

Segmentasi Pelanggan Alfagift Jambi Menggunakan K-Means *Clustering* Berdasarkan Model RFM

Maulidya Nazlita Az-Zahara¹, Ulfa Khaira², Pradita Eko Prasetyo Utomo³, Akhiyar Waladi⁴

^{1,2,3,4}Informatika, Fakultas Sains dan Teknologi, Universitas Jambi, Jambi

¹maulidyanazlita@gmail.com, ²ulfa.ilkom@gmail.com, ³pradita.eko@unja.ac.id, ⁴akhiyar.waladi@unja.ac.id

Abstrak

Transformasi digital di sektor ritel mendorong kebutuhan analisis perilaku pelanggan berbasis data. Penelitian ini mengusulkan pendekatan segmentasi pelanggan aplikasi Alfagift di wilayah Jambi menggunakan kombinasi model *Recency, Frequency, Monetary* (RFM) dan algoritma *K-Means Clustering*. Permasalahan utama yang diangkat adalah belum tersedianya analisis segmentasi yang terukur pada *platform* ritel digital skala besar yang mempertimbangkan konteks musiman. Data yang digunakan mencakup 1.529.558 transaksi selama tiga bulan sebelum dan sesudah Hari Raya Idul Fitri 2025. Tahapan penelitian diawali dengan pra-pemrosesan data yang cermat, dilanjutkan dengan deteksi serta penanganan outlier menggunakan *Isolation Forest* guna menjaga kestabilan sentroid. Setelah itu, dilakukan perhitungan skor RFM dan penentuan jumlah kluster optimal ($K=3$). Proses klasterisasi pelanggan menghasilkan tiga segmen yang memiliki karakteristik perilaku berbeda secara signifikan. Segmen pertama adalah *frequent low-spending customers* (20,6%), segmen kedua *high-value but infrequent buyers* (22,5%), dan segmen dominan ketiga adalah *promotion-driven customers* (57,0%). Evaluasi kualitas kluster menunjukkan *Silhouette Score* 0,2362 serta *Calinski-Harabasz Index* 1257,17, yang mengonfirmasi bahwa struktur kluster terbentuk dengan baik, stabil, dan dapat diinterpretasikan secara jelas. Kontribusi utama dari penelitian ini terletak pada penerapan integrasi model RFM dan *K-Means* pada *dataset* ritel digital skala besar di Indonesia yang mempertimbangkan konteks musiman Idul Fitri, sebuah area yang belum banyak dikaji secara komprehensif sebelumnya. Temuan ini memberikan dasar ilmiah yang kuat bagi manajemen ritel dalam merancang strategi pemasaran yang lebih tersegmentasi, personal, dan efisien.

Kata kunci: segmentasi pelanggan, RFM, K-means, *alfagift*, *data mining*

1. Pendahuluan

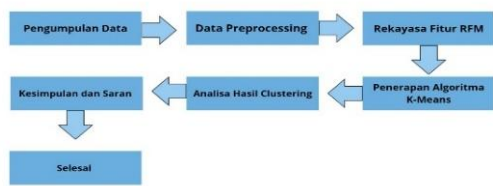
Perkembangan teknologi digital telah mendorong transformasi besar pada industri ritel di Indonesia. Pergeseran perilaku konsumen dari belanja konvensional ke *platform* digital menuntut pelaku usaha untuk menghadirkan pengalaman belanja yang efisien, personal, dan berbasis data (Nurnadya *et al.*, 2024). Alfamart sebagai salah satu jaringan ritel terbesar di Indonesia merespons tren ini dengan meluncurkan Alfagift, aplikasi yang menggabungkan sistem *e-commerce* dan *loyalty* program dalam satu ekosistem digital. Melalui Alfagift, pelanggan dapat melakukan pembelian, mengumpulkan poin, hingga mendapatkan promosi yang disesuaikan dengan preferensi mereka. Kehadiran *platform* ini tidak hanya memperkuat daya saing Alfamart dalam era ritel modern, tetapi juga menghasilkan volume data pelanggan yang masif dan potensial untuk dianalisis.

Permasalahan utama yang muncul adalah bagaimana memahami perilaku pelanggan Alfagift secara terukur agar strategi promosi dan retensi pelanggan menjadi lebih efektif. Setiap pengguna memiliki karakteristik pembelian yang berbeda, baik

dari segi waktu transaksi, frekuensi belanja, maupun jumlah pengeluaran. Data yang beragam ini sulit dianalisis tanpa pendekatan sistematis berbasis analisis data. Oleh karena itu, segmentasi pelanggan menjadi langkah penting untuk mengelompokkan konsumen berdasarkan kesamaan perilaku sehingga strategi pemasaran dapat lebih tepat sasaran (Nurnadya *et al.*, 2024).

Model RFM (*Recency, Frequency, Monetary*) merupakan salah satu pendekatan yang paling umum dan efektif digunakan dalam analisis perilaku pelanggan (Ho *et al.*, 2023). Model ini mengevaluasi pelanggan berdasarkan tiga dimensi: waktu transaksi terakhir (*recency*), frekuensi pembelian (*frequency*), dan total nilai transaksi (*monetary*). Meski demikian, model RFM tradisional bersifat deskriptif dan belum sepenuhnya mampu menangkap keterkaitan kompleks antar variabel. Untuk mengatasi keterbatasan tersebut, berbagai penelitian telah mengintegrasikan RFM dengan algoritma klasterisasi.

Penelitian sebelumnya oleh Arthanugraha *et al.*, (2024) menerapkan model RFM untuk segmentasi pelanggan toko *e-commerce* Shopee, menghasilkan



Gambar 1. Alur Penelitian

tiga segmen dengan karakteristik berbeda namun terbatas pada skala data kecil. Nugraha *et al.*, (2025) melakukan segmentasi berbasis RFM K-Means pada perusahaan *payment gateway* dengan *Davies-Bouldin Index* 0.1089, namun konteks penelitiannya adalah industri *fintech*, bukan ritel fisik digital. Fadillah *et al.*, (2023) menganalisis perilaku pelanggan Ralali.com menggunakan RFM, tetapi tanpa integrasi algoritma klasterisasi yang optimal. Yusak *et al.*, (2024) berhasil mengidentifikasi tiga segmen pelanggan ritel daring menggunakan K-Means berbasis RFM, namun tidak mempertimbangkan faktor musiman seperti Hari Raya Idul Fitri yang sangat relevan di Indonesia. Dari tinjauan tersebut, terdapat research gap yang jelas: belum ada penelitian yang mengintegrasikan *RFM K-Means* pada *platform* ritel digital skala besar di Indonesia dengan mempertimbangkan variabel musiman Idul Fitri.

Penelitian ini bertujuan untuk mengisi gap tersebut dengan melakukan segmentasi pelanggan Alfagift di wilayah Jambi menggunakan *K-Means Clustering* berbasis model RFM pada *dataset* 1.5 juta transaksi periode Idul Fitri 2025. Kontribusi ilmiah penelitian ini meliputi: (1) penerapan model RFM *K-Means* pada *dataset* ritel digital skala besar di Indonesia, (2) analisis segmentasi dalam konteks musiman Idul Fitri yang belum banyak dikaji, dan (3) validasi kluster menggunakan *multiple metrics* (*Silhouette Score*, *Calinski-Harabasz Index*) untuk memastikan kualitas segmentasi. Melalui hasil segmentasi ini, diharapkan dapat diperoleh klasifikasi pelanggan yang bermanfaat bagi manajemen ritel dalam menyusun strategi yang lebih tepat, termasuk pengembangan program loyalitas, penyesuaian promosi, serta peningkatan kepuasan pelanggan secara menyeluruh.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan *data mining* kuantitatif dengan metode *K-Means Clustering* berbasis model RFM. Pemilihan metode ini didasarkan pada kemampuan *K-Means* dalam menangani *dataset* berskala besar secara efisien, serta keunggulan model RFM dalam mengkuantifikasi perilaku transaksi pelanggan secara multidimensional (Parmar & Panchal, 2024). Tahapan penelitian dilaksanakan secara sistematis dan berurutan, sebagaimana ditampilkan pada Gambar 1.

--- Missing Values ---		
	Missing Count	Percentage (%)
TRO_PONTA_ID	591	0.038639

Gambar 2. Hasil Penanganan *Missing Value* Data

2.1 Deskripsi Dataset

Penelitian ini menggunakan *dataset* internal dari jaringan ritel Alfagift Provinsi Jambi yang mencakup transaksi selama tiga bulan sebelum dan sesudah Hari Raya Idul Fitri 2025. Pemilihan periode ini dilakukan secara sengaja (*purposeful*) untuk menangkap dinamika perilaku belanja musiman yang merupakan salah satu kontribusi utama penelitian ini. *Dataset* terdiri dari 1.529.558 baris transaksi dengan 16 kolom variabel yang merepresentasikan atribut transaksional pelanggan. Variabel utama yang digunakan dalam penelitian ini meliputi: TRO_DATE (tanggal dan waktu transaksi), TRO_QTY (jumlah item yang dibeli per transaksi), TRO_NET (nilai total transaksi dalam rupiah), TRO_PAYMENT_METHOD (jenis pembayaran), TRO_NO_ORDER (nomor unik transaksi untuk perhitungan frekuensi), TRO_PONTA_ID (ID pelanggan sebagai kunci pengelompokan), dan TRO_PLU (kode produk yang dibeli).

2.2 Pra-Pemrosesan Data

Pra-proses data dilakukan untuk memastikan kualitas data dan mempersiapkan *dataset* untuk langkah selanjutnya. Proses ini mencakup:

a. Penanganan Nilai yang Hilang

Sebagian kecil (0.039%) kolom TRO_PONTA_ID terdeteksi tidak lengkap. Kolom ini berfungsi sebagai identitas pelanggan unik sehingga berpengaruh langsung terhadap perhitungan RFM. Mengingat proporsi data yang hilang sangat kecil, baris dengan nilai kosong pada kolom ini dihapus untuk menjaga kualitas analisis tanpa mengorbankan representativitas *dataset*, sebagaimana pada Gambar 2.

b. Penghapusan Duplikat Data

Transaksi duplikat dihapus untuk menjaga kualitas dan keakuratan analisis, menghasilkan 1.528.957 baris data unik setelah menghilangkan record dengan *timestamp* dan detail transaksi identik. Sebagaimana ditampilkan pada Gambar 3.

c. Standardisasi dan Filterisasi

Deskripsi produk distandardisasi guna meningkatkan konsistensi dan deskripsi produk distandardisasi ke format huruf kapital untuk menghi-

```
print("Jumlah baris sebelum menghapus duplikat:", len(df_cleaned))
# Identifikasi dan hapus baris duplikat
df_cleaned_no_duplicates = df_cleaned.drop_duplicates().copy()
print("Jumlah baris setelah menghapus duplikat:", len(df_cleaned_no_duplicates))
# Periksa apakah masih ada duplikat
print("\nJumlah duplikat setelah dihapus:", df_cleaned_no_duplicates.duplicated().sum())
Jumlah baris sebelum menghapus duplikat: 1528967
Jumlah baris setelah menghapus duplikat: 1528736
Jumlah duplikat setelah dihapus: 0
```

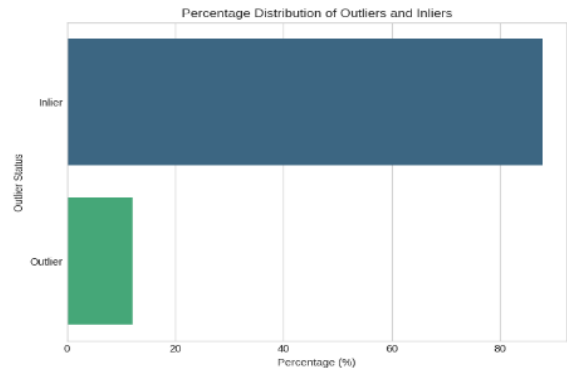
Gambar 3. Penghapusan Data Duplikat

langkah inkonsistensi akibat variasi pengetikan. Selain itu, dilakukan filterisasi pada kolom kode stok untuk menghapus kode non-produk seperti "POST" yang merepresentasikan transaksi administratif, bukan transaksi penjualan nyata. Langkah ini memastikan *dataset* hanya memuat transaksi pembelian yang relevan untuk analisis perilaku pelanggan.

d. Deteksi dan Penanganan *Outliers* dengan *Isolation Forest*

Metode *Isolation Forest* diterapkan untuk mendeteksi dan menghapus data yang menyimpang secara signifikan dari pola umum transaksi. Pemilihan *Isolation Forest* didasarkan pada efisiensi komputasi tingginya $O(n \log n)$ yang ideal untuk *dataset* skala besar (>1,5 juta baris), sifatnya yang non-parametrik (tidak memerlukan asumsi distribusi data), serta keunggulannya dalam mengisolasi anomali pada ruang multidimensi tanpa terpengaruh oleh masking effect (Vashisth & Goyal, 2024). Justifikasi utama penggunaan metode ini terletak pada sensitivitas algoritma K-Means terhadap nilai ekstrem; K-Means menghitung sentroid berbasis rata-rata (mean), sehingga keberadaan outlier bernilai sangat tinggi akan menggeser sentroid secara drastis dan merusak batas kluster harian yang wajar. Pembersihan outlier dengan *Isolation Forest* terbukti menjaga kestabilan posisi sentroid, meningkatkan kepadatan internal kluster, dan mengoptimalkan metrik evaluasi (*Silhouette Score*).

Algoritma tersebut diinisialisasi menggunakan parameter kontaminasi sebesar 0,05 sebagai ambang batas estimasi anomali teoretis guna mencegah penghapusan data valid secara berlebihan. Pelatihan model dilakukan pada tingkat transaksi individu (1.528.736 baris) menggunakan fungsi `fit_predict()`, yang memberikan label -1 untuk anomali dan +1 untuk data normal. Proses deteksi mengidentifikasi 185.894 baris transaksi yang dikategorikan sebagai anomali, yang secara matematis menghasilkan tingkat outlier sebesar 12,16% ($(185.894 / 1.528.736) \times 100\%$), sebagaimana ditampilkan pada Gambar 4. Persentase ini melebihi ambang batas kontaminasi awal (0,05) akibat akumulasi transaksi dengan nilai ekstrem selama periode musiman Idul Fitri. Transaksi outlier ini kemudian dikecualikan dari *dataset* utama sebelum proses agregasi fitur RFM per pelanggan untuk memastikan bahwa pengelompokan akhir secara

Gambar 4. Persentase *Inliers* dan *Outliers* berdasarkan *Isolation Forest*

akurat mencerminkan segmen perilaku belanja yang representative.

2.3 Rekayasa Fitur RFM

Pendekatan model *Recency, Frequency, Monetary* (RFM) pertama kali diperkenalkan oleh Arthur Hughes pada tahun 1994 dan telah berkembang menjadi salah satu teknik segmentasi pelanggan paling efektif dalam analisis perilaku konsumen (Parmar & Panchal, 2024). Model ini dipilih karena kemampuannya mengkuantifikasi tiga dimensi kunci perilaku pembelian secara simultan:

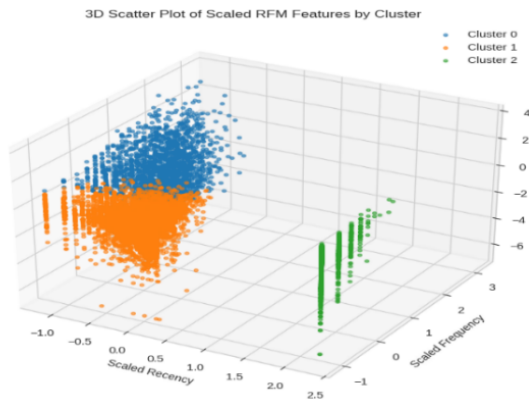
1. *Recency* (R): jumlah hari sejak transaksi terakhir pelanggan hingga tanggal referensi (*cut-off date*), di mana nilai *recency* rendah menandakan pelanggan masih aktif;
2. *Frequency* (F): jumlah transaksi unik yang dilakukan pelanggan dalam periode analisis, di mana nilai tinggi mencerminkan loyalitas transaksi.
3. *Monetary* (M): total nilai uang yang dibelanjakan pelanggan, dihitung dari agregasi nilai TRO_NET seluruh transaksi.

Nilai RFM dihitung untuk masing-masing pelanggan berdasarkan data historis transaksi, menghasilkan tabel agregat dengan TRO_PONTA_ID sebagai indeks. Seluruh nilai RFM kemudian dinormalisasi menggunakan *StandardScaler* sebelum proses klusterisasi untuk memastikan setiap dimensi memiliki kontribusi yang seimbang. Nilai negatif pada kolom *Monetary* merupakan hasil dari proses standarisasi ini, bukan indikasi kerugian transaksi. Sampel hasil perhitungan RFM ditampilkan pada Tabel 1.

Berdasarkan sampel pada Tabel 1 hasil pengolahan data RFM menampilkan variasi nilai pada masing-masing indikator pelanggan. Nilai **Recency** menunjukkan jarak waktu sejak transaksi terakhir pelanggan, **Frequency** menggambarkan jumlah transaksi yang dilakukan dalam periode tertentu, sedangkan **Monetary** merepresentasikan total nilai transaksi yang telah dinormalisasi. Berdasarkan tabel,

Tabel 1. Hasil Perhitungan RFM

Tro_ponta_id	Recency	Frequency	Monetary
0	47	6	-10,634648
1	33	6	-1,685266
2	1	41	-9,783591
3	45	3	-3,709809
4	11	10	-13,828152



Gambar 5. Visualisasi Distribusi Pelanggan dalam Ruang 3D RFM

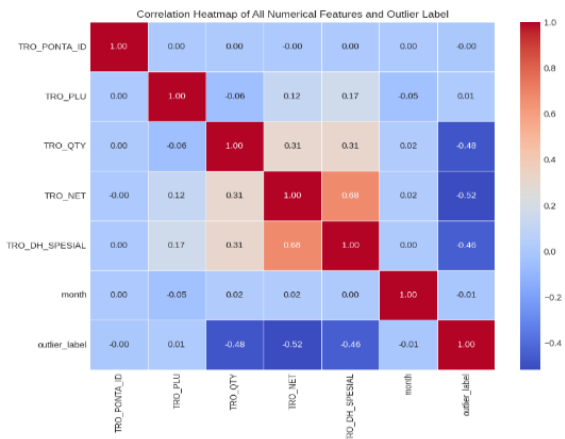
pelanggan memiliki rentang *Recency* antara 1 hingga 47 hari, *Frequency* antara 3 hingga 41 kali transaksi, dan *Monetary* yang berkisar antara $-1,68$ hingga $-13,82$. Nilai negatif pada kolom *Monetary* muncul akibat proses standardisasi data untuk menyeimbangkan skala antarvariabel sebelum proses klusterisasi dilakukan, bukan karena adanya nilai transaksi yang merugi.

Untuk memahami hubungan antar variabel RFM, dilakukan analisis korelasi yang divisualisasikan menggunakan *heatmap*. Analisis ini memungkinkan identifikasi variabel-variabel yang saling terkait, misalnya korelasi antara frekuensi pembelian dengan nilai transaksi, sehingga membantu interpretasi profil klaster yang lebih (Miranda-Guerra *et al.*, 2024). Sebagaimana ditampilkan pada Gambar 5, derajat korelasi antar variabel RFM dapat diamati melalui gradasi warna pada *heatmap*.

Gambar di atas menunjukkan visualisasi tiga dimensi (3D *scatter plot*) menampilkan distribusi pelanggan dalam tiga klaster berbeda, yang masing-masing direpresentasikan oleh warna biru (Cluster 0), oranye (Cluster 1), dan hijau (Cluster 2). Setiap titik dalam grafik menggambarkan pelanggan berdasarkan nilai *scaled recency*, *scaled frequency*, dan *scaled monetary*. Pola pemisahan yang jelas antarwarna menunjukkan bahwa algoritma berhasil mengelompokkan pelanggan berdasarkan kesamaan perilaku transaksi mereka.

2.4 Analisis Korelasi

Analisis korelasi merupakan teknik statistik yang digunakan untuk menilai kekuatan dan arah hubungan linier antar-variabel dalam suatu *dataset*.



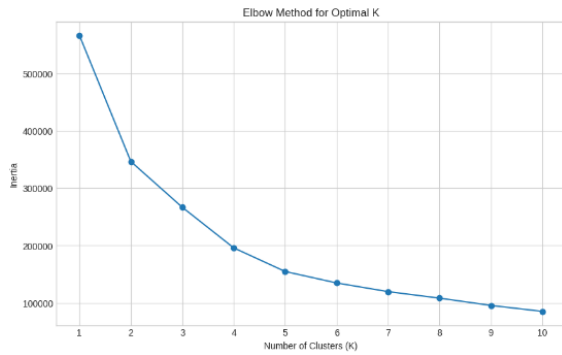
Gambar 6. Heatmap Correlation Antar Variabel

Koefisien korelasi mengukur sejauh mana dua variabel berubah bersama dengan nilai berkisar antara -1 hingga $+1$ di mana nilai positif menunjukkan bahwa kedua variabel cenderung bergerak searah, sedangkan nilai negatif mengindikasikan bahwa satu variabel meningkat saat variabel lainnya menurun. Semakin mendekati nilai absolut 1, semakin kuat keterkaitan linier antar-variabel, sementara nilai yang mendekati nol menunjukkan hubungan yang lemah atau tidak bermakna secara linier (Regasa, 2023).

Dalam konteks pengolahan data pelanggan, analisis korelasi memungkinkan peneliti untuk mengidentifikasi variabel-variabel apa yang saling terkait, misalnya frekuensi pembelian dengan nilai transaksi atau *recency* dengan loyalitas pelanggan. Dengan mengetahui pola hubungan ini, perusahaan dapat menyusun strategi pemasaran yang lebih tepat sasaran, sekaligus memahami struktur data pelanggan secara lebih menyeluruh dan menghindari asumsi hubungan yang hanya bersifat (Miranda-Guerra *et al.*, 2024). Untuk memvisualisasikan hubungan tersebut, digunakan *heatmap* yang menampilkan koefisien korelasi melalui gradasi warna. Sebagaimana ditampilkan pada Gambar 6, warna terang menggambarkan hubungan yang kuat, sedangkan warna gelap menunjukkan korelasi yang lemah. Adapun sel berwarna putih menandakan tidak adanya hubungan antar variabel.

2.5 Algoritma Klusterisasi: K-Means Clustering

Algoritma *K-Means Clustering* dipilih sebagai metode segmentasi utama karena kemampuannya menangani *dataset* berskala besar secara komputasi efisien dan menghasilkan klaster yang mudah diinterpretasikan (Refialy *et al.*, 2021). *K-Means* bekerja secara iteratif: dimulai dengan pemilihan *K* titik *centroid* secara acak, setiap data ditempatkan pada klaster dengan *centroid* terdekat berdasarkan jarak *Euclidean*, kemudian posisi *centroid* diperbarui hingga tidak ada perubahan signifikan dalam keanggotaan klaster. Tujuannya adalah meminimalkan *Within-Cluster Sum of Squares* (WCSS) atau *Sum of*



Gambar 7. Elbow Method Graphic

Squared Errors (SSE).

Dalam konteks bisnis digital, metode ini banyak diterapkan untuk menganalisis perilaku pelanggan dan membentuk segmen yang lebih terarah. Seperti dijelaskan oleh (Arthanugraha *et al.*, 2024), pengelompokan berbasis RFM (*Recency, Frequency, Monetary*) yang dikombinasikan dengan algoritma *K-Means* memungkinkan perusahaan memahami pola transaksi pelanggan, mengidentifikasi pelanggan setia, serta menyusun strategi pemasaran yang sesuai dengan karakteristik tiap segmen. Dengan demikian, penerapan algoritma *K-Means* tidak hanya efektif dalam menganalisis data numerik, tetapi juga memberikan dasar ilmiah yang kuat bagi pengambilan keputusan strategis dalam manajemen pelanggan. Berikut langkah-langkahnya:

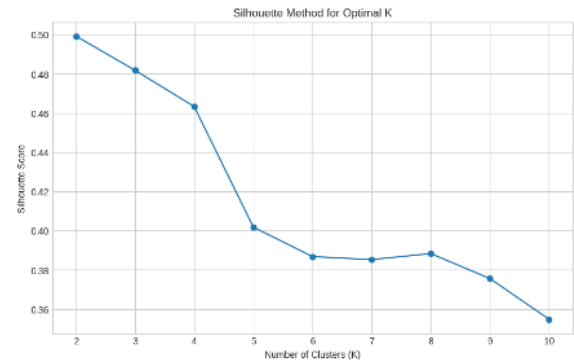
a. Metode Siku (*Elbow Method*)

Penentuan jumlah kluster optimal (K) merupakan langkah kritis yang dilakukan menggunakan dua metode komplementer. Metode Elbow menganalisis perubahan nilai WCSS seiring meningkatnya K; titik di mana penurunan WCSS mulai melandai menandakan K optimal (Vania & Sari, 2023). Berdasarkan Gambar 7, grafik Elbow mengindikasikan K optimal berada antara 3 dan 7 karena titik siku yang khas tidak terlihat jelas.

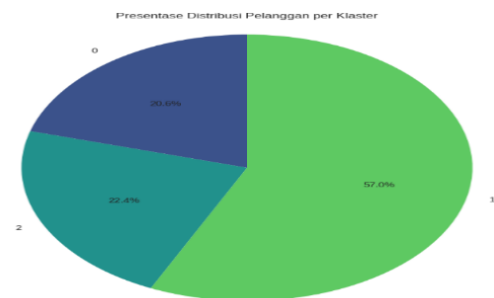
Grafik tersebut mengusulkan 5 sebagai jumlah kluster optimal (K). Namun, "titik siku" yang khas yang biasanya terlihat pada grafik semacam ini tidak terlihat jelas dalam kasus ini. Nilai inersia, yang mengukur dispersi dalam kluster, terus menurun hingga K=5, menunjukkan bahwa K optimal mungkin terletak antara 3 dan 7. Untuk menyempurnakan perkiraan ini, analisis siluet, metode evaluasi komplementer, dapat diterapkan (Sagala & Gunawan, 2022).

b. Metode Siluet / *Silhouette Method*

Untuk memvalidasi hasil *Elbow Method*, digunakan *Silhouette Method* yang menghitung skor siluet rata-rata pada setiap nilai K. Skor siluet mengukur seberapa baik setiap data point lebih cocok dengan klasternya sendiri dibandingkan kluster lain, dengan nilai mendekati +1 menunjukkan pemisa-



Gambar 8. Silhouette Method Graphic



Gambar 9. Pie Chart Cluster

han antar kluster yang baik (Myagmarsuren, 2024). Berdasarkan hasil perhitungan, K = 3 memberikan rata-rata skor siluet tertinggi dibandingkan nilai K lainnya, sehingga ditetapkan sebagai jumlah kluster optimal. Kombinasi kedua metode ini meningkatkan keandalan penentuan K, sebagaimana dinyatakan oleh Sagala & Gunawan, (2022). Hasil *Silhouette Method* ditampilkan pada Gambar 8.

3. Hasil dan Pembahasan

3.1 Formasi Kluster dan Distribusi Pelanggan

Setelah proses klusterisasi dengan K=3 selesai dilakukan, setiap pelanggan diberi label kluster (0, 1, atau 2) berdasarkan kemiripan profil RFM-nya. Formasi kluster yang dihasilkan ditampilkan pada Tabel 2.

Distribusi kluster menunjukkan bahwa Kluster 2 mendominasi dengan proporsi 57,0% dari total pelanggan, diikuti Kluster 1 (22,4%) dan Kluster 0 (20,6%). Perbedaan proporsi yang signifikan ini mengindikasikan bahwa mayoritas pelanggan Alfagift Jambi berorientasi pada promosi, sementara segmen pelanggan aktif berfrekuensi tinggi dan pelanggan bernilai tinggi memiliki ukuran yang relatif setara. Hal ini merupakan temuan yang relevan secara strategis karena menunjukkan bahwa program promosi memiliki peran dominan dalam menggerakkan aktivitas transaksi pada platform Alfagift di wilayah Jambi terbentuk berdasarkan Gambar 9.

Tabel 2. Formasi dan *Distribusi Kluster Pelanggan Alfagift*

Cluster_Label	Customer_Count	%	Karakteristik Utama
0	1956	20,567823	<i>Frequent low-spending customers</i>
1	2132	22,418507	<i>High-value but infrequent buyers</i>
2	5422	57,013670	<i>Promotion-driven customers</i>

Tabel 3. Formasi Radar Chart

	Cluster 0	Cluster 1	Cluster 2
TRO_QTY	-0,171686	-0,19715	-0,20158
TRO_NET	-0,367287	0,668213	0,102518
TRO_DH_SPESIAL	-0,332638	-0,30703	0,576743

3.2 Matriks Evaluasi Kinerja Klasterisasi

Kualitas klasterisasi dievaluasi menggunakan tiga metrik validasi internal yang saling melengkapi untuk memastikan objektivitas penilaian.

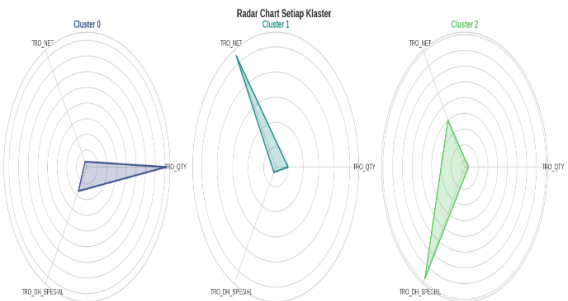
Silhouette Score sebesar 0.2362 menunjukkan adanya pemisahan antar kluster, meskipun belum optimal. Nilai ini mengindikasikan bahwa sebagian besar pelanggan telah terdistribusi ke kluster yang tepat, namun terdapat beberapa pelanggan dengan profil RFM yang berada di batas antar kluster. Kondisi ini dapat dijelaskan oleh karakteristik perilaku belanja yang bersifat kontinu, di mana transisi antar segmen tidak selalu memiliki batas yang tajam (Myagmarsuren, 2024). *Calinski-Harabasz Index* sebesar 1257,17 merupakan indikasi positif kualitas klasterisasi; nilai yang tinggi ini menunjukkan bahwa kluster-kluster terbentuk dengan baik, yakni antar kluster jauh terpisah sementara anggota dalam kluster memiliki kesamaan tinggi (Kossakov *et al.*, 2024). Perpaduan antara *Silhouette Score* yang moderat dan *Calinski-Harabasz* yang tinggi mengindikasikan bahwa klasterisasi menghasilkan segmen yang bermakna secara bisnis, meskipun batas geometris antar kluster tidak sepenuhnya tegas.

3.3 Analisis *Profiling* Kluster dan Interpretasi Perilaku Pelanggan

Analisis *profiling* dilakukan untuk memperoleh pemahaman mendalam tentang karakteristik perilaku tiap kluster. Visualisasi Radar Chart pada Gambar 10 menampilkan nilai centroid rata-rata dari tiga variabel utama (TRO_QTY, TRO_NET, TRO_DH_SPESIAL) untuk setiap kluster, sebagaimana dirangkum pada Tabel 3.

Gambar 10 di atas menampilkan hasil visualisasi Radar Chart untuk masing-masing kluster pelanggan yang diperoleh melalui algoritma *K-Means Clustering* berbasis model RFM. Setiap kluster digambarkan melalui bentuk area yang merepresentasikan nilai centroid rata-rata dari tiga variabel utama, yaitu:

- **TRO_QTY** (jumlah transaksi atau frekuensi pembelian),
- **TRO_NET** (total nilai transaksi atau pengelu-



Gambar 10. Radar Chart Visualisasi Profil Kluster Pelanggan

aran pelanggan), dan

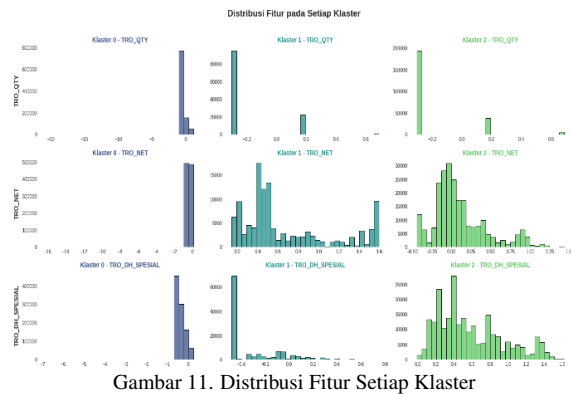
- **TRO_DH_SPESIAL** (aktivitas pelanggan terhadap transaksi khusus atau promosi tertentu).

Radar Chart digunakan untuk memperjelas profil perilaku pelanggan antar kluster dan membandingkan kekuatan relatif tiap variabel secara visual. Perbedaan pola bentuk area antar kluster mengindikasikan variasi karakteristik perilaku belanja pelanggan pada *platform* Alfagift Jambi.

Kluster 0: *Frequent Low-Spending Customers*: Berdasarkan nilai centroid, Kluster 0 memiliki nilai TRO_NET negatif (-0,367) yang paling rendah dibandingkan kluster lain, namun TRO_QTY relatif lebih tinggi. Ini mengindikasikan pelanggan yang sering bertransaksi tetapi dengan nominal kecil setiap kali pembelian. Interpretasinya, segmen ini merupakan pelanggan loyal harian yang berbelanja kebutuhan rutin dengan nilai kecil. Strategi yang tepat untuk segmen ini adalah **program loyalitas** berbasis frekuensi, seperti **reward poin per transaksi** atau **cashback bertingkat** yang meningkat seiring akumulasi pembelian.

Kluster 1: *High-Value but Infrequent Buyers*: Kluster 1 menunjukkan nilai TRO_NET tertinggi (0,668) dengan TRO_QTY yang relatif rendah. Pola ini mencerminkan pelanggan yang jarang bertransaksi namun menghabiskan jumlah besar dalam setiap pembelian. Secara perilaku, segmen ini mungkin merepresentasikan pelanggan yang melakukan pembelian bulk atau pembelian produk premium dalam frekuensi rendah. Strategi yang direkomendasikan adalah **program VIP membership** atau penawaran eksklusif untuk meningkatkan frekuensi kunjungan tanpa mengurangi nilai transaksi per kunjungan.

Kluster 2: *Promotion-Driven Customers*: Kluster 2 menonjol pada variabel TRO_DH_SPESIAL (0,577) yang paling dominan di antara ketiga kluster. Ini mengindikasikan pelanggan yang cenderung aktif bertransaksi ketika terdapat program promosi atau event tertentu, namun



Gambar 11. Distribusi Fitur Setiap Kluster

kurang aktif pada periode normal. Fakta bahwa klaster ini merupakan yang **terbesar (57,0%)** memiliki implikasi strategis signifikan: mayoritas pelanggan Alfagift Jambi adalah *price-sensitive* dan sangat responsif terhadap promosi. Oleh karena itu, penjadwalan kampanye promosi yang tepat waktu, terutama menjelang periode musiman seperti Idul Fitri, dapat memaksimalkan aktivitas transaksi pada segmen dominan ini.

Sebagai langkah validasi terhadap profil pelanggan yang diperoleh melalui visualisasi Radar chart, dilakukan pula analisis menggunakan *Histogram Chart* untuk setiap fitur utama yang telah dikelompokkan berdasarkan label klaster. Pendekatan ini memungkinkan peneliti untuk meninjau distribusi nilai tiap variabel (fitur) dalam setiap klaster secara visual seperti pada Gambar 11.

Validasi visual melalui *Histogram Chart* pada Gambar 11 memperkuat *profiling* klaster di atas. Distribusi nilai fitur dalam setiap klaster menunjukkan pola yang konsisten dengan nilai centroid Radar Chart: Klaster 2 menunjukkan distribusi TRO_DH_SPESIAL yang condong ke kanan dibandingkan klaster lain, mengkonfirmasi dominasi karakteristik promosi. Sementara itu, distribusi TRO_NET pada Klaster 1 memiliki ekor kanan yang lebih panjang, mengkonfirmasi keberadaan pelanggan bernilai tinggi dalam klaster tersebut (Syahra *et al.*, 2025).

3.4 Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini dibandingkan dengan studi terdahulu yang menerapkan metode K-Means *Clustering* untuk segmentasi pelanggan, guna mengidentifikasi kesamaan serta perbedaan yang signifikan dengan temuan penelitian lainnya.

Nurnadya *et al.* (2024) meneliti penerapan ritel modern di era digital melalui aplikasi Alfagift dengan fokus pada peningkatan kepuasan pelanggan. Hasil penelitian mereka menunjukkan bahwa fitur digital seperti *coin rewards*, *voucher*, dan layanan pengantaran mampu meningkatkan loyalitas pengguna. Kesamaan dengan penelitian ini terletak pada objek yang sama, yaitu Alfagift, serta tujuan untuk memahami perilaku konsumen di *platform* digital.

Namun, perbedaan utamanya terletak pada metode dan pendekatan penelitian: Nurnadya menggunakan pendekatan kualitatif deskriptif berbasis wawancara, sedangkan penelitian ini mengombinasikan model RFM dan algoritma K-Means *Clustering* berbasis data transaksi historis berskala masif, sehingga menghasilkan temuan segmentasi yang lebih kuantitatif dan objektif.

Arthanugraha *et al.* (2024) melakukan segmentasi pelanggan Pabuaran Store di Shopee menggunakan model RFM untuk mengidentifikasi kelompok pelanggan aktif, pasif, dan potensial. Temuan mereka menunjukkan adanya kelompok pelanggan berfrekuensi tinggi dan bernilai transaksi besar sebagai pelanggan inti. Hasil ini memiliki kemiripan dengan karakteristik Klaster 1 (*High-Value Buyers*) pada penelitian ini. Meski demikian, perbedaannya terletak pada skala data dan ruang lingkup: penelitian Arthanugraha *et al.* berfokus pada satu toko individu dalam *marketplace*, sedangkan penelitian ini mengevaluasi ekosistem ritel digital skala besar dengan cakupan lebih dari 1,5 juta transaksi serta mempertimbangkan faktor musiman Idul Fitri.

Fadillah *et al.* (2023) menganalisis perilaku konsumen pada Ralali.com menggunakan model RFM untuk memetakan pelanggan potensial. Kesamaan dengan penelitian ini terdapat pada penggunaan variabel RFM sebagai indikator utama perilaku transaksi. Namun, penelitian ini memperkuat analisis tersebut dengan mengintegrasikan algoritma K-Means *Clustering* dan menguji kualitas pemisahan klaster menggunakan kombinasi metrik *Silhouette Score* serta *Calinski-Harabasz Index*. Selain itu, *dataset* yang digunakan dalam penelitian ini memiliki skala yang jauh lebih masif.

Nugraha *et al.* (2025) melakukan penelitian tentang segmentasi pelanggan berbasis RFM menggunakan algoritma K-Means *Clustering* pada perusahaan *payment gateway*, menghasilkan lima klaster pelanggan dengan validasi *Davies-Bouldin Index* (DBI) sebesar 0,1089. Persamaan dengan penelitian ini terletak pada penggunaan kerangka kerja RFM K-Means serta penentuan jumlah klaster awal menggunakan *Elbow Method*. Perbedaan mendasar terletak pada domain industri dan pengujian validasi: penelitian Nugraha *et al.* berfokus pada transaksi keuangan digital di sektor fintech, sedangkan penelitian ini berfokus pada transaksi belanja ritel fisik-digital (*omnichannel*). Lebih lanjut, penelitian ini menggunakan kombinasi *Silhouette Score* dan *Calinski-Harabasz Index* untuk mengukur kerapatan serta pemisahan klaster dalam konteks dinamika transaksi musiman.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan kombinasi model RFM dan algoritma K-Means *Clustering* untuk segmentasi pelanggan Alfagift di wilayah Jambi pada *dataset* 1.529.558 transaksi

periode tiga bulan sebelum dan sesudah Hari Raya Idul Fitri 2025. Tiga segmen pelanggan yang terbentuk mencerminkan heterogenitas perilaku transaksi yang bermakna secara bisnis: *frequent low-spending customers* (20,6%), *high-value but infrequent buyers* (22,4%), dan *promotion-driven customers* (57,0%). Temuan bahwa mayoritas pelanggan bersifat *promotion-driven* mengisyaratkan pentingnya optimasi strategi promosi musiman, khususnya menjelang Idul Fitri. Meskipun *Silhouette Score* (0,2362) mengindikasikan batas klaster yang kurang tegas ($<0,25$), model tetap dinilai handal secara operasional bisnis karena didukung nilai *Calinski-Harabasz Index* yang sangat tinggi (1257,17) mengonfirmasi pemisahan antar-sentroid yang signifikan, serta sifat alami data transaksi ritel yang bersifat kontinu, di mana validasi *Radar Chart* dan Histogram membuktikan validitas yang terbentuk kontras dan dapat ditindaklanjuti.

Untuk penelitian lanjutan, beberapa arah pengembangan direkomendasikan: (1) penambahan variabel non-transaksional (engagement aplikasi) untuk memperdalam profil klaster; (2) eksplorasi algoritma DBSCAN atau Gaussian Mixture Models untuk menguji konsistensi segmentasi pada batas beririsan; (3) validasi longitudinal lintas periode musiman untuk menguji stabilitas klaster; serta (4) integrasi hasil segmentasi ke dalam sistem rekomendasi berbasis *machine learning* untuk personalisasi promosi secara *real-time* pada platform Alfagift.

Daftar Pustaka

- Arthanugraha, A., Azzuhri, M. B., Ramadhan, Y. W., & Heikal, J. (2024): *Customer segmentation of Pabuaran Store on Shopee e-commerce using RFM model analysis (Case study of H&M brand sales products)*, BUDGETING: Journal of Business, Management and Accounting, 5(2), 829–834. <https://doi.org/10.31539/budgeting.v5i2.9724>
- Fadillah, A., Mogi, I. K. A., & Darmawan, I. D. M. B. A. (2023): *Analisis data Recency, Frequency, dan Monetary (RFM) guna melakukan pendekatan kepada konsumen Ralali.com*, Jurnal Pengabdian Informatika, 2, 83–88.
- Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D. S., & Le, T. G. (2023): *An extended RFM model for customer behaviour and demographic analysis in retail industry*, Business Systems Research, 14(1), 26–53. <https://doi.org/10.2478/bsrj-2023-0002>
- Kossakov, M., Mukasheva, A., Balbayev, G., Seidazimov, S., Mukammejanova, D., & Sydybayeva, M. (2024): *Quantitative comparison of machine learning clustering methods for tuberculosis data analysis*, Engineering Proceedings, 60(1). <https://doi.org/10.3390/engproc2024060020>
- Miranda-Guerra, M., Cevallos, S., & Torres, K. D. (2024): *Relationship marketing and customer loyalty: A quantitative analysis in the retail sector*, IBIMA Business Review. <https://doi.org/10.5171/2024.710234>
- Myagmarsuren, S. (2024): *Exploring the use of silhouette score in K-means clustering for image segmentation*, International Journal of Engineering Research & Technology (IJERT), 13(4). <http://www.ijert.org>
- 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 (JUK-TISI), 4(2), 1361–1369. <https://doi.org/10.62712/juktisi.v4i2.641>
- Nurnadya, R. A., Bany, M. N., Laura, M., Cantika, G., Abidin, M. N., Samudra, T., Hidayat, R., Veronica, L., & Rihidima, C. (2024): *Analisis penerapan ritel modern di era digital dalam meningkatkan kepuasan konsumen pada aplikasi Alfagift*, Economic and Business Management International Journal, 6(3), 58–64.
- Parmar, N. N., & Panchal, S. S. (2024): *Enhancing customer segmentation in virtual shopping through RFM analysis*, International Research Journal of Engineering and Technology (IRJET), 2164–2168. <http://www.irjet.net>
- Refialy, L. P., Maitimu, H., & Pesulima, M. S. (2021): *Perbaikan kinerja clustering K-Means pada data ekonomi nelayan dengan perhitungan Sum of Square Error (SSE) dan optimasi nilai K cluster*, Techno.Com, 20(2), 321–329. <https://doi.org/10.33633/tc.v20i2.4572>
- Regasa, E. (2023): *Factors affecting loan repayment performance of smallholder farmers: A case of Sinqe Bank, Jima Arjo Woreda of East Wallaga Zone, Oromia Region*, Research Journal of Finance and Accounting, 14(9), 21–29. <https://doi.org/10.7176/rjfa/14-9-03>
- Sagala, N. T. M., & Gunawan, A. A. S. (2022): *Discovering the optimal number of crime cluster using Elbow, Silhouette, Gap Statistics, and NbClust methods*, ComTech: Computer, Mathematics and Engineering Applications, 13(1), 1–10. <https://doi.org/10.21512/comtech.v13i1.7270>
- Syakra, Y., Fadlil, A., & Yuliansyah, H. (2025): *Customer segmentation using RFM and K-Means clustering to support CRM in retail industry*, Sinkron, 9(3), 1120–1131. <https://doi.org/10.33395/sinkron.v9i3.14907>
- Vashisth, S., & Goyal, A. (2024): *Dynamic anomaly detection using robust random cut forests in resource-constrained IoT environments*, Informatica (Slovenia), 48(23), 107–120. <https://doi.org/10.31449/inf.v48i23.6862>
- Yusak, A., Rumapea, N., Pratiwi, D., & Sari, S. (2024): *Analisis segmentasi pelanggan ri-*

*tel online menggunakan K-Means clustering
berdasarkan model Recency, Frequency, Mone-*

*tary (RFM), Jurnal Sains dan Teknologi, 6(3),
292–299.*

Halaman ini sengaja dikosongkan