kini, serta mendukung pengambilan keputusan bisnis yang lebih tepat sasaran.

Kata Kunci: Segmentasi Pelanggan, Fuzzy C-Means, Recency, Frequency, Monetary, E-Commerce, Data Mining.

INTRODUCTION

Perkembangan teknologi informasi, big data analytics, dan machine learning telah mendorong transformasi signifikan dalam strategi pemasaran digital. Dalam ranah e-commerce, analisis perilaku pelanggan menjadi semakin penting untuk meningkatkan efektivitas pengambilan keputusan

berbasis data [1], [2]. Model Recency-Frequency-Monetary (RFM) banyak digunakan karena kesederhanaannya dalam memetakan nilai pelanggan, namun pendekatan ini masih terbatas dalam menangkap dinamika perilaku konsumen modern [3], [4].

Penelitian terdahulu menunjukkan bahwa model RFM tradisional sering menghasilkan

segmentasi yang statis dan kurang adaptif terhadap perubahan perilaku, terutama dalam data transaksi yang cepat berubah [5], [6]. Selain itu, metode hard clustering seperti K-Means belum mampu menggambarkan tumpang-tindih karakteristik pelanggan yang berada antara dua segmen, sehingga menurunkan akurasi segmentasi [7], [8].

Untuk mengatasi keterbatasan tersebut, algoritma Fuzzy C-Means (FCM) menawarkan pendekatan soft clustering yang memungkinkan setiap pelanggan memiliki derajat keanggotaan ganda pada beberapa segmen. Pendekatan fuzzy ini terbukti meningkatkan fleksibilitas, interpretabilitas, dan ketepatan segmentasi dalam konteks perilaku pelanggan yang dinamis [9], [10]. Namun, tantangan seperti penentuan jumlah kluster optimal dan sensitivitas terhadap inisialisasi masih perlu diperhatikan [11], [12].

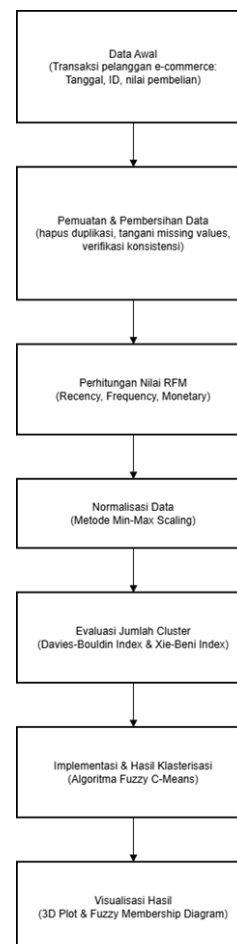
Meskipun kombinasi RFM dan FCM mulai banyak dikaji, sebagian besar penelitian terdahulu belum menekankan evaluasi validitas fuzzy menggunakan indeks seperti Davies-Bouldin Index (DBI) dan Xie-Beni Index (XBI), serta belum banyak diuji pada konteks e-commerce Indonesia [13], [14], [15]. Celah ini menunjukkan perlunya pendekatan segmentasi yang lebih komprehensif dan empiris berbasis data transaksi nyata.

Berdasarkan eksperimen menggunakan data Urbanmart, integrasi model RFM dengan algoritma FCM menghasilkan segmentasi yang lebih adaptif dan representatif dibandingkan metode konvensional. Evaluasi berbasis DBI dan XBI menunjukkan peningkatan compactness dan separation kluster, sekaligus menggambarkan keanggotaan fuzzy yang lebih realistis terhadap variasi perilaku pelanggan [14], [16]. Temuan ini menegaskan keunggulan pendekatan soft clustering dalam menangani ambiguitas data transaksi e-commerce.

Penelitian ini bertujuan mengembangkan metode segmentasi pelanggan berbasis integrasi RFM-FCM untuk menghasilkan struktur kluster yang lebih akurat dan informatif. Secara teoretis, penelitian ini memperkuat penerapan fuzzy logic dalam analitik pelanggan, sedangkan secara praktis, hasilnya berpotensi membantu industri e-commerce dalam personalisasi pemasaran, program loyalitas, dan identifikasi pelanggan bernilai tinggi atau berisiko churn [17], [18]. Dengan demikian, penelitian ini memberikan kontribusi penting bagi pengembangan teknik segmentasi pelanggan yang relevan di era ekonomi digital.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk menghasilkan segmentasi pelanggan berbasis kombinasi Recency, Frequency, Monetary (RFM) dan algoritma Fuzzy C-Means (FCM). Proses penelitian dilaksanakan melalui serangkaian tahapan sistematis mulai dari pemilihan data, pra-pemrosesan, pembentukan fitur, normalisasi, evaluasi jumlah kluster, penerapan FCM, hingga analisis hasil klusterisasi. Pendekatan ini dipilih karena algoritma fuzzy mampu menangkap variasi dan tumpang tindih perilaku pelanggan yang tidak dapat ditangani oleh metode klusterisasi keras seperti K-Means. Seluruh tahapan dilakukan menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Scikit-Fuzzy, serta Matplotlib untuk memastikan proses replikasi yang terukur dan transparan.



Gambar 1. Tahapan Penelitian

1. Data Awal

Data awal berasal dari transaksi pelanggan e-commerce yang mencatat tanggal transaksi, identitas pelanggan, dan jumlah nilai pembelian. Data ini kemudian dimuat ke dalam lingkungan analisis untuk diproses lebih lanjut. Validitas data dijaga

dengan memastikan kelengkapan informasi yang relevan dan penghapusan data duplikat atau tidak konsisten.

2. Pemuatan dan Pembersihan Data

Pemuatan dan Pembersihan Data Langkah awal setelah pengumpulan data adalah memuat data ke dalam lingkungan pemrograman Python. Kemudian dilakukan proses pembersihan data untuk menghapus duplikasi, menangani nilai kosong (missing values), serta memverifikasi integritas dan konsistensi data. Tahap ini penting untuk menjamin kualitas data sebelum dilakukan analisis lebih lanjut. Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan segmentasi pelanggan berbasis algoritma Fuzzy C-Means (FCM) dan model RFM (Recency, Frequency, Monetary). Prosedur penelitian dirancang dalam enam tahap utama berikut:

3. Perhitungan Nilai RFM

Dalam tahap ini, tiga dimensi utama perilaku pelanggan dihitung:

1. Recency dihitung berdasarkan selisih antara tanggal analisis dan tanggal pembelian terakhir.
2. Frequency dihitung sebagai jumlah transaksi oleh pelanggan dalam periode tertentu.
3. Monetary dihitung sebagai total pengeluaran dari pelanggan selama periode tersebut. Ketiga indikator ini menggambarkan intensitas dan nilai hubungan pelanggan dengan platform.

4. Normalisasi Data

Setelah nilai RFM diperoleh, langkah berikutnya adalah normalisasi data menggunakan metode Min-Max Scaling. Proses ini dilakukan untuk menyamakan skala antar variabel sehingga tidak ada satu variabel yang mendominasi hasil klusterisasi [10], [19], [20].

5. Evaluasi Jumlah Cluster dengan DBI dan XBI

Penentuan jumlah kluster optimal dilakukan melalui evaluasi indeks validitas internal seperti Davies-Bouldin Index (DBI) dan Xie-Beni Index (XBI). Indeks ini mengukur kompaksi dan separasi antar kluster, sehingga memudahkan pemilihan jumlah kluster terbaik untuk segmentasi [2], [10]

6. Implementasi dan Hasil Klusterisasi

Setelah jumlah kluster ditentukan, algoritma Fuzzy C-Means diterapkan pada data RFM yang telah dinormalisasi. Proses ini melibatkan iterasi pembaruan pusat kluster dan matriks keanggotaan hingga konvergen, dengan mempertimbangkan parameter seperti tingkat fuzziness dan metrik jarak [10], [21]

7. Visualisasi Hasil

Hasil klusterisasi divisualisasikan menggunakan grafik tiga dimensi dan diagram keanggotaan fuzzy untuk memudahkan interpretasi segmen. Visualisasi ini memberikan wawasan tentang distribusi dan karakteristik masing-masing kluster.

RESULTS AND DISCUSSION

Penelitian ini menghasilkan segmentasi pelanggan e-commerce berbasis model RFM dan algoritma Fuzzy C-Means (FCM). Proses eksperimen dilakukan menggunakan Python dalam lingkungan Jupyter Notebook. Seluruh proses eksperimen dijalankan secara sistematis, mulai dari akuisisi data hingga interpretasi hasil klusterisasi, untuk menghasilkan segmentasi pelanggan yang representatif dan dapat diimplementasikan dalam strategi pemasaran.

Data Awal

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul UrbanMart Transactions Dataset, yang berisi catatan transaksi pelanggan pada platform e-commerce UrbanMart selama periode tahun 2023 hingga 2024. Dataset ini digunakan sebagai sumber utama dalam proses segmentasi pelanggan berbasis model RFM (Recency, Frequency, Monetary) dan algoritma Fuzzy C-Means (FCM).

Data terdiri atas beberapa atribut penting yang merepresentasikan perilaku transaksi pelanggan, antara lain:

1. TransactionID : Nomor unik untuk setiap transaksi.
2. CustomerID : Identitas unik pelanggan yang digunakan untuk mengelompokkan transaksi per individu.
3. TransactionDate : Tanggal terjadinya transaksi yang digunakan dalam perhitungan Recency.
4. TransactionValue : Total nilai pembelian pada setiap transaksi yang digunakan untuk menghitung Monetary.

5. ProductCategory : Kategori produk yang dibeli pelanggan (misalnya Fashion, Sport, Beauty, Home, dan Electronics).
6. PaymentMethod : Jenis metode pembayaran yang digunakan, seperti Credit Card, Bank Transfer, atau E-Wallet.
7. CustomerGender dan CustomerAgeGroup : Atribut demografis pelanggan yang dapat digunakan sebagai variabel tambahan dalam analisis perilaku.
8. Region : Lokasi atau wilayah tempat pelanggan melakukan transaksi (misalnya Jakarta, Bandung, Medan, dan Surabaya).

Sebelum dilakukan proses analisis lebih lanjut, data dimuat ke dalam lingkungan Python menggunakan pustaka Pandas dan diperiksa untuk memastikan kelengkapan serta konsistensi formatnya. Total terdapat 50.000 baris transaksi yang merepresentasikan aktivitas pembelian dari berbagai pelanggan di beberapa wilayah Indonesia. Data awal bisa dilihat di tabel 1

Table 1 Data Awal

Transac tionID	Custo merID	Transacti onDate	Transacti onValue	...	Regio n
TRANSO 00000	CUSTO 1574	2023-04- 09	467907	...	Surab aya
TRANSO 00001	CUSTO 1836	2024-08- 05	2713789	...	Surab aya
TRANSO 00002	CUSTO 1291	2024-04- 07	4539248	...	Bandu ng
TRANSO 00003	CUSTO 0245	2024-06- 12	3058426	...	Jakart a
TRANSO 00004	CUSTO 0163	2023-12- 18	3304874	...	Jakart a
TRANSO 00005	CUSTO 0331	2024-11- 24	1556151	...	Bandu ng
TRANSO 00006	CUSTO 1956	2023-03- 10	1664460	...	Jakart a
TRANSO 00007	CUSTO 1713	2023-04- 10	4194384	...	Meda n
...	...	...	...	...	...
TRANSO 50000	CUSTO 1537	2024-02- 29	1089046	...	Yogya karta

Dataset ini memperlihatkan variasi transaksi pelanggan dari berbagai kategori produk, metode pembayaran, serta wilayah geografis yang berbeda. Informasi ini menjadi fondasi dalam proses analisis perilaku pelanggan karena setiap atribut berpotensi memberikan kontribusi terhadap pemahaman pola pembelian dan preferensi pelanggan.

Pemuatan dan Pembersihan Data

Langkah pertama dalam eksperimen ini adalah memuat dataset transaksi pelanggan ke dalam lingkungan pemrograman Python sebagai dasar untuk seluruh proses analisis dan klusterisasi. Proses pemuatan data dilakukan dengan menggunakan pustaka seperti Pandas dan NumPy yang mendukung pengolahan data berukuran besar secara efisien. Dataset yang digunakan berisi catatan transaksi pelanggan dari platform e-commerce dengan beberapa atribut penting, seperti CustomerID, tanggal transaksi, jumlah transaksi, dan nilai pembelian (total amount).

Setelah data berhasil dimuat, tahap selanjutnya adalah pembersihan data (data cleaning). Tahap ini merupakan bagian krusial dalam proses data preprocessing karena kualitas data yang digunakan akan berpengaruh langsung terhadap hasil analisis. Proses pembersihan dimulai dengan menghapus entri duplikat yang dapat menyebabkan distorsi pada hasil perhitungan Recency, Frequency, dan Monetary. Selanjutnya, dilakukan penanganan nilai kosong (missing values) dengan dua pendekatan, yaitu penghapusan baris yang tidak relevan atau pengisian nilai kosong menggunakan metode imputasi sesuai konteks atribut. Misalnya, jika terdapat pelanggan tanpa catatan transaksi lengkap, data tersebut dihapus karena tidak memberikan informasi yang valid bagi model segmentasi.

Selain itu, dilakukan juga koreksi terhadap format data yang tidak konsisten, seperti kesalahan pada tipe data tanggal (datetime format) atau angka yang masih disimpan dalam bentuk teks. Semua atribut numerik, seperti total pembelian dan jumlah transaksi, dikonversi ke format numerik yang sesuai untuk memudahkan proses perhitungan selanjutnya.

Tujuan utama dari serangkaian tahapan ini adalah untuk memastikan bahwa hanya data yang valid, konsisten, dan relevan yang digunakan dalam proses analisis. Dengan memastikan kebersihan dan integritas data sejak awal, penelitian ini dapat menghasilkan model segmentasi yang lebih akurat, stabil, dan dapat diandalkan, sehingga setiap pola yang diidentifikasi benar-benar merepresentasikan perilaku pelanggan yang sesungguhnya. Untuk melihat statistik data bisa dilihat dari tabel 2.

Table 2 Statistik Data

Statistik	TransactionValue
count	50,000.00
mean	2,534,081.16
std	1,429,537.72
min	50,060.00

25%	1,304,192.00
50%	2,537,391.00
75%	3,768,264.75
max	4,999,929.00

Perhitungan Nilai RFM

Setelah proses pembersihan dan penyaringan data transaksi dilakukan, tahap selanjutnya adalah menghitung nilai RFM (Recency, Frequency, Monetary) untuk setiap pelanggan sebagai dasar segmentasi. Ketiga komponen ini digunakan untuk menggambarkan perilaku pembelian pelanggan dari tiga dimensi utama, yaitu waktu terakhir bertransaksi, intensitas pembelian, dan nilai ekonomi transaksi.

1. Recency (R) dihitung sebagai selisih antara tanggal analisis dan tanggal transaksi terakhir yang dilakukan pelanggan. Nilai Recency yang lebih kecil menunjukkan bahwa pelanggan tersebut baru-baru ini melakukan pembelian, sehingga cenderung masih aktif.
2. Frequency (F) mencerminkan jumlah transaksi yang dilakukan oleh pelanggan selama periode observasi tertentu. Pelanggan dengan nilai Frequency tinggi menunjukkan tingkat interaksi yang lebih sering dengan platform e-commerce dan berpotensi memiliki loyalitas yang lebih tinggi.
3. Monetary (M) merepresentasikan total nilai atau jumlah uang yang dikeluarkan pelanggan dalam seluruh transaksi selama periode analisis. Nilai Monetary yang tinggi menunjukkan pelanggan dengan kontribusi ekonomi besar bagi perusahaan.

Perhitungan ketiga variabel ini menghasilkan vektor karakteristik unik untuk setiap pelanggan, yang akan menjadi dasar utama dalam proses klusterisasi menggunakan algoritma Fuzzy C-Means (FCM). Dengan kata lain, setiap pelanggan diproyeksikan ke dalam ruang tiga dimensi (R, F, M) sehingga hubungan antarperilaku pembelian dapat dianalisis secara komputasional.

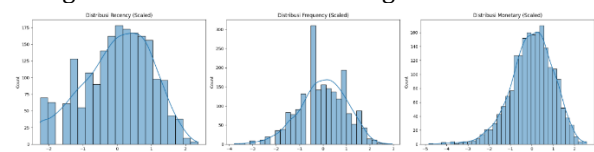
Hasil perhitungan nilai RFM ditampilkan pada Tabel 3, yang menunjukkan sebagian contoh data pelanggan setelah proses kalkulasi. Terlihat bahwa setiap pelanggan memiliki nilai Recency, Frequency, dan Monetary yang berbeda-beda, mencerminkan variasi pola transaksi antarindividu. Data setelah dilakukan perhitungan RFM bisa dilihat di Tabel 3.

Table 3 Data RFM

No	CustomerID	Recency	Frequency	Monetary
1	CUST00001	1	22	63882371
2	CUST00002	18	23	57227590
3	CUST00003	22	28	76991379
4	CUST00004	10	33	87579881
5	CUST00005	21	33	96718054
6	CUST00006	41	24	66248621
7	CUST00007	12	25	59011079
8	CUST00008	44	31	80814082
9	CUST00009	87	17	46563039
.....				
2000	CUST02000	14	22	61695792

Normalisasi Data

Setelah proses perhitungan nilai Recency (R), Frequency (F), dan Monetary (M) untuk masing-masing pelanggan, tahap berikutnya adalah melakukan transformasi skala data agar seluruh variabel memiliki kontribusi yang seimbang dalam proses klusterisasi. Perbedaan satuan dan rentang nilai antarvariabel dapat menyebabkan bias dalam pembentukan kluster apabila tidak dilakukan penyetaraan terlebih dahulu. Misalnya, nilai Monetary yang biasanya memiliki skala jauh lebih besar dibanding Recency dan Frequency dapat mendominasi hasil perhitungan jarak antar data. Untuk mengatasi hal tersebut, penelitian ini menerapkan dua tahap transformasi, yaitu transformasi logaritmik (log transformation) dan normalisasi skala menggunakan metode Min-Max Scaling. Transformasi logaritmik digunakan untuk mengurangi efek skewness atau kemencengan distribusi data, khususnya pada variabel Monetary yang cenderung memiliki nilai ekstrem. Selanjutnya, metode Min-Max Scaling digunakan untuk mengubah rentang nilai setiap variabel ke dalam interval [0,1]. Proses ini bertujuan agar seluruh dimensi memiliki bobot yang sama dan algoritma Fuzzy C-Means (FCM) dapat menghitung jarak antar data secara proporsional tanpa dipengaruhi perbedaan skala numerik. Data setelah di log transformasi bisa dilihat di gambar 1



Gambar 1 Log Transformasi

Hasil proses normalisasi ditunjukkan pada Tabel 4, yang menampilkan nilai logaritmik (Log) dan nilai hasil normalisasi (Scaled) dari setiap variabel R, F, dan M untuk sebagian sampel pelanggan. Terlihat bahwa nilai setiap variabel telah berada dalam rentang yang seragam setelah dilakukan scaling, menandakan bahwa proses normalisasi berhasil dilakukan dengan baik. Hasil proses normalisasi bisa dilihat di Tabel 4

Table 4 Hasil Proses Normalisasi

No	CustID	R_Log	F_Log	M_Log	R_Scaled	F_Scaled	M_Scaled
1	CUST0001	0,69 314 7	3,13 549 4	17,9 725 5	- 2,24 648	- 0,51 664	0,15 067 7
2	CUST0002	2,94 443 9	3,17 805 4	17,8 625 5	- 0,03 102	- 0,30 346	- 0,30 324
3	CUST0003	3,13 549 4	3,36 729 6	18,1 592	0,15 699 8	0,64 447 8	0,92 084 5
4	CUST0004	2,39 789 5	3,52 636 1	18,2 880 6	- 0,56 886	1,44 125 1	1,45 254 6
5	CUST0005	3,09 104 2	3,52 636 1	18,3 873 1	0,11 325 4	1,44 125 1	1,86 207 3
6	CUST0006	3,73 767 6	3,21 887 6	18,0 089 3	0,74 959 3	- 0,09 898	0,30 075 5
7	CUST0007	2,56 494 9	3,25 809 7	17,8 932 4	- 0,40 447	0,09 748 6	- 0,17 661
8	CUST0008	3,80 666 2	3,46 573 6	18,2 076 6	0,81 748 4	1,13 757 5	1,12 079 4
9	CUST0009	4,47 733 7	2,89 037 2	17,6 563 2	1,47 748 5	- 1,74 449	- 1,15 42
...	...	...	...	...	...	...	...
2000	CUST2000	2,70 805 4	3,13 549 4	17,9 377 3	- 0,26 364	- 0,51 664	0,00 696 9

Dengan adanya tahap normalisasi ini, seluruh variabel RFM dapat berkontribusi secara seimbang dalam proses klusterisasi. Langkah ini menjadi krusial untuk memastikan bahwa segmentasi yang dihasilkan oleh algoritma FCM benar-benar mencerminkan perilaku pelanggan secara proporsional, bukan hasil dari perbedaan skala antar variabel. Tahapan ini juga meningkatkan stabilitas perhitungan fungsi objektif pada FCM, sehingga mempercepat proses konvergensi menuju pembentukan kluster yang optimal.

Evaluasi Jumlah Kluster

Untuk Penentuan jumlah kluster yang optimal merupakan tahap penting dalam proses klusterisasi, karena jumlah kluster yang tepat akan menghasilkan segmentasi yang representatif dan memiliki interpretasi yang kuat. Pada penelitian ini digunakan dua metrik validitas internal, yaitu Davies-Bouldin Index (DBI) dan Xie-Beni Index

(XBI), untuk mengukur kualitas hasil klusterisasi dan menentukan jumlah kluster yang paling sesuai.

1. Davies-Bouldin Index (DBI) digunakan untuk menilai rasio antara jarak antar kluster dan kekompakan di dalam kluster. Nilai DBI yang lebih rendah menunjukkan struktur kluster yang lebih baik, karena menandakan bahwa jarak antar kluster cukup besar sementara penyebaran data di dalam kluster relatif kecil.
2. Xie-Beni Index (XBI) digunakan untuk mengukur rasio antara kepadatan intra-kluster terhadap jarak minimum antar pusat kluster (centroid). Sama halnya dengan DBI, nilai XBI yang lebih rendah mengindikasikan kualitas kluster yang lebih optimal dengan pemisahan yang jelas antarsegmen.

Eksperimen dilakukan dengan menguji jumlah kluster dari $k = 2$ hingga $k = 8$. Hasil pengujian disajikan pada Tabel 5, yang menunjukkan nilai minimum DBI dan XBI untuk masing-masing konfigurasi kluster. Hasil menentukan jumlah kluster bisa dilihat di tabel 5.

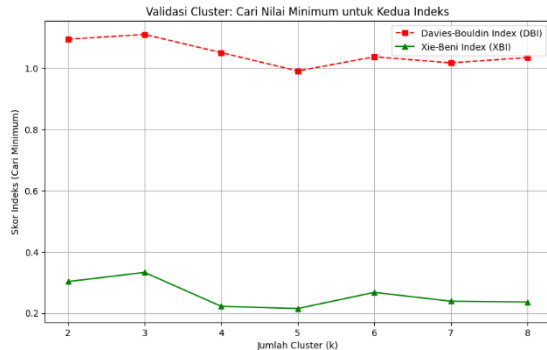
Table 5 Hasil Kluster

Jumlah	DBI	XBI
Cluster (k)	(Minimum)	(Minimum)
2	1.093995438	0.3026738312
3	1.109539194	0.3326017661
4	1.050236148	0.2219743339
5	0.9903240671	0.2146351176
6	1.036619795	0.2673756821
7	1.016670767	0.2385763239
8	1.03419503	0.2359506013

Dari tabel tersebut terlihat bahwa nilai DBI terendah dicapai pada $k = 5$ dengan skor 0.9903, sementara nilai XBI terendah juga ditemukan pada $k = 5$ dengan skor 0.2146. Hal ini menunjukkan bahwa struktur kluster pada jumlah lima kelompok memiliki keseimbangan terbaik antara jarak antar kluster dan kepadatan internalnya.

Hasil perbandingan antarindeks divisualisasikan pada Gambar 2, yang menampilkan grafik hubungan antara jumlah kluster dan nilai indeks validitas. Garis merah menunjukkan tren nilai DBI, sedangkan garis hijau merepresentasikan nilai XBI. Terlihat bahwa pada $k = 5$, kedua garis mencapai titik terendah secara bersamaan, menandakan titik optimal dalam proses klusterisasi. Setelah nilai

tersebut, baik DBI maupun XBI cenderung meningkat kembali, yang mengindikasikan menurunnya kualitas pemisahan antar kluster jika jumlah kluster ditambah. Grafik Validasi Cluster bisa dilihat di Gambar 2.



Gambar 2 Grafik Kluster

Dengan demikian, dapat disimpulkan bahwa jumlah kluster lima ($k = 5$) merupakan konfigurasi paling ideal untuk dataset penelitian ini. Pemilihan nilai tersebut tidak hanya didukung oleh perhitungan matematis dari kedua metrik validitas, tetapi juga sesuai dengan interpretasi perilaku pelanggan yang ditemukan pada tahap analisis hasil klusterisasi berikutnya. Struktur lima kluster ini memberikan keseimbangan terbaik antara heterogenitas antarsegmen dan homogenitas dalam segmen, sehingga mendukung keakuratan segmentasi pelanggan secara keseluruhan.

Implementasi dan Hasil Klusterisasi

Dengan jumlah kluster yang ditetapkan sebanyak lima, algoritma Fuzzy C-Means (FCM) diterapkan pada data hasil perhitungan nilai Recency, Frequency, dan Monetary (RFM) yang sebelumnya telah dinormalisasi untuk menghilangkan perbedaan skala antarvariabel. Proses klusterisasi dilakukan secara iteratif dengan memperbarui pusat kluster (centroid) dan derajat keanggotaan setiap data hingga mencapai kondisi konvergensi, yaitu ketika perubahan nilai fungsi objektif berada di bawah ambang batas tertentu. Melalui pendekatan fuzzy, setiap pelanggan tidak hanya dikategorikan secara tegas ke dalam satu kelompok, tetapi memiliki tingkat keanggotaan relatif terhadap beberapa kluster. Hal ini mencerminkan bahwa perilaku pelanggan sering kali tumpang tindih dan tidak bersifat absolut, sehingga model FCM mampu menangkap kompleksitas tersebut secara lebih realistis dibandingkan metode klusterisasi konvensional seperti K-Means. Jumlah cluster bisa dilihat di tabel 6.

Table 6 Jumlah Kluster

Cluster_Fuzzy	Jumlah Pelanggan	Persentase (%)
0	338	16.9
1	402	20.1
2	504	25.2
3	303	15.15
4	453	22.65

Lima pusat kluster (centroid) yang dihasilkan menggambarkan segmentasi perilaku pelanggan dengan karakteristik yang berbeda-beda. Cluster 0 mencakup pelanggan dengan nilai Frequency dan Monetary sedang namun Recency tinggi, menunjukkan kelompok pelanggan yang pernah aktif tetapi sudah cukup lama tidak bertransaksi. Segmen ini penting untuk strategi reaktivasi, seperti pengiriman email promosi atau penawaran eksklusif untuk mendorong transaksi ulang. Cluster 1 menampilkan pelanggan dengan Frequency dan Monetary rendah serta Recency tinggi, yang diidentifikasi sebagai pelanggan dorman atau hampir hilang. Mereka membutuhkan pendekatan pemasaran yang lebih agresif untuk menarik kembali minat, misalnya melalui program diskon atau kampanye win-back.

Sementara itu, Cluster 2 berisi pelanggan dengan Frequency dan Monetary tinggi serta Recency rendah, menggambarkan pelanggan yang sangat aktif dan loyal. Segmen ini merupakan target utama dalam strategi retensi, karena kontribusi mereka terhadap pendapatan perusahaan cenderung signifikan. Cluster 3 mencerminkan pelanggan dengan Frequency rendah namun Monetary tinggi, yakni mereka yang jarang bertransaksi tetapi melakukan pembelian dalam jumlah besar—potensi untuk strategi premium loyalty. Terakhir, Cluster 4 terdiri atas pelanggan dengan Recency rendah serta Frequency dan Monetary sedang, menggambarkan pelanggan baru yang mulai aktif berbelanja.

Hasil segmentasi ini menunjukkan keberagaman pola perilaku pelanggan yang berhasil diidentifikasi melalui pendekatan fuzzy clustering. Dengan pemahaman ini, perusahaan dapat merancang strategi pemasaran yang lebih terarah dan personal, menyesuaikan pesan promosi serta penawaran produk berdasarkan karakteristik spesifik setiap segmen pelanggan. Pelabelan cluster untuk Pelanggan bisa dilihat di gambar 3.

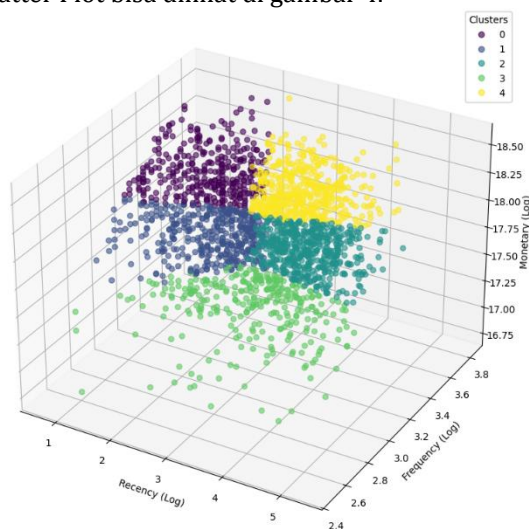
	CustomerID	Recency	Frequency	Monetary	Cluster_Fuzzy
0	CUST000001	1	22	63882371	1
1	CUST000002	18	23	57227590	1
2	CUST000003	22	28	76991379	4
3	CUST000004	10	33	87579881	0
4	CUST000005	21	33	96718054	4

Gambar 3 Label Kluster Pelanggan

Visualisasi Hasil

Visualisasi segmentasi pelanggan dilakukan menggunakan dua pendekatan utama, yaitu plot tiga dimensi (3D scatter plot) dan boxplot distribusi nilai RFM. Visualisasi ini bertujuan untuk memberikan pemahaman yang lebih intuitif terhadap hasil segmentasi yang diperoleh melalui algoritma Fuzzy C-Means (FCM).

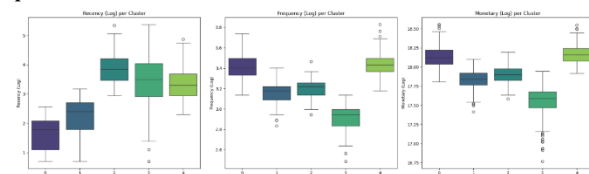
Gambar 4.5 menampilkan distribusi pelanggan dalam ruang tiga dimensi berdasarkan nilai logaritmik dari Recency, Frequency, dan Monetary. Setiap titik merepresentasikan satu pelanggan, sementara perbedaan warna menunjukkan keanggotaan terhadap lima kluster yang telah terbentuk. Terlihat bahwa kluster 2 (kuning) memiliki posisi dominan pada area dengan nilai Frequency dan Monetary tinggi serta Recency rendah, menggambarkan kelompok pelanggan yang aktif dan loyal. Sementara itu, kluster 1 dan 0 (warna biru tua dan ungu) cenderung berada pada area dengan Recency tinggi, menunjukkan pelanggan dorman atau tidak aktif dalam periode waktu terkini. Adapun kluster 3 (hijau muda) terletak pada bagian dengan Frequency rendah tetapi Monetary relatif tinggi, mengindikasikan pelanggan dengan pembelian bernilai besar namun jarang dilakukan. Kluster 4 (hijau terang) menempati area dengan Recency rendah serta Frequency sedang, yang merepresentasikan pelanggan baru yang mulai aktif. Diagram 3D Scatter Plot bisa dilihat di gambar 4.



Gambar 4 Visualisasi Hasil

Selain itu, distribusi statistik dari masing-masing variabel RFM terhadap setiap kluster

divisualisasikan melalui boxplot seperti ditunjukkan pada Gambar 5. Diagram ini memperlihatkan sebaran nilai Recency, Frequency, dan Monetary dalam skala logaritmik pada setiap kluster, sehingga dapat diamati perbedaan karakteristik antarsegmen secara lebih detail. Terlihat bahwa nilai Recency tertinggi terdapat pada kluster 0 dan 1, sedangkan nilai Frequency dan Monetary tertinggi dimiliki oleh kluster 2. Pola ini mendukung hasil numerik pada centroid dan memperjelas hubungan antara intensitas pembelian dan loyalitas pelanggan. Gambar persebaran nilai RFM bisa dilihat di Gambar 5.



Gambar 5 Persebaran Nilai RFM

Visualisasi keanggotaan fuzzy juga memperlihatkan bahwa batas antarsegmen tidak bersifat tegas, melainkan tumpang tindih, sesuai dengan karakteristik algoritma FCM yang memungkinkan setiap pelanggan memiliki derajat keanggotaan pada beberapa kluster sekaligus. Hal ini memperkuat kemampuan interpretatif model dalam menggambarkan perilaku pelanggan yang kompleks dan dinamis. Secara keseluruhan, kedua visualisasi tersebut menegaskan bahwa hasil segmentasi yang diperoleh tidak hanya relevan secara analitis, tetapi juga dapat diimplementasikan secara praktis oleh tim pemasaran e-commerce dalam merancang strategi retensi pelanggan, personalisasi penawaran, serta peningkatan loyalitas secara berkelanjutan.

CONCLUSION

Penelitian ini berhasil menghasilkan segmentasi pelanggan e-commerce yang komprehensif dengan mengintegrasikan model Recency, Frequency, Monetary (RFM) dan algoritma Fuzzy C-Means (FCM). Dataset UrbanMart yang terdiri atas 50.000 transaksi dianalisis melalui rangkaian tahapan mulai dari pemuatan dan pembersihan data, perhitungan nilai RFM, transformasi logaritmik, normalisasi menggunakan Min-Max Scaling, hingga proses penentuan jumlah kluster optimal menggunakan Davies-Bouldin Index (DBI) dan Xie-Beni Index (XBI). Hasil evaluasi menunjukkan bahwa jumlah kluster paling ideal adalah lima, dengan nilai DBI dan XBI terendah masing-masing 0,9903 dan 0,2146, sehingga memberikan keseimbangan terbaik antara kompaksi dan separasi kluster.

Implementasi algoritma FCM menghasilkan lima kelompok pelanggan dengan karakteristik yang berbeda-beda, yaitu pelanggan dorman, pelanggan

berisiko churn, pelanggan loyal dan sangat aktif, pelanggan premium dengan nilai pembelian tinggi namun jarang bertransaksi, serta pelanggan baru yang mulai menunjukkan aktivitas pembelian. Segmentasi fuzzy ini memungkinkan setiap pelanggan memiliki derajat keanggotaan pada lebih dari satu segmen sehingga mencerminkan kompleksitas perilaku pembelian yang tidak dapat ditangkap oleh metode klusterisasi konvensional. Visualisasi melalui 3D scatter plot, boxplot distribusi RFM, dan diagram keanggotaan fuzzy memperkuat interpretasi hasil klusterisasi dan menunjukkan bahwa pendekatan FCM berhasil menggambarkan pola perilaku pelanggan secara lebih realistis dan informatif. Dengan demikian, penelitian ini membuktikan bahwa kombinasi RFM–FCM merupakan pendekatan yang efektif untuk segmentasi pelanggan berbasis data transaksi e-commerce, memberikan dasar yang kuat bagi pengembangan strategi pemasaran berbasis personalisasi, retensi pelanggan, program loyalitas, serta kampanye reaktivasi yang lebih tepat sasaran. Secara keseluruhan, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan metode segmentasi yang lebih adaptif dan representatif, tetapi juga menawarkan nilai praktis yang signifikan bagi pengambil keputusan dalam industri e-commerce untuk memahami struktur perilaku pelanggan dan mengoptimalkan strategi pemasaran berbasis data.

REFERENCE

- [1] A. A. Alsmadi, A. Shuhaiber, M. Al Okaily, A. Al Gasaymeh, and N. Alrawashdeh, "Big data analytics and innovation in e-commerce: current insights and future directions," *Journal of Financial Services Marketing*, vol. 29, no. 4, pp. 1635–1652, 2024, doi: 10.1057/s41264-023-00235-7.
- [2] M. A. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, 2023, doi: 10.1007/s10257-023-00640-4.
- [3] A. Handojo, N. Pujawan, B. Santosa, and M. L. Singgih, "A multi-layer recency-frequency-monetary method for customer priority segmentation in online transaction," *Cogent Engineering*, vol. 10, no. 1, 2023, doi: 10.1080/23311916.2022.2162679.
- [4] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, D. S. Man, and T. G. Le, "An extended RFM model for customer behaviour and demographic analysis in the retail industry," *Business Systems Research*, vol. 14, no. 1, pp. 26–53, 2023, doi: 10.2478/bsrj-2023-000227.
- [5] M. Y. Smaili and H. Hachimi, "New RFM D classification model for improving customer analysis and response prediction," *Ain Shams Engineering Journal*, vol. 14, p. 102254, 2023, doi: 10.1016/j.asej.2023.102254.
- [6] C. Wang, M.-C. Zhou, W. Pedrycz, and Z. W. Li, "Comparative study on noise-estimation-based fuzzy C-means clustering for image segmentation," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 241–253, 2024, doi: 10.1109/TCYB.2022.3217897.
- [7] B. Beiranvand, "Comparison of K-means and Fuzzy C-Means algorithms to optimize clustering results," *International Journal of Data Science and Analytics*, 2024, doi: 10.1016/j.jdsa.2024.00000.
- [8] J. Pérez-Ortega, "Hybrid Fuzzy C-Means Clustering Algorithm, improving solution quality and reducing computational complexity," *Axioms*, vol. 13, no. 9, p. 592, 2024, doi: 10.3390/axioms13090592.
- [9] S.-C. Chang, W.-C. Chuang, and J.-T. Jeng, "New Interval Improved Fuzzy Partitions Fuzzy C-Means Clustering Algorithms under Different Distance Measures for Symbolic Interval Data Analysis," *Applied Sciences*, vol. 13, no. 22, p. 12531, 2023, doi: 10.3390/app132212531.
- [10] G. Marín Díaz, "A Fuzzy XAI Framework for Customer Segmentation and Risk Detection: Integrating RFM, 2 Tuple Modeling, and Strategic Scoring," *Mathematics*, vol. 13, no. 13, p. 2141, 2025, doi: 10.3390/math13132141.
- [11] D. Krasnov, D. Davis, K. Malott, Y. Chen, X. Shi, and A. F. Wong, "Fuzzy CMeans Clustering: A Review of Applications in Breast Cancer Detection," *Entropy*, vol. 25, no. 7, 2023, doi: 10.3390/e25071021.
- [12] C. Song, "Performance study of fuzzy C-means clustering algorithm in discrete data processing," *Journal of Intelligent & Fuzzy Systems*, vol. 47, no. 3, pp. 3211–3223, 2024, doi: 10.3233/JIFS-235082.
- [13] C. Rungruang, P. Riyapan, A. Intarasit, and J. Muangprathub, "RFM model customer segmentation based on hierarchical approach using formal concept analysis," *Expert Systems with Applications*, vol. 237, p. 121449, 2024, doi: 10.1016/j.eswa.2023.121449.

- [14] A. Wasilewski, K. Juszczyszyn, and V. Suryani, "Multi-factor evaluation of clustering methods for e-commerce customer segmentation," *Egyptian Informatics Journal*, vol. 28, p. 100562, 2024, doi: 10.1016/j.eij.2024.100562.
- [15] L. Abednego, C. E. Nugraheni, and A. Salsabina, "Customer segmentation: Transformation from data to marketing strategy BT - IAIC SEMNASTIK 2023," 2023. doi: 10.34306/conferenceseries.v4i1.645.
- [16] M. I. J. Putra, V. Alexander, A. Chusyairi, and R. A. Abdurrahman, "Enhancing e-commerce customer segmentation with fuzzy C-means soft clustering probabilities," *Journal of Artificial Intelligence and Computational Intelligence*, vol. 9, no. 5, 2025, doi: 10.30871/jaic.v9i5.10652.
- [17] A. Griva, E. Papaioannou, A. Kokkinis, and K. Konstantoulaki, "A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data," *Journal of Marketing Analytics*, vol. 12, no. 1, pp. 41–59, 2024.
- [18] M. A. Uddin, R. Ahmed, and S. Alam, "Data-driven strategies for digital native markets: consumer insights and segmentation in online retail," *Journal of Retailing and Consumer Services*, vol. 71, p. 103800, 2024, doi: 10.1016/j.jretconser.2024.103800.
- [19] A. M. Ifrim, "A Hybrid Numerical–Semantic Clustering Algorithm Based on RFM and Semantic Data," *Algorithms*, vol. 18, no. 10, 2025, doi: 10.3390/a18100607.
- [20] Y. Syahra, A. Fadlil, and H. Yuliansyah, "Customer Segmentation Using RFM and KMeans: Preprocessing with MinMax Normalization," *Sinkron: Jurnal Teknik Informatika*, vol. 8, no. 1, 2025, doi: 10.14710/sinkron.v8i1.14907.
- [21] K. Zhou and S. Yang, "Intelligent customer segmentation: unveiling consumer patterns with RFM and fuzzy Cmeans clustering," 2025. doi: [DOI pending].