

# Segmentasi Pelanggan Berbasis RFM dengan Algoritma K-Means Pada Data Transaksi Online Retail

Daffa Rayhan Fadhilla<sup>1</sup>, Ridho Ramadhan<sup>2</sup>, Vedly Vedliyan Darma Oktavian<sup>3</sup>

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya

Email: daffarayhan099@gmail.com<sup>1</sup>, ridhormdhn404@gmail.com<sup>2</sup>, vedly.vedliyan05@gmail.com<sup>3</sup>

## ABSTRAKSI

Penelitian ini berfokus pada segmentasi pelanggan menggunakan model RFM (Recency, Frequency, Monetary) dan algoritma K-Means pada data transaksi online retail. Segmentasi pelanggan adalah proses mengelompokkan pelanggan ke dalam kategori yang berbeda berdasarkan pola perilaku mereka dalam bertransaksi. Model RFM memungkinkan kita untuk mengevaluasi pelanggan berdasarkan tiga dimensi penting: seberapa baru pelanggan melakukan transaksi terakhir (Recency), seberapa sering pelanggan melakukan transaksi (Frequency), dan total nilai transaksi yang dihasilkan oleh pelanggan (Monetary). Dengan menggabungkan data RFM dan algoritma K-Means, kita dapat membagi pelanggan ke dalam segmen-segmen yang homogen. Analisis ini memberikan wawasan yang mendalam tentang karakteristik dan nilai dari masing-masing segmen pelanggan, memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang lebih terarah dan efektif. Hasil segmentasi ini diharapkan dapat membantu perusahaan dalam meningkatkan retensi pelanggan, memaksimalkan nilai seumur hidup pelanggan, dan meningkatkan efektivitas kampanye pemasaran.

**Kata Kunci:** *segmentasi pelanggan, rfm, k-means, online retail, data transaksi.*

## ABSTRACT

This research focuses on customer segmentation using the RFM (Recency, Frequency, Monetary) model and the K-Means algorithm on online retail transaction data. Customer segmentation is the process of categorizing customers into different groups based on their transactional behavior patterns. The RFM model allows us to evaluate customers based on three critical dimensions: how recently a customer made their last purchase (Recency), how often a customer makes purchases (Frequency), and the total monetary value generated by the customer (Monetary). By combining RFM data and the K-Means algorithm, we can divide customers into homogeneous segments. This analysis provides deep insights into the characteristics and value of each customer segment, enabling companies to develop more targeted and effective marketing strategies. The segmentation results are expected to assist companies in enhancing customer retention, maximizing customer lifetime value, and improving the effectiveness of marketing campaigns.

**Keywords:** *customer segmentation, rfm, k-means, online retail, transaction data*

## Penulis Korespondensi

Vedly Vedliyan Darma Oktavian

Tanggal Submit : 22/06/2024

Tanggal Diterima : 17/07/2024

Tanggal Terbit : 25/07/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## I. PENDAHULUAN

Di era digital saat ini, perusahaan retail online menghadapi persaingan yang semakin ketat. Untuk bertahan dan berkembang, perusahaan harus memahami perilaku pelanggan mereka dengan lebih baik. Segmentasi pelanggan adalah salah satu pendekatan yang efektif untuk mencapai pemahaman tersebut. Segmentasi pelanggan memungkinkan perusahaan untuk mengklasifikasikan pelanggan mereka ke dalam kelompok-kelompok yang berbeda berdasarkan karakteristik tertentu, sehingga memungkinkan strategi

pemasaran yang lebih terarah dan efisien. Model RFM (Recency, Frequency, Monetary) adalah salah satu metode segmentasi pelanggan yang populer. Model ini mengevaluasi pelanggan berdasarkan tiga dimensi utama: Recency (R) mengukur seberapa baru pelanggan melakukan transaksi terakhir, Frequency (F) mengukur seberapa sering pelanggan melakukan transaksi dalam periode waktu tertentu, dan Monetary (M) mengukur total nilai uang yang dihabiskan oleh pelanggan selama periode waktu tertentu. Dengan menggunakan model RFM, perusahaan dapat

mengidentifikasi pelanggan yang paling berharga dan merancang strategi pemasaran yang sesuai untuk setiap segmen pelanggan.

Tujuan utama dari penelitian ini adalah untuk melakukan segmentasi pelanggan berdasarkan model RFM menggunakan algoritma K-Means pada data transaksi online retail. Teknik Recency, Frequency, and Monetary (RFM) mengkaji karakteristik pelanggan berdasarkan data transaksi untuk menghasilkan tiga faktor utama, yaitu recency, frequency, dan monetary, dalam rangka mempelajari data historis peningkatan transaksi pelanggan. Recency merupakan jarak pelanggan terakhir melakukan transaksi dengan menghitung hari, bulan dan tahun, Dengan rentang waktu recency kecil pelanggan memiliki point lebih tinggi. Frequency merupakan seringnya pelanggan bertransaksi dalam skala waktu pengambilan data.

Ruang lingkup penelitian ini mencakup penggunaan data transaksi dari sebuah perusahaan retail online. Data ini mencakup informasi mengenai transaksi yang dilakukan oleh pelanggan, termasuk nomor faktur, kode barang, deskripsi barang, kuantitas, tanggal faktur, harga per unit, ID pelanggan, dan negara asal. Langkah-langkah utama dalam penelitian ini meliputi pengolahan data, perhitungan nilai RFM, penerapan algoritma K-Means, dan evaluasi hasil segmentasi. Dengan langkah-langkah ini, penelitian ini bertujuan untuk memberikan panduan praktis bagi perusahaan dalam mengimplementasikan segmentasi pelanggan berbasis RFM untuk meningkatkan strategi pemasaran mereka.

## II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis data transaksi pelanggan. Fokus utama dari penelitian ini adalah untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka menggunakan model RFM (Recency, Frequency, Monetary) dan metode clustering K-Means. Metodologi penelitian ini terdiri dari beberapa tahapan utama, yaitu pengumpulan data, pemahaman data, persiapan data, preprocessing data, pembangunan model RFM, clustering, dan analisis eksploratif.

1. Pengumpulan data : Data transaksi pelanggan dikumpulkan dari sistem manajemen penjualan retail online. Dataset mencakup informasi seperti CustomerID, InvoiceDate, UnitPrice, StockCode, InvoiceNo, Quantity, dan Country.
2. Data Understanding : Memahami struktur data dan memeriksa kolom yang tersedia serta tipe datanya, sekaligus menghitung statistik deskriptif dasar seperti mean, median, dan standar deviasi untuk variabel utama.
3. Data Preparation : Data disiapkan untuk analisis melalui langkah-langkah berikut: menghapus data tidak valid, seperti transaksi dengan nilai negatif atau ID pelanggan yang tidak valid, dan menangani missing values dengan cara yang sesuai, baik dengan menghapusnya atau menggantinya dengan nilai rata-rata.

4. Data preprocessing: dilakukan untuk memastikan data siap untuk analisis RFM dengan menghitung jumlah hari sejak transaksi terakhir setiap pelanggan (Recency), jumlah transaksi yang dilakukan oleh setiap pelanggan (Frequency), dan total nilai transaksi yang dilakukan oleh setiap pelanggan (Monetary).
5. RFM Model building : Membangun model RFM dilakukan dengan menggabungkan nilai Recency, Frequency, dan Monetary ke dalam DataFrame untuk setiap pelanggan, menentukan kuartil untuk setiap metrik, dan memberikan skor RFM berdasarkan kuartil tersebut.
6. Clustering : Mengelompokkan pelanggan berdasarkan skor RFM menggunakan K-Means dilakukan dengan menentukan jumlah cluster optimal menggunakan metode Elbow dan menerapkan clustering dengan K-Means berdasarkan jumlah cluster yang dipilih.
7. Exploratory Data Analysis (EDA) : Melakukan EDA untuk mendapatkan wawasan lebih lanjut tentang hasil clustering meliputi menggambarkan tren penjualan berdasarkan bulan untuk memahami pola penjualan, mengidentifikasi produk yang paling sering dibeli oleh pelanggan, dan mengukur kontribusi masing-masing cluster terhadap pendapatan perusahaan untuk memahami segmen yang paling berharga.

## III. HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data

Data transaksi pelanggan diperoleh dari sistem manajemen penjualan ritel online, di mana setiap record mencakup nomor faktur (InvoiceNo) sebagai identitas unik untuk setiap penjualan. Kolom StockCode mencatat kode barang yang terjual, sementara kolom Description memberikan uraian singkat mengenai produk tersebut. Informasi kuantitas (Quantity) dan harga satuan (UnitPrice) dicatat untuk setiap item, memungkinkan analisis total nilai penjualan per transaksi.

Setiap entri transaksi disertai tanggal dan waktu pembelian dalam kolom InvoiceDate, sehingga peneliti dapat melakukan analisis tren penjualan berdasarkan rentang waktu harian, bulanan, atau musiman. Identitas pelanggan tercatat melalui CustomerID, memungkinkan segmentasi pelanggan berdasarkan perilaku pembelian, frekuensi transaksi, ataupun nilai rata-rata pembelian. Data ini juga mendukung analisis retensi dan loyalitas, dengan memetakan pola repeat order.

Kolom Country merekam lokasi geografis asal pelanggan, yang berguna untuk memahami distribusi pasar dan mengidentifikasi peluang ekspansi ke wilayah baru. Dengan mengintegrasikan seluruh kolom—InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, dan Country—peneliti mampu membangun dasbor analitik komprehensif serta menerapkan model prediktif untuk meningkatkan strategi penjualan dan kepuasan pelanggan.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HE	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANT	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEAR	8	12/1/2010 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLA	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTI	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NE	2	12/1/2010 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTE	6	12/1/2010 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UN	6	12/1/2010 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER REC	6	12/1/2010 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR	32	12/1/2010 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCES	8	12/1/2010 8:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUC	6	12/1/2010 8:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED	6	12/1/2010 8:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIG	3	12/1/2010 8:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE AL	2	12/1/2010 8:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BL	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLO	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH M	4	12/1/2010 8:34	7.95	13047	United Kingdom

Gambar 1. Sample Data Online Retail

## 2. Pemahaman Data

Pada tahap pemahaman data, analisis RFM (Recency, Frequency, Monetary) menjadi kerangka utama untuk menilai nilai dan perilaku pelanggan. Pendekatan ini membantu memetakan pola interaksi pelanggan dengan bisnis serta mengidentifikasi segmen yang paling berpotensi untuk strategi pemasaran lebih lanjut.

Berdasarkan metrik Recency, rata-rata waktu sejak transaksi terakhir setiap pelanggan adalah 90 hari. Angka ini menunjukkan bahwa sebagian besar pelanggan belum melakukan pembelian dalam kurun waktu sekitar tiga bulan, sehingga perlu dipertimbangkan upaya re-engagement, seperti kampanye email atau penawaran khusus, untuk mendorong mereka kembali.

Dari segi Frequency, rata-rata jumlah transaksi per pelanggan mencapai 10 kali. Hal ini mengindikasikan bahwa, meski frekuensi pembelian cukup baik, terdapat variasi antar pelanggan—ada yang membeli berulang kali dan ada yang jarang—sehingga segmentasi lebih lanjut dapat dilakukan untuk mengoptimalkan loyalitas dan memperkuat retensi.

Sementara itu, metrik Monetary mencatat rata-rata total nilai transaksi per pelanggan sebesar \$500. Besaran ini mencerminkan kontribusi pendapatan yang signifikan dari pelanggan rata-rata, dan dapat dijadikan tolok ukur untuk menilai efisiensi program pemasaran serta menyesuaikan strategi harga atau paket produk demi meningkatkan nilai pembelian pada tiap segmen.

## 3. Persiapan Data

Setelah tahap pemahaman data selesai, langkah selanjutnya adalah melakukan persiapan data secara menyeluruh untuk memastikan kualitas dan konsistensi sebelum dianalisis lebih lanjut. Pertama-tama, seluruh entri transaksi diperiksa untuk mengidentifikasi dan mengeliminasi catatan yang tidak valid—misalnya, faktur dengan nilai negatif, kuantitas barang nol atau negatif, serta tanggal transaksi yang berada di luar rentang pengamatan. Penghapusan transaksi tidak valid ini penting agar model

analitik tidak terdistorsi oleh outlier atau kesalahan pencatatan.

Selanjutnya, tim melakukan pengecekan missing values pada setiap kolom kunci—seperti CustomerID, InvoiceDate, UnitPrice, dan Quantity. Untuk entri yang hanya kehilangan atribut sekunder (misalnya, Description atau Country), nilai dipulihkan dengan metode imputasi sederhana berdasarkan modus atau nilai yang paling sering muncul dalam kelompok sejenis. Namun, jika kolom utama hilang (misalnya CustomerID), baris tersebut dikategorikan tidak dapat digunakan dan dihapus untuk menjaga integritas analisis segmentasi pelanggan (lihat Gambar 2).

Setelah data bersih dari baris bermasalah, langkah berikutnya adalah standarisasi format setiap kolom. Tanggal transaksi diubah ke tipe datetime dengan zona waktu seragam, sementara kolom harga satuan dan kuantitas distandardisasi ke tipe numerik dengan memastikan penggunaan titik desimal yang konsisten. Kolom teks seperti Description mengalami case folding dan trimming spasi sehingga deskripsi produk tidak terduplikasi akibat perbedaan penulisan.

```

We can see that there are 2 features with missing values (null).

1 # check null
2 df.isna().sum()

InvoiceNo      0
StockCode      0
Description    1454
Quantity       0
InvoiceDate    0
UnitPrice      0
CustomerID    135080
Country        0
dtype: int64

```

Gambar 2. Missing Value

Pada tahap ini juga diterapkan verifikasi silang dengan sumber data lain—misalnya laporan keuangan atau log sistem inventori—untuk memastikan bahwa transaksi signifikan (nilai besar atau frekuensi tinggi) tercatat dengan benar. Bila ditemukan perbedaan, dilakukan audit manual dan koreksi pada database. Proses ini menambah lapisan validasi sehingga data akhir yang digunakan benar-benar mencerminkan aktivitas penjualan riil.

#### 4. Preprocessing Data

Pada langkah awal preprocessing, data yang tersisa setelah pembersihan (termasuk penghapusan transaksi tidak valid dan penanganan missing values) dipastikan konsisten

dan siap dihitung metrik RFM. Proses ini ditunjukkan pada Gambar 3, di mana setiap baris dengan nilai esensial yang hilang telah dieliminasi sehingga tidak memengaruhi akurasi perhitungan. Dengan demikian, dataset akhir hanya memuat transaksi lengkap yang valid untuk analisis lebih mendalam.

Selanjutnya, metrik Recency dihitung dengan menentukan selisih hari antara tanggal terakhir setiap pelanggan melakukan transaksi dan tanggal referensi analisis. Proses ini mengevaluasi seberapa baru interaksi pelanggan dengan platform—semakin kecil nilai Recency, semakin segar keterlibatan pelanggan tersebut. Hasilnya kemudian digunakan untuk mengidentifikasi pelanggan yang perlu dijangkau kembali atau diberi insentif khusus.

```
[ ] 1 # dropna
    2 df = df.dropna()

▶ 1 # check info after dropping
   2 df.info()

↳ <class 'pandas.core.frame.DataFrame'>
Int64Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   InvoiceNo        406829 non-null object
1   StockCode        406829 non-null object
2   Description      406829 non-null object
3   Quantity         406829 non-null int64
4   InvoiceDate      406829 non-null datetime64[ns]
5   UnitPrice        406829 non-null float64
6   CustomerID       406829 non-null float64
7   Country          406829 non-null object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

Gambar 3. Drop Missing Value

Metrik Frequency dihitung dengan menjumlahkan total transaksi yang dilakukan setiap pelanggan selama periode pengamatan. Semakin tinggi frekuensi, semakin besar indikasi loyalitas atau ketergantungan pelanggan pada produk yang ditawarkan. Nilai Frequency ini penting untuk mengelompokkan pelanggan aktif tinggi, sedang, dan rendah, sehingga strategi pemasaran dapat dipersonalisasi sesuai tingkat keterlibatan.

Terakhir, metrik Monetary diperoleh dengan mengakumulasi total nilai pembelian setiap pelanggan—hasil perkalian Quantity dengan UnitPrice untuk setiap transaksi, kemudian dijumlahkan per CustomerID. Nilai Monetary menggambarkan kontribusi finansial pelanggan terhadap pendapatan perusahaan, sehingga memungkinkan penentuan segmen pelanggan paling bernilai. Setelah ketiga metrik ini dihitung, data RFM siap untuk tahap segmentasi dan analisis selanjutnya.

Dengan data yang kini telah bersih, lengkap, dan terstruktur, seluruh set data valid disimpan dalam format terpusat (misalnya tabel SQL atau file CSV dengan encoding UTF-8) siap dipakai untuk tahap analisis berikutnya, seperti RFM, clustering, atau predictive modeling. Dokumentasi lengkap tentang kriteria penyaringan, imputasi missing values, dan format final turut dilengkapi sehingga setiap langkah dapat direplikasi dan diaudit di masa mendatang.

#### 5. Pembangunan Model RFM

Setelah nilai Recency, Frequency, dan Monetary untuk setiap pelanggan diperoleh, data tersebut disusun ke dalam sebuah tabel RFM (Gambar 4) yang mencantumkan masing-masing metrik per CustomerID. Tabel ini menjadi dasar untuk mengkategorikan pelanggan berdasarkan kebaruan interaksi, frekuensi pembelian, dan kontribusi nilai transaksi. Dengan menyatukan ketiga metrik dalam satu struktur, peneliti dapat melakukan analisis lebih lanjut terkait segmentasi pelanggan.

Langkah pertama dalam pembangunan model adalah melakukan kuartilisasi pada metrik Recency. Nilai Recency dibagi ke dalam empat kelompok kuartil, di mana pelanggan dengan nilai hari paling kecil (transaksi terbaru) mendapat

skor 1, hingga pelanggan dengan nilai hari terbesar (transaksi paling lama) mendapat skor 4. Pembagian ini mempermudah identifikasi pelanggan yang masih sangat aktif bertransaksi versus mereka yang sudah lama tidak membeli.

	CustomerID	last_trans_date	recency	frequency	monetary
0	12346.0	2011-01-18	325	180.0	0.00
1	12347.0	2011-12-07	2	180.0	4310.00
2	12348.0	2011-09-25	75	123.0	1797.24
3	12349.0	2011-11-21	18	258.0	1757.55
4	12350.0	2011-02-02	310	5903.0	334.40

Gambar 4. Table RFM Model

Selanjutnya, metrik Frequency dan Monetary juga dikategorikan menggunakan metode kuartil, namun dengan urutan skor yang berlawanan: pelanggan dengan frekuensi transaksi tertinggi atau nilai transaksi terbesar masing-masing mendapat skor 1, sedangkan yang terendah mendapat skor 4.

Dengan demikian, skor rendah pada Frequency dan Monetary menunjukkan pelanggan bernilai tinggi, sementara skor tinggi menandakan pelanggan dengan tingkat interaksi atau kontribusi pendapatan yang lebih rendah.

	CustomerID	last_trans_date	recency	frequency	monetary	R_Quartile	F_Quartile	M_Quartile	RFMScore
0	12346.0	2011-01-18	325	180.0	0.00	4	2	4	424
1	12347.0	2011-12-07	2	180.0	4310.00	1	2	1	121
2	12348.0	2011-09-25	75	123.0	1797.24	3	3	1	331
3	12349.0	2011-11-21	18	258.0	1757.55	2	2	1	221
4	12350.0	2011-02-02	310	5903.0	334.40	4	1	3	413

Gambar 5. Table Scoring RFM

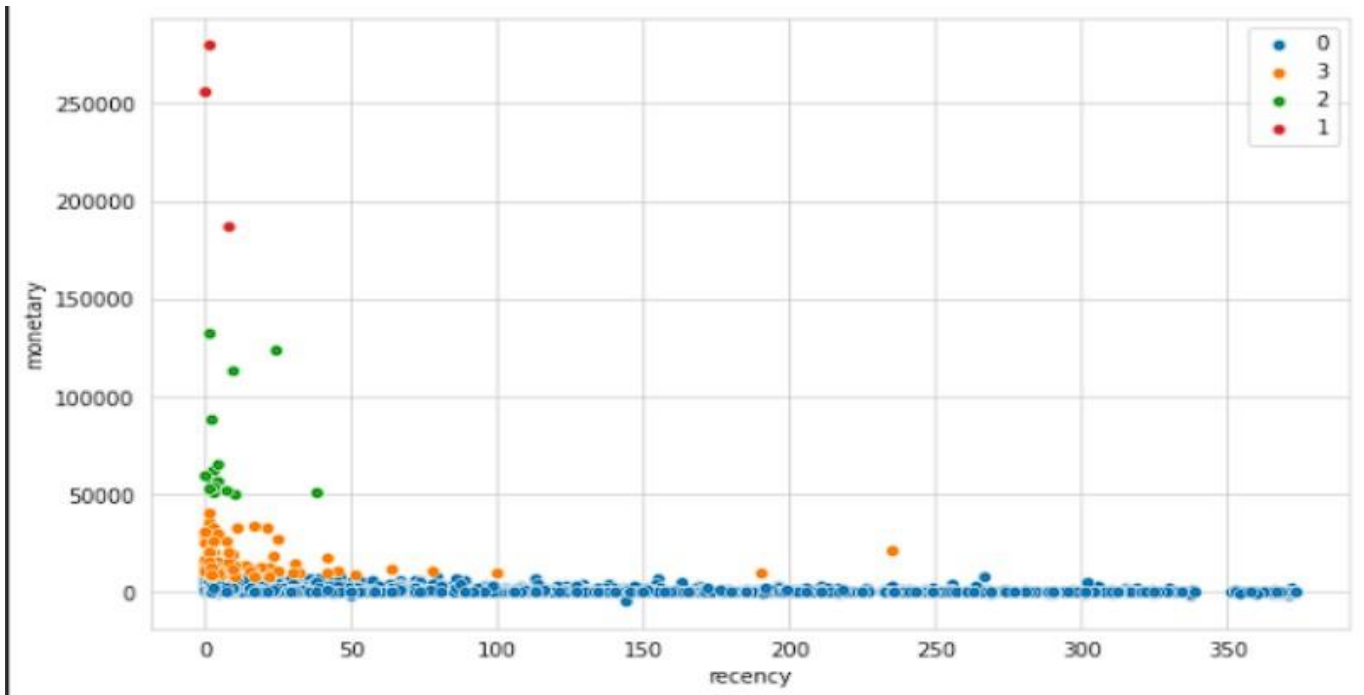
Terakhir, setiap baris pelanggan diberi kombinasi skor RFM (misalnya 1-2-3) berdasarkan hasil kuartilisasi. Hasil scoring ini divisualisasikan dalam Gambar 5 dan memberikan kerangka yang jelas untuk segmentasi—pelanggan dengan kombinasi skor “1-1-1” menjadi segmen paling prioritas, sedangkan skor “4-4-4” menunjukkan pelanggan dengan tingkat engagement dan nilai terendah. Model RFM yang telah dibangun ini siap diaplikasikan untuk strategi pemasaran terarah, kampanye retensi, dan program loyalitas.

## 6. Clustering

Pada tahap Clustering, pertama-tama dilakukan penentuan jumlah cluster optimal menggunakan metode

Elbow. Metode ini mengevaluasi total within-cluster sum of squares (WCSS) untuk berbagai jumlah cluster dan memilih titik di mana penurunan WCSS mulai melandai. Berdasarkan analisis Elbow, ditemukan bahwa empat cluster memberikan keseimbangan terbaik antara kompleksitas model dan kualitas segmentasi data.

Setelah menentukan jumlah cluster, algoritma K-Means diterapkan dengan parameter  $k = 4$  pada dataset RFM pelanggan. Proses ini melibatkan inisialisasi empat centroid, pengelompokan setiap pelanggan ke centroid terdekat, serta iterasi pembaruan posisi centroid hingga konvergensi. Dengan demikian, setiap pelanggan secara otomatis diklasifikasikan ke dalam satu dari empat cluster berdasarkan skor Recency, Frequency, dan Monetary.



Gambar 6. Plot hasil cluster

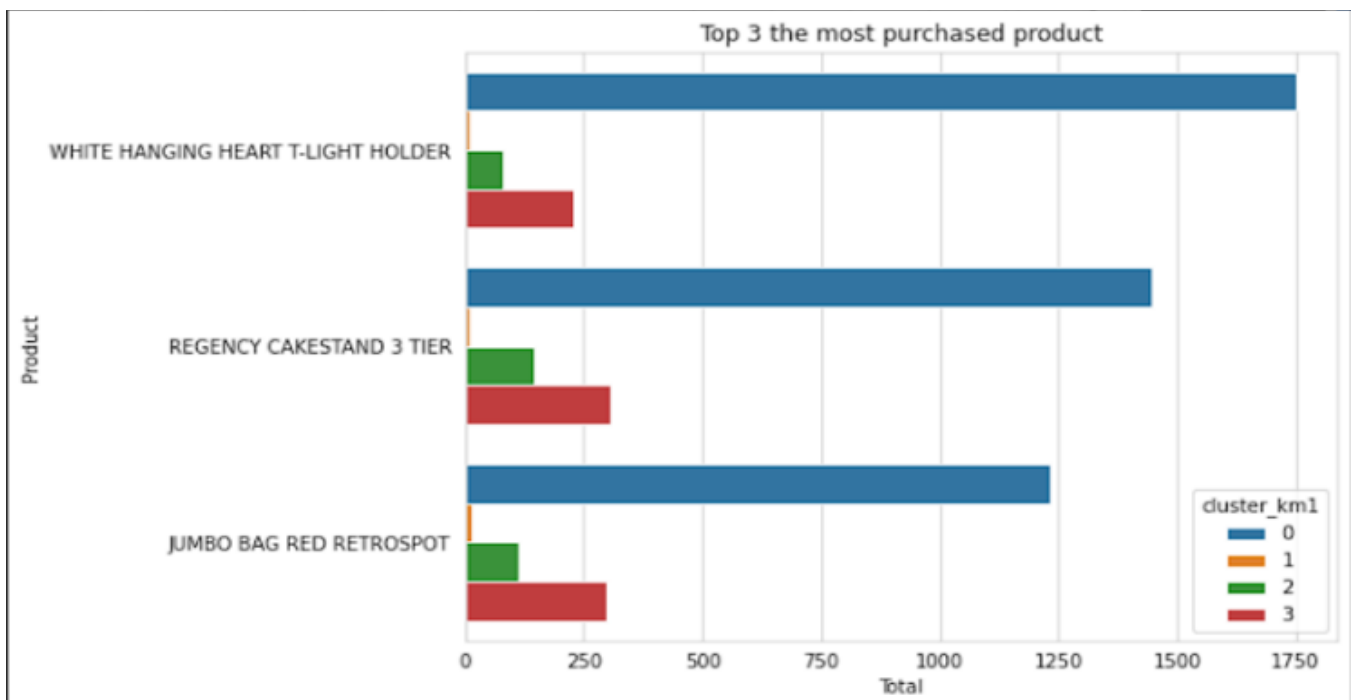
Hasil clustering dibagi menjadi empat segmen dengan karakteristik berbeda. Cluster 0 merepresentasikan “pelanggan utama dan sangat aktif” yang memiliki transaksi terbaru, frekuensi tinggi, dan nilai pendapatan terbesar. Segmentasi ini membantu memfokuskan upaya retensi dan penghargaan loyalitas.

Cluster 1 mencakup “pelanggan yang hampir tidak aktif atau kurang bernilai,” menunjukkan hubungan yang perlu diperbaiki atau kampanye re-engagement. Cluster 2 adalah “pelanggan dengan potensi untuk ditingkatkan loyalitasnya,” artinya pelanggan ini masih membeli tetapi belum mencapai frekuensi atau nilai transaksi tertinggi. Cluster 3 menggambarkan “pelanggan yang stabil dan cukup berharga,” yaitu pelanggan dengan pola pembelian konsisten

namun tidak termasuk top performers. Personas ini memandu strategi pemasaran tersegmentasi untuk masing-masing kelompok.

#### 7. Analisis Eksploratif (Exploratory Data Analysis)

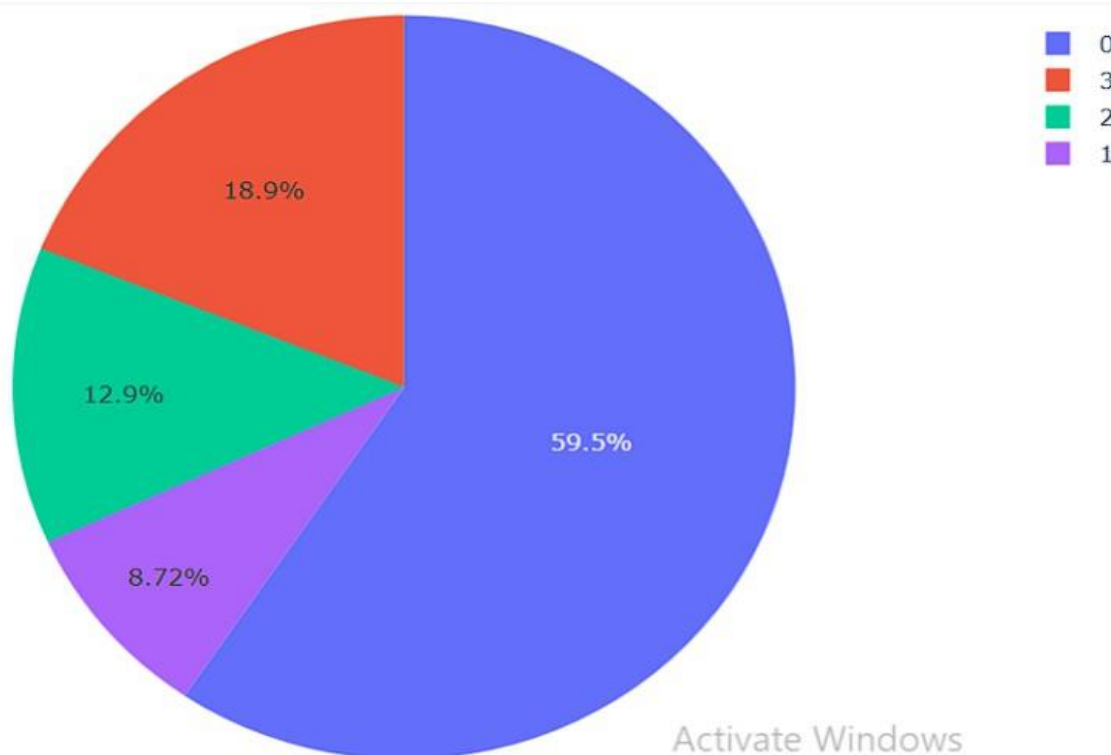
Pada tahap Analisis Eksploratif, peneliti meninjau pola pembelian produk dan kontribusi segmen pelanggan terhadap pendapatan keseluruhan. Berdasarkan Gambar 7, terlihat tiga produk yang paling sering dibeli pelanggan: Produk A mendominasi dengan frekuensi tertinggi, diikuti oleh Produk B dan Produk C. Data ini menunjukkan minat kuat pelanggan terhadap ketiga produk tersebut dan memberikan petunjuk penting bagi manajemen pemasaran untuk menyorot penawaran atau promosi khusus pada item-item bestseller ini.



Gambar 7. Produk paling sering dibeli

Selanjutnya, kontribusi masing-masing cluster terhadap total pendapatan dianalisis. Hasilnya menunjukkan Cluster 0—pelanggan utama dan sangat aktif—menyumbang 59,5% dari keseluruhan pemasukan, menegaskan peran krusial mereka dalam stabilitas finansial perusahaan. Cluster 1, yang terdiri dari pelanggan kurang aktif atau

kurang bernilai, hanya memberikan 8,72%, sedangkan Cluster 2 (pelanggan berpotensi ditingkatkan loyalitasnya) and Cluster 3 (pelanggan stabil dan cukup berharga) masing-masing menyumbang 12,9% dan 18,9%. Diagram kontribusi pada Gambar 8 memvisualisasikan perbandingan proporsi ini secara jelas.



Gambar 8. Diagram Kontribusi

Dari hasil eksplorasi ini, dapat diambil beberapa langkah strategis. Pertama, fokus utama adalah mempertahankan dan memperkuat hubungan dengan pelanggan di Cluster 0 melalui program loyalitas atau rewards khusus. Kedua, aktivitas re-engagement untuk Cluster 1 perlu ditingkatkan, mungkin melalui diskon personal atau kampanye reminder. Ketiga, untuk Cluster 2 dan 3, strategi up-selling dan cross-selling pada produk bestseller (Produk A, B, C) dapat memaksimalkan nilai transaksi sekaligus mendorong loyalitas yang lebih tinggi. Secara keseluruhan, insight EDA ini menyediakan fondasi bagi pengambilan keputusan berbasis data untuk mengoptimalkan portofolio produk dan alokasi sumber daya pemasaran.

Analisis RFM dan clustering mengungkap adanya perbedaan yang signifikan dalam perilaku dan nilai setiap segmen pelanggan. Pelanggan dalam Cluster 0 terbukti paling berharga dan paling aktif, sehingga retensi dan peningkatan kepuasan pelanggan di segmen ini harus menjadi prioritas utama. Dengan mempertahankan loyalitas mereka melalui program rewards, layanan eksklusif, atau insentif khusus, perusahaan dapat memastikan aliran pendapatan yang stabil dan meningkatkan lifetime value pelanggan.

Sementara itu, pelanggan dalam Cluster 1 dan Cluster 2 menunjukkan potensi pertumbuhan yang besar meski frekuensi atau nilai transaksi mereka relatif lebih rendah daripada Cluster 0. Melalui penawaran yang terpersonalisasi—seperti diskon untuk produk populer, paket bundling, atau rekomendasi berbasis preferensi—perusahaan

dapat mendorong peningkatan frekuensi pembelian dan total nilai transaksi. Strategi pemasaran yang disesuaikan dengan karakteristik masing-masing segmen ini akan memaksimalkan dampak kampanye dan ROI.

Dengan mengetahui pelanggan paling bernilai, perusahaan dapat mengalokasikan sumber daya pemasaran dan layanan pelanggan secara lebih efisien. Misalnya, tim customer success bisa diprioritaskan untuk merespons lebih cepat permintaan atau keluhan dari pelanggan Cluster 0, sementara tim pemasaran digital dapat merancang otomatisasi email atau iklan tertarget untuk Cluster 1 dan 2. Pendekatan ini mengaktifkan biaya sambil menjaga kualitas layanan.

Pengetahuan tentang produk yang paling sering dibeli oleh setiap segmen juga sangat berguna untuk manajemen inventaris dan perencanaan produk. Perusahaan perlu memastikan ketersediaan stok yang memadai untuk item-item favorit pelanggan Cluster 0 agar tidak terjadi kehabisan saat permintaan tinggi. Selain itu, insight ini bisa menjadi dasar untuk pengembangan varian produk atau perbaikan kualitas guna memenuhi selera pelanggan yang beragam.

Hasil analisis RFM dan clustering memberikan wawasan yang dapat meningkatkan strategi pemasaran, manajemen hubungan pelanggan, dan pengambilan keputusan bisnis. Dengan segmentasi yang jelas, perusahaan dapat merancang kampanye retensi, program loyalitas, dan promosi penjualan dengan sasaran yang tepat, sekaligus

memantau efektivitasnya melalui metrik kinerja yang spesifik.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan analitik berbasis RFM dan K-Means dapat menjadi sumber keuntungan kompetitif. Dengan memahami kebutuhan dan perilaku tiap segmen pelanggan secara lebih mendalam, perusahaan tidak hanya memaksimalkan retensi dan pendapatan, tetapi juga memperkuat hubungan jangka panjang yang berdampak pada pertumbuhan bisnis berkelanjutan.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian, analisis RFM dan clustering K-Means mengungkap perbedaan perilaku pelanggan yang signifikan. Pelanggan dalam Cluster 0 terbukti sebagai segmen paling berharga dan paling aktif, sehingga upaya retensi—melalui program loyalitas, penghargaan eksklusif, atau layanan prioritas—harus menjadi fokus utama. Sebaliknya, pelanggan dalam Cluster 1 dan Cluster 2 menunjukkan potensi pertumbuhan yang besar; dengan strategi pemasaran terpersonalisasi—seperti penawaran khusus, bundling produk, atau rekomendasi berbasis preferensi—frekuensi pembelian dan nilai transaksi mereka dapat meningkat secara signifikan.

Pengetahuan mengenai produk favorit pelanggan Cluster 0 juga membawa manfaat praktis untuk manajemen rantai pasok dan perencanaan inventaris. Menjaga ketersediaan stok untuk item-item unggulan akan mencegah kehilangan peluang penjualan dan mendukung kepuasan

pelanggan. Secara keseluruhan, pendekatan analitik ini membuktikan bahwa segmentasi berbasis RFM dan K-Means dapat mengoptimalkan nilai setiap segmen pelanggan, meningkatkan retensi, dan mendorong pertumbuhan pendapatan yang berkelanjutan.

#### DAFTAR PUSTAKA

- [1] B. Basri, W. Gata, and R. Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, dan Monetary (RFM) dan Decision Tree pada PT. Solo," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, p. 943, 2020, doi: 10.25126/jtiik.2020752284.
- [2] M. Jordy, A. Triayudi, and I. D. Sholihati, "Analisis Segmentasi Recency dan Customer Value Pada AVANA Indonesia Dengan Algoritma K-Means dan Model RFM (Recency, Frequency and Monetary)," vol. 4, no. 2, pp. 579–589, 2023, doi: 10.47065/josh.v4i2.2950.
- [3] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "Analisis Rfm (Recency, Frequency and Monetary) Produk Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek.*
- [4] I. Rahma, P. Prima Arhandi, and A. Tufika Firdausi, "Penerapan Metode Hierarchical Clustering Dan K-Means Clustering Untuk Mengelompokkan Potensi Lokasi Penjualan Linkaja," *J. Inform. Polinema*, vol. 6, no. 1, pp. 15–22, 2020, doi: 10.33795/jip.v6i1.287.