

# Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering

I Komang Yosua Triantara<sup>a1</sup>, Made Agung Raharja<sup>a2</sup>, Ida Bagus Gede Sarasvananda<sup>a3</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>triantara.2308561140@student.unud.ac.id  
<sup>2</sup>made.agung@unud.ac.id  
<sup>3</sup>sarasvananda@unud.ac.id

## Abstract

*The rapid growth of online retail has generated vast transactional data, creating significant opportunities for advanced customer segmentation. While the standard RFM (Recency, Frequency, Monetary) model is widely used in Customer Relationship Management (CRM), it possesses a key limitation by not capturing the temporal dynamics between customer purchases. This research addresses that gap by proposing an RFM-T model, which enhances the traditional framework with Interpurchase Time (IPT) to provide a more holistic view of customer behavior. Using a dual-clustering methodology on an online retail dataset, the K-Means algorithm is first applied for broad segmentation, followed by Hierarchical Clustering to explore deeper sub-segments within high-value groups. The process yielded four primary clusters, and the model's robustness was systematically validated through a strong Silhouette Score, a low Davies-Bouldin Index, and a high Calinski-Harabasz Index. This detailed analysis successfully identified distinct customer personas, such as 'Consistent Loyalists' (low IPT) and 'Periodic Premium Buyers' (high monetary value), which are crucial for developing targeted strategies. The findings demonstrate that this integrated RFM-T framework provides a quantitatively validated with Silhouette Score 0.410, Davies-Bouldin Index 0.720, and Calinski-Harabasz Index 1365.14 this score show actionable model for personalized marketing and effective customer retention.*

**Keywords:** Customer Segmentation, K-Means, Hierarchical Clustering, RFM, E-commerce

## 1. Pendahuluan

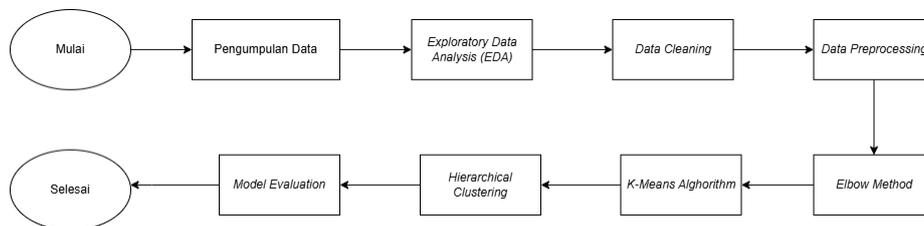
Pertumbuhan industri ritel daring (*e-commerce*) telah membawa perubahan signifikan terhadap cara perusahaan membangun hubungan dengan pelanggan. Salah satu dampak utamanya adalah meningkatnya volume data transaksi digital yang dapat dimanfaatkan untuk memahami perilaku pelanggan. Menurut laporan *We Are Social*, pada Januari 2024 ada sekitar 56,1% pengguna internet di dunia yang terbiasa belanja secara online setiap minggunya [1]. Analisis yang tepat terhadap data tersebut dapat memberikan pengetahuan mengenai pola perilaku pelanggan bagi perusahaan. Pemahaman mengenai pola perilaku pelanggan penting bagi perusahaan untuk mempertahankan pelanggan yang loyal daripada hanya menarik pelanggan baru [2]. *Customer Relationship Management (CRM)* merupakan pendekatan penting dalam mengelola interaksi perusahaan dengan pelanggan. Salah satu teknik utama dalam CRM adalah segmentasi pelanggan, yaitu proses mengelompokkan pelanggan berdasarkan pola perilaku mereka [3]. Segmentasi ini memungkinkan perusahaan mengidentifikasi kelompok pelanggan bernilai tinggi, pelanggan baru, maupun pelanggan yang berisiko untuk meninggalkan layanan. Salah satu model segmentasi pelanggan yang biasa digunakan yaitu RFM yang dapat mengidentifikasi pelanggan berdasarkan metrik *recency*, *frequency*, dan *monetary* [4]. *Recency* (R) mengukur waktu sejak pelanggan terakhir melakukan pembelian, nilai yang lebih kecil menunjukkan bahwa pelanggan baru-baru ini berbelanja. *Frequency* (F) menghitung jumlah total pembelian yang dilakukan oleh pelanggan dalam periode waktu tertentu, semakin tinggi nilainya, semakin sering pelanggan melakukan pembelian. *Monetary* (M) menentukan total pengeluaran

pelanggan selama periode yang dianalisis [5]. Namun demikian, metode RFM memiliki keterbatasan karena hanya menggambarkan perilaku pelanggan dan tidak mempertimbangkan dinamika waktu antar pembelian. Penelitian sebelumnya telah mengusulkan pengayaan model menjadi RFTMS dengan menambahkan elemen waktu dan kepuasan pelanggan, serta menerapkan algoritma seperti *K-Medoid* [6]. Namun, pendekatan tersebut belum secara spesifik mengevaluasi pengaruh fitur interval waktu antar transaksi terhadap kualitas segmentasi. Selain itu, kombinasi evaluasi clustering dengan metrik kuantitatif seperti *Silhouette Score* atau *Davies-Bouldin Index* masih jarang digunakan secara sistematis dalam penelitian serupa. Penelitian ini mengusulkan pendekatan segmentasi pelanggan berbasis RFM-T, yaitu model RFM yang menambahkan fitur interval waktu antar transaksi pelanggan. Dataset bersumber dari platform *kaggle* bernama *Online Retail* dan akan dianalisis menggunakan algoritma *K-Means* untuk segmentasi awal, lalu dianalisis lebih lanjut menggunakan *Hierarchical Clustering* untuk mengeksplorasi struktur kluster yang lebih dalam [7]. Kualitas kluster dievaluasi menggunakan metrik evaluasi *clustering* agar segmentasi yang dihasilkan terukur secara kuantitatif.

## 2. Metode Penelitian

### 2.1. Alur Penelitian

Pada penelitian ini, akan dimulai dengan pengumpulan data, kemudian mengeksplorasi data di dalam tahap *Exploratory Data Analysis (EDA)* untuk mengetahui karakteristik data. Setelah itu, akan dilakukan *data cleaning* untuk membersihkan data sebelum dilakukan *data preprocessing*. Pemodelan akan menggunakan *K-Means Clustering* sebagai tahap awal dengan *elbow method* untuk optimasi jumlah kluster. Setelah terbentuk kluster dari algoritma *K-Means*, kembali dilakukan *clustering* dengan algoritma *Hierarchical Clustering* untuk lebih memahami segmentasi pelanggan di tiap klusternya. Kemudian setiap kluster akan dievaluasi dengan *multi-metrik* dan menginterpretasikan hasil segmentasi setiap kluster.



Gambar 1. Alur Penelitian

### 2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder *Online Retail* yang diambil dari platform *Kaggle*. Dataset ini berisi 541910 data yang merupakan informasi mengenai tanggal invoice, harga barang, jumlah barang, dan deskripsi dari barang yang dibeli pelanggan. Untuk memudahkan tahap analisis data ini kemudian dibuatkan data frame melalui library *pandas* yang tersedia di *python*

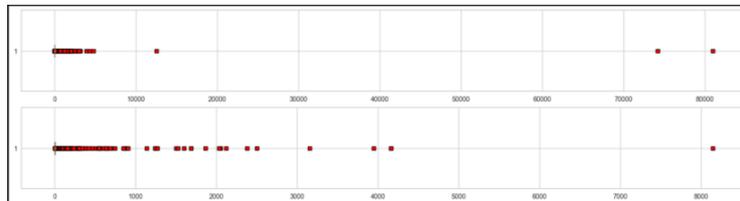
|   | InvoiceNo | StockCode | Description                         | Quantity | InvoiceDate    | UnitPrice | CustomerID   | Country        |
|---|-----------|-----------|-------------------------------------|----------|----------------|-----------|--------------|----------------|
| 0 | 536365    | 85123A    | WHITE HANGING HEART T-LIGHT HOLDER  | 6        | 12/1/2010 8:26 | 2.550000  | 17850.000000 | United Kingdom |
| 1 | 536365    | 71053     | WHITE METAL LANTERN                 | 6        | 12/1/2010 8:26 | 3.390000  | 17850.000000 | United Kingdom |
| 2 | 536365    | 84406B    | CREAM CUPID HEARTS COAT HANGER      | 8        | 12/1/2010 8:26 | 2.750000  | 17850.000000 | United Kingdom |
| 3 | 536365    | 84029G    | KNITTED UNION FLAG HOT WATER BOTTLE | 6        | 12/1/2010 8:26 | 3.390000  | 17850.000000 | United Kingdom |
| 4 | 536365    | 84029E    | RED WOOLLY HOTTIE WHITE HEART.      | 6        | 12/1/2010 8:26 | 3.390000  | 17850.000000 | United Kingdom |

Gambar 2. Data Head Online Retail

### 2.3. Exploratory Data Analysis

Tahap awal dari penelitian ini dimulai dengan eksplorasi data untuk memahami karakteristik dataset. EDA dilakukan untuk identifikasi struktur data, serta mendeteksi potensi anomali dan

outlier. Pada tahap ini ditemukan bahwa data memiliki 541909 baris dan 8 kolom. Pada kolom *CustomerID* dan *Description* ditemukan missing values yang akan ditangani dalam proses *preprocessing*. Serta, beberapa duplikasi juga ditemukan dari kolom *InvoiceNo* yang kemungkinan merupakan data dari produk yang tidak jadi dibeli. Pengecekan *outlier* dilakukan dengan menggunakan *boxplot horizontal* dari kolom *Quantity* dan *Price*. Outlier ditandai dengan kotak merah dan ditemukan bahwa distribusi data tidak seimbang serta banyak *outlier* di kedua kolom seperti dapat dilihat pada *Gambar 3*.



Gambar 3. Outliers

## 2.4. Data Cleaning

Berdasarkan hasil eksplorasi awal, dilakukan tahap pra-pemrosesan data untuk meningkatkan kualitas dataset sebelum analisis lebih lanjut. Pada tabel pertama terlihat adanya nilai kosong yang cukup besar pada kolom *CustomerID* (135.080 nilai kosong) dan *Description* (1.454 nilai kosong). Untuk menjaga konsistensi data, seluruh baris dengan *CustomerID* kosong dihapus, sedangkan data duplikasi dihilangkan agar tidak menimbulkan bias dalam analisis. Selain itu, perubahan tipe data juga dilakukan sebagaimana terlihat pada tabel kedua, di mana *InvoiceDate* yang semula bertipe *object* dikonversi menjadi *datetime64[ns]* agar dapat diproses secara kronologis, serta *Quantity* yang awalnya *int64* diubah menjadi *float64* sebagai penyesuaian setelah proses normalisasi dan penanganan outlier. Penanganan outlier sendiri diterapkan dengan pendekatan *winsorization-style*, yaitu mengganti nilai ekstrem berdasarkan batas atas dan bawah yang dihitung dari persentil ke-1 (0.01) dan ke-99 (0.99), serta diperluas menggunakan  $1,5 \times IQR$ . Strategi ini dipilih karena beberapa kolom, khususnya *Quantity* dan *UnitPrice*, menunjukkan variasi yang sangat tinggi sehingga berpotensi mendominasi hasil segmentasi. Dengan pendekatan ini, distribusi data menjadi lebih seimbang tanpa kehilangan informasi penting yang dapat merepresentasikan perilaku pelanggan secara lebih akurat.

| index | Column      | Dtype   | Missing | Unique | Freq   |
|-------|-------------|---------|---------|--------|--------|
| 0     | InvoiceNo   | object  | 0       | 25000  | 1114   |
| 1     | StockCode   | object  | 0       | 4070   | 2313   |
| 2     | Description | object  | 1454    | 4223   | 2369   |
| 3     | Quantity    | int64   | 0       | 722    | 148227 |
| 4     | InvoiceDate | object  | 0       | 23260  | 1114   |
| 5     | UnitPrice   | float64 | 0       | 1630   | 50496  |
| 6     | CustomerID  | float64 | 135080  | 4372   | 7983   |
| 7     | Country     | object  | 0       | 38     | 495478 |

| index | Column      | Dtype          | Missing | Unique | Freq   |
|-------|-------------|----------------|---------|--------|--------|
| 0     | InvoiceNo   | object         | 0       | 18538  | 542    |
| 1     | StockCode   | object         | 0       | 3665   | 2023   |
| 2     | Description | object         | 0       | 3877   | 2016   |
| 3     | Quantity    | float64        | 0       | 165    | 69805  |
| 4     | InvoiceDate | datetime64[ns] | 0       | 17296  | 542    |
| 5     | UnitPrice   | float64        | 0       | 353    | 45102  |
| 6     | CustomerID  | float64        | 0       | 4339   | 7676   |
| 7     | Country     | object         | 0       | 37     | 349227 |

Gambar 4. Data Sebelum Cleaning, Data Setelah Cleaning

## 2.5. Data Preprocessing

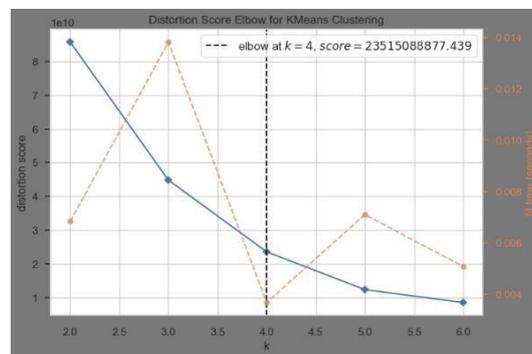
Pada *data preprocessing* dilakukan *feature engineering* untuk membuat metrik RFM-T sebagai fitur yang akan digunakan untuk analisis segmentasi pelanggan. Pertama, fitur *Revenue* dibuat dengan menghitung hasil perkalian *Quantity* dan *UnitPrice* dalam setiap transaksi pelanggan. Fitur RFMT akan dihitung dengan mengelompokkan data berdasarkan *CustomerID*. Atribut *recency* dihitung berdasarkan jumlah hari sejak transaksi terakhir pelanggan hingga tanggal acuan. Atribut *frequency* diperoleh dari jumlah unik transaksi (*InvoiceNo*) yang dilakukan oleh setiap pelanggan. Sedangkan atribut *monetary* dihitung dari total nilai pembelian (*Revenue*) yang merupakan hasil perkalian antara jumlah produk dan harga satuan untuk setiap transaksi, yang kemudian dijumlahkan per pelanggan.

| CustomerID   | Recency | Frequency | Monetary    | Interpurchase Time |
|--------------|---------|-----------|-------------|--------------------|
| 12347.000000 | 2       | 7         | 4310.000000 | 52                 |
| 12348.000000 | 75      | 4         | 1770.780000 | 70                 |
| 12352.000000 | 36      | 8         | 1756.340000 | 32                 |
| 12356.000000 | 22      | 3         | 2811.430000 | 100                |
| 12358.000000 | 1       | 2         | 1150.420000 | 74                 |

Gambar 5. Hasil Data Preprocessing

## 2.6. Elbow Method

*Elbow Method* adalah sebuah teknik dalam analisis kluster untuk menentukan jumlah kluster optimal (nilai k) dalam algoritma *K-Means*. Metode ini bekerja dengan menghitung nilai distorsi atau jumlah kuadrat jarak antar titik data dengan pusat klasternya[8]. Pemilihan jumlah kluster dilakukan menggunakan metode Elbow, yang menunjukkan titik optimal pada k=4. Pada titik ini, distortion score sekitar  $2,35 \times 10^{10}$ , yang menandakan bahwa data dalam masing-masing kluster cukup rapat terhadap centroid-nya. Penurunan skor setelah k=4 tidak lagi signifikan, sehingga 4 kluster dapat dianggap jumlah yang paling representatif untuk segmentasi data.

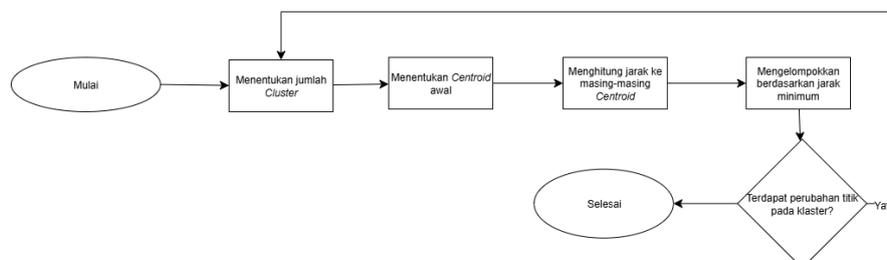


Gambar 6. Elbow Method

## 2.7. K-Means Algorithm

Proses segmentasi dilakukan dengan menggunakan algoritma *K-Means* dengan jumlah kluster sebanyak lima sesuai dengan hasil dari *Elbow Method*. Algoritma *K-Means* bekerja dengan cara memilih secara acak sejumlah titik data sebagai pusat kluster awal (*centroid*) [9]. Kemudian setiap data dihitung jaraknya ke seluruh centroid menggunakan rumus 1. *Centroid* diperbarui sebagai rata-rata dari anggota klasternya dan proses ini diulang hingga posisi *centroid stabil*. Pada penelitian ini, jumlah kluster yang digunakan yaitu empat dan *K-Means* digunakan untuk mengelompokkan pelanggan berdasarkan fitur RFMT untuk mengidentifikasi pola perilaku yang serupa.

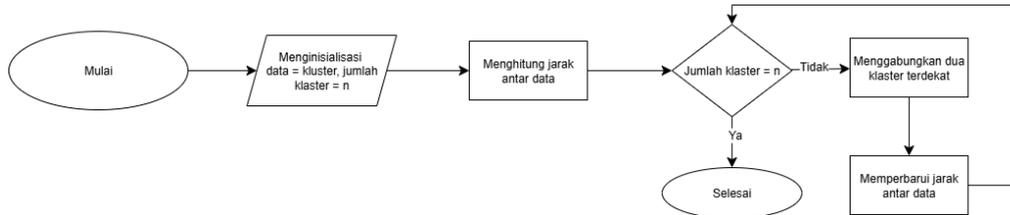
$$d_{ij} = \sqrt{[(x_1 - x_2)^2 + (y_1 - y_2)^2]} \quad (1)$$



Gambar 7. Flowchart K-Means Algorithm

## 2.8. Hierarchical Algorithm

*Hierarchical Alghorithm* merupakan metode pengelompokan data yang membentuk struktur hirarki dalam bentuk pohon atau dendrogram. Tujuan dari algoritma ini yaitu untuk mengelompokkan data berdasarkan kemiripan antar titik data dari yang paling umum hingga ke kelompok kecil yang sangat mirip [10]. *Hierarchical Clustering* digunakan setelah *K-Means* untuk menganalisis lebih dalam struktur tiap kluster, kluster yang dianalisis lebih dalam merupakan kluster yang memiliki karakteristik pelanggan paling bernilai.



Gambar 8. Flowchart *Hierarchical Algorithm*

## 2.9. Model Evaluation

Pengujian performa algoritma dalam melakukan *clustering* pada dataset ini dinilai berdasarkan perbandingan nilai *cluster validation metrics*. *Silhouette Score* mengukur seberapa mirip suatu data terhadap klasternya sendiri dibandingkan dengan kluster lain. *Davies-Bouldin Indeks (DBI)* untuk mengukur rata-rata rasio antara jarak dalam kluster dan jarak antar kluster. Sedangkan *Calinski-Harabasz Indeks (CHI)* membandingkan variasi antar kluster dengan variasi dalam kluster [11].

$$Silhoutte = \frac{b-a}{\max(a,b)} \quad (2)$$

$$DBI = rata - rata \left( \frac{S_i + S_j}{M_{ij}} \right) \quad (3)$$

$$CHI = \frac{Between-Cluster Dispersion}{Within-Cluster Dispersion} \times \frac{n-k}{k-1} \quad (4)$$

## 3. Hasil dan Pembahasan

### 3.1. Hasil Analisis Segmentasi *K-Means*

Berdasarkan hasil klasterisasi menggunakan algoritma *K-Means*, didapatkan empat segmen pelanggan dengan karakteristik yang berbeda-beda berdasarkan *nilai Recency, Frequency, Monetary, dan Interpurchase Time* seperti pada Gambar 9.

| Tabel Rata-rata RFMT per Kluster |         |           |          |                    |
|----------------------------------|---------|-----------|----------|--------------------|
|                                  | Recency | Frequency | Monetary | Interpurchase_Time |
| Clusters                         |         |           |          |                    |
| 0                                | 197.29  | 2.95      | 923.90   | 30.85              |
| 1                                | 46.54   | 2.90      | 1048.33  | 90.94              |
| 2                                | 27.29   | 7.24      | 2876.04  | 30.93              |
| 3                                | 5.82    | 74.41     | 78848.06 | 5.95               |

Gambar 9. Hasil *Clustering K-Means*

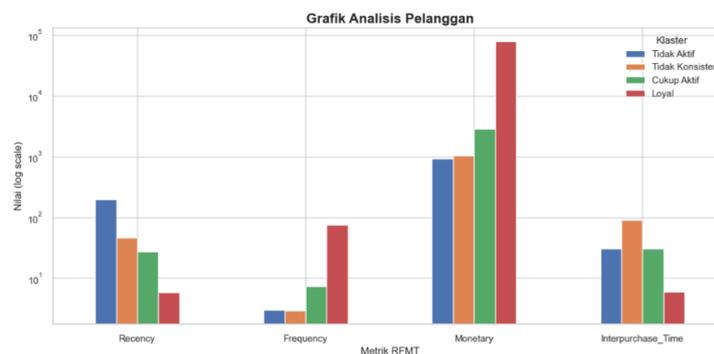
Kluster 0 terdiri dari pelanggan yang memiliki nilai *recency* yang sangat tinggi (197 hari), artinya mereka sudah cukup lama tidak melakukan transaksi. Nilai *frequency* dan *monetary* mereka juga relatif rendah, menandakan bahwa mereka adalah pelanggan pasif dengan aktivitas belanja yang minim. Segmentasi ini dapat dikategorikan sebagai pelanggan tidak aktif yang berpotensi *churn*.

Oleh karena itu, strategi yang dapat diterapkan pada segmen ini adalah mengirimkan kampanye reaktivasi seperti penawaran diskon khusus atau promosi terbatas untuk menghidupkan kembali minat belanja mereka.

Klaster 1 mencerminkan pelanggan dengan *recency* yang sedang (46 hari), namun memiliki frekuensi pembelian yang rendah dan jarak waktu antar pembelian (*interpurchase time*) yang cukup tinggi, yaitu sekitar 90 hari. Ini menunjukkan bahwa pelanggan dalam klaster ini bersifat tidak konsisten dan jarang melakukan pembelian ulang. Strategi yang dapat diterapkan adalah mendorong pelanggan untuk meningkatkan frekuensi transaksi dengan memberikan penawaran berkala atau program loyalitas agar mereka lebih terikat pada layanan atau produk yang ditawarkan.

Klaster 2 terdiri dari pelanggan yang cukup aktif, terlihat dari *recency* yang rendah (27 hari) dan frekuensi pembelian yang sedang hingga tinggi (7.24). Mereka juga memiliki nilai *monetary* yang cukup besar, yakni sekitar 2876.04, menandakan bahwa pelanggan ini tergolong bernilai bagi perusahaan. Untuk mempertahankan keterlibatan mereka, strategi seperti personalisasi penawaran, pemberian rekomendasi produk berdasarkan histori pembelian, serta peningkatan pengalaman pelanggan dapat diterapkan.

klaster 3 menunjukkan karakteristik pelanggan paling bernilai. Mereka memiliki *recency* yang sangat rendah (5.82), frekuensi pembelian yang sangat tinggi (74.41), dan nilai *monetary* besar mencapai 78,848.06. Pelanggan dalam klaster ini merupakan pelanggan super loyal dan memberikan kontribusi signifikan terhadap pendapatan. Strategi yang paling tepat untuk segmen ini adalah memberikan perlakuan khusus seperti program VIP, penawaran *eksklusif*, atau *early access* terhadap produk baru.



**Gambar 10.** Grafik Analisis Pelanggan

Secara keseluruhan, *Gambar 10* memperlihatkan perbedaan yang jelas antar klaster berdasarkan dimensi RFMT. Klaster tidak aktif ditandai dengan nilai *recency* tinggi serta *frequency* dan *monetary* paling rendah. Klaster tidak konsisten memiliki pola pembelian tidak teratur dengan *frequency* rendah dan *interpurchase time* relatif tinggi. Klaster cukup aktif menunjukkan perilaku belanja lebih stabil dengan *recency* rendah, *frequency* sedang dan *monetary* cukup tinggi. Sementara klaster loyalk merupakan segmen paling bernilai dengan *recency* rendah, *frequency* tinggi, *monetary* yang paling besar dan *interpurchase time* yang paling singkat

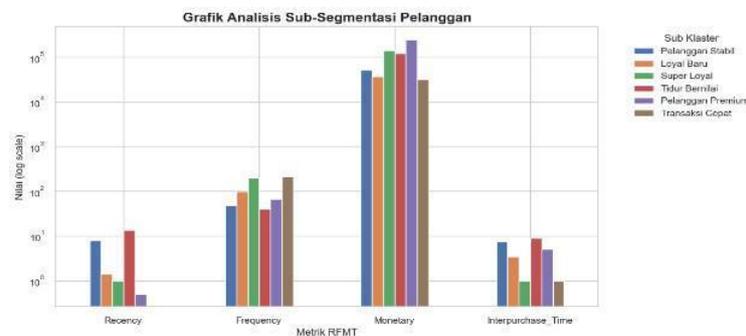
### 3.2. Hasil Analisis Segmentasi *Hierarchical Clustering*

Analisis segmentasi *Hierarchical Clustering* dilakukan pada klaster 3 karena menunjukkan karakteristik pelanggan paling bernilai berdasarkan *K-Means Clustering*. Dari proses *sub-segmentasi* ini diperoleh enam sub-klaster yang masing-masing menunjukkan pola perilaku yang berbeda berdasarkan metrik RFMT (*Recency*, *Frequency*, *Monetary*, dan *Interpurchase Time*) seperti pada *Gambar 11*.

| Sub_Cluster | Recency | Frequency | Monetary  | Interpurchase_Time |
|-------------|---------|-----------|-----------|--------------------|
| 0           | 7.80    | 47.90     | 51612.55  | 7.50               |
| 1           | 1.40    | 98.20     | 36567.46  | 3.40               |
| 2           | 1.00    | 201.00    | 138919.84 | 1.00               |
| 3           | 13.67   | 40.67     | 122437.64 | 9.00               |
| 4           | 0.50    | 67.00     | 248873.94 | 5.00               |
| 5           | 0.00    | 210.00    | 31713.87  | 1.00               |

**Gambar 11.** Hasil Sub-Segmentasi *Hierarchical Clustering* pada Kluster 3

Beberapa sub-kluster, seperti sub-kluster 2 dan 4, memperlihatkan perilaku pelanggan yang sangat aktif dengan nilai transaksi yang tinggi, sehingga dapat dikategorikan sebagai pelanggan prioritas atau VIP yang sangat loyal dan bernilai strategis. Sementara itu, sub-kluster 5 dan 1 menunjukkan frekuensi pembelian yang sangat tinggi namun dengan nilai transaksi yang relatif lebih rendah, menandakan adanya potensi peningkatan nilai pembelian melalui strategi seperti upselling atau bundling. Di sisi lain, sub-kluster 0 dan 3 menunjukkan nilai recency dan interpurchase time yang lebih tinggi, mengindikasikan bahwa pelanggan dalam kelompok ini cenderung kurang aktif dan memiliki risiko churn yang lebih besar.



**Gambar 12.** Hasil Analisis *Sub-Segmentasi Pelanggan* pada Kluster 3

Pada *Gambar 12* terdapat gambaran yang lebih detail tentang perbedaan perilaku pelanggan dalam enam *sub-kluster*. Kluster super loyal dan pelanggan premium merupakan segmen paling bernilai karena memiliki *frequency* tinggi dan *monetary* cukup besar, sedangkan kluster tidak bernilai memiliki kontribusi rendah karena memiliki *recency* tinggi, *frequency* paling rendah. Sementara itu, kluster transaksi cepat meskipun memiliki tidak sebesar premium, mereka memiliki pola pembelian yang sering dan singkat jaraknya. Analisis ini menunjukkan bahwa sub-segmentasi RFMT mampu menangkap variasi perilaku pelanggan secara lebih granular, sehingga memudahkan dalam penentuan strategi pemasaran yang tepat untuk setiap segmen.

### 3.3. Evaluasi Performa Model Clustering

Pengujian performa algoritma dalam melakukan *clustering* pada dataset ini dinilai berdasarkan perbandingan nilai *cluster validation metrics* dengan hasil pengujian yang ditunjukkan pada Tabel 1.

**Tabel 1.** Performa *Clustering*

| <b>Cluster Validation Metrics K-Means</b> |           |
|---|-----------|
| Nilai <i>Silhouette</i>                   | 0.410     |
| <i>Davies-Bouldin Indeks</i>              | 0.720     |
| <i>Calinski-Harabasz Indeks</i>           | 1365.1363 |

Hasil evaluasi *K-Means* menunjukkan bahwa kluster yang terbentuk cukup baik, dengan *Silhouette Score* 0.410 yang menandakan pemisahan antar kluster sudah baik, *Davies-Bouldin*

*Index* 0.720 yang menunjukkan kluster relatif terpisah, dan *Calinski-Harabasz Index* 1365.14 yang menandakan distribusi antar kluster cukup optimal. Nilai-nilai ini menunjukkan bahwa *K-Means* mampu mengelompokkan pelanggan dengan performa yang layak.

#### 4. Kesimpulan

Penelitian ini berhasil mengelompokkan pelanggan secara efektif menggunakan pendekatan RFM yang diperkaya metrik *Interpurchase Time (RFMT)*. Penambahan metrik *Interpurchase Time* dapat membedakan antara pelanggan loyal harian dari pembeli musiman bernilai tinggi. Segmentasi utama dengan *K-Means* menghasilkan empat kluster utama, yang keandalannya divalidasi oleh *Silhouette Score* sebesar 0,410, nilai *Davies-Bouldin Index* yang rendah, dan *Calinski-Harabasz Index* yang tinggi menandakan bahwa kluster terbentuk dengan baik. Sub-segmentasi menggunakan *hierarchical clustering* pada kluster dengan karakteristik pelanggan paling bernilai berhasil mengidentifikasi persona pelanggan yang lebih spesifik. Penelitian ini membuktikan bahwa kombinasi metode *RFMT* dan pendekatan *dual-clustering* efektif untuk memetakan intensitas dan konsistensi pembelian pelanggan. Untuk pengembangan ke depan, penelitian dapat menggunakan algoritma klusterisasi lain seperti *DBSCAN* atau *GMM*, serta menambahkan data demografi atau perilaku lain untuk meningkatkan akurasi segmentasi dan penerapan strategisnya

#### Daftar Pustaka

- [1] We Are Social, "Digital 2024: 5 billion social media users," <https://wearesocial.com/id/blog/2024/01/digital-2024-5-billion-social-media-users/>.
- [2] A. Yuda, "Komparasi Algoritma K-Means dan Hierarchical Clustering Untuk Mengetahui Data Customer Dalam Layanan Internet," *Informatika Sains Teknologi*, vol. 2, no. 2, pp. 9–14, Jul. 2024, doi: 10.34005/insit.v2i2.4119.
- [3] V. Guerola-Navarro, H. Gil-Gomez, R. Oltra-Badenes, and P. Soto-Acosta, "Customer relationship management and its impact on entrepreneurial marketing: a literature review," *International Entrepreneurship and Management Journal*, vol. 20, no. 2, pp. 507–547, Jun. 2024, doi: 10.1007/s11365-022-00800-x.
- [4] I. YUNITA, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, "Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran," *Jurnal Media Teknik dan Sistem Industri*, vol. 9, no. 1, p. 58, Mar. 2025, doi: 10.35194/jmtsi.v9i1.4452.
- [5] M. A. Raharja, I. Kadek, A. Surya, I. Komang, and A. Mogi, "Clustering Customer For Determine Market Strategy Using K-Means and Topsis: Case Study," vol. 2, 2022, [Online]. Available: <http://eprosiding.idbbali.ac.id/index.php/imade>
- [6] P. A. Windjaya, B. Siregar, and K. Kunci, "(RFMTS) Menggunakan Algoritma K-Medoids Clustering," *Multidisciplinary Scientific Journal*, vol. 2.
- [7] A. Romanazzi, D. Scocciolini, M. Savoia, and N. Buratti, "Iterative hierarchical clustering algorithm for automated operational modal analysis," *Autom Constr*, vol. 156, Dec. 2023, doi: 10.1016/j.autcon.2023.105137.
- [8] V. A. Permadi, S. P. Tahalea, and R. P. Agusdin, "K-Means and Elbow Method for Cluster Analysis of Elementary School Data," *Progres Pendidikan*, vol. 4, no. 1, pp. 50–57, Jan. 2023, doi: 10.29303/prospek.v4i1.328.
- [9] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, "Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 5, no. 4, pp. 354–364, Feb. 2024, doi: 10.35746/jtim.v5i4.434.
- [10] B. Moseley and J. R. Wang, "Approximation Bounds for Hierarchical Clustering: Average Linkage, Bisecting K-means, and Local Search," 2023. [Online]. Available: <http://jmlr.org/papers/v24/18-080.html>.
- [11] 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, Apr. 2025, doi: 10.25126/jtiik.2025128619.