

Segmentasi Pelanggan E-Commerce Menggunakan Algoritma K-Means Clustering Berbasis Analisis RFM

Gaten Isnan Abdurrozaq^{1*}, Riky Ardiva Dwitama Putra², Afriansyah Septyarno³, Wasis Yoga Wicaksono⁴

¹Teknik Informatika/Fakultas
Ilmu Komputer
Universitas Duta Bangsa Surakarta
^{1*}230103102@mhs.udb.ac.id

²Teknik Informatika/Fakultas
Ilmu Komputer
Universitas Duta Bangsa Surakarta
²230103117@mhs.udb.ac.id

³Teknik Informatika/Fakultas
Ilmu Komputer
Universitas Duta Bangsa Surakarta
³230103084@mhs.udb.ac.id

⁴Teknik Informatika/Fakultas
Ilmu Komputer
Universitas Duta Bangsa Surakarta
⁴230103121@mhs.udb.ac.id

Abstrak— Sebagian besar perusahaan e-commerce masih memanfaatkan data transaksi sebatas laporan operasional tanpa segmentasi pelanggan yang memadai. Penelitian ini bertujuan menghasilkan model segmentasi pelanggan e-commerce yang objektif beserta interpretasi setiap segmen sebagai dasar strategi pemasaran. Metode yang digunakan mengombinasikan analisis Recency, Frequency, Monetary (RFM) dan algoritma K-Means Clustering pada dataset Online Retail (UCI Machine Learning Repository), melalui preprocessing, normalisasi StandardScaler, penentuan cluster optimal menggunakan Elbow Method, serta evaluasi dengan Silhouette Score, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI). Dari 541.909 data mentah diperoleh 397.884 transaksi valid milik 4.338 pelanggan. Hasil menunjukkan $k = 5$ sebagai cluster optimal (Silhouette Score 0,617; DBI 0,716; CHI 3.437,64), menghasilkan lima segmen yaitu Potential Customer (70,29%), Dormant Customer (24,48%), High-Value Active (4,91%), Loyal Customer (0,18%), dan Champions (0,14%). Integrasi RFM dan K-Means menghasilkan segmentasi pelanggan yang valid, objektif, mendukung penyusunan strategi pemasaran yang lebih tepat sasaran, serta mudah diterapkan.

Kata kunci— segmentasi pelanggan, RFM, K-Means Clustering, e-commerce, strategi pemasaran

Abstract— Most e-commerce companies still use transaction data solely for operational reporting without adequate customer segmentation. This study aims to develop an objective e-commerce customer segmentation model, along with an interpretation of each segment, to serve as the basis for marketing strategies. The method used combines Recency, Frequency, Monetary (RFM) analysis and the K -Means Clustering algorithm on the Online Retail dataset (UCI Machine Learning Repository), involving preprocessing, StandardScaler normalization, determination of the optimal number of clusters using the Elbow Method, and evaluation using the Silhouette Score, the Davies-Bouldin Index (DBI), and the Calinski-Harabasz Index (CHI). From 541,909 raw data points, 397,884 valid transactions belonging to 4,338 customers were obtained. The results indicate that $k = 5$ is the optimal number of clusters (Silhouette Score 0.617; DBI 0.716; CHI 3,437.64), resulting in five segments: Potential Customers (70.29%), Dormant Customers (24.48%), High-Value Active Customers (4.91%), Loyal Customers (0.18%), and Champions (0.14%). The integration of RFM and K-Means produces a valid and objective customer segmentation that supports the development of more targeted marketing strategies and is easy to implement.

Keywords— customer segmentation, RFM, K-Means Clustering, e-commerce, marketing strategy

V. PENDAHULUAN

Data transaksi pelanggan merupakan aset informasi strategis bagi perusahaan *e-commerce* dalam mendukung pengambilan keputusan berbasis data. Berdasarkan Statistik *E-Commerce* 2024 (BPS), nilai transaksi *e-commerce* Indonesia mencapai sekitar Rp1.288,93 triliun, meningkat 17,08% dari tahun sebelumnya [1]. Volume data yang besar ini menyulitkan analisis manual sehingga diperlukan pendekatan analitis

yang efektif [2]. Namun, sebagian besar perusahaan masih memanfaatkan data transaksi sebatas laporan operasional, bukan sebagai dasar strategi pemasaran, retensi, maupun loyalitas pelanggan [2], sehingga cenderung menerapkan pemasaran seragam meskipun nilai dan pola transaksi tiap pelanggan berbeda [3].

Data mining, khususnya segmentasi berbasis clustering, mampu mengubah data transaksi menjadi informasi pendukung keputusan yang

lebih objektif dengan mengidentifikasi kelompok pelanggan berkarakteristik serupa [3]. Model *Recency, Frequency, Monetary* (RFM) banyak digunakan karena merepresentasikan perilaku pelanggan berdasarkan waktu, frekuensi, dan nilai transaksi [4], dan efektif sebagai atribut masukan K-Means Clustering yang komputasinya sederhana dan efisien [5]. Kombinasi keduanya terbukti menghasilkan segmentasi yang lebih informatif dan homogen [4][6][7][8].

Beberapa studi terdahulu menerapkan kombinasi RFM dan K-Means pada data e-commerce, namun masing-masing memiliki keterbatasan yaitu Sharyanto dan Lestari [9] hanya mengevaluasi cluster dengan Davies-Bouldin Index tanpa indeks validasi tambahan. Fauzan dan Alfian [10] berfokus pada perbandingan algoritma tanpa rekomendasi strategi per segmen. Lukas dkk. [11] menentukan cluster optimal via Elbow Method tanpa verifikasi indeks lain, dan Hafidz Ardana dkk. [12], meski memakai Elbow Method dan Silhouette Score, tetap berhenti pada tahap segmentasi tanpa interpretasi strategi operasional. Keempat studi ini menunjukkan pola yang sama yaitu clustering berhenti pada pengelompokan tanpa validasi ganda yang konsisten dan tanpa interpretasi strategi pemasaran per segmen [9][10][11][12], serta umumnya menggunakan dataset internal yang sulit direplikasi dan dibandingkan antarstudi [11][13][14]. Dengan demikian, research gap penelitian ini adalah belum tersedianya studi RFM–K-Means yang sekaligus menerapkan validasi ganda Elbow Method dan Silhouette Score, interpretasi cluster ke dalam strategi pemasaran per segmen, dan dataset publik yang replikatif.

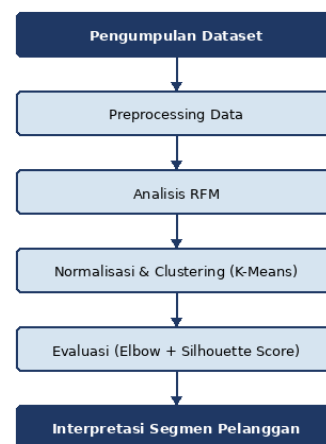
Penelitian ini bertujuan menghasilkan model segmentasi pelanggan *e-commerce* yang objektif beserta interpretasi strategi pemasaran per segmen, menggunakan RFM dan K-Means pada dataset publik Online Retail (UCI) dengan evaluasi Elbow Method dan Silhouette Score secara bersamaan. Berbeda dengan Febri dkk. [15] yang memetakan segmentasi ke strategi pemasaran namun terbatas pada studi kasus satu

perusahaan (Auto2000 Kenjeran) dengan data internal non-publik, penelitian ini menguji kerangka segmentasi pada dataset publik berskala besar (541.909 baris, 4.338 pelanggan) yang replikatif, disertai validasi ganda. Kontribusi (*novelty*) penelitian ini adalah segmentasi objektif yang dipetakan ke rekomendasi strategi pemasaran aplikatif, didukung validasi ganda dan dataset publik replikatif sebagai pembeda dari studi-studi sebelumnya [9][10][11][12][15].

VI. METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian secara sistematis, mulai dari pengumpulan dataset hingga interpretasi hasil segmentasi pelanggan menggunakan analisis RFM dan algoritma K-Means Clustering.

A. Tahapan Penelitian



Gambar 15. Alur Tahapan Penelitian

Alur tahapan penelitian ditunjukkan pada Gambar 1. Tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Pengumpulan Dataset, yaitu mengumpulkan data transaksi penjualan ritel daring dari Online Retail Dataset (UCI Machine Learning Repository) periode Desember 2010–Desember 2011.
2. Preprocessing Data, yaitu membersihkan data melalui penghapusan nilai kosong, pemfilteran nilai tidak valid, perhitungan TotalPrice, konversi format tanggal, dan penentuan Reference Date.

3. Analisis RFM, yaitu menghitung nilai *Recency*, *Frequency*, dan *Monetary* setiap pelanggan berdasarkan agregasi data per CustomerID [4].
4. Normalisasi dan Clustering (K-Means), yaitu menstandarisasi nilai RFM menggunakan StandardScaler, kemudian mengelompokkan pelanggan menggunakan algoritma K-Means Clustering [5].
5. Evaluasi, yaitu menentukan jumlah cluster optimal menggunakan Elbow Method serta mengevaluasi kualitas hasil clustering menggunakan Silhouette Score, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI)..
6. Interpretasi Segmen Pelanggan, yaitu menganalisis karakteristik RFM tiap cluster dan menyusun profil segmen sebagai dasar rekomendasi strategi pemasaran.

B. Dataset Penelitian

Dataset yang digunakan adalah Online Retail Dataset yang bersumber dari UCI Machine Learning Repository. Dataset ini berisi data transaksi penjualan ritel daring di Inggris selama periode Desember 2010 hingga Desember 2011 dengan 8 atribut, yaitu *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, *CustomerID*, dan *Country*. Karakteristik dataset disajikan pada Tabel 1.

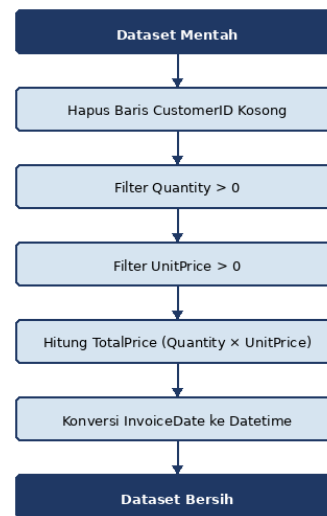
Tabel 3. Karakteristik Dataset Penelitian

Karakteristik	Nilai
Nama Dataset	Online Retail
Sumber	UCI Machine Learning Repository
Periode Data	Desember 2010 – Desember 2011
Jumlah Transaksi	541.909 baris
Pelanggan Unik	4.372 CustomerID
Jumlah Atribut	8 atribut
Format File	CSV (Comma-Separated Values)

C. Tahap Preprocessing Data

Preprocessing dilakukan untuk memastikan kualitas dan konsistensi data sebelum dilakukan analisis. Tahapan preprocessing meliputi penghapusan baris dengan CustomerID kosong, pemfilteran nilai tidak valid, perhitungan TotalPrice, konversi format tanggal, serta

penentuan Reference Date. Alur preprocessing ditunjukkan pada Gambar 2, sedangkan rincian setiap tahap disajikan pada Tabel 2.



Gambar 16. Alur Preprocessing Data

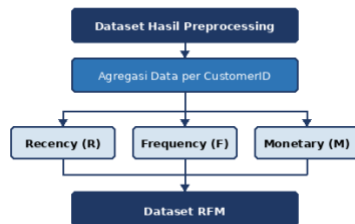
Tabel 4. Tahapan Preprocessing Data

Tahapan	Keterangan
Penghapusan Nilai Kosong	Menghapus baris dengan CustomerID kosong.
Filter Nilai Tidak Valid	Menghapus data dengan Quantity ≤ 0 atau UnitPrice ≤ 0 .
Perhitungan TotalPrice	Menambahkan atribut TotalPrice = Quantity x UnitPrice.
Konversi InvoiceDate	Mengubah InvoiceDate ke format datetime.
Penentuan Reference Date	Menetapkan satu hari setelah transaksi terakhir sebagai acuan perhitungan Recency.

D. Tahap Analisis RFM

Data dikelompokkan berdasarkan CustomerID untuk menghitung nilai *Recency*, *Frequency*, dan *Monetary* (RFM) setiap pelanggan [4]. Recency dihitung dari selisih antara *Reference Date* dan tanggal transaksi terakhir pelanggan, dengan *Reference Date* ditetapkan satu hari setelah tanggal transaksi terakhir agar perhitungan *Recency* dilakukan secara konsisten. *Frequency* dihitung dari jumlah invoice unik setiap pelanggan menggunakan fungsi `nunique()` pada atribut *InvoiceNo*, sedangkan *Monetary* dihitung dari total TotalPrice setiap pelanggan. Alur pembentukan data RFM ditunjukkan pada Gambar 3,

sedangkan perhitungan setiap atribut dijelaskan pada Tabel 3.



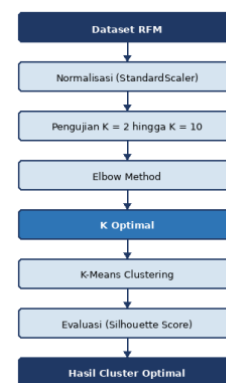
Gambar 17. Alur Pembentukan Data RFM

Tabel 5. Perhitungan Atribut RFM

Atribut	Simbol	Keterangan
Recency (R)	R	Menunjukkan seberapa lama pelanggan terakhir melakukan transaksi (dalam hari).
Frequency (F)	F	Jumlah invoice unik setiap pelanggan
Monetary (M)	M	Menunjukkan total nilai pengeluaran pelanggan.

E. Clustering dan Evaluasi

Atribut RFM yang terdiri atas *Recency*, *Frequency*, dan *Monetary* dinormalisasi menggunakan StandardScaler sebelum proses clustering. Mengingat atribut *Monetary* memiliki rentang nilai yang lebar dan berpotensi mengandung outlier, dilakukan pengujian menggunakan transformasi logaritmik ($\log_{1p}$), sedangkan hasil evaluasinya dibahas pada Bab III. Jumlah cluster dievaluasi pada rentang $k = 2$ hingga $k = 10$ menggunakan Elbow Method. Selanjutnya, algoritma K-Means Clustering [5] diterapkan menggunakan jumlah cluster yang dipilih, kemudian hasil clustering dievaluasi menggunakan Silhouette Score, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI) untuk mendukung pemilihan jumlah cluster secara lebih objektif. Alur proses clustering dan evaluasi ditunjukkan pada Gambar 4.



Gambar 18. Alur Clustering & Evaluasi

Persamaan (1) Standarisasi Data:

$$z = (x - \mu) / \sigma$$

Keterangan: x = nilai data; μ = rata-rata atribut; σ = standar deviasi atribut.

Persamaan (2) Silhouette Score

$$S(i) = [b(i) - a(i)] / \max[a(i), b(i)]$$

Keterangan: $a(i)$ = rata-rata jarak data ke anggota dalam cluster yang sama; $b(i)$ = rata-rata jarak data ke cluster terdekat lainnya.

F. Interpretasi Karakteristik Segmen Pelanggan

Interpretasi ini mencakup penentuan label segmen (misalnya Champions, High-Value Active, Potential Customer, atau Dormant Customer), proporsi jumlah pelanggan pada tiap cluster, serta pemetaan setiap segmen ke dalam rekomendasi strategi pemasaran. Hasil lengkap interpretasi tersebut disajikan pada Bab III (Hasil dan Pembahasan).

VII. HASIL DAN PEMBAHASAN

Hasil penelitian ditulis secara jelas dan padat, dengan pembahasan yang menguraikan arti penting temuan.

A. Hasil Preprocessing Data

Tabel 6. Ringkasan Proses Preprocessing Data

Tahapan	Jumlah
Total data awal	541.909
Total data dihapus	144.025 (26,58%)
Total data bersih	397.884
Customer akhir	4.338

Tahap preprocessing menghasilkan 397.884 transaksi valid dari 541.909 data awal setelah menghapus data yang tidak memenuhi kriteria analisis (Tabel 4). Sebagian besar data yang dieliminasi merupakan transaksi tanpa CustomerID sebanyak 135.080 baris (24,9%), sedangkan 8.945 transaksi lainnya dihapus karena memiliki nilai $Quantity \leq 0$ atau $UnitPrice \leq 0$ yang mengindikasikan transaksi pembatalan atau kesalahan pencatatan. Proses ini mengurangi jumlah pelanggan yang dapat dianalisis dari 4.372 menjadi 4.338 pelanggan. Dengan demikian, dataset hasil preprocessing telah memenuhi kebutuhan pembentukan atribut *Recency*, *Frequency*, dan *Monetary* (RFM) sehingga layak digunakan pada clustering. Hasil preprocessing ditunjukkan pada Gambar 5.

Initial Data (Before Preprocessing)		Cleaned Data (After Preprocessing)	
Customer Count	541,909	Customer Count	4338
Transaction Count	135,080	Transaction Count	397,884
Customer Segments	4372	Customer Segments	4338
Customer Segments	34	Customer Segments	34

Gambar 19. Hasil Preprocessing Data

B. Analisis RFM

Tabel 7. Statistik Deskriptif Nilai RFM (n = 4.338 pelanggan)

Atribut RFM	Min	Max	Mean	Std Dev
Recency (hari)	1	374,0	92,54	100,01
Frequency (transaksi)	1	209,0	4,27	7,7
Monetary (£)	3,75	280.206,02	2.054,27	8.989,23

Analisis RFM dilakukan terhadap 4.338 pelanggan menggunakan *Reference Date* 10 Desember 2011. Berdasarkan Tabel 5, atribut *Recency* menunjukkan sebagian pelanggan telah lama tidak melakukan transaksi, *Frequency* mengindikasikan mayoritas pelanggan masih memiliki intensitas pembelian yang relatif rendah. Di sisi lain, atribut *Monetary* memiliki variasi nilai paling tinggi, ditunjukkan oleh standar deviasi yang jauh melebihi nilai rata-ratanya. Kondisi ini menunjukkan adanya perbedaan nilai transaksi yang cukup besar antar

pelanggan, sehingga penggunaan analisis RFM menjadi relevan untuk mengidentifikasi kelompok pelanggan dengan karakteristik yang berbeda. Distribusi nilai setiap atribut RFM ditunjukkan pada Gambar 6.

RFM Analysis Summary	
Analisis Customer	4338
Nilai rata-rata Recency	92.54 hari
Nilai rata-rata Monetary	2.054.27

CustomerID	Recency	Frequency	Monetary
12345.0	105	1	7790.16
12345.0	2	7	4702.0
12345.0	75	4	1797.29
12345.0	99	1	1797.09
12345.0	192	1	2344
12345.0	98	8	2084.94
12345.0	204	1	89.0
12345.0	232	1	1079.4
12345.0	274	1	4294
12345.0	27	1	2011.43
12345.0	33	1	4207.67
12345.0	1	1	1200.00

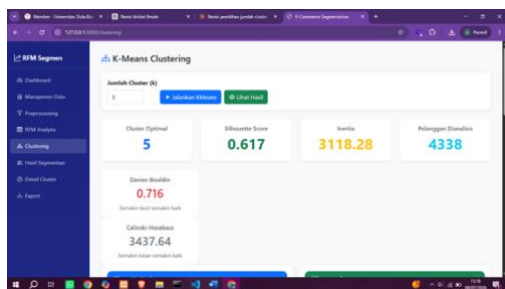
Gambar 20. Hasil Perhitungan Nilai RFM

C. Penentuan Jumlah Cluster Optimal

Tabel 8. Perbandingan Inertia, Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index per Jumlah Cluster

k	Inertia	Silhouette	DBI	CHI
2	9012,64	0,896	0,745	1925,06
3	5439,3	0,594	0,71	3018,43
4	4092,14	0,616	0,753	3149,72
5	3118,28	0,617	0,716	3437,64
6	2472,57	0,596	0,629	3693,76
7	2022,41	0,516	0,633	3923,09
8	1742,37	0,517	0,701	4001,62
9	1445,63	0,478	0,711	4330,25
10	1303,39	0,427	0,732	4320,91

Elbow Method menunjukkan bahwa penurunan nilai inertia mulai melandai pada rentang $k = 4-5$, sehingga jumlah cluster optimal perlu dikonfirmasi menggunakan indeks validasi kuantitatif. Berdasarkan Tabel 6, $k = 5$ memberikan kombinasi hasil evaluasi yang lebih baik dibandingkan $k = 4$, ditunjukkan oleh Silhouette Score yang sedikit lebih tinggi, Davies-Bouldin Index (DBI) yang lebih rendah, serta Calinski-Harabasz Index (CHI) yang lebih tinggi. Kombinasi ketiga metrik tersebut menunjukkan bahwa hasil clustering pada $k = 5$ memiliki tingkat kekompakan dan pemisahan cluster yang lebih baik. Oleh karena itu, $k = 5$ dipilih sebagai jumlah cluster optimal. Hasil evaluasi ditampilkan pada Gambar 7.



Gambar 21. Penentuan Jumlah Cluster Menggunakan Elbow Method dan Silhouette Score

D. Analisis Pengaruh Outlier pada Atribut Monetary

Atribut Monetary memiliki variasi nilai yang jauh lebih besar dibandingkan atribut RFM lainnya (Tabel 5), sehingga berpotensi dipengaruhi oleh outlier. Oleh karena itu, dilakukan pengujian menggunakan transformasi logaritmik ($\log_{10}$) untuk mengevaluasi pengaruhnya terhadap kualitas clustering. Hasil evaluasi menunjukkan bahwa transformasi log justru menurunkan kualitas clustering, ditandai dengan penurunan Silhouette Score dan peningkatan Davies-Bouldin Index dibandingkan model tanpa transformasi. Hasil tersebut menunjukkan bahwa StandardScaler telah mampu menormalkan perbedaan skala antaratribut tanpa menghilangkan informasi penting pada nilai *Monetary*, sehingga penelitian ini tetap menggunakan nilai *Monetary* asli pada proses clustering.

E. Hasil Segmentasi Pelanggan

Tabel 9. Karakteristik Rata-Rata Setiap Cluster

Cl.	Label Segmen	Recency (hari)	Frequency (transaksi)	Monetary (£)
0	Potential Customer	43,95	3,65	1.339,07
1	Dormant Customer	248,56	1,55	478,11
2	High-Value Active	15,67	21,29	12.831,74
3	Loyal Customer	6,5	120,5	55.312,69
4	Champions	7,67	42,83	190.863,46

K-Means dengan $k=5$ mengelompokkan 4.338 pelanggan ke dalam lima segmen dengan karakteristik RFM yang berbeda (Tabel 7). Visualisasi pada Gambar 8 menunjukkan

pemisahan cluster yang cukup jelas, terutama pada dimensi *Frequency* dan *Monetary*, sehingga mengindikasikan bahwa atribut RFM mampu membedakan karakteristik pelanggan secara efektif. Dibandingkan konfigurasi $k=4$, penggunaan $k=5$ menghasilkan pemisahan pelanggan bernilai tinggi menjadi dua segmen, yaitu Loyal Customer dan Champions, sehingga interpretasi karakteristik pelanggan menjadi lebih spesifik.

Potential Customer merupakan segmen terbesar dengan 3.049 pelanggan (70,29%). Karakteristiknya menunjukkan aktivitas transaksi yang masih relatif rendah sehingga pelanggan pada segmen ini memiliki potensi untuk ditingkatkan melalui peningkatan frekuensi pembelian dan nilai transaksi.

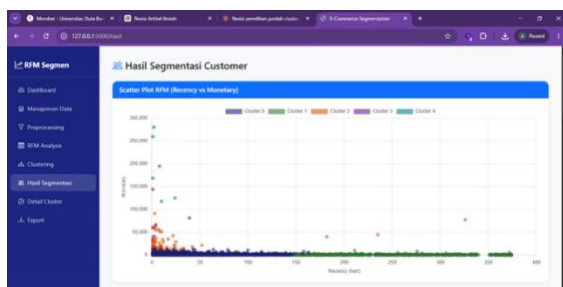
Dormant Customer terdiri atas 1.062 pelanggan (24,48%) dengan karakteristik pelanggan yang sudah lama tidak melakukan transaksi dan memiliki aktivitas pembelian paling rendah. Kondisi tersebut menunjukkan tingginya risiko pelanggan tidak kembali melakukan pembelian tanpa adanya upaya reaktivasi.

Segmen High-Value Active mencakup 213 pelanggan (4,91%) yang menunjukkan aktivitas transaksi tinggi dengan nilai pembelian yang relatif besar. Karakteristik ini menunjukkan bahwa segmen tersebut merupakan sumber pendapatan yang stabil dan berpotensi berkembang menjadi pelanggan bernilai lebih tinggi.

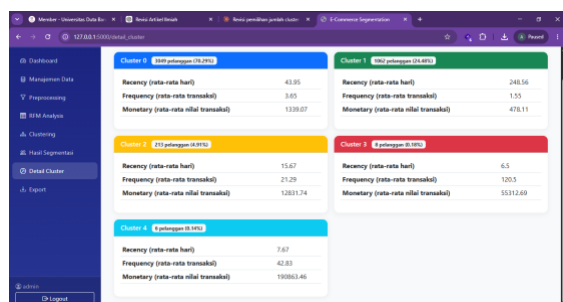
Segmen Loyal Customer hanya terdiri atas 8 pelanggan (0,18%), namun memiliki intensitas transaksi tertinggi dibandingkan segmen lainnya. Meskipun nilai pembeliannya masih berada di bawah segmen Champions, kelompok ini menunjukkan loyalitas yang kuat melalui frekuensi pembelian yang sangat tinggi.

Segmen Champions merupakan kelompok terkecil dengan 6 pelanggan (0,14%), tetapi memiliki nilai transaksi tertinggi di antara seluruh segmen. Meskipun jumlahnya sangat sedikit, kelompok ini memberikan kontribusi pendapatan yang jauh lebih besar dibandingkan segmen lainnya sehingga menjadi pelanggan dengan nilai bisnis tertinggi.

Secara keseluruhan, distribusi kelima segmen menunjukkan bahwa sebagian besar pelanggan masih berada pada kategori bernilai rendah hingga menengah, sedangkan pelanggan bernilai tinggi hanya mencakup sebagian kecil populasi. Temuan ini menunjukkan bahwa karakteristik pelanggan bersifat heterogen sehingga penerapan strategi pemasaran yang berbeda untuk setiap segmen menjadi lebih efektif dibandingkan pendekatan yang seragam. Distribusi proporsi pelanggan pada setiap segmen ditunjukkan pada Gambar 9.



Gambar 22. Visualisasi Hasil Segmentasi Pelanggan Menggunakan K-Means



Gambar 23. Distribusi Proporsi Pelanggan pada Setiap Cluster

F. Implikasi Strategi Pemasaran

Tabel 10. Proporsi Pelanggan per Segmen

Cl.	Label	Pelanggan (n)	Proporsi
0	Potential Customer	3.049	70,29%
1	Dormant Customer	1.062	24,48%
2	High-Value Active	213	4,91%
3	Loyal Customer	8	0,18%
4	Champions	6	0,14%
Total	---	4.338	100%

Tabel 11. Rekomendasi Strategi Pemasaran per Segmen

Cl.	Label	Strategi Utama
0	Potential Customer	Onboarding, rekomendasi produk personal, email nurturing
1	Dormant Customer	Kampanye reaktivasi (diskon eksklusif, notifikasi produk baru)
2	High-Value Active	Cross-selling dan up-selling berbasis riwayat transaksi
3	Loyal Customer	Program loyalitas berbasis frekuensi (reward poin, apresiasi repeat order) untuk mempertahankan intensitas transaksi
4	Champions	Layanan VIP, dedicated account manager, dan early access produk baru untuk mempertahankan pelanggan bernilai tertinggi

Karakteristik RFM pada setiap segmen menjadi dasar penyusunan strategi pemasaran yang berbeda (Tabel 9). Segmen Potential Customer yang mendominasi populasi pelanggan difokuskan pada peningkatan frekuensi pembelian melalui *onboarding* dan rekomendasi produk yang dipersonalisasi. Dormant Customer diprioritaskan untuk program reaktivasi melalui insentif dan promosi yang relevan guna mendorong pelanggan kembali bertransaksi. High-Value Active diarahkan pada strategi cross-selling dan up-selling untuk meningkatkan nilai pembelian. Sementara itu, Loyal Customer dan Champions, meskipun hanya mencakup sebagian kecil pelanggan (Tabel 8), memerlukan strategi retensi seperti program loyalitas dan layanan eksklusif karena memberikan nilai bisnis yang lebih tinggi. Temuan ini menunjukkan bahwa karakteristik setiap segmen memerlukan pendekatan pemasaran yang berbeda sehingga rekomendasi strategi dapat disesuaikan dengan profil masing-masing pelanggan.

G. Keterbatasan Penelitian dan Saran Pengembangan

Penelitian ini menggunakan dataset Online Retail yang merepresentasikan transaksi pelanggan di Inggris pada periode Desember 2010 hingga Desember 2011, sehingga hasil segmentasi belum tentu menggambarkan karakteristik *e-commerce* Indonesia saat ini yang memiliki pola transaksi, metode pembayaran, dan perilaku belanja yang berbeda. Oleh karena

itu, penelitian selanjutnya disarankan memvalidasi kerangka RFM dan K-Means pada dataset transaksi *e-commerce* Indonesia dengan menerapkan tahapan analisis dan evaluasi yang sama. Validasi tersebut diharapkan dapat menilai konsistensi karakteristik segmen serta meningkatkan relevansi praktis hasil penelitian pada konteks pasar lokal.

VIII. KESIMPULAN

Penelitian ini menghasilkan model segmentasi pelanggan *e-commerce* dengan analisis RFM dan algoritma K-Means Clustering pada Online Retail Dataset. Melalui proses preprocessing, diperoleh 397.884 transaksi valid yang merepresentasikan 4.338 pelanggan untuk dianalisis. Berdasarkan evaluasi menggunakan Silhouette Score sebesar 0,617, Davies-Bouldin Index (DBI) sebesar 0,716, dan Calinski-Harabasz Index (CHI) sebesar 3.437,64, diperoleh $k = 5$ sebagai jumlah cluster optimal yang menghasilkan lima segmen pelanggan dengan karakteristik yang berbeda.

Hasil segmentasi menunjukkan bahwa segmen Champions hanya mencakup 0,14% dari total pelanggan, namun merupakan kelompok dengan nilai transaksi tertinggi sehingga menjadi pelanggan yang paling bernilai bagi perusahaan. Selain itu, pengujian transformasi logaritmik pada atribut Monetary menunjukkan bahwa penggunaan nilai asli dengan StandardScaler memberikan kualitas clustering yang lebih baik. Temuan tersebut menunjukkan bahwa integrasi RFM, K-Means, dan validasi multi-indeks mampu menghasilkan segmentasi pelanggan yang objektif sebagai dasar penyusunan strategi pemasaran yang lebih terarah. Penelitian selanjutnya dapat memvalidasi pendekatan ini menggunakan dataset *e-commerce* Indonesia untuk menilai konsistensi hasil pada konteks pasar lokal.

REFERENSI

- Badan Pusat Statistik, Statistik E-Commerce 2024. Jakarta, Indonesia: Badan Pusat Statistik, 2025.
- Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper

- Indonesia Global). KONSTELASI: Konvergensi Teknologi dan Sistem Informasi, 1(1), 204-215.
- Sulistyo, G. B., Hasan, N., Kiswati, S., & Muningsih, E. (2025). Segmentasi pelanggan dan optimalisasi penjualan pada data retail online berbasis model RFM. *CONTEN: Computer and Network Technology*, 5(1), 16-22.
- Simanjuntak, W., Hombing, N. M. B., & Wijaya, A. (2025). Analisis Segmentasi Pelanggan Menggunakan K-Means Untuk Personalisasi Penawaran Produk Dalam Konteks E-Commerce. *Jurnal Sains Dan Teknologi E-ISSN: 3063-9980*, 2(2), 111-117.
- Warianti, D. T., Astagina, P., Julianto, R., & Arini, F. Y. (2024). Optimasi K-Means Menggunakan Algoritma Firefly Untuk Segmentasi Pelanggan pada E-commerce. *JURNAL FASILKOM*, 14(3), 775-785.
- Mado, P. M. K., & Hendry, H. (2025). Implementasi algoritma clustering K-Means untuk segmentasi pelanggan di e-commerce. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 6(3), 1680-1686.
- Wiguno, T. C., & Nataliani, Y. (2022). Penerapan K-Means clustering berdasarkan analisis RFM terhadap segmentasi pembeli untuk meningkatkan strategi CRM (Disertasi doctoral). *J. Media Inform. Budidarma*, 6(4), 1871.
- Geni, Siti & Safitri, Hersa & Merry, Felisitas & Wati, Masna & H., Haviluddin. (2025). Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian. *JUKI : Jurnal Komputer dan Informatika*. 7. 89-99. 10.53842/juki.v7i1.1049.
- Sharyanto, Sharyanto & Lestari, Dwi. (2022). Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce. *JURIKOM (Jurnal Riset Komputer)*. 9. 866. 10.30865/jurikom.v9i4.4525.
- Fauzan, R. M., & Alfian, G. (2024). Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma Klasterisasi K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 170-177.
- Lukas, I., Okmayura, F., Irani, A. T., Juliastuti, E., Amirulhaq, M., Ardiansyah, R., & Fillia, S. (2025). PEMODELAN RFM & K-MEANS CLUSTERING UNTUK SEGMENTASI PELANGGAN DALAM PENJUALAN ONLINE. *Journal of Software Engineering and Information System (SEIS)*, 49-58.
- Hafidz Ardana, C., Khoyum, A. A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1-9.
- Andini, M., Catra, R., Homausyah, W., Aurelia, H., Meiriza, A., Tania, K., & Yamani, Z. (2026). Analisis Segmentasi Pelanggan Menggunakan RFM dan K-Means Clustering sebagai Dasar Penyusunan Aturan Pendukung Keputusan. *Building of Informatics, Technology and Science (BITS)*, 7(4), 2752-2760. <https://doi.org/10.47065/bits.v7i4.9511>
- Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-means clustering pada transaksi online retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122-137.
- Febri, A., Ningsih, N., & Lemantara, J. (2021). Aplikasi Analisis Segmentasi Pelanggan untuk Menentukan Strategi Pemasaran Menggunakan Kombinasi Metode K-Means dan Model RFM. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 139-151.