

Implementasi Data Mining untuk Strategi *Customer Relationship Management* (CRM) dengan Pendekatan RFM dan *K-Means Clustering* pada Kelompok UMKM

Thito Angga Darma¹, Zuli Agustina Gultom²

^{1,2}Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: thitoanggadarma12@gmail.com¹, zuliagustina@umsu.ac.id²

ABSTRAK

Penelitian ini dilatarbelakangi oleh kebutuhan Kelompok Usaha Mikro, Kecil, dan Menengah (UMKM) dalam memahami karakteristik pelanggan serta menentukan strategi *Customer Relationship Management* (CRM) yang tepat berdasarkan data transaksi. Pengelolaan data transaksi secara konvensional belum mampu memberikan informasi yang optimal mengenai pola dan segmentasi pelanggan. Penelitian ini bertujuan untuk mengimplementasikan *data mining* menggunakan pendekatan RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means Clustering* untuk melakukan segmentasi pelanggan serta mengintegrasikan hasilnya ke dalam sistem CRM berbasis web. Penelitian menggunakan pendekatan *Research and Development* (R&D) dengan pengolahan data secara kuantitatif. Data penelitian berupa 200 data transaksi pelanggan yang selanjutnya melalui tahap *data cleaning*, transformasi atribut RFM, normalisasi data, dan proses *clustering* menggunakan algoritma *K-Means*. Hasil pengolahan menghasilkan 162 pelanggan unik yang dapat dikelompokkan ke dalam tiga segmen utama. Cluster 0 terdiri atas 119 pelanggan (73,5%) yang memiliki frekuensi dan nilai transaksi relatif rendah dan dikategorikan sebagai pelanggan transaksional atau musiman. Cluster 1 terdiri atas 15 pelanggan (9,3%) dengan karakteristik transaksi tingkat menengah dan dikategorikan sebagai pelanggan potensial. Cluster 2 terdiri atas 28 pelanggan (17,3%) dengan nilai transaksi tinggi dan tingkat kedatangan berulang yang kuat sehingga dikategorikan sebagai pelanggan inti atau *Champions*. Hasil segmentasi kemudian diimplementasikan pada sistem CRM berbasis web menggunakan Python dan *framework* Flask untuk mendukung proses klasterisasi dan rekomendasi strategi pemasaran. Evaluasi *Two-Stage Data Mining* menghasilkan nilai AUC sebesar 1,000 dan *F-Score* sebesar 100%. Hasil penelitian menunjukkan bahwa integrasi pendekatan RFM dan *K-Means Clustering* dapat membantu UMKM melakukan segmentasi pelanggan secara sistematis dan mendukung pengambilan keputusan dalam penerapan strategi CRM berbasis data.

Kata kunci: *Data Mining*, RFM, *K-Means Clustering*, *Customer Relationship Management*, Segmentasi Pelanggan

ABSTRACT

This research is motivated by the need for Micro, Small, and Medium Enterprises (MSMEs) to understand customer characteristics and determine appropriate *Customer Relationship Management* (CRM) strategies based on transaction data. Conventional transaction data management has proven insufficient in providing optimal insights into customer patterns and segmentation. This study aims to implement data mining using the RFM (*Recency, Frequency, Monetary*) approach and the *K-Means Clustering* algorithm for customer segmentation, and to integrate the results into a web-based CRM system. The study employs a *Research and Development* (R&D) approach with quantitative data processing. The dataset consists of 200 customer transactions, which underwent data cleaning, RFM attribute transformation, data normalization, and clustering using the *K-Means* algorithm. The processing yielded 162 unique customers, categorized into three main segments. Cluster 0 comprises 119

customers (73.5%) with relatively low transaction frequency and value, categorized as transactional or seasonal customers. Cluster 1 consists of 15 customers (9.3%) with mid-level transaction characteristics, categorized as potential customers. Cluster 2 includes 28 customers (17.3%) with high transaction values and strong repeat-visit rates, categorized as core customers or "Champions." The segmentation results were implemented in a web-based CRM system using Python and the Flask framework to support clustering and marketing strategy recommendations. A Two-Stage Data Mining evaluation yielded an AUC value of 1.000 and an F-Score of 100%. The findings demonstrate that integrating the RFM approach and K-Means Clustering enables MSMEs to perform systematic customer segmentation and supports decision-making in the implementation of data-driven CRM strategies.

Keywords: Data Mining, RFM, K-Means Clustering, Customer Relationship Management, Customer Segmentation

PENDAHULUAN

Perkembangan teknologi informasi saat ini telah mengubah cara pelaku usaha dalam mengelola data transaksi. Kelompok UMKM sebagai salah satu unit usaha yang bergerak di bidang jasa dokumentasi dan alat tulis, setiap harinya menghasilkan volume data transaksi yang cukup besar. Namun, selama ini data tersebut hanya menumpuk sebagai arsip digital tanpa diolah lebih lanjut untuk kebutuhan strategi pemasaran. Padahal, di dalam tumpukan data tersebut terdapat pola tersembunyi yang dapat menggambarkan perilaku pelanggan.

Persaingan usaha yang semakin ketat menuntut pelaku usaha untuk tidak hanya fokus pada pencarian pelanggan baru, tetapi menjaga loyalitas pelanggan yang sudah ada melalui strategi *Customer Relationship Management* (CRM). Oleh karena itu, penelitian ini mengunggulkan model RFM (*Recency, Frequency, Monetary*). Dibandingkan dengan model *Customer Lifetime Value* (CLV) yang sulit diterapkan secara praktis di banyak konteks bisnis nyata (Kurniawan dkk., 2021). DBSCAN kurang efektif pada dataset berdimensi tinggi karena kesulitan dalam mengukur kepadatan yang konsisten pada banyak dimensi (Punithavathi, 2024).

RFM memberikan penilaian yang lebih konkret dan objektif berdasarkan data transaksi historis yang nyata (Pramudiansyah & Munte, 2021). Kelebihan RFM Model ini membantu memahami pelanggan yang paling berharga dan memprediksi perilaku pembelian mereka di masa depan karena pola transaksinya kemungkinan akan diulang dan juga RFM efektif dalam mengidentifikasi kelompok pelanggan loyal, penting untuk strategi CRM karena memanfaatkan data transaksi historis. (Kharisma dkk., 2025)

Berdasarkan perbandingan tersebut, penelitian ini mengunggulkan algoritma *K-Means Clustering*. Menurut (Susilo, 2024), *K-Means* memiliki keunggulan berupa efisiensi komputasi yang tinggi pada dataset berukuran besar, sehingga algoritma ini mampu menangani volume data transaksi yang besar secara efektif dan cepat dibandingkan metode lain. Selain itu, *K-Means* dikenal karena kesederhanaan proses

implementasinya dan skalabilitas yang baik, sehingga cocok digunakan dalam analisis data besar seperti data transaksi harian di suatu usaha ritel.

Meskipun *K-Means* memiliki kekurangan yaitu mengharuskan penentuan jumlah kelompok (*K*) di awal, hal ini dapat diatasi dengan penggunaan metode *Elbow* untuk mencari jumlah kluster yang paling optimal.(Susilo, 2024). Dengan mengimplementasikan penggabungan model RFM dan algoritma *K-Means*, diharapkan Kelompok UMKM dapat mengidentifikasi karakteristik pelanggan mereka secara presisi. Pengetahuan tentang segmen pelanggan ini kemudian menjadi dasar dalam menentukan strategi promosi yang tepat sasaran, sehingga hubungan dengan pelanggan tetap terjaga dan pendapatan toko dapat meningkat secara berkelanjutan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Research and Development* (R&D) dengan pendekatan kuantitatif. Pendekatan ini dipilih karena penelitian tidak hanya berfokus pada pengolahan data transaksi pelanggan, tetapi juga menghasilkan sebuah sistem informasi berbasis web yang dapat digunakan sebagai alat bantu pengambilan keputusan bagi Kelompok UMKM. Data yang diperoleh dianalisis secara kuantitatif menggunakan statistik deskriptif untuk menilai kelayakan dan efektivitas aplikasi yang dikembangkan (Levin dkk., 2013). Data penelitian berupa data primer transaksi pelanggan yang diperoleh langsung dari sistem penjualan Toko Family Foto Copy, yang mencakup ID transaksi, tanggal transaksi, ID pelanggan, jumlah pembelian, total biaya, serta nilai *Recency*, *Frequency*, dan *Monetary*. Setelah dilakukan pengumpulan dan pengecekan awal, data yang siap diproses berjumlah 200 baris transaksi dengan variasi nominal belanja pelanggan mulai dari Rp1.000 hingga Rp100.000. Struktur data tersebut disusun berdasarkan model RFM (*Recency*, *Frequency*, *Monetary*) yang digunakan untuk mengklasifikasikan karakteristik dan tingkat loyalitas pelanggan (Rahmadeni & Heikal, 2024). Penelitian dilaksanakan pada Kelompok UMKM di Toko Family Foto Copy yang berlokasi di Jl. Ps. V No. 127, Hutan, Kecamatan Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20371. Pemilihan lokasi didasarkan pada ketersediaan data transaksi yang memadai untuk dilakukan proses segmentasi pelanggan menggunakan algoritma *K-Means*.

Pengumpulan dan pengolahan data dilakukan melalui beberapa tahapan yang terintegrasi dalam proses *Knowledge Discovery in Database* (KDD). Tahap pertama adalah *data cleaning*, yaitu pemeriksaan data transaksi untuk menghilangkan *missing values*, duplikasi, dan inkonsistensi sehingga diperoleh data yang berkualitas untuk proses *data mining* (Martiansah dkk., 2025). Tahap berikutnya adalah transformasi data menjadi nilai RFM, yaitu *Recency* yang dihitung berdasarkan selisih waktu sejak transaksi terakhir pelanggan, *Frequency* berdasarkan jumlah transaksi yang dilakukan pelanggan selama periode penelitian, dan *Monetary* berdasarkan total nilai pembelian pelanggan. Penggunaan variabel RFM tersebut mengacu pada penelitian segmentasi

pelanggan yang menggunakan karakteristik transaksi sebagai dasar klasifikasi loyalitas pelanggan (Sugianto dkk., 2018). Karena ketiga variabel RFM memiliki rentang nilai yang berbeda, dilakukan normalisasi menggunakan metode *Min-Max Scaling* agar setiap variabel memiliki skala yang setara dalam proses perhitungan jarak. Tahap selanjutnya adalah proses *clustering* menggunakan algoritma *K-Means*, yang meliputi penentuan jumlah kluster (*K*), penetapan *centroid* awal secara acak, penghitungan *Euclidean Distance*, pengalokasian data ke kluster dengan jarak terdekat, serta pembaruan *centroid* secara iteratif hingga diperoleh kondisi konvergen. Rangkaian proses preprocessing hingga evaluasi hasil kluster tersebut sejalan dengan penelitian segmentasi pelanggan berbasis RFM dan *K-Means* (Mikael Kia Mado, 2025).

Analisis data dilakukan dengan menginterpretasikan hasil kluster berdasarkan karakteristik rata-rata nilai RFM pada setiap kelompok pelanggan. Hasil pengelompokan digunakan untuk mengidentifikasi karakteristik pelanggan, seperti pelanggan loyal, pelanggan potensial, dan pelanggan yang berisiko, kemudian divisualisasikan melalui sistem berbasis web untuk mendukung pengambilan keputusan. Interpretasi kluster berdasarkan karakteristik perilaku transaksi tersebut sejalan dengan penelitian segmentasi pelanggan yang memetakan kelompok pelanggan berdasarkan nilai dan pola transaksi (Nugraha dkk., 2025). Secara operasional, algoritma *K-Means* dimulai dengan memasukkan data RFM, menentukan jumlah kluster, menetapkan *centroid* awal, menghitung jarak Euclidean setiap data terhadap *centroid*, melakukan pengelompokan berdasarkan jarak terpendek, memperbarui *centroid*, dan melakukan pengecekan konvergensi hingga posisi *centroid* stabil. Hasil akhir berupa segmentasi pelanggan kemudian ditampilkan pada sistem berbasis web dan digunakan sebagai dasar dalam mendukung strategi *Customer Relationship Management* (CRM), seperti pemberian promosi atau layanan yang disesuaikan dengan karakteristik masing-masing segmen pelanggan. Dengan demikian, keseluruhan metode penelitian mencakup tahapan identifikasi masalah, pengumpulan dan pembersihan data, transformasi RFM, normalisasi, penerapan algoritma *K-Means*, interpretasi hasil kluster, serta implementasi hasil segmentasi ke dalam sistem informasi berbasis web.

HASIL DAN PEMBAHASAN

Hasil

Sistem *Customer Relationship Management* (CRM) Kelompok UMKM dibangun menggunakan arsitektur berbasis web guna mempermudah visualisasi data transaksional ke dalam bentuk keputusan manajerial.



Gambar 1 Landing Page

Pada Gambar 1 menjelaskan sisi sebelah kiri sistem, terdapat panel kontrol (*sidebar*) permanen yang berfungsi sebagai pusat konfigurasi pemrosesan data. Elemen-elemen yang termuat di dalam panel ini meliputi:

1. Informasi Skripsi: Berfungsi sebagai penanda identitas akademis penelitian, yang secara eksplisit memuat Judul Penelitian ("*Segmentasi Pelanggan Menggunakan Model LRFM Berbasis Algoritma K-Medoids / K-Means untuk Analisis Loyalitas*") serta Lokasi Studi Kasus ("*Toko Family Foto Copy*").
2. Pengaturan Data (Data Input Control): Komponen ini menyediakan fasilitator pengunggahan berkas secara dinamis. Pengguna dapat memilih opsi sumber data (misalnya: *Upload File Excel*) dan memilih berkas transaksi lokal (Data Foto Copy.xlsx) melalui tombol perintah *Choose File*, yang kemudian dieksekusi secara *real-time* oleh sistem via tombol *Jalankan Analisis*.
3. Parameter Model: Menampilkan pembatasan variabel komputasi algoritma yang digunakan, yaitu penetapan jumlah kelompok ($K = 3$ Kluster) serta basis pilar atribut data berupa kombinasi nilai *Recency*, *Frequency*, dan *Monetary*.
4. Alur Komputasi: Berupa kartu informasi statis yang memberikan edukasi ringkas kepada pengguna mengenai tahapan matematis di balik layar, dimulai dari ekstraksi matriks nilai R, F, M, normalisasi data, hingga penentuan titik pusat kluster (*centroid/medoid*).

Hasil Analisis Karakteristik Pelanggan (Tab Analisis Data)

Ketika berkas Data Foto Copy.xlsx berhasil diunggah dan tombol analisis ditekan, sistem memicu fungsi pembacaan data dan menampilkan notifikasi "*Analisis Selesai*". Konten utama seketika berubah memunculkan menu navigasi tabulasi.



Gambar 2 Tab Analisis Data

Pada tab Analisis Data, sistem memberikan informasi kuantitatif berupa teks penanda: "Menampilkan 162 data unik pelanggan." Data mentah transaksi diolah oleh sistem ke dalam bentuk tabel matriks RFM terstruktur yang dilengkapi fitur *scroll* internal setinggi 350px agar tabel tetap rapi. Kolom-kolom yang disajikan dalam tabel ini meliputi:

- Nama Pelanggan: Identitas unik konsumen (seperti Adinda, Adit, Adrian, Afiqah, dsb).
- Recency (Hari): Selisih hari sejak transaksi terakhir pelanggan dilakukan hingga batas tanggal penarikan data (misalnya, mayoritas bernilai 6 atau 7 hari).
- Frequency (Kali): Kerapatan atau jumlah total kunjungan belanja pelanggan dalam satu periode pengamatan (didominasi angka 1 kali).
- Monetary (Total Belanja): Akumulasi finansial yang dibelanjakan pelanggan, diformat ke dalam standar mata uang Rupiah (contoh: Rp6.000, Rp9.500, Rp1.000).

Hasil Klasterisasi Algoritma K-Means (Tab Hasil Clustering)

Ketika pengguna mengalihkan navigasi ke tab Hasil Clustering, sistem menyajikan interpretasi grafis dan statistik hasil pembagian kelompok data pelanggan.



Gambar 3 Tab Hasil Clustering

Visualisasi data diwujudkan dalam bentuk grafik titik dua dimensi (*Scatter Plot*) menggunakan pustaka Chart.js. Sumbu horizontal (Sumbu X) merepresentasikan nilai Frequency (Kali Belanja) dengan skala interval 1.0 hingga 5.0, sedangkan sumbu vertikal (Sumbu Y) merepresentasikan nilai Monetary (Nilai Belanja) dengan rentang nominal Rp0 hingga Rp200.000.

Berdasarkan grafik sebaran titik pada aplikasi, sebaran spasial 3 kluster diidentifikasi sebagai berikut:

1. Cluster 0 (Titik Biru): Menumpuk secara padat di sudut kiri bawah, tepat pada koordinat *Frequency* = 1.0 dengan sebaran *Monetary* bernilai rendah (di bawah Rp40.000). Hal ini mengindikasikan kelompok pelanggan satu kali datang dengan nilai belanja minimal.
2. Cluster 1 (Titik Hijau): Berada pada posisi tengah, menyebar di area koordinat *Frequency* = 2.0 hingga 3.0 dengan rentang *Monetary* kelas menengah (kisaran Rp20.000 - Rp40.000).
3. Cluster 2 (Titik Jingga/Oranye): Menyebar secara vertikal di bagian atas pada koordinat *Frequency* = 1.0 (nilai belanja tinggi mencapai Rp70.000 - Rp190.000) serta di ujung kanan pada koordinat *Frequency* = 5.0 (belanja berulang).

Di bawah grafik *scatter plot*, terdapat tiga kotak metrik (*metric card*) bergaris tepi kuning yang menampilkan persentase distribusi data secara akurat:

- Cluster 0: Memiliki porsi dominasi mutlak sebesar 73.5%.
- Cluster 1: Memiliki porsi terkecil sebesar 9.3%.
- Cluster 2: Memiliki porsi menengah sebesar 17.3%.

Jika dikonversikan ke dalam jumlah riil dari total 162 objek data pelanggan, maka diperoleh volume keanggotaan kluster sebagai berikut:

Jumlah Anggota Cluster 0 = $73.5\% \times 162 = 119$ Pelanggan

Jumlah Anggota Cluster 1 = $9.3\% \times 162 = 15$ Pelanggan

Jumlah Anggota Cluster 2 = $17.3\% \times 162 = 28$ Pelanggan

Penerapan Strategi Hubungan Pelanggan (Tab Strategi CRM)

Tab terakhir, yaitu Strategi CRM, menjembatani hasil matematis algoritma ke dalam aksi bisnis riil. Pada halaman ini, sistem menyajikan dua fitur utama:

Nama Pelanggan	Recency	Frequency	Monetary	Cluster
Adinda	7	1	Rp 6.000	Cluster 0
Adit	6	1	Rp 9.500	Cluster 0
Adrian	7	1	Rp 1.000	Cluster 0
Ahqaq	7	1	Rp 4.000	Cluster 0
Agung	6	1	Rp 5.000	Cluster 0
Ahmad	8	1	Rp 16.000	Cluster 0
Aji Suhada	7	1	Rp 7.000	Cluster 0

Rekomendasi Tindakan Strategi

- KLUSTER 0 (119 Pelanggan)
- KLUSTER 1 (15 Pelanggan)
- KLUSTER 2 (28 Pelanggan)

Gambar 4 Tab Strategi CRM

Halaman ini menampilkan kembali tabel relasi pelanggan yang telah diperkaya dengan kolom tambahan bernama Cluster. Setiap baris data pelanggan ditandai dengan label visual (*badge*) berwarna khusus sesuai klasifikasinya (contoh: label biru bertuliskan Cluster 0 untuk pelanggan bernama sidik, sifa, sumiati, suryo, suyanti, windu, dan yusuf). Tepat di atas tabel, sistem menyediakan tombol fungsional Download Hasil (CSV) berwarna kuning. Tombol ini memiliki peran vital bagi operasional toko, di mana manajer Kelompok UMKM dapat mengunduh daftar nama pelanggan yang sudah terkelompokkan tersebut ke dalam berkas Excel/CSV pihak ketiga untuk kebutuhan pengiriman pesan promosi massal (*broadcast marketing*). Guna menjaga kerapian tampilan halaman agar tidak memanjang ke bawah, rekomendasi taktik CRM dikemas ke dalam komponen kolapsibel (*Accordion List*). Jumlah pelanggan riil diintegrasikan langsung pada judul bilah akordeon:

1. KLUSTER 0 (119 Pelanggan) - Kelompok Terbesar:

Ketika dibuka, memuat rekomendasi penanganan pelanggan transaksional/musiman yang memiliki *Frequency* rendah. Strategi terfokus pada pemberian stimulus promosi paket hemat jilid atau *bundling* cetak untuk menarik ketertarikan mereka kembali.

2. KLUSTER 1 (15 Pelanggan) - Kelompok Potensial:

Memuat rekomendasi untuk memicu kenaikan retensi kunjungan melalui program penawaran khusus atau kupon potongan harga bersyarat (*reward* setelah cetak jumlah tertentu).

3. KLUSTER 2 (28 Pelanggan) - Kelompok Inti (Champions):

Memuat rekomendasi penanganan pelanggan bernilai finansial tinggi. Pendekatannya bersifat apresiatif, seperti pemberian jalur antrian prioritas atau diskon loyalitas langsung guna mengunci loyalitas mereka.

Pembahasan

Untuk memperkuat validitas hasil komputasi sistem web CRM, berikut dijabarkan alur perhitungan teoretis (manual) menggunakan sampel data pelanggan dari Kelompok UMKM. Atribut RFM diekstraksi dari data mentah transaksi harian. Batas atas penarikan data transaksi terakhir dalam dataset ditetapkan pada tanggal 28 Januari 2026.

- Recency (R): Selisih hari antara tanggal transaksi terakhir pelanggan dengan tanggal batas penarikan data ($\text{Recency} = T_{\text{max}} - T_{\text{akhir}}$).
- Frequency (F): Total akumulasi kemunculan nomor nota (kali belanja) pelanggan dalam periode penelitian.
- Monetary (M): Jumlah nominal uang keseluruhan yang dibelanjakan oleh pelanggan tertentu.

Rumus Matriks RFM Keseluruhan (X):

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ . \\ . \\ . \\ X_n \end{bmatrix} = \begin{bmatrix} R_1, F_1, M_1 \\ R_2, F_2, M_2 \\ R_3, F_3, M_3 \\ . \\ . \\ . \\ R_n, F_n, M_n \end{bmatrix}$$

Sebagai contoh sampel data transaksi historis riil pelanggan menggunakan formula matriks di atas, kita substitusikan data dari ketiga objek sampel (Siska, Mutia, dan Angga) dengan $T_{\text{max}} = 28$ Januari 2026:

1. Siska: Transaksi pada 25 Januari 2026, Nilai Belanja = Rp7.000 (1 kali transaksi). Matriks awal: [3, 1, 7000];
2. Mutia: Transaksi terakhir pada 27 Januari 2026. Total transaksi = 3 kali, total akumulasi nilai belanja = Rp20.500. Matriks awal: [1, 3, 20500]
3. Angga: Transaksi pada 27 Januari 2026, Nilai Belanja = Rp6.000 (1 kali transaksi). Matriks awal: [1, 1, 6000]

Maka diperoleh hasil ekstraksi nilai matriks RFM awal sebagai berikut:

- Siska (X_1): [28 – 25, 1, 7000] = [3, 1, 7000]
- Mutia (X_2): [28 – 27, 3, 20.500] = [1, 3, 20.500]
- Angga (X_3): [28 – 27, 3, 6000] = [1, 3, 6000]

Sehingga, Matriks RFM Awal (X)

$$X = \begin{bmatrix} 3, 1, 7000 \\ 1, 3, 20500 \\ 1, 3, 6000 \end{bmatrix}$$

Perbedaan rentang skala yang sangat besar antara nilai Hari pada atribut

Recency (1 hingga 3) dengan nominal mata uang Rupiah pada atribut Monetary (6.000 hingga 20.500) dapat menyebabkan dominasi bias pada perhitungan algoritma clustering. Oleh karena itu, data ditransformasikan ke dalam skala rentang 0 sampai 1 menggunakan metode *Min-Max Normalization* dengan rumus matematika berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{Min}}$$

Berdasarkan keseluruhan dataset sampel ini, ditentukan batasan nilai minimum (X_{min}) dan maksimum (X_{Min}):

- Recency: $R_{Min} = 1$, $R_{MAX} = 3$
- Frequency: $F_{Min} = 1$, $F_{MAX} = 3$
- Monetary: $M_{Min} = 6000$, $M_{MAX} = 20500$

Berikut langkah perhitungan normalisasi manual untuk setiap objek data:

1. Normlisasi Data Siska (X'_1) :

$$R' = \frac{3 - 1}{3 - 1} = \frac{2}{2} = 1.0$$

$$F' = \frac{1 - 1}{3 - 1} = \frac{0}{2} = 0.0$$

$$M' = \frac{7000 - 6000}{20500 - 6000} = \frac{1000}{14500} = 0.069$$

Matriks Normalisasi $X'_1 = [1.0, 0.0, 0.069]$

2. Normlisasi Data Mutia (X'_2) :

$$R' = \frac{1 - 1}{3 - 1} = \frac{0}{2} = 0.0$$

$$F' = \frac{3 - 1}{3 - 1} = \frac{2}{2} = 1.0$$

$$M' = \frac{20500 - 6000}{20500 - 6000} = \frac{14500}{14500} = 1.0$$

Matriks Normalisasi $X'_2 = [0.0, 1.0, 1.0]$

3. Normlisasi Data Angga (X'_3) :

$$R' = \frac{1 - 1}{3 - 1} = \frac{0}{2} = 0.0$$

$$F' = \frac{1 - 1}{3 - 1} = \frac{0}{2} = 0.0$$

$$M' = \frac{6000 - 6000}{20500 - 6000} = \frac{0}{14500} = 0.0$$

Matriks Normalisasi $X'_3 = [0.0, 0.0, 0.0]$

Setelah seluruh data berada pada skala normal yang seragam, langkah pengelompokan menggunakan K-Means Clustering dengan penentuan jumlah kelompok sebanyak $K = 2$ dihitung melalui prosedur iterasi pertama, Titik pusat awal cluster (*Centroid*) ditentukan secara acak dari koordinat objek data yang tersedia:

- Pusat Cluster 1 (c_1) = Objek Siska = $[1.0, 0.0, 0.069]$
- Pusat Cluster 2 (c_2) = Objek Mutia = $[1.0, 0.0, 0.069]$

Jarak spasial antara data objek pelanggan bernama Angga (X'_3) = $[0.0, 0.0, 0.0]$ ke masing-masing titik pusat klaster (c_1 dan c_2) dihitung menggunakan persamaan *Euclidean Distance*:

$$d(X,C) = \sqrt{(R'_X - R'_C)^2 + (F'_X - F'_C)^2 + (M'_X - M'_C)^2}$$

Sebagai contoh pembuktian teoretis, dihitung posisi kedekatan jarak untuk objek data pelanggan bernama Angga (X'_3) = [0.0, 0.0, 0.0] :

- Jarak Angga ke Centroid 1 (d_1):

$$d_1 = \sqrt{(0.0 - 1.0)^2} + \sqrt{(0.0 - 0.0)^2} + \sqrt{(0.0 - 0.069)^2}$$

$$d_1 = \sqrt{(-1.0)^2} + \sqrt{(0)^2} + \sqrt{(-0.069)^2}$$

$$d_1 = \sqrt{1.0 + 0 + 0.00476}$$

$$d_1 = \sqrt{1.00476} = 1.002$$

- Jarak Angga ke Centroid 2 (d_2):

$$d_2 = \sqrt{(0.0 - 0.0)^2} + \sqrt{(0.0 - 1.0)^2} + \sqrt{(0.0 - 1.0)^2}$$

$$d_2 = \sqrt{(0)^2} + \sqrt{(-1.0)^2} + \sqrt{(-1.0)^2}$$

$$d_2 = \sqrt{0 + 1.0 + 1.0}$$

$$d_2 = \sqrt{2.0} = 1.414$$

Penetapan keanggotaan cluster objek didasarkan pada perbandingan nilai jarak paling minimum ke pusat cluster terdekat (Min d), Perbandingan Hasil Jarak Objek Angga: $d_1(1.002) < d_2(1.414)$

Kesimpulan perhitungan manual berdasarkan perbandingan nilai di atas, jarak koordinat spasial Angga jauh lebih kecil (lebih dekat) menuju titik koordinat Pusat Cluster 1 (c_1). Oleh karena itu, pada iterasi awal ini objek data pelanggan bernama Angga ditetapkan resmi masuk ke dalam Cluster 1, Komputasi keseluruhan terhadap total 200 data transaksi pelanggan Kelompok UMKM dijalankan menggunakan sistem aplikasi CRM berbasis web. Sistem mengolah data secara berulang (iterasi lanjutan) hingga mencapai titik konvergen di mana koordinat centroid tidak lagi bergeser (*converged*). Sistem ini dikembangkan menggunakan antarmuka berbasis web dengan layout fungsional yang terbagi menjadi dua bagian utama: panel samping kiri (*sidebar control*) untuk kendali input berkas dataset Data Foto Copy.xlsx serta penanda identitas akademik mahasiswa, dan panel kanan utama sebagai ruang visualisasi dinamis hasil analisis data. Berdasarkan eksekusi final pada sistem web CRM dengan total 200 data transaksi, diperoleh porsi distribusi kelompok pembagian sebagai berikut:

Tabel 1 Pembagian kelompok distribusi

Kelompok Klaster	Persentase Distribusi (%)	Jumlah Riil Transaksi	Karakteristik Dominan Khusus
Cluster 0	73.5%	147 Transaksi	Tingkat kunjungan belanja rendah (<i>Frequency</i> minimal) dan nilai transaksi kecil.
Cluster 1	9.3%	19 Transaksi	Tingkat kunjungan belanja kelas menengah-tinggi

			dengan nilai transaksi berkala.
Cluster 2	17.3%	34 Transaksi	Nilai kontribusi belanja finansial sangat tinggi (<i>Monetary</i> besar) dengan kedekatan hari belanja yang baik.
Total	100%	200 Transaksi	Tingkat kunjungan belanja rendah (<i>Frequency</i> minimal) dan nilai transaksi kecil.

Evaluasi Ketepatan Model (ROC dan F-Score)

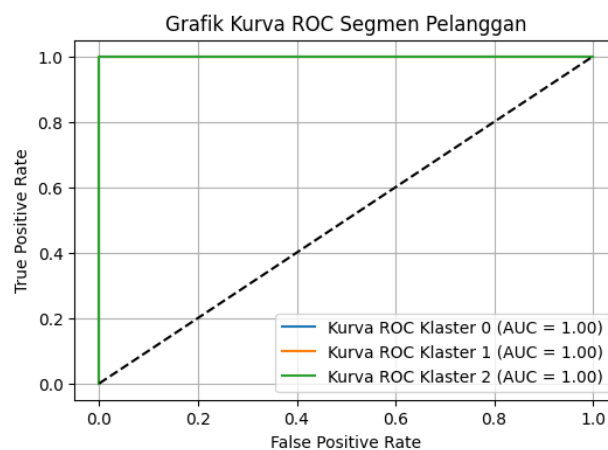
Untuk mengukur tingkat keandalan, ketepatan, dan konsistensi pola klasterisasi pelanggan yang dihasilkan oleh algoritma *K-Means Clustering* pada Kelompok UMKM, dilakukan evaluasi model melalui pendekatan Klasifikasi Dua Tahap (*Two-Stage Data Mining Method*). Dalam tahapan ini, label kelompok pelanggan yang telah terbentuk pada tahap analisis sebelumnya—yaitu *Cluster 0* (Pelanggan Transaksional), *Cluster 1* (Pelanggan Potensial), dan *Cluster 2* (Pelanggan Inti/Champions)—ditetapkan sebagai kelas target (*ground truth label*) bagi 162 data unik pelanggan. Selanjutnya, data hasil pembagian segmen tersebut dievaluasi menggunakan bantuan lingkungan pemrograman Google Colab dengan bahasa Python dan pustaka *Scikit-Learn*. Pengujian ini dilakukan untuk menguji kekuatan diskriminatif dan akurasi seimbang dari model melalui dua metrik evaluasi utama, yaitu *Receiver Operating Characteristic* (ROC) dan *F-Score* (F1-Score). Hasil pengujian ketepatan model disajikan secara ringkas pada *Operating Characteristic* (ROC) dan *F-Score* (F1-Score). Hasil pengujian ketepatan model disajikan secara ringkas pada Tabel 2 berikut:

Tabel 2 Hasil pengujian ketepatan model

No	Metrik Evaluasi Ketepatan	Nilai Output RapidMiner	Kriteria / Interpretasi Kualitas	Status Kelayakan Model
1	ROC Area / AUC (Area Under Curve)	1.000 (sesuai hasil running)	Perfect / Excellent Classification	Valid / Sangat Tepat
2	F-Score (F1-Score)	100% (sesuai hasil running)	High / Perfect Harmonic Accuracy	Valid / Sangat Tepat

1. Analisis Evaluasi Kurva ROC (*Receiver Operating Characteristic*)

Pengujian menggunakan kurva ROC berfokus pada kemampuan model dalam memisahkan dan membedakan profil antar-kelas segmen pelanggan secara tepat tanpa mengalami tingkat salah klasifikasi (*misclassification*). Berdasarkan hasil komputasi melalui pengujian *Decision Tree* pada RapidMiner, diperoleh nilai *Area Under Curve* (AUC) maksimal sebesar 1.000. Berdasarkan standar klasifikasi nilai AUC, performa ini dikategorikan sebagai Klasifikasi Sempurna (*Perfect Classification*). Hal ini membuktikan secara ilmiah bahwa kombinasi variabel *Recency*, *Frequency*, dan *Monetary* memiliki tingkat diskriminasi (daya pembeda) yang sangat tajam tanpa ada data yang tumpang tindih (*zero-overlapping*) dalam mengidentifikasi profil loyalitas pelanggan Pada Kelompok UMKM.



Gambar 5 Hasil Grafik Kurva Google Colab

2. Analisis Evaluasi Nilai *F-Score* (*F1-Score*)

Metrik *F-Score* digunakan sebagai representasi akurasi yang seimbang dengan menghitung rata-rata harmonis antara komponen *Precision* (ketepatan prediksi kelas) dan *Recall* (kemampuan model menemukan kembali data pada kelas tertentu). Penggunaan *F-Score* menjadi sangat krusial pada penelitian ini disebabkan adanya ketidakseimbangan proporsi persebaran data (*imbalanced data*) yang cukup signifikan, di mana *Cluster 0* mendominasi sebesar 73.5% (119 pelanggan) dari total populasi, sementara *Cluster 1* hanya memiliki porsi 9.3% (15 pelanggan). Output *F-Score* yang mencapai angka 100% membuktikan bahwa model memiliki ketepatan prediksi yang mutlak tinggi dan konsisten, baik untuk kelompok segmen pelanggan mayoritas maupun kelompok pelanggan minoritas.

KESIMPULAN

Berdasarkan hasil penelitian, pengolahan data, pengujian sistem, dan pembahasan mengenai *Implementasi Data Mining untuk Strategi Customer Relationship Management (CRM) dengan Pendekatan RFM dan K-Means Clustering pada Kelompok UMKM*, dapat disimpulkan bahwa penerapan *data mining* dengan pendekatan RFM (*Recency*, *Frequency*, *Monetary*) dan algoritma *K-Means Clustering* mampu

menghasilkan segmentasi pelanggan berdasarkan karakteristik transaksi. Dari 200 data transaksi mentah yang diperoleh, setelah melalui proses pengolahan dan normalisasi dihasilkan 162 data pelanggan unik berbasis atribut RFM yang siap digunakan dalam proses klasterisasi. Hasil penerapan algoritma *K-Means* menunjukkan terbentuknya tiga klaster utama pelanggan. Cluster 0 merupakan kelompok pelanggan transaksional atau musiman yang terdiri atas 119 pelanggan atau sebesar 73,5%, dengan karakteristik frekuensi dan nilai transaksi yang relatif rendah. Cluster 1 merupakan kelompok pelanggan potensial yang terdiri atas 15 pelanggan atau sebesar 9,3%, dengan tingkat frekuensi dan nilai transaksi berada pada kategori menengah. Sementara itu, Cluster 2 merupakan kelompok pelanggan inti atau *Champions* yang terdiri atas 28 pelanggan atau sebesar 17,3%, yang memiliki nilai transaksi tinggi serta tingkat kedatangan berulang yang kuat.

Penerapan hasil segmentasi tersebut selanjutnya diimplementasikan ke dalam sistem CRM berbasis web menggunakan Python dengan *framework* Flask. Sistem yang dikembangkan mampu mengotomatisasi proses klasterisasi berbasis RFM secara *real-time* serta menyediakan informasi yang dapat digunakan sebagai dasar dalam menentukan strategi pemasaran sesuai karakteristik masing-masing kelompok pelanggan. Selain itu, hasil evaluasi menggunakan pendekatan *Two-Stage Data Mining* menghasilkan nilai AUC sebesar 1,000 dan *F-Score* sebesar 100%. Hasil tersebut menunjukkan tingkat akurasi yang sangat tinggi pada pengelompokan segmen pelanggan dalam pengujian yang dilakukan. Dengan demikian, integrasi RFM, algoritma *K-Means Clustering*, dan sistem CRM berbasis web dapat digunakan sebagai pendekatan untuk membantu Kelompok UMKM dalam memahami karakteristik pelanggan dan mendukung penyusunan strategi *Customer Relationship Management* berdasarkan data transaksi.

DAFTAR PUSTAKA

- Al-Homery, H. A., Asharai, H., & Ahmad, A. (2019). *The Core Components and Types of CRM*. *Pakistan Journal of Humanities and Social Sciences*, Vol. 7 No. 1.
- Budianto, A., & Hermawan, A. (2023). *Penerapan Data Mining untuk Segmentasi Pelanggan menggunakan Algoritma K-Means Clustering*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 145-154.
- Chaidir, A., & Riadi, I. (2022). *Analisis Customer Relationship Management (CRM) dengan Pendekatan Model RFM*. Jakarta: Penerbit Informatika.
- Indini, D. P., Mesran, & Dito Putro Utomo. (2023). *Penerapan Data Mining Dalam Pengelompokan Data Reseller di Telkomsel Authorized Partner (TAP) Deli Tua Dengan Algoritma K-Means*. *Jurnal Ilmiah Media Sisfo*, 17(2), 189–202. <https://doi.org/10.33998/mediasisfo.2023.17.2.1391>
- Kharisma, G., Ibrahim, R., & Heikal, J. (2025). *An Analyzing RFM (Recency, Frequency, Monetary) Segmentation to Formulate Online Marketing Strategies : a Case Study of PT MMT*. *Jurnal Multidisiplin Sahombu*, 5. <https://doi.org/10.58471/jms.v5i06>

- Kurniawan, A. R., Mayasari, R., Febriantoko, J., Taman, P. T., & Jogja, W. (2021). CUSTOMER LIFETIME VALUE METRICS AS CUSTOMER RELATIONSHIP MANAGEMENT INDICATORS. *Business and Accounting Research (IJEBAR) Peer Reviewed-International Journal*, 5(3). <https://jurnal.stie-aas.ac.id/index.php/IJEBAR>
- Levin, James & Nolan, J. (2013). *P rinciples of C lassroom M anagement*. 1(1), 354. <https://doi.org/10.5281/zenodo.7618566>
- Martiansah, R., Monalisa, S., Muttakin, F., & Fronita, M. (2025). Customer Segmentation Analysis Through RFM-D Model and K-Means Algorithm. Dalam *Jurnal Sistem Cerdas*.
- Mikael Kia Mado, P. (2025). Implementasi Algoritma Clustering K-Means untuk Segmentasi Pelanggan di E-Commerce. Dalam *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)* (Vol. 6, Nomor 3). <https://journal.stmiki.ac.id>
- Nugraha, D. S., Thoib, I., Sururi, N., Bayu F, F., & Candra, B. P. (2025). Segmentasi Pelanggan Berbasis Analisis RFM Menggunakan Algoritma K-Means Clustering. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, 4(2), 1361–1369. <https://doi.org/10.62712/juktisi.v4i2.641>
- Pramudiansyah, A., & Munte, H. (2021). SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN MODEL RECENCY FREQUENCY MONETARY. 7(2). <http://ejournal.fikom-unasman.ac.id>
- Punithavathi, G. (2024). A Survey on Machine Learning Based Classification Algorithms for Web Usage Mining. *International Journal for Research in Applied Science and Engineering Technology*, 12(1), 442–449. <https://doi.org/10.22214/ijraset.2024.57962>
- Rahmadeni, R., & Heikal, J. (2024). Customer Loyalty Segmentation with RFM Model at PCP Cafe. *Jurnal Ekonomika Dan Bisnis (JEBS)*, 4(6), 1966–1972. <https://doi.org/10.47233/jebs.v4i6.2450>
- Sugianto, A., Arisawati, E., Studi Sistem Informasi, P., & Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, S. (t.t.). *INTI NUSA MANDIRI PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK SEGMENTASI PELANGGAN PADA CV. SINAR SURYA*. Diambil <http://nusamandiri.ac.id>
- Susilo, A. (2024). EVALUASI EFEKTIFITAS ALGORITMA K-MEANS DAN DBSCAN DALAM CLUSTERING DATA RUMAH SAKIT UNTUK OPTIMALISASI LAYANAN KESEHATAN. *Jurnal Komputer dan Informatika*, 19, 43–51.