



Analisis Segmentasi Pelanggan E-Commerce menggunakan Metode RFM dan K-MEANS Clustering untuk Strategi Bisnis

Efrain Yaphet Obert*, Mulia Rahmayu

Fakultas Teknik Informatika, Teknologi Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1,*} 17220543@bsi.ac.id, ² mulia.mlh@bsi.ac.id

(* : coresponding author)

Abstrak-Pesatnya perkembangan industri *e-commerce* menghasilkan data transaksi pelanggan dalam jumlah besar. Namun, sebagian besar perusahaan menghadapi permasalahan berupa keterbatasan dalam mengolah data mentah tersebut menjadi informasi bisnis yang strategis. Akibatnya, pihak manajemen kesulitan memahami variasi perilaku transaksi, memprediksi preferensi pembeli, dan mempertahankan loyalitas pelanggan secara personal. Kondisi ini berpotensi meningkatkan risiko hilangnya pelanggan (*customer churn*) serta memicu inefisiensi biaya operasional pemasaran akibat promosi yang tidak tepat sasaran. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berbasis data transaksi guna memberikan rekomendasi strategi pemasaran yang efektif dan terukur. Metode yang diterapkan adalah integrasi model RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means Clustering*. Dataset sekunder yang bersumber dari Kaggle diproses melalui tahap *Exploratory Data Analysis* (EDA), pembersihan data, dan transformasi tipe data. Penentuan jumlah kluster optimal dilakukan menggunakan *Elbow Method* dan divalidasi dengan metrik evaluasi kuantitatif *Silhouette Score*. Hasil pengujian menunjukkan bahwa pembagian empat kluster merupakan struktur terbaik dengan nilai *Silhouette Score* sebesar 0,43. Empat kluster utama yang terbentuk meliputi *Lost Customer* sebesar 39,2%, *Low-Value Customer* sebesar 37,2%, *Big Spender* sebesar 20,4%, dan *Loyal Customer* sebesar 3,2%. Hasil segmentasi ini kemudian divisualisasikan secara interaktif menggunakan Looker Studio. Implementasi *dashboard* ini memberikan kontribusi praktis berupa *insight* operasional bagi manajemen dalam mengeksekusi strategi promosi tersegmentasi, mengoptimalkan alokasi anggaran pemasaran, serta menekan risiko hilangnya pelanggan potensial secara sistematis.

Kata Kunci: Segmentasi Pelanggan; RFM; K-Means Clustering; E-Commerce; Data Mining.

Abstract-The rapid development of the *e-commerce* industry has generated vast amounts of customer transaction data. However, most companies face challenges in processing this raw data into strategic business insights. Consequently, management struggles to understand variations in transactional behavior, predict buyer preferences, and maintain personalized customer loyalty. This condition risks increasing customer churn and driving promotional cost inefficiencies due to misdirected marketing efforts. This study aims to perform customer segmentation based on transaction data to provide targeted and measurable marketing strategy recommendations. The methodology integrates the RFM (*Recency, Frequency, Monetary*) model with the *K-Means Clustering* algorithm. A secondary dataset sourced from Kaggle was processed through *Exploratory Data Analysis* (EDA), data cleaning, and data type transformation. The optimal number of clusters was determined using the *Elbow Method* and validated with a quantitative evaluation metric, the *Silhouette Score*. The test results indicated that a four-cluster solution provided the best structure with a *Silhouette Score* of 0.43. The four main identified clusters include *Lost Customers* (39.2%), *Low-Value Customers* (37.2%), *Big Spender Customers* (20.4%), and *Loyal Customers* (3.2%). These segmentation results were subsequently visualized through an interactive Looker Studio dashboard. The implementation of this dashboard provides practical contributions by delivering operational insights for executing segmented promotional strategies, optimizing marketing budget allocation, and systematically mitigating customer churn risks.

Keywords: Customer Segmentation; RFM; K-Means Clustering; E-Commerce; Data Mining.

1. PENDAHULUAN

Perkembangan teknologi informasi, khususnya dalam industri *e-commerce*, telah menghasilkan pertumbuhan data transaksi yang sangat pesat. Setiap transaksi pelanggan menghasilkan rekaman data historis yang berpotensi digunakan untuk memahami perilaku konsumen guna mendukung strategi bisnis. Namun, sebagian besar perusahaan *e-commerce* ritel saat ini masih menghadapi permasalahan berupa keterbatasan dalam mengolah data mentah tersebut menjadi sebuah informasi strategis. Indikasi masalah yang timbul di lapangan adalah perlakuan strategi pemasaran yang masih bersifat massal (*generic marketing*), di mana seluruh pelanggan mendapatkan penawaran promosi yang sama tanpa adanya pertimbangan nilai atau loyalitas masing-masing. Hal ini berdampak pada tingginya biaya operasional promosi yang tidak tepat sasaran serta meningkatnya risiko kehilangan pelanggan potensial (*customer churn*).

Dalam lanskap bisnis digital saat ini, implementasi *Customer Relationship Management* (CRM) modern menjadi pilar utama untuk mempertahankan daya saing perusahaan. Pengolahan *big data* transaksi *e-commerce* yang sangat dinamis menyajikan tantangan tersendiri, di mana ekstraksi pola belanja bernilai tinggi memerlukan pendekatan analitis yang presisi agar dapat diubah menjadi strategi keputusan bisnis yang responsif. Salah satu pendekatan yang dapat digunakan untuk memahami karakteristik pelanggan adalah segmentasi pelanggan. Dengan melakukan segmentasi, perusahaan dapat mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan kesamaan karakteristik sehingga strategi pemasaran dapat disesuaikan dengan kebutuhan masing-masing segmen. Pada penelitian ini digunakan metode RFM (*Recency, Frequency, Monetary*) untuk mengukur nilai pelanggan, kemudian dilakukan proses pengelompokan menggunakan algoritma *K-Means Clustering*.

Pada penelitian sebelumnya, Fadhilah et al. menggunakan metode RFM untuk mengevaluasi perilaku pelanggan berdasarkan nilai *Recency, Frequency*, dan *Monetary* sehingga mampu menghasilkan segmentasi



pelanggan yang lebih efektif [1]. Selanjutnya, Syahra et al. mengombinasikan metode RFM dan algoritma *K-Means Clustering* untuk mengidentifikasi pola perilaku pelanggan dan menghasilkan segmentasi yang lebih terstruktur [2]. Selain itu, Saumendra et al. menegaskan bahwa teknik *data mining* berbasis *clustering* terbukti efektif dalam menghasilkan segmentasi pelanggan yang presisi guna mendukung pengambilan keputusan bisnis berbasis data [3]. Algoritma ini bekerja secara iteratif dengan meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS) melalui perhitungan jarak Euclidean.

Penggunaan metode *clustering* dalam segmentasi terbukti mampu mengidentifikasi pola perilaku pelanggan secara lebih akurat dibandingkan metode konvensional [4]. Dalam lingkup *data mining*, metode RFM (*Recency, Frequency, Monetary*) menjadi pendekatan yang sangat efektif untuk mengekstraksi nilai dari perilaku transaksi pelanggan sebelum dilakukan proses pengelompokan [5]. Algoritma *K-Means Clustering* diimplementasikan untuk membagi pelanggan ke dalam segmen-segmen yang homogen berdasarkan skor RFM tersebut guna menghasilkan *insight* bisnis yang aplikatif [6]. Tahapan ini bertujuan untuk mengurangi *noise*, menangani data yang tidak konsisten, serta memastikan data siap digunakan dalam proses *clustering* [7].

Selain itu, *preprocessing* seperti normalisasi dan standarisasi terbukti dapat meningkatkan performa algoritma *K-Means* dalam menghasilkan *cluster* yang lebih optimal [8]. Efektivitas sebuah *platform e-commerce* sangat bergantung pada integrasi sistem pembayaran dan pengiriman yang otomatis. Hal ini berimplikasi pada ketersediaan data transaksi yang terstruktur, yang kemudian dapat dimanfaatkan lebih lanjut untuk analisis perilaku pelanggan melalui teknik *data mining* [9]. Teknik pengumpulan data dalam penelitian ini menggunakan metode studi dokumentasi melalui ekstraksi dataset sekunder. Penggunaan data sekunder dari platform publik yang valid merupakan metode pengumpulan data yang sah dalam penelitian kuantitatif berbasis komputasi tanpa memerlukan wawancara lapangan [10]. Kluster dengan rata-rata nilai *Frequency* dan *Monetary* tertinggi dipetakan sebagai kelompok pelanggan utama (*Loyal Customer*) [11]. Nilai *Frequency* menunjukkan tingkat loyalitas atau kebiasaan belanja, sedangkan *Monetary* merepresentasikan besarnya nilai finansial yang diberikan pelanggan. Pelanggan yang memiliki skor *Frequency* dan *Monetary* di atas rerata secara akademis dikategorikan ke dalam kelompok *Loyal Customer* (Pelanggan Loyal) karena secara konsisten memberikan kontribusi retensi dan pendapatan terbesar bagi bisnis [12].

Pemilihan algoritma *K-Means* didasarkan pada keunggulannya dalam mengelompokkan data numerik berukuran besar secara efisien berdasarkan perhitungan jarak Euclidean terdekat menuju titik centroid kluster [13]. Tahap interpretasi hasil kluster pada algoritma partisi dilakukan dengan menganalisis nilai rata-rata (centroid) pada setiap atribut untuk menentukan profil atau label kelompok yang terbentuk [14]. Algoritma *K-Means Clustering* dapat diterapkan untuk melakukan segmentasi pelanggan pada data *e-commerce* guna menghasilkan pengelompokan berdasarkan karakteristik pelanggan yang berbeda [15]. Penamaan kategori segmen ini mengacu pada kerangka segmentasi perilaku pelanggan (*Customer Relationship Management / CRM*) menurut [16]. Penelitian yang dilakukan mengkaji penerapan model *Recency, Frequency, Monetary* (RFM) dan algoritma *K-Means Clustering* dalam melakukan segmentasi pelanggan berdasarkan data transaksi *online retail* periode 2010–2011.

Tahapan analisis diawali dengan menghitung matriks nilai RFM dari tiap pelanggan, dilanjutkan dengan normalisasi data dan proses klusterisasi. Hasil penelitian menunjukkan terbentuknya tiga segmen pelanggan utama dengan karakteristik perilaku transaksi yang berbeda. Temuan ini membuktikan bahwa kombinasi analisis RFM dan *K-Means Clustering* efektif membantu perusahaan dalam memetakan profil konsumen serta menyusun strategi pemasaran terpersonalisasi untuk meningkatkan loyalitas pelanggan [17]. Penerapan teknik *data mining* menggunakan algoritma *K-Means Clustering* dengan normalisasi *Min-Max Normalization* terbukti efektif dalam membagi pelanggan ke dalam kelompok-kelompok homogen berdasarkan kesamaan pola perilaku transaksi, di mana pembentukan 4 *cluster* optimal mampu memandu perusahaan dalam pengambilan keputusan strategi pemasaran secara presisi [18].

Integrasi model RFM dan algoritma *K-Means* yang dievaluasi menggunakan metrik *Within-Cluster Sum of Squares* (WSS) serta *Silhouette Score* terbukti mampu mengoptimalkan strategi pemasaran ritel melalui pembentukan segmen pelanggan bernilai tinggi (*high-value*) secara presisi untuk perancangan promosi terarah dan program retensi [19]. Penerapan transformasi logaritma pada model RFM sebelum klusterisasi *K-Means* yang dievaluasi dengan *Silhouette Score* (0,503) terbukti mampu menghasilkan tiga segmen pelanggan yang terstruktur (*Loyal Customers*, *Need Attention*, dan *At Risk*) guna menekan risiko *customer churn* melalui strategi retensi yang lebih tepat sasaran [20]. Berdasarkan penelitian-penelitian tersebut, penelitian ini menerapkan metode RFM dan *K-Means Clustering* pada data transaksi *e-commerce* untuk menghasilkan segmentasi pelanggan sebagai dasar penyusunan strategi bisnis.

Meskipun berbagai penelitian terdahulu telah mengaplikasikan algoritma pemodelan RFM dan *K-Means* untuk segmentasi pelanggan, keterbatasan utama dari studi-studi sebelumnya terletak pada penyampaian hasil analisis yang masih bersifat statis dan terisolasi dari operasional bisnis harian. Sebagian besar penelitian berhenti pada tahap pembentukan kluster tanpa menyajikan visualisasi data yang mudah diakses oleh pihak manajemen. Penelitian ini mengisi celah tersebut dengan mengintegrasikan hasil segmentasi algoritma *K-Means* berbasis RFM ke dalam *dashboard* interaktif Looker Studio. Melalui visualisasi data yang dapat dipantau secara langsung,



pemangku kepentingan dapat mengeksekusi strategi promosi yang lebih tepat sasaran, mengoptimalkan alokasi anggaran pemasaran, serta menekan risiko *customer churn* secara sistematis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Secara keseluruhan, alur dan tahapan pelaksanaan penelitian ini disajikan secara sistematis pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian diawali dengan pengumpulan data sekunder berupa transaksi penjualan e-commerce yang diperoleh melalui platform Kaggle. Dataset tersebut terdiri dari 5.000 data transaksi dengan 14 atribut awal. Sebelum memasuki tahap segmentasi, data melalui proses preprocessing yang meliputi verifikasi missing value, pemeriksaan duplikasi data, serta konversi tipe data atribut Order Date ke dalam format *datetime64*. Selanjutnya, dilakukan ekstraksi fitur berbasis metode RFM (Recency, Frequency, Monetary) dengan memfokuskan analisis pada 4 atribut utama (Customer Name, Order ID, Order Date, dan Sales). Fitur RFM yang telah distandardisasi kemudian diolah menggunakan algoritma K-Means Clustering untuk mengelompokkan pelanggan ke dalam beberapa segmen. Tahap akhir penelitian diakhiri dengan visualisasi serta interpretasi karakteristik kluster untuk merumuskan rekomendasi strategi bisnis yang relevan.

2.2 Kajian Algoritma dan Metode

2.2.1 Model *Recency, Frequency, dan Monetary* (RFM)

Model *Recency, Frequency, dan Monetary* (RFM) merupakan metode analisis perilaku pelanggan kuantitatif yang digunakan untuk melakukan segmentasi berbasis riwayat transaksi. Analisis RFM mengelompokkan pelanggan berdasarkan tiga indikator utama:

a. *Recency*

Merupakan jangka waktu (dalam satuan hari) sejak transaksi terakhir yang dilakukan oleh pelanggan hingga tanggal referensi penutupan data (*anchor date*). Nilai *Recency* yang semakin kecil mengindikasikan bahwa pelanggan baru saja melakukan transaksi.

b. *Frequency*

Merupakan total akumulasi frekuensi atau jumlah transaksi yang dilakukan oleh seorang pelanggan dalam periode waktu yang diamati.

c. *Monetary*

Merupakan akumulasi total nominal nilai uang yang dikeluarkan oleh pelanggan untuk seluruh transaksi yang dilakukan.

Pengelompokan data transaksi menggunakan variabel RFM bertujuan untuk memisahkan objek berdasarkan derajat kemiripan karakteristiknya. Kluster dengan rata-rata nilai *Frequency* dan *Monetary* tertinggi dipetakan sebagai kelompok pelanggan utama (*Loyal Customer*) [11].

2.2.2 Algoritma *K-Means Clustering*

K-Means Clustering merupakan algoritma analisis kluster berbasis *unsupervised learning* yang mengelompokkan data ke dalam K kelompok berdasarkan tingkat kemiripan fitur numerik. Algoritma ini bekerja secara iteratif dengan meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS) melalui perhitungan jarak Euclidean. Penggunaan metode *clustering* dalam segmentasi terbukti mampu mengidentifikasi pola perilaku pelanggan secara lebih akurat dibandingkan metode konvensional [4]. Langkah-langkah utama dalam algoritma K-Means meliputi:

a. Inisialisasi pusat kluster (*centroid*) awal secara acak sebanyak K titik.

b. Perhitungan jarak setiap titik data (x) ke setiap *centroid* (c_i) menggunakan fungsi jarak Euclidean yang dirumuskan sebagai berikut:

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

c. Pengelompokkan setiap data ke centroid terdekat.

d. Pembaruan nilai centroid baru berdasarkan nilai rata-rata dari seluruh anggota data yang masuk ke dalam kluster tersebut.



e. Pengulangan proses iterasi hingga posisi centroid tidak lagi mengalami perubahan konvergen.

2.2.3 Standardisasi Data (Z-Score)

Sebelum melakukan pengelompokan dengan algoritma *K-Means*, variabel RFM perlu disetarakan skalanya untuk mencegah bias kalkulasi akibat perbedaan satuan dan domain nilai. Sebelum tahap scaling/standardisasi diterapkan, pemeriksaan sebaran nilai variabel RFM dilakukan terlebih dahulu untuk mengidentifikasi keberadaan pencilan (*outlier*). Dalam penelitian ini, nilai-nilai ekstrem tetap dipertahankan (*retained*) tanpa pemotongan maupun penghapusan, karena data tersebut merepresentasikan perilaku transaksi alami pelanggan (seperti kelompok *Big Spender* dengan nilai *Monetary* tinggi). Metode standardisasi *Z-Score* kemudian digunakan untuk menyetarakan skala variabel tanpa menghilangkan karakteristik asli sebaran data. Transformasi ini mengubah sebaran data mentah sehingga memiliki nilai rata-rata (μ) bernilai 0 dan standar deviasi (σ) bernilai 1, Penerapan *preprocessing* berupa standardisasi terbukti dapat meningkatkan performa algoritma *K-Means* dalam menghasilkan kluster yang lebih optimal. Persamaan transformasi *Z-Score* dirumuskan sebagai berikut:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

Keterangan penjelas variabel rumus, z : Nilai fitur setelah distandardisasi, x : Nilai mentah data awal, μ : Nilai rata-rata dari keseluruhan sampel data pada variabel tersebut, σ : Nilai standar deviasi (*standard deviation*) dari keseluruhan sampel data pada variabel tersebut

2.2.4 Penentuan Jumlah Kluster Optimal (Elbow Method)

Elbow Method merupakan salah satu teknik heuristik yang digunakan untuk menentukan jumlah kluster K paling optimal dalam algoritma *K-Means*. Metode ini mengukur total variansi dalam kluster menggunakan nilai *inertia* atau *Within-Cluster Sum of Squares* (WCSS). Persamaan perhitungan WCSS dirumuskan sebagai berikut:

$$WCSS = \sum (x - \mu)^2 \quad (3)$$

Keterangan penjelas variabel rumus: WCSS: Total jarak kuadrat data terhadap pusat kluster (*inertia*), x : Titik data mentah sampel dalam kluster, μ : Nilai rata-rata pusat kluster (*centroid*). Jumlah kluster optimal ditentukan oleh titik sudut "siku" (*elbow*) pada grafik di mana penurunan nilai WCSS mulai melandai secara signifikan. Algoritma *K-Means Clustering* diimplementasikan untuk membagi pelanggan ke dalam segmen-segmen yang homogen berdasarkan skor RFM tersebut guna menghasilkan *insight* bisnis yang aplikatif [6].

2.2.5 Validasi Kluster (Silhouette Score)

Silhouette Score digunakan untuk mengukur kualitas dan tingkat pemisahan antar-kluster yang dihasilkan oleh algoritma *K-Means*. Nilai metrik ini mengevaluasi seberapa dekat suatu objek dengan klusternya sendiri dibandingkan dengan kluster lainnya. Nilai Silhouette Score berkisar antara -1 hingga +1, di mana nilai yang mendekati +1 mengindikasikan bahwa data telah terkelompokkan secara optimal pada kluster yang tepat. Persamaan Silhouette Coefficient s_i untuk sebuah sampel data i dirumuskan sebagai berikut:

$$s_i = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

$a(i)$: Rata-rata jarak Euclidean antara sampel i dengan seluruh titik data lain dalam kluster yang sama (*intra-cluster distance*). $b(i)$: Rata-rata jarak Euclidean terkecil antara sampel i dengan titik data pada kluster terdekat lainnya (*nearest-cluster distance*).

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Pada bagian ini disajikan hasil penelitian dan pemrosesan data transaksi *e-commerce* periode 2024–2025 yang terdiri dari 5.000 baris data dan 14 kolom variabel yang diperoleh melalui platform Kaggle. Pada penelitian ini, dataset tersebut dikelola menggunakan lingkungan komputasi Google Colab dengan memfokuskan analisis pada 4 kolom utama (*Customer Name*, *Order ID*, *Order Date*, dan *Sales*) dari 5.000 baris data mentah berbentuk csv.

3.1.1 Struktur Data

Tahapan ini bertujuan untuk mengevaluasi tipe data, memastikan tidak ada nilai kosong (*missing value*), serta memahami sebaran dan karakteristik data transaksi sebelum data ditransformasikan ke dalam variabel RFM dan diolah dengan algoritma *K-Means*. Evaluasi variabel utama menunjukkan bahwa atribut *Customer Name*, *Order Date*, dan *Sales* merupakan variabel krusial untuk perhitungan model RFM. Namun, pada tipe variabel *Order Date*, dilakukan perubahan tipe data dari *object* menjadi *datetime64* sesuai kebutuhan kalkulasi selisih hari pada model



RFM. Penampilan struktur data dan statistik deskriptif pada bagian awal Bab III ini merupakan penerapan tahap *Data Understanding* sesuai kerangka kerja CRISP-DM. Tahapan ini bertujuan untuk mengevaluasi tipe data, memastikan tidak ada nilai kosong (*missing value*), serta memahami sebaran dan karakteristik data transaksi (seperti nilai rata-rata *Sales* sebesar 106.733,20) sebelum data ditransformasikan ke dalam variabel RFM dan diolah dengan algoritma K-Means.

3.1.2 Statistik Deskriptif

Untuk memahami persebaran nilai pada dataset, dilakukan perhitungan statistik deskriptif pada variabel numerik. Hasil kalkulasi statistik deskriptif disajikan pada Gambar 2:

	Order ID	Quantity	Unit Price	Discount	Sales	Profit
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	12500.500000	2.992600	39760.904600	10.051000	106733.204870	15941.746962
std	1443.520003	1.413133	22831.783946	7.084662	85108.208202	14897.684916
min	10001.000000	1.000000	222.000000	0.000000	264.100000	19.120000
25%	11250.750000	2.000000	20312.250000	5.000000	39766.537500	4892.295000
50%	12500.500000	3.000000	39459.500000	10.000000	83080.325000	11108.525000
75%	13750.250000	4.000000	59721.750000	15.000000	156968.587500	22467.967500
max	15000.000000	5.000000	79998.000000	20.000000	398485.000000	89688.440000

Gambar 2. Statistik Deskriptif Dataset

Berdasarkan Gambar 2, dapat diperoleh beberapa informasi penting dari statistik deskriptif *dataset* sebagai berikut:

- Variabel *Sales* memiliki rata-rata nilai transaksi sebesar 106.733,20 dengan nilai maksimum mencapai 398.485. Angka 106.733,20 pada variabel *Sales* merupakan nilai rata-rata transaksi (*mean*) dari 5.000 data transaksi yang dianalisis. Artinya, secara rata-rata, setiap transaksi pembelian pelanggan pada *e-commerce* ini menghasilkan nilai penjualan sebesar 106.733,20.
- Perhitungan nilai rata-rata *Sales*. $\text{Mean Sales} = \text{Total penjualan seluruh transaksi} / \text{Total jumlah transaksi (count)} = \text{sales} / 5000 = 106,733.20$.
- Jumlah data yang konsisten sebanyak 5.000 baris menunjukkan bahwa dataset tidak memiliki *missing value* pada variabel utama yang digunakan dalam penelitian.

3.1.3 Data Cleaning dan Preprocessing

Data Cleaning dan *Preprocessing* adalah tahap awal dimana data yang akan diteliti harus dibersihkan dari *duplicate*, *missing value*, dan *nan* untuk memberikan hasil akhir yang relevan.

a. Missing Value

Pengecekan *missing value* dilakukan menggunakan fungsi `isnull().sum()` pada bahasa pemrograman Python. Hasil pengecekan menunjukkan nilai 0 pada seluruh variabel utama, sehingga dapat disimpulkan bahwa dataset tidak memiliki data yang hilang dan layak untuk diproses lebih lanjut.

b. Konversi Tipe Data

Selanjutnya dilakukan konversi tipe data pada variabel *Order Date* dari tipe *object* menjadi *datetime64* menggunakan fungsi `pd.to_datetime()`. Proses ini dilakukan agar data tanggal dapat diproses dalam perhitungan *Recency* pada model RFM. Sebelumnya pada Gambar 1. *Dtype* kolom *Order Date* adalah *object*, untuk membuat metode (RFM) *Recency* dapat menghitung kapan transaksi awal dan kapan transaksi akhir dilakukan, *Dtype* perlu diubah menjadi *datetime64[ns]*, dengan begitu *recency* dapat menghitung jumlah hari kapan pelanggan tersebut bertransaksi.

c. Data Duplikat

Pengecekan data duplikat dilakukan untuk memastikan tidak terdapat data transaksi yang tercatat lebih dari satu kali. Proses ini dilakukan menggunakan fungsi `duplicated().sum()`.

Jumlah duplikat seluruh kolom: 0

Order ID	Order Date	Customer Name	Region	City	Category	Sub-Category	Product Name	Quantity	Unit Price	Discount	Sales	Profit	Payment Mode
----------	------------	---------------	--------	------	----------	--------------	--------------	----------	------------	----------	-------	--------	--------------

Gambar 3. Duplikat

Berdasarkan hasil pengecekan pada Gambar 3, tidak ditemukan data duplikat (jumlah duplikat seluruh kolom 0) pada *dataset* sehingga data dinyatakan layak untuk digunakan pada proses analisis selanjutnya.

3.1.4 Pembentukan Model RFM

Setelah proses *preprocessing* selesai dilakukan, tahap berikutnya adalah pembentukan model RFM (*Recency*, *Frequency*, *Monetary*). Model ini digunakan untuk mengukur perilaku pelanggan berdasarkan waktu transaksi



terakhir, frekuensi transaksi, dan total nilai transaksi pelanggan. Proses pembentukan model RFM dilakukan dengan melakukan agregasi data berdasarkan *Customer Name* sehingga diperoleh nilai *Recency*, *Frequency*, dan *Monetary* untuk setiap pelanggan.

a. *Recency*

Recency digunakan untuk mengukur jarak waktu (dalam satuan hari) sejak transaksi terakhir yang dilakukan oleh seorang pelanggan hingga tanggal referensi penutupan data (*anchor date*). Tanggal referensi yang digunakan dalam penelitian ini ditetapkan berdasarkan tanggal transaksi paling akhir yang terekam dalam dataset, yaitu 3 Oktober 2025. Rumus perhitungan *Recency* untuk setiap pelanggan (*Customer Name*) adalah sebagai berikut:

$\text{Recency (Hari)} = \text{Tanggal Referensi (3 Oktober 2025)} - \text{Tanggal Transaksi Terakhir Pelanggan}$, Nilai *Recency* Kecil (Mendekati 0 Hari): Menandakan bahwa selisih waktu transaksi terakhir pelanggan dengan tanggal referensi sangat dekat.

b. *Frequency*

Frequency digunakan untuk mengukur seberapa sering pelanggan melakukan transaksi berdasarkan jumlah transaksi yang dilakukan.

c. *Monetary*

Monetary digunakan untuk mengukur total akumulasi nilai pengeluaran/pembelian dari seorang pelanggan selama periode penelitian. Angka *Monetary* diperoleh dari penjumlahan seluruh nilai pada kolom Sales untuk setiap entitas pelanggan (*Customer Name*).

Proses pembentukan model RFM dilakukan dengan melakukan agregasi data berdasarkan *Customer Name* sehingga diperoleh nilai *Recency*, *Frequency*, dan *Monetary* untuk setiap pelanggan. Hasil perhitungan dan ringkasan statistik dari ketiga variabel RFM tersebut disajikan pada Gambar 4.

	count	mean	std	min	25%	50%	75%	max
Recency	4844.0	363.326177	209.182899	0.0	184.0000	361.5	541.0000	730.0
Frequency	4844.0	1.032205	0.178885	1.0	1.0000	1.0	1.0000	3.0
Monetary	4844.0	110170.525258	88632.704703	264.1	40757.2375	85848.5	160902.3375	650151.9

Gambar 4. RFM describe

Berdasarkan Gambar 4, model RFM yang terbentuk mampu merepresentasikan distribusi perilaku pelanggan secara kuantitatif. Evaluasi terhadap variabel *Recency*, *Frequency*, dan *Monetary* ini selanjutnya dijadikan sebagai variabel masukan (*input features*) dalam proses segmentasi pelanggan menggunakan algoritma K-Means Clustering.

3.1.5 Standardisasi Data

Setelah pembentukan model RFM selesai dilakukan, tahap berikutnya adalah standardisasi data menggunakan metode *StandardScaler* dari *library Scikit-Learn*. Proses standardisasi dilakukan karena variabel *Recency*, *Frequency*, dan *Monetary* memiliki rentang nilai yang berbeda secara signifikan. Perbedaan skala tersebut dapat memengaruhi proses perhitungan jarak pada algoritma K-Means Clustering. tahapan *feature scaling* berbasis standardisasi merupakan prasyarat mutlak dalam implementasi *machine learning* modern untuk menyamakan kontribusi setiap variabel numerik dan menghindari bias kalkulasi akibat perbedaan dimensi data mentah.

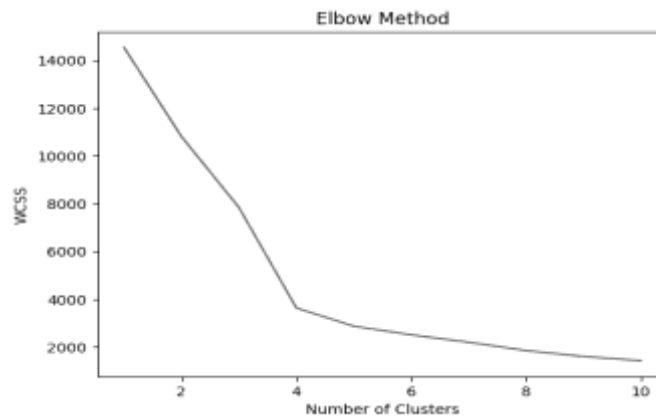
Standardisasi dilakukan menggunakan metode Z-Score sehingga data memiliki rata-rata (mean) mendekati 0 dan standar deviasi (*standard deviation*) sebesar 1. Secara matematis transformasi setiap titik data mentah (x) menjadi nilai terskala (Z) dihitung menggunakan persamaan sebagai berikut:

$$Z = \frac{x - \mu}{\sigma}$$

z : Nilai fitur setelah distandardisasi, x : Nilai mentah data awal, μ : Nilai rata-rata dari keseluruhan sampel data pada variabel tersebut, σ : Nilai standar deviasi (*standard deviation*) dari keseluruhan sampel data pada variabel tersebut. Dengan dilakukan standardisasi, setiap variabel RFM memiliki kontribusi yang seimbang dalam proses pembentukan *cluster*.

3.1.6 Penentuan Jumlah Cluster Menggunakan Elbow Method

Setelah proses standardisasi selesai dilakukan, tahap berikutnya adalah menentukan jumlah *cluster* optimal sebelum melakukan proses *clustering* menggunakan algoritma K-Means. Penentuan jumlah *cluster* dilakukan menggunakan metode *Elbow Method* dengan melihat nilai *Within Cluster Sum of Squares* (WCSS) atau *inertia* pada setiap jumlah *cluster*. Nilai *inertia* menunjukkan total jarak kuadrat antara data dengan *centroid cluster*. Semakin kecil nilai *inertia* menunjukkan bahwa *cluster* yang terbentuk semakin kompak. Grafik pengujian *Elbow Method* ditunjukkan pada Gambar 5.



Gambar 5. Elbow Method

Berdasarkan Gambar 5, terlihat bahwa penurunan nilai *inertia* terjadi cukup signifikan hingga nilai $K = 4$. Setelah titik tersebut, penurunan grafik mulai melandai sehingga penambahan jumlah *cluster* tidak memberikan peningkatan yang signifikan. Oleh karena itu, jumlah *cluster* optimal yang digunakan dalam penelitian ini adalah sebanyak 4 kluster. Untuk memperkuat pembuktian secara ilmiah dan objektif bahwa $K = 4$ merupakan skema terbaik, dilakukan pengujian validasi kluster menggunakan metrik *Silhouette Score* untuk rentang $K = 2$ hingga $K = 6$. Nilai *Silhouette Score* mengukur tingkat kemiripan suatu objek dengan klasternya sendiri dibandingkan dengan kluster lain (rentang nilai -1 hingga 1). Hasil evaluasi disajikan pada Tabel 1.

Tabel 1. Evaluasi Nilai *Silhouette Score* untuk menentukan Jumlah Kluster

Jumlah Kluster (K)	<i>Silhouette Score</i>	Evaluasi Kualitas Pemisahan
K = 2	0,33	Pemisahan kluster kurang spesifik
K = 3	0,38	Struktur kelompok mulai terbentuk
K = 4	0,43	Nilai tertinggi (Pemisahan Kluster Optimal)
K = 5	0,42	Terjadi <i>overlapping</i> antar sub-kluster
K = 6	0,38	Terdapat kluster yang terlalu kecil / <i>redundant</i>

Berdasarkan Tabel 1, hasil pengujian mengonfirmasi bahwa pengelompokan dengan $K = 4$ menghasilkan nilai *Silhouette Score* tertinggi yaitu sebesar 0,43. Hal ini membuktikan secara empiris bahwa pembagian menjadi 4 segmen pelanggan memberikan derajat kerapatan internal (*cohesion*) dan pemisahan antar-kluster (*separation*) yang paling optimal dibanding pilihan jumlah kluster lainnya. Oleh karena itu, jumlah cluster optimal yang digunakan dalam penelitian ini ditetapkan sebanyak 4 cluster

3.1.7 Hasil Clustering K-Means

Tahapan *clustering* menggunakan algoritma K-Means merupakan proses inti dalam penelitian ini untuk mengelompokkan data pelanggan berdasarkan kesamaan karakteristik *Recency*, *Frequency*, dan *Monetary* (RFM).

a. Hasil Segmentasi Cluster

Setelah jumlah *cluster* optimal ditentukan, proses *clustering* dilakukan menggunakan algoritma K-Means dengan jumlah *cluster* sebanyak 4. Hasil *clustering* yang menampilkan rata-rata nilai *Recency*, *Frequency*, dan *Monetary* pada tiap kelompok pelanggan disajikan pada Tabel 2.

Tabel 2. Rata-rata Nilai *Recency* (R), *Frequency* (F), dan *Monetary* (M) per cluster

Cluster	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
0	181,25	1,00	67.777,16
1	342,80	1,00	240.469,67
2	555,49	1,00	73.947,53
3	256,84	2,01	219.432,55

Berdasarkan Tabel 2, nilai pusat kluster (*centroid*) yang dihasilkan dari algoritma K-Means pada variabel *Recency*, *Frequency*, dan *Monetary*, dilakukan pengelompokan penamaan segmen pelanggan. Penamaan kategori segmen ini mengacu pada kerangka segmentasi perilaku pelanggan (*Customer Relationship Management* / CRM) yang disesuaikan dengan karakteristik distribusi nilai RFM sebagaimana tercantum pada Tabel 3.

Tabel 3. Segmentasi Pelanggan

Cluster	Karakteristik	Kategori
Cluster 0	<i>Frequency</i> dan <i>Monetary</i> rendah	Low Value Customer



Cluster	Karakteristik	Kategori
Cluster 1	<i>Monetary</i> tinggi namun <i>frequency</i> rendah	Big Spender
Cluster 2	<i>Recency</i> tinggi dan <i>monetary</i> rendah	Lost Customer
Cluster 3	<i>Frequency</i> dan <i>Monetary</i> tinggi	Loyal Customer

Berdasarkan Tabel 2 dan Tabel 3, terlihat bahwa *cluster Lost Customer* memiliki nilai *Recency* tertinggi dan *Monetary* rendah dibandingkan dengan *cluster Loyal Customer* yang memiliki *Monetary* tinggi dan *Frequency* tinggi. Selanjutnya, *cluster Big Spender* memiliki *Monetary* tinggi namun *Frequency* transaksi rendah, sedangkan *cluster Low Value Customer* memiliki *Recency* paling rendah tetapi jumlah *Monetary* transaksinya rendah. Hal ini menunjukkan bahwa pelanggan pada masing-masing *cluster* memiliki karakteristik yang berbeda-beda sehingga perusahaan dapat memfokuskan strategi bisnisnya sesuai kebutuhan.

3.2 Implementasi

Berdasarkan hasil pengelompokan pelanggan menggunakan algoritma *K-Means Clustering* serta evaluasi profil *Recency*, *Frequency*, dan *Monetary* (RFM), dirumuskan rekomendasi strategi bisnis yang spesifik dan *actionable* untuk masing-masing segmen pelanggan. Pemetaan karakteristik perilaku beserta usulan strategi *Customer Relationship Management* (CRM) disajikan pada Tabel 4.

Tabel 4. Rekomendasi Strategi bisnis

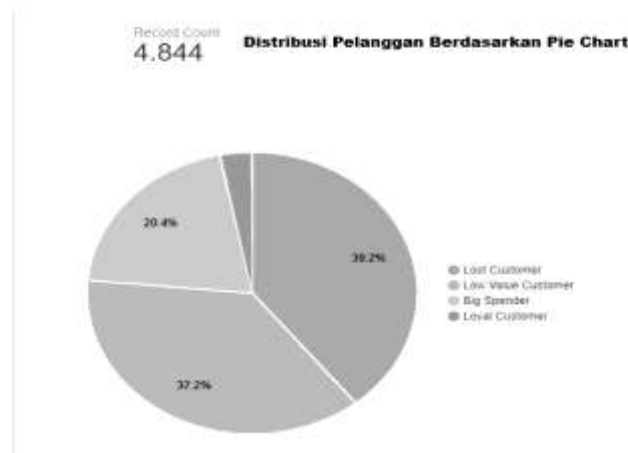
Segmen Pelanggan (Cluster)	Karakteristik Perilaku	Rekomendasi Strategi Bisnis (Actionable Insights)
Cluster 0: <i>Low Value Customer</i>	Transaksi sering, tapi total belanja kecil.	Berikan promosi <i>upselling</i> produk murah, dan Bundle Promo.
Cluster 1: <i>Big Spender</i>	Belanja nominal besar, tapi transaksi jarang.	Berikan <i>personalized offer</i> dan <i>cross Selling</i> untuk transaksi berikutnya.
Cluster 2: <i>Lost Customer</i>	Sudah lama tidak transaksi (<i>Recency</i> tinggi), belanja rendah.	Kampanye Reactivation Campaign melalui pemicu pesan (email/push notification) khusus + <i>special comeback promo</i> .
Cluster 3: <i>Loyal Customer</i>	Baru saja belanja, sering bertransaksi, total belanja besar.	<i>VIP Program</i> , <i>Loyalty Reward</i> , Exclusive Promo.

- Segmen *Low Value Customer* (Cluster 0)**
Kelompok pelanggan ini memiliki frekuensi belanja yang cukup tinggi namun dengan total pengeluaran (*Monetary*) yang relatif rendah. Strategi CRM difokuskan pada peningkatan nilai transaksi (*Average Order Value*) melalui program *upselling* produk bernilai lebih tinggi serta skema *Bundle Promo* (misalnya: "Beli 2 Lebih Hemat").
- Segmen *Big Spender* (Cluster 1)**
Pelanggan pada segmen ini melakukan transaksi dengan rata-rata nominal belanja yang sangat besar, namun memiliki jeda waktu transaksi yang cukup panjang. Skenario kampanye yang diterapkan adalah *personalized cross-selling* melalui penawaran eksklusif berupa rekomendasi produk pelengkap premium berdasarkan riwayat belanja terdahulu.
- Segmen *Lost Customer* (Cluster 2)**
Segmen ini mencakup pelanggan yang sudah lama tidak melakukan transaksi (*Recency* tinggi) dengan kontribusi nilai belanja yang rendah. Untuk mencegah *churn* permanen, dieksekusi skenario *Reactivation Campaign* berupa pengiriman pesan otomatis melalui *email marketing* atau *push notification* berisi voucher promosi khusus "We Miss You - Diskon 20% untuk Transaksi Berikutnya".
- Segmen *Loyal Customer* (Cluster 3)**
Merupakan kelompok pelanggan paling bernilai (*most valuable*) yang baru saja bertransaksi, sering berbelanja, dan menghasilkan total pengeluaran tertinggi. Strategi retensi diimplementasikan melalui *VIP Member Program*, apresiasi poin *Loyalty Reward*, serta akses prapenjualan eksklusif untuk produk baru guna menjaga loyalitas jangka panjang.

3.2.1 Visualisasi Distribusi Cluster

Bentuk visualisasi proporsi jumlah pelanggan pada setiap *cluster* ditampilkan melalui grafik *Pie Chart* pada

Gambar 6.

**Gambar 6.** Pie Chart untuk distribusi ke-4 *cluster*.

Berdasarkan visualisasi pada Gambar 6, segmen *Lost Customer* memiliki 1.900 (39,2%) pelanggan, *Low Value Customer* sebanyak 1.804 (37,2%) pelanggan, *Big Spender* sebanyak 986 (20,4%) pelanggan, dan *Loyal Customer* sebanyak 154 (3,2%) pelanggan. Dibandingkan *cluster* lainnya, *Lost Customer* merupakan *cluster* dengan jumlah paling banyak. Hal ini menunjukkan bahwa sebagian besar pelanggan sudah lama tidak melakukan transaksi kembali.

3.3 Pembahasan

Hasil segmentasi pelanggan *e-commerce* menggunakan integrasi model RFM dan algoritma K-Means Clustering menghasilkan empat klaster utama dengan karakteristik perilaku yang terdistribusi secara spesifik. Sebagaimana ditampilkan pada Gambar 6, segmen *Lost Customer* mencakup proporsi terbesar yaitu 39,2%, disusul *Low Value Customer* (37,2%), *Big Spender* (20,4%), dan *Loyal Customer* (3,2%). Sesuai arahan strategi *Customer Relationship Management* (CRM) dan rekomendasi tindakan pada Tabel 3, analisis mendalam serta pembahasan hasil penelitian diuraikan pada 3.3.1.

3.3.1 Insight dan Strategi Bisnis

a. Segmentasi *Loyal Customer* (3,2%)

Klaster ini memiliki tingkat retensi paling tinggi dengan rata-rata *Recency* sebesar 256,84 hari, *Frequency* rata-rata sebesar 2,01 kali, dan nilai *Monetary* mencapai 219.432,55. Tingginya angka *Frequency* dan *Monetary* mengindikasikan bahwa kelompok ini adalah pilar utama pendapatan perusahaan. Temuan ini sejalan dengan kajian penelitian [2], serta [1], yang menegaskan bahwa pelanggan dengan nilai *Frequency* dan *Monetary* di atas rata-rata merupakan aset krusial yang wajib dipertahankan melalui program *Customer Loyalty*, insentif eksklusif, serta akses produk VIP untuk menjaga tingkat retensi.

b. Segmentasi *Big Spender* (20,4%)

Klaster *Big Spender* ditandai dengan nilai *Monetary* rata-rata tertinggi sebesar 240.469,67 per transaksi, namun dengan *Frequency* rata-rata 1,00 kali dan *Recency* 342,80 hari. Pola ini menunjukkan bahwa pelanggan melakukan pembelian produk ber-margin tinggi namun tidak rutin melakukan transaksi ulang. Selain itu, penelitian sebelumnya menegaskan bahwa teknik *data mining* berbasis *clustering* terbukti efektif dalam menghasilkan segmentasi pelanggan yang presisi guna mendukung pengambilan keputusan bisnis berbasis data [3], dan [5], pelanggan tipe ini membutuhkan pendekatan *personalized offer*, promosi *bundling* produk nilai tinggi, serta strategi *cross-selling* untuk memicu peningkatan frekuensi transaksi pada masa mendatang.

c. Segmentasi *Low Value Customer* (37,2%)

Klaster ini memiliki nilai *Monetary* rata-rata paling rendah yaitu 67.777,16 dan *Frequency* 1,00 kali, tetapi memiliki nilai *Recency* relatif baru yaitu 181,25 hari. Tingginya angka transaksi baru pada nominal kecil menunjukkan bahwa pelanggan masih dalam tahap eksplorasi produk *entry-level*. Untuk meningkatkan nilai ekonomis segmen ini, strategi *up-selling* dan penetapan ambang batas minimum transaksi (misalnya voucher diskon dengan batas minimal belanja) sangat direkomendasikan.

d. Segmentasi *Lost Customer* (39,2%)

Dominasi segmen *Lost Customer* dengan *Recency* rata-rata mencapai 555,49 hari menjadi peringatan akan tingginya tingkat *customer churn*. Angka *Recency* yang lama mengindikasikan bahwa pelanggan telah pasif atau beralih ke kompetitor. Hasil ini mendukung temuan penelitian sebelumnya [4], serta [6] yang menyatakan bahwa pemetaan RFM efektif mengisolasi segmen berisiko tinggi sehingga perusahaan dapat mengalokasikan



anggaran pemasaran secara efisien melalui kampanye reaktivasi (*win-back campaign*), *push notification*, dan kupon khusus penarikan kembali (*comeback promo*).

Berdasarkan hasil visualisasi dan interpretasi *cluster*, metode RFM dan K-Means Clustering mampu menghasilkan segmentasi pelanggan yang dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan bisnis dan strategi pemasaran yang lebih tepat sasaran.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai segmentasi pelanggan *e-commerce*, dapat disimpulkan bahwa penerapan metode *Recency, Frequency, Monetary* (RFM) dan algoritma *K-Means Clustering* berhasil menjawab seluruh permasalahan penelitian. Pertama, proses ekstraksi data transaksi *e-commerce* yang berukuran besar berhasil mentransformasikan data mentah menjadi tiga variabel kuantitatif utama (RFM), sehingga informasi strategis mengenai pola perilaku belanja pelanggan dapat diekstraksi secara objektif. Kedua, masalah pemasaran yang bersifat massal (*generic marketing*) berhasil diatasi melalui pengelompokan pelanggan menggunakan *K-Means Clustering* yang dievaluasi dengan *Elbow Method* dan *Silhouette Score* sebesar 0,43, menghasilkan empat kluster optimal sebagai dasar perumusan strategi *Customer Relationship Management* (CRM) yang spesifik. Ketiga, risiko hilangnya pelanggan potensial (*customer churn*) dapat ditekan karena algoritma berhasil mengidentifikasi segmen *Lost Customer* dan *Low Value Customer* secara presisi untuk pelaksanaan kampanye *re-engagement*. Keempat, permasalahan media analisis dijawab melalui pembangunan *dashboard* interaktif Google Looker Studio yang memudahkan pihak manajemen dalam memahami distribusi profil pelanggan dan tren belanja secara *real-time*. Keberhasilan integrasi ini memberikan panduan praktis yang sangat berguna bagi pemangku kepentingan dalam mengoptimalkan alokasi anggaran promosi perusahaan. Meskipun penelitian ini terbatas pada analisis data transaksi historis, kombinasi metode ini terbukti memberikan nilai tambah strategis bagi efisiensi pemasaran. Untuk penelitian selanjutnya, disarankan agar mengintegrasikan variabel demografi serta membandingkan algoritma *clustering* lain guna meningkatkan presisi segmentasi.

REFERENCES

- [1] M. F. Fadhilah, A. L. A. Suyoso, and I. Puspitasari, "Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency, Frequency dan Monetary (RFM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 575–584, 2024, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.57152/malcom.v5i1.1491>
- [2] Y. Syahra, A. Fadlil, and H. Yuliansyah, "Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry," *Jurnal Dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1120–1131, 2025, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.33395/sinkron.v9i3.14907>
- [3] D. Saumendra and J. Nayak, "Customer Segmentation via Data Mining Techniques: State-of-the-Art Review," vol. 281, pp. 489–507, 2022, Accessed: Apr. 22, 2026. [Online]. Available: 10.1007/978-981-16-9447-9_38
- [4] D. A. Imanuel and G. Alfian, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, vol. 12, no. 2, pp. 283–292, 2025, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.25126/jtiik.2025128619>
- [5] R. M. Fauzan and G. Alfian, "Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma Klusterisasi K-Means," *Jurnal Informatika Sunan Kalijaga*, vol. 9, no. 3, 2024, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.14421/jiska.2024.9.3.170-177>
- [6] N. Nuraini, Y. Umaidah, and I. Maulana, "Penerapan K-Means Clustering Berbasis Model LRFM Untuk Segmentasi Pelanggan Pada PT Fazri Putra Mandiri," *Jurnal Ilmiah Wahana Pendidikan*, vol. 12, no. 3, pp. 92–103, 2026, Accessed: Apr. 22, 2026. [Online]. Available: <https://jurnal.peneliti.net/index.php/JIWP/article/view/12742>
- [7] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 16, no. 2, pp. 105–112, 2022, Accessed: Apr. 22, 2026. [Online]. Available: 10.32815/jitika.v16i2.761
- [8] R. Alamsyah and N. Rokhman, "Preprocessing Algorithm for K-Means Anomaly Detection on Payment Logs," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 19, no. 2, 2025, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.22146/ijccs.105290>
- [9] F. Maulana, E. R. Nainggolan, and M. Rahmayu, "Pemanfaatan API Midtrans Dan RajaOngkir Dalam Sistem Penjualan Online," *Reputasi: Jurnal Rekayasa Perangkat Lunak*, vol. 5, no. 1, pp. 77–87, 2024, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.31294/reputasi.v5i1.3456>
- [10] P. S. Mustafa, H. Gusdiyanto, A. M. N. K. & Victoria, and Nurika Dyah Lestariningsih., *METODOLOGI PENELITIAN KUANTITATIF*. Mojokerto: Insight Mediatama, 2020. Accessed: Aug. 18, 2026. [Online]. Available: <https://repository.insightmediatama.co.id/books/article/view/21>
- [11] D. Jollyta, W. & Ramdhan, and H. Mawengkang, *Konsep Data Mining dan Penerapan*. Yogyakarta: Deepublish Publisher, 2020. Accessed: Aug. 18, 2026. [Online]. Available: <https://books.google.co.id/books?hl=id&lr=&id=vjhWEQAAQBAJ&oi=fnd&pg=PP1&dq=Konsep+Data+Mining+dan+Penerapan&ots=HsgIrnBTC1&sig=rUvty3qYXQgO>



- _W1wq89eL2aRvh8&redir_esc=y#v=onepage&q=Konsep%20Data%20Mining%20dan%20Penerapan&f=false
- [12] G. A. Pradnyana, I. G. M. & Darmawiguna, and I. N. S. W. Wijaya, *DATA MINING: Menemukan Pengetahuan dalam Data - Rajawali Pers*. Depok, 2020. Accessed: Aug. 18, 2026. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=7nBpEQAAQBAJ&oi=fnd&pg=PA51&dq=Data+Mining:+Menemukan+Pengetahuan+dalam+Data&ots=M86iZLgeHd&sig=0EGsHynH8NdRj_CNvfzBrkrzElE&redir_esc=y#v=onepage&q&f=false
- [13] N. Purwati, H. & Kurniawan, and S. Karnila, *Data Mining*, vol. 1. Purwokerto: Zahira Media Publisher, 2021. Accessed: Aug. 18, 2026. [Online]. Available: [https://books.google.co.id/books?hl=id&lr=&id=Q3NHEAAAQBAJ&oi=fnd&pg=PR5&dq=Purwati,+N.,+Kurniawan,+H.,+%26+Karnila,+S.+\(2021\).+Data+Mining.+Purwokerto:+Zahira+Media+Publis her&ots=cN7eUJnHD_&sig=u45qtpaZkTTO8CEZunKnMuXsV0&redir_esc=y#v=onepage&q&f=false](https://books.google.co.id/books?hl=id&lr=&id=Q3NHEAAAQBAJ&oi=fnd&pg=PR5&dq=Purwati,+N.,+Kurniawan,+H.,+%26+Karnila,+S.+(2021).+Data+Mining.+Purwokerto:+Zahira+Media+Publis her&ots=cN7eUJnHD_&sig=u45qtpaZkTTO8CEZunKnMuXsV0&redir_esc=y#v=onepage&q&f=false)
- [14] M. & Arhami and M. Nasir, *Data Mining: Algoritma dan Implementasi*. Yogyakarta: Andi, 2020. Accessed: Aug. 18, 2026. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=AtcCEAAAQBAJ&oi=fnd&pg=PP1&dq=Data+Mining:+Algoritma+da n+Implementasi&ots=hDpqGK4Wq7&sig=8O3GpLULNeOZVdSqu3G9Jq1cyol&redir_esc=y#v=onepage&q=Data%20 Mining%3A%20Algoritma%20dan%20Implementasi&f=false
- [15] A. G. Ramadhan, "DATA MINING UNTUK SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS: STUDI KASUS PADA DATA PELANGGAN DI TOKO RETAIL," *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 8, no. 10, 2023, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.36418/syntax-literate.v8i10>
- [16] Efraim Turban, Carol Pollard, and Gregory Wood, *Information Technology for Management: On-Demand Strategies for Performance, Growth and Sustainability (11th Edition)*, 11th ed. John Wiley & Sons, Inc, 2018. Accessed: Aug. 24, 2026. [Online]. Available: https://www.researchgate.net/publication/349702108_Information_Technology_for_Management_On-Demand_Strategies_for_Performance_Growth_and_Sustainability_11th_edition
- [17] Adzriel Yusak Noah Rumapea, Dian Pratiwi, and Syandra Sari, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM)," Dec. 2024, Accessed: Sep. 17, 2026. [Online]. Available: <https://ejournal.sisfokomtek.org/index.php/saintek/article/view/4607>
- [18] H. Kishendrian, N. Hanum, C. Prianto, and W. I. Rahayu, "Penerapan Metode Clustering Dalam Segmentasi Pelanggan Perusahaan Logistik," *SINTECH (Science and Information Technology)*, no. 3, pp. 137–146, 2023, Accessed: Apr. 22, 2026. [Online]. Available: <https://doi.org/10.31598/sintechjournal.v6i3.1432>
- [19] B. Al Hamid and T. Wiradinata, "Customer Segmentation Strategy Using RFM Model and K-Means Clustering for Marketing Strategy Optimization in Retail Industry," *Journal of Applied Informatics and Computing*, pp. 45–56, 2026, Accessed: Sep. 29, 2026. [Online]. Available: <https://doi.org/10.30871/jaic.v10i2.12302>
- [20] Rifat Naufal and Alamsyah, "Customer Segmentation Using RFM Analysis and the K-Means Clustering Algorithm to Support Data-Driven Marketing Strategies," *Information Technology Education Journal*, vol. 5, no. 3, pp. 78–95, Aug. 2026, Accessed: Sep. 29, 2026. [Online]. Available: <https://doi.org/10.59562/intec.v5i3.12821>