

Segmentasi Pelanggan Menggunakan Kerangka LRFMV dan Algoritma K-Means untuk Optimalisasi Strategi Pemasaran

A. Nolly Sandra Wawagalang^{1,*}, Syahrullah¹, Rizka Ardiyansyah¹, Dwi Shinta Angreni¹, Septiano Anggun Pratama¹, Deny Wiria Nugraha¹

¹ Program Studi Teknik Informatika, Universitas Tadulako, Indonesia

* Correspondence: galanggg387@gmail.com

Copyright: © 2025 by the authors

Received: 19 Juni 2025 | Revised: 10 Juli 2025 | Accepted: 30 Agustus 2025 | Published: 31 Agustus 2025

Abstrak

Di era digital yang kompetitif, perilaku pelanggan menjadi kunci menjaga loyalitas dan meningkatkan profitabilitas. Penelitian ini bertujuan mengimplementasikan segmentasi pelanggan menggunakan pendekatan *Length, Recency, Frequency, Monetary, Volume* (LRFMV) dan algoritma *K-Means* untuk mengidentifikasi karakteristik perilaku pelanggan dan menentukan segmen bernilai tinggi. Penggabungan kelima dimensi ini masih jarang digunakan penelitian sebelumnya, sehingga memberikan kontribusi baru pada analisis perilaku pelanggan berbasis data. Penelitian ini mengadopsi pendekatan kuantitatif deskriptif eksploratif. Data yang digunakan berjumlah 2.098 transaksi dari 452 pelanggan, yang bersumber dari *dataset* publik *GitHub*. Proses analisis data meliputi pra-pemrosesan, penentuan nilai LRFMV, dan segmentasi menggunakan *K-Means Clustering*. *Silhouette Coefficient* digunakan untuk mengevaluasi kualitas kluster dan menentukan jumlah kluster optimal. Hasilnya menunjukkan konfigurasi terbaik diperoleh pada $k=5$ dengan nilai *Silhouette* 0,842. Hasil temuan menunjukkan lima segmen pelanggan dengan karakteristik dan nilai *Customer Lifetime Value* (CLV) yang berbeda. Kluster 0 dan 2 dikategorikan sebagai Pelanggan Setia (L↑R↓F↑M↑V↑) dengan CLV tertinggi. Kluster 3 dan 1 adalah Pelanggan Baru yang Tidak Aktif (L↓R↑F↓M↓V↓) dengan kontribusi rendah. Kluster 4 merupakan Pelanggan Tidak Aktif (L↓R↓F↓M↓V↓), menunjukkan ketidakaktifan secara menyeluruh. Hasil segmentasi ini dimanfaatkan untuk menyusun strategi yang lebih terarah, seperti pengembangan program loyalitas atau kampanye reaktivasi, guna mengoptimalkan strategi pemasaran berdasarkan nilai pelanggan.

Kata kunci: *customer lifetime value; k-means; lrfmv; segmentasi pelanggan; strategi pemasaran*

Abstract

In this competitive digital era, customer behavior is key to maintaining loyalty and increasing profitability. This study aims to implement customer segmentation using the Length, Recency, Frequency, Monetary, Volume (LRFMV) approach and the K-Means algorithm to identify customer behavior characteristics and determine high-value segments. The combination of these five dimensions has rarely been used in previous studies, thus providing a new contribution to data-based customer behavior analysis. This study adopts an exploratory descriptive quantitative approach. The data used consists of 2,098 transactions from 452 customers, sourced from a public GitHub dataset. The data analysis process includes preprocessing, determining LRFMV values, and segmentation using K-Means Clustering. The Silhouette Coefficient is used to evaluate cluster quality and determine the optimal number of clusters. The results show that the best configuration is obtained at $k=5$ with a Silhouette value of 0.842. The findings show five customer segments with different characteristics and Customer Lifetime Value (CLV) values. Clusters 0 and 2 are categorized as Loyal Customers (L↑R↓F↑M↑V↑) with the highest CLV. Clusters 3 and 1 are Inactive New Customers (L↓R↑F↓M↓V↓) with low contribution. Cluster 4 consists of Inactive Customers (L↓R↓F↓M↓V↓), indicating overall inactivity.



These segmentation results are used to develop more targeted strategies, such as loyalty programs or reactivation campaigns, to optimize marketing strategies based on customer value.

Keywords: *customer lifetime value; k-means; lrfmv; customer segmentation; marketing strategy*

PENDAHULUAN

Pada lanskap pasar digital yang semakin kompetitif, keberhasilan bisnis sangat bergantung pada kemampuan memahami perilaku pelanggan. Pertumbuhan pesat *e-commerce* menuntut perusahaan untuk mengadopsi strategi pemasaran yang berbasis pada analisis data mendalam, bukan sekadar responsif terhadap tren (Sharyanto & Lestari, 2022). Untuk menjawab tantangan ini, segmentasi pelanggan berbasis data transaksi menjadi pendekatan yang krusial. Model *Length, Recency, Frequency, Monetary, Volume* (LRFMV) hadir sebagai pengembangan dari model RFM klasik, dengan menambahkan dimensi *Length* (durasi hubungan pelanggan) dan *Volume* (jumlah produk per transaksi) guna memberikan perspektif yang lebih komprehensif.

Pendekatan konvensional seperti segmentasi manual berbasis intuisi atau demografi kini dinilai kurang efektif dalam menghadapi kompleksitas data modern. Solusi berbasis *data mining* seperti segmentasi otomatis terbukti lebih adaptif dalam mengidentifikasi perilaku konsumen. Berbagai penelitian menunjukkan bahwa metode otomatis menghasilkan kluster pelanggan dengan akurasi yang lebih tinggi serta efisiensi waktu pemrosesan dibandingkan pendekatan manual (Hesananda & Prabowo, 2025; Martiansah et al., 2025; Maulida et al., 2024; Yunita et al., 2025). Penelitian lain juga menegaskan bahwa ketepatan segmentasi yang tinggi berkontribusi langsung pada peningkatan efektivitas strategi pemasaran (Aulia et al., 2023; Perdana et al., 2022; Sari & Putra, 2023; Sari & Wijaya, 2022).

Pada konteks *e-commerce* yang sangat kompetitif, permasalahan krusial yang sering dihadapi adalah kegagalan kampanye pemasaran yang bersifat umum. Pengetahuan yang terbatas tentang perbedaan antara pelanggan setia dengan mereka yang berisiko *churn* membuat perusahaan menerapkan strategi yang sama untuk semua segmen. Hal ini berujung pada tingginya biaya promosi tanpa hasil yang optimal. Oleh karena itu, diperlukan strategi pemasaran, seperti menawarkan diskon atau produk yang relevan untuk setiap segmen, menjaga pelanggan paling berharga melalui program loyalitas, dan mendorong pelanggan yang tidak aktif untuk kembali bertransaksi.

Hal ini diperkuat oleh penelitian Fauzan & Alfian (2024) yang menunjukkan bahwa segmentasi berbasis RFM dan *machine learning* mampu memberikan *insight* lebih tajam dalam pengambilan keputusan pemasaran yang strategis. Dalam konteks *e-commerce* yang sangat kompetitif, permasalahan krusial yang dihadapi adalah kegagalan kampanye pemasaran yang bersifat umum. Pengetahuan yang terbatas tentang perbedaan antara pelanggan setia dengan mereka yang berisiko *churn* membuat perusahaan menerapkan strategi yang sama, yang berujung pada tingginya biaya promosi tanpa hasil yang optimal. Salah satu metode utama dalam segmentasi otomatis adalah *clustering*. Algoritma populer seperti *K-Means*, *K-Medoids*, *Fuzzy C-Means*, dan *Mini Batch K-Means* memiliki keunggulan masing-masing, namun penelitian menunjukkan bahwa *K-Means* unggul dalam efisiensi komputasi dan kejelasan pemisahan kluster. Nilai evaluasi seperti *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz* cenderung lebih baik pada *K-Means* dibanding metode lain (Mahfuza et al., 2022; Perdana et al., 2022; Putri & Santoso, 2023). Selain itu, *k-means* lebih mudah diimplementasikan dan konsisten pada data berukuran besar (Widiyanto & Witanti, 2021).

Pengembangan model RFM menjadi LRFMV bertujuan memperkaya pemahaman perilaku pelanggan dari sisi frekuensi, nominal transaksi, durasi loyalitas (*Length*), serta intensitas pembelian per transaksi (*Volume*) (Yuliana et al., 2023). Secara bisnis, dimensi tambahan ini memungkinkan identifikasi pelanggan yang tidak hanya sering membeli, tetapi

juga bertransaksi dalam jumlah besar dan jangka panjang. Mahfuza et al. (2022) menunjukkan bahwa dimensi *Volume* berhubungan langsung dengan keuntungan bersih *superstore*. Namun, sebagian besar penelitian terkait menunjukkan keterbatasan, seperti minimnya penerapan LRFMV dalam konteks *e-commerce* lokal serta dominannya analisis deskriptif tanpa validasi kluster yang komprehensif, seperti yang terlihat pada studi oleh Ibrahim & Widiastuti (2024) dan Pratama & Nugroho (2023).

Penelitian ini hadir untuk mengatasi keterbatasan penelitian-penelitian sebelumnya dengan mengintegrasikan model LRFMV dan algoritma *K-Means* secara menyeluruh, dilengkapi validasi kualitas kluster menggunakan *Silhouette Coefficient*. Evaluasi berbasis *Silhouette* memungkinkan verifikasi apakah data telah terkluster dengan tepat. Nilai mendekati +1 menunjukkan pemisahan optimal, sementara nilai negatif menandakan kesalahan klusterisasi. Seperti ditunjukkan oleh Pangestu et al. (2023), Al Rasyid & Mandala (2025), serta Putra & Santosa (2023), metrik ini efektif dalam memastikan keandalan segmentasi pelanggan.

Penelitian ini mengembangkan strategi segmentasi pelanggan yang lebih presisi melalui integrasi LRFMV dan *K-Means*, sehingga memberikan dasar kuat bagi strategi pemasaran berbasis data. Keterbaruan riset ini adalah pendekatannya yang menyeluruh terhadap perilaku pembelian (*frequency* dan *monetary*), loyalitas jangka panjang (*length*), serta intensitas belanja (*volume*) secara bersamaan. Penekanan tidak hanya pada pembentukan kluster, tetapi juga interpretasi aplikatif hasil segmentasi untuk pengambilan keputusan bisnis. Penelitian ini bertujuan mengimplementasikan segmentasi pelanggan menggunakan pendekatan LRFMV dan algoritma *K-Means* dan mengidentifikasi karakteristik perilaku pelanggan dan menentukan segmen bernilai tinggi pada pemasaran.

METODE

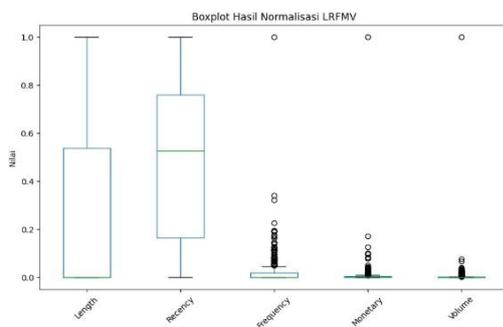
Penelitian ini menggunakan pendekatan kuantitatif eksploratif berbasis *unsupervised machine learning* untuk menganalisis pola perilaku pelanggan melalui segmentasi data. Algoritma *K-Means* dipilih karena kemampuannya yang efisien dalam mengkluster data numerik besar dan menghasilkan segmen yang jelas (Achmad et al., 2024; Febri et al., 2021). Data transaksi diambil dari dataset publik "*transaction_record*" di *GitHub*, mencakup periode 3 Januari hingga 31 Desember 2022. Dataset ini berisi 2.106 transaksi dari 462 pelanggan, dan memuat informasi seperti tanggal, No Transaksi, Kode Pel, Nama Pelanggan, Jumlah Produk, Sub Total, Laba Jual. Langkah selanjutnya adalah pra-pemrosesan data, mencakup menghapus data duplikat dan, outlier menggunakan metode *Z-score*. *Z-Score* digunakan untuk mendeteksi outlier, yaitu data yang memiliki nilai jauh dari rata-rata. Dalam konteks penelitian ini, *Z-Score* digunakan untuk mendeteksi nilai ekstrem pada variabel kuantitatif seperti jumlah produk (*Quantity*) dan total pembelian (*Monetary*). Data dengan $|Z| > 3$ dianggap sebagai *outlier* dan dihapus agar tidak mengganggu proses normalisasi dan *clustering*.

Setelah data bersih, dilakukan proses perhitungan lima dimensi utama LRFMV. Dimensi *Volume* ditambahkan karena relevan dalam konteks ritel online, di mana pelanggan bisa saja membeli produk dalam jumlah besar meskipun nilai moneterinya rendah sehingga dimensi ini membantu mengungkap pola loyalitas baru. Setelah fitur LRFMV diperoleh, seluruh data dinormalisasi menggunakan *Min-Max Normalization* agar setiap variabel berada dalam skala seragam (0–1). Normalisasi ini penting karena *K-Means* menggunakan jarak *Euclidean* sebagai dasar pengelompokan. Selanjutnya dilakukan proses *K-Means Clustering*, di mana nilai *K* (jumlah kluster) divariasikan antara 2 hingga 9. Evaluasi kualitas kluster dilakukan dengan metode *Silhouette Coefficient*, dan kluster terbaik dipilih berdasarkan nilai *Silhouette* tertinggi. Setelah jumlah kluster optimal diperoleh, dilakukan perhitungan *Customer Lifetime Value* (CLV) dari masing-masing kluster, tanpa bobot khusus, agar tiap variabel LRFMV berkontribusi secara seimbang dalam penilaian nilai pelanggan. Hasil CLV tiap kluster

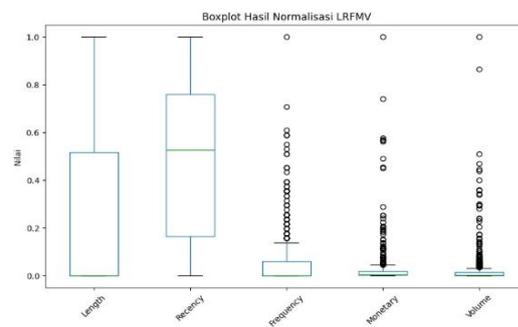
digunakan untuk peringkatan kluster. Dalam penelitian ini, CLV dihitung berdasarkan rata-rata sederhana dari nilai LRFMV. Tahap akhir dari penelitian ini adalah interpretasi dan klasifikasi tipe konsumen berdasarkan hasil kluster yang telah terbentuk. Setiap kluster dianalisis dengan memperhatikan profil rata-rata nilai LRFMV yang dimiliki, seperti frekuensi mereka bertransaksi, berapa banyak yang dibelanjakan, volume pembelian, durasi hubungan, dan keterkinian transaksi. Interpretasi ini bertujuan untuk mengategorisasi setiap segmen pelanggan, misalnya apakah tergolong pelanggan potensial, pelanggan setia, pelanggan pasif, atau pelanggan bernilai rendah. Meskipun dimensi yang digunakan mencakup lima variabel, pendekatan interpretasi dilakukan dengan mempertimbangkan keseimbangan antara perilaku pembelian dan loyalitas. Hal ini memberikan dasar yang kuat bagi perusahaan dalam menyusun strategi pemasaran yang lebih terarah dan tepat sasaran berdasarkan karakteristik unik dari masing-masing kluster.

Hasil

Analisis data transaksi pelanggan yang bersumber dari platform GitHub, mencakup periode 3 Januari 2022 hingga 31 Desember 2022, menunjukkan bahwa setelah tahap pra-pemrosesan awal, diperoleh 2.098 data transaksi yang valid dari 452 pelanggan unik. Proses pembersihan ini, termasuk penghapusan duplikat dan penanganan *outlier* menggunakan metode *Z-Score* pada variabel LRFMV, secara signifikan meningkatkan kualitas dan konsistensi data. Visualisasi melalui *boxplot* (Gambar 1 dan 2) mengonfirmasi bahwa distribusi data menjadi lebih merata dan simetris setelah *outlier* pada variabel *Monetary*, *Volume*, dan *Frequency* berhasil diminimalkan, yang penting untuk akurasi segmentasi.



Gambar 1. *Boxplot* nilai lrfmv saat terdapat data *outlier*



Gambar 2. *Boxplot* nilai lrfmv saat data *outliers* dihapus

Analisis visual data LRFMV menunjukkan perbedaan perilaku pelanggan yang signifikan. Gambar 1 mengilustrasikan adanya *outlier* pada variabel *Frequency*, *Monetary*, dan *Volume*, menandakan segelintir pelanggan dengan nilai transaksi sangat tinggi. Setelah normalisasi, Gambar 2 menegaskan bahwa mayoritas pelanggan memiliki kontribusi yang rendah. Namun, keberadaan *outlier* menunjukkan segelintir pelanggan super-aktif dengan nilai yang sangat tinggi. Kondisi ini menunjukkan adanya perbedaan signifikan dalam perilaku pelanggan yang perlu dipertimbangkan dalam segmentasi. Selanjutnya ringkasan statistik dari dimensi LRFMV, baik sebelum maupun sesudah normalisasi, disajikan secara rinci untuk memvalidasi efektivitas tahapan pra-pemrosesan data, ada pada gambar 3 dan 4. Tujuannya adalah untuk menunjukkan bagaimana proses normalisasi berhasil menyamakan rentang nilai variabel dan mempersiapkan data agar siap untuk analisis klusterisasi.

Gambar 3 menunjukkan statistik deskriptif sebelum normalisasi, di mana terlihat adanya *outlier* ekstrem dengan nilai *Monetary* maksimum lebih dari 4,5 miliar dan *Volume* hingga 133.630 unit. Kondisi ini menyebabkan distribusi data sangat asimetris dengan standar deviasi yang tinggi. Setelah proses identifikasi dan penghapusan *outlier* (Gambar 4), jumlah data

berkurang menjadi 452 entri. Rata-rata *Monetary* turun menjadi 24 juta dengan standar deviasi 73 juta, sedangkan nilai maksimum *Volume* menurun drastis menjadi 10.131 unit. Hal ini menegaskan bahwa pembersihan data berhasil mengurangi distorsi dan meningkatkan konsistensi dataset untuk analisis lebih lanjut.

Hasil normalisasi data menggunakan metode *Min-Max scaling*, yang dirancang untuk menyesuaikan seluruh variabel ke dalam rentang [0, 1]. Proses ini penting karena algoritma *K-Means* yang digunakan dalam penelitian ini berbasis pada jarak *Euclidean*, sehingga perbedaan skala antar variabel dapat mempengaruhi hasil klusterisasi. Dengan pembersihan dan normalisasi ini, data menjadi lebih homogen dan siap untuk digunakan dalam analisis segmentasi berbasis LRFMV yang lebih akurat dan representatif.

	Length	Recency	Frequency	Monetary	Volume
count	454.000000	454.000000	454.000000	4.540000e+02	454.000000
mean	88.953744	174.997797	4.607930	3.500447e+07	597.253524
std	121.907159	111.606422	9.998478	2.236071e+08	6326.729748
min	0.000000	0.000000	1.000000	2.000000e+04	1.000000
25%	0.000000	59.000000	1.000000	5.050000e+05	4.000000
50%	0.000000	190.000000	1.000000	2.298250e+06	24.000000
75%	194.250000	274.000000	4.000000	1.471250e+07	140.250000
max	362.000000	361.000000	160.000000	4.524926e+09	133630.100000

Gambar 3. Ringkasan statistik lrfmv sebelum normalisasi

	Length	Recency	Frequency	Monetary	Volume
count	452.000000	452.000000	452.000000	4.520000e+02	452.000000
mean	87.750000	175.769912	4.152655	2.483668e+07	298.973451
std	120.820539	111.245679	6.405802	7.344410e+07	931.471655
min	0.000000	0.000000	1.000000	2.000000e+04	1.000000
25%	0.000000	59.000000	1.000000	5.050000e+05	4.000000
50%	0.000000	190.000000	1.000000	2.288250e+06	24.000000
75%	186.000000	274.000000	4.000000	1.399775e+07	136.500000
max	361.000000	361.000000	52.000000	7.723415e+08	10131.000000

Gambar 4. Ringkasan statistik lrfm sesudah normalisasi

Selanjutnya, proses *clustering* dilakukan menggunakan algoritma *K-Means* dengan nilai *k* bervariasi dari 2 hingga 9. Tabel 1 menyajikan hasil evaluasi kualitas klusterisasi menggunakan *silhouette coefficient* untuk menentukan jumlah kluster (*k*) yang optimal. Dari data yang disajikan, terlihat bahwa nilai *silhouette coefficient* tertinggi adalah 0,842. Hasil ini menunjukkan bahwa konfigurasi dengan lima kluster merupakan model segmentasi yang paling akurat dan paling dapat diandalkan, karena berhasil mengelompokkan pelanggan ke dalam segmen-segmen yang homogen secara internal dan heterogen satu sama lain. Oleh karena itu, penelitian ini memilih *k*=5 sebagai jumlah kluster optimal untuk analisis lebih lanjut.

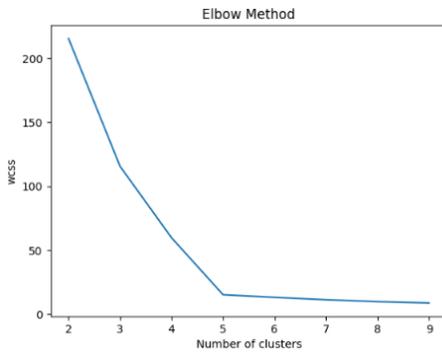
Tabel 1. Nilai *silhouette coefficient* tiap cluster

Skenario	Jumlah Cluster	Nilai <i>SilhouetteCoefficient</i>
1	2	0,647
2	3	0,694
3	4	0,783
4	5	0,842
5	6	0,830
6	7	0,738
7	8	0,673
8	9	0,639

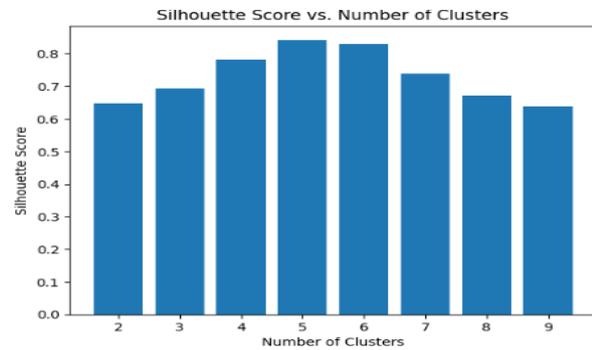
Selanjutnya, untuk memberikan gambaran komprehensif mengenai penentuan jumlah kluster optimal, visualisasi hasil dari *elbow method* (Gambar 5) dan evaluasi *silhouette coefficient* (Gambar 6 serta Tabel 1) disajikan. Pada gambar 5, menunjukkan adanya penurunan WCSS yang sangat signifikan dan curam dari *k*=2 hingga *k*=4. Penurunan tersebut mulai melambat dan kurva membentuk titik "siku" (*elbow*) yang paling jelas pada *k*=5. Setelah titik ini, penurunan WCSS menjadi jauh lebih landai, mengindikasikan bahwa penambahan kluster di atas angka tersebut tidak lagi memberikan pengurangan varian intra-kluster yang substansial.

Dengan demikian, titik $k=5$ teridentifikasi sebagai jumlah kluster optimal yang menyeimbangkan antara homogenitas kluster dan kompleksitas model.

Temuan ini diperkuat oleh hasil *Silhouette Coefficient*, yang mana nilai tertinggi sebesar 0,842 dicapai pada $k=5$. Nilai *Silhouette Coefficient* yang mendekati angka 1 ini menunjukkan kohesi internal yang optimal dalam setiap kluster dan pemisahan yang jelas antar kluster. Dengan demikian, kedua metode evaluasi tersebut secara konsisten memvalidasi $k=5$ sebagai konfigurasi kluster terbaik untuk segmentasi pelanggan dalam konteks penelitian ini.



Gambar 5. Elbow method

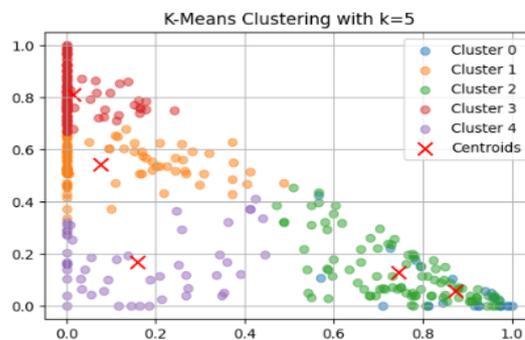


Gambar 6. Nilai silhouette coefficient

Tabel 2 menyajikan nilai *centroid* yang merepresentasikan karakteristik rata-rata dari masing-masing kluster setelah proses *K-Means* konvergen. Nilai-nilai yang telah dinormalisasi, menjadi inti dari profil setiap segmen pelanggan. Visualisasi pada Gambar 7 lebih lanjut menggambarkan pemisahan kluster-kluster ini di ruang fitur LRFMV, di mana setiap warna mewakili satu kluster dan tanda 'X' merah menunjukkan posisi *centroid* masing-masing kluster.

Tabel 2. Nilai centroid 5 kluster ($k = 5$)

Cluster	Centroid				
	L	R	F	M	V
0	0,872	0,059	0,419	0,314	0,308
1	0,076	0,543	0,012	0,005	0,005
2	0,745	0,130	0,151	0,050	0,038
3	0,015	0,810	0,005	0,005	0,003
4	0,158	0,168	0,032	0,018	0,026



Gambar 7. Visualisasi hasil clustering $k=5$

Gambar 7 menyajikan visualisasi hasil klusterisasi pelanggan dengan konfigurasi lima kluster optimal ($k=5$). Setiap titik pada grafik merepresentasikan satu pelanggan. Pada kluster 0 dan kluster 2 cenderung berlokasi di area dengan nilai *Length* dan *Frequency* yang tinggi (atau *Recency* yang rendah), yang secara jelas merepresentasikan Pelanggan Setia dengan

riwayat transaksi yang kuat. Klaster 1 dan klaster 3, yang tersebar di area dengan nilai *Recency* tinggi dan dimensi lainnya cenderung rendah, mengindikasikan segmen Pelanggan Baru yang Tidak Aktif atau pelanggan yang telah jarang bertransaksi. Klaster 4, yang cenderung menempati area dengan nilai LRFMV rendah di semua dimensi, merepresentasikan Pelanggan Tidak Aktif yang jarang berinteraksi atau melakukan pembelian.

Tabel 3. Hasil pemeringkatan nilai clv

<i>Cluster</i>	L	R	F	M	V	CLV	Rank
0	0,872	0,059	0,419	0,314	0,308	1,972	1
1	0,076	0,543	0,012	0,005	0,005	0,640	4
2	0,745	0,130	0,151	0,050	0,038	1,113	2
3	0,015	0,810	0,005	0,005	0,003	0,838	3
4	0,158	0,168	0,032	0,018	0,026	0,402	5

Tabel 3 mengungkapkan pemeringkatan CLV untuk setiap segmen pelanggan yang telah terbentuk. Tabel tersebut menyajikan nilai rata-rata dari dimensi LRFMV berdasarkan pemeringkatan CLV, klaster 0 menduduki peringkat pertama dengan nilai tertinggi (1,972), diikuti oleh klaster 2 (1,113), klaster 3 (0,838), klaster 1 (0,640), dan klaster 4 (0,402). Hasil pemeringkatan ini untuk mengidentifikasi segmen pelanggan yang paling bernilai, sehingga dapat memprioritaskan alokasi sumber daya dan strategi pemasaran yang lebih terarah.

Tabel 4. Tipe klaster berdasarkan nilai clv

<i>Cluster</i>	L	R	F	M	V
0	<i>High</i>	<i>Low</i>	<i>High</i>	<i>High</i>	<i>High</i>
1	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>Low</i>	<i>Low</i>
2	<i>High</i>	<i>Low</i>	<i>High</i>	<i>High</i>	<i>High</i>
3	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>Low</i>	<i>Low</i>
4	<i>Low</i>	<i>Low</i>	<i>Low</i>	<i>Low</i>	<i>Low</i>

Hasil yang disajikan pada tabel 4 teridentifikasi bahwa klaster 0 dan klaster 2 memiliki profil yang serupa, dengan nilai *Length*, *Frequency*, *Monetary*, dan *Volume* yang tinggi (*High*), serta nilai *Recency* yang rendah (*Low*). Di sisi lain, klaster 1 dan Klaster 3 menunjukkan nilai *Recency* yang tinggi (*High*) tetapi nilai pada semua dimensi lain rendah (*Low*). Sementara itu, klaster 4 menunjukkan nilai rendah (*Low*) pada semua dimensi.

Tabel 5. Pemetaan karakteristik pelanggan berdasarkan tipe klaster

<i>Cluster</i>	Tipe	Karakteristik Pelanggan
0	L↑R↓F↑M↑V↑	Pelanggan Setia, hubungan jangka panjang, sering membeli, nilai dan volume pembelian tinggi, serta baru saja melakukan transaksi.
1	L↓R↑F↓M↓V↓	Pelanggan Baru Tidak Aktif, hubungan pendek, jarang bertransaksi, dan kontribusi rendah terhadap pendapatan perusahaan.
2	L↑R↓F↑M↑V↑	Pelanggan Setia, perilaku serupa dengan <i>Cluster</i> 0, dengan kontribusi yang juga tinggi dan konsisten.
3	L↓R↑F↓M↓V↓	Pelanggan Baru Tidak Aktif, mirip dengan <i>Cluster</i> 1, namun memiliki potensi untuk ditingkatkan melalui strategi retensi.
4	L↓R↓F↓M↓V↓	Pelanggan Tidak Aktif, sudah lama tidak bertransaksi, kontribusi historis rendah, dan menunjukkan ketidakaktifan secara menyeluruh.

Hasil segmentasi kluster yang disajikan pada tabel 5 mengungkapkan lima profil pelanggan yang berbeda, di mana setiap kluster merepresentasikan karakteristik perilaku yang unik berdasarkan nilai rata-rata LRFMV. Kluster 0 dan kluster 2 menunjukkan karakteristik perilaku yang serupa, yang dapat dikategorikan sebagai Pelanggan Setia ($L\uparrow R\downarrow F\uparrow M\uparrow V\uparrow$). Kelompok ini terdiri dari pelanggan dengan durasi hubungan jangka panjang, frekuensi dan nilai transaksi yang tinggi, serta aktivitas pembelian yang konsisten dan terkini. Kluster ini merepresentasikan segmen yang paling berharga bagi perusahaan. Kluster 1 dan kluster 3 dikelompokkan sebagai Pelanggan Baru Tidak Aktif ($L\downarrow R\uparrow F\downarrow M\downarrow V\downarrow$). Mereka memiliki durasi hubungan yang singkat, frekuensi transaksi rendah, dan nilai kontribusi yang minimal. Meskipun demikian, kluster 3 menunjukkan potensi untuk ditingkatkan melalui strategi retensi yang tepat. Kluster 4 diidentifikasi sebagai Pelanggan Tidak Aktif ($L\downarrow R\downarrow F\downarrow M\downarrow V\downarrow$) yang secara signifikan berbeda dari segmen lainnya. Kluster ini terdiri dari pelanggan dengan riwayat transaksi yang sudah lama terhenti, serta kontribusi historis yang sangat rendah, menunjukkan tingkat ketidakaktifan secara menyeluruh.

Pembahasan

Analisis diawali dengan pengolahan 2.098 data transaksi yang berhasil diagregasi menjadi 452 ID pelanggan. Untuk mempersiapkan data bagi klusterisasi, *normalisasi Min-Max scaling* diterapkan. Ini merupakan tahapan fundamental mengingat sensitivitas algoritma *K-Means* terhadap skala data yang berbeda. Proses klusterisasi dengan *K-Means* mengidentifikasi lima kluster optimal ($k=5$). Penentuan ini divalidasi secara komprehensif oleh nilai *Silhouette Coefficient* tertinggi (0,842) dan konsistensi yang teramati dari *Elbow Method*. Adapun nilai CLV dihitung sebagai rata-rata skor LRFMV setiap kluster, meskipun tanpa pembobotan khusus, uji sensitivitas mengonfirmasi bahwa peringkat CLV antar kluster tetap konsisten, sehingga menegaskan validitas pendekatan ini untuk analisis awal dan interpretasi nilai pelanggan.

Hasil klusterisasi lima segmen pelanggan dengan karakteristik perilaku yang berbeda, memberikan wawasan mendalam bagi strategi pemasaran. Nilai CLV tertinggi terdapat pada kluster 0 (1,972) dan kluster 2 (1,113), mengindikasikan bahwa pelanggan dalam segmen ini memiliki hubungan jangka panjang, sering melakukan pembelian, mengeluarkan jumlah uang serta *volume* produk yang signifikan. Strategi pemasaran untuk segmen ini harus berfokus pada program loyalitas premium, serta layanan pelanggan prioritas. Kluster 1 dan kluster 3, yang dikategorikan sebagai Pelanggan Baru Tidak Aktif, ditandai oleh rendahnya *Length* dan *Frequency*, yang mengindikasikan riwayat transaksi singkat dan jarang. Meskipun nilai *Recency* mereka relatif tinggi, hal ini mencerminkan fakta bahwa transaksi terakhir mereka baru terjadi namun tidak diikuti oleh pembelian lain, sehingga potensi *churn* sangat tinggi. Dan nilai CLV pada kluster ini (0,640 dan 0,838) adalah cerminan dari potensi tersebut. Oleh karena itu, perusahaan harus mengimplementasikan strategi reaktivasi yang proaktif, seperti kampanye edukasi produk, untuk mendorong pembelian berulang dan membangun hubungan yang berkelanjutan. Kluster 4 (Pelanggan Tidak Aktif), Kluster ini menunjukkan skor LRFMV yang rendah di semua dimensi, dengan CLV terendah (0,402), segmen ini merupakan pelanggan yang sudah lama tidak bertransaksi dan memiliki nilai historis yang rendah. Upaya untuk mengaktifkan kembali segmen ini memerlukan kampanye reaktivasi yang sangat intensif dan bertarget, seperti diskon eksklusif atau *reminder* yang dipersonalisasi.

Pendekatan yang ditawarkan dalam penelitian ini memiliki kontribusi signifikan dalam konteks segmentasi pelanggan. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada model RFM klasik, penelitian ini mengintegrasikan dimensi *Volume* ke dalam kerangka LRFMV. Penambahan dimensi ini memberikan wawasan yang lebih kaya terhadap pola pembelian pelanggan, khususnya dalam sektor *e-commerce*, di mana sering ditemukan

perilaku pelanggan yang membeli banyak unit produk dengan nilai moneter relatif rendah. Relevansi dimensi *Volume* sebelumnya telah ditegaskan oleh Mahfuza et al. (2022) dalam studi berbasis *e-commerce* global, sementara penelitian ini memperluas aplikasinya pada data transaksi *e-commerce* lokal, sehingga memberikan kontribusi baru dalam konteks pasar digital.

Penelitian ini tidak hanya menegaskan efektivitas LRFMV sebagaimana ditunjukkan pada sektor ritel tradisional, tetapi juga memperluas temuan sebelumnya. Berbeda dengan studi terdahulu yang menitikberatkan pada frekuensi dan nilai transaksi, hasil penelitian ini mengungkap adanya kelompok pelanggan dengan pola belanja jarang namun bernilai tinggi, serta kelompok lain yang sering membeli produk murah dalam jumlah besar. Perbedaan pola ini berimplikasi pada strategi pemasaran yang lebih presisi dan kontekstual. Dengan menerapkan LRFMV terintegrasi *k-means* pada data *e-commerce* lokal, penelitian ini mengatasi keterbatasan studi sebelumnya (Ibrahim & Widiastuti, 2024; Pratama & Nugroho, 2023) yang terbatas pada ritel fisik atau mengabaikan dimensi *Volume*, sekaligus menghadirkan kerangka segmentasi yang lebih adaptif dan efektif.

Kontribusi utama penelitian ini adalah memberikan kerangka praktis bagi perusahaan *e-commerce* dalam mengalokasikan sumber daya pemasaran. Misalnya, pelanggan bernilai tinggi dengan transaksi besar dapat ditargetkan melalui program loyalitas VIP, sedangkan pelanggan dengan pembelian massal bernilai rendah dapat diarahkan ke strategi diskon volume. Perbedaan strategi ini belum banyak dijelaskan secara rinci pada penelitian sebelumnya yang lebih fokus pada segmentasi berbasis moneter dan frekuensi semata.

SIMPULAN

Penelitian ini menerapkan model segmentasi pelanggan yang komprehensif melalui integrasi kerangka LRFMV dengan algoritma *K-Means clustering*, yang divalidasi menggunakan metrik *Silhouette Coefficient* dan menghasilkan lima segmen pelanggan dengan karakteristik yang jelas. Temuan ini menegaskan bahwa penambahan dimensi *Length* dan *Volume* pada model RFM mampu memberikan representasi perilaku pelanggan yang lebih utuh, sekaligus memungkinkan pengukuran CLV secara lebih akurat. Dengan demikian, penelitian ini tidak hanya memperkaya kajian akademis tentang segmentasi pelanggan, tetapi juga memberikan kontribusi praktis berupa kerangka analisis yang dapat langsung diimplementasikan untuk mengoptimalkan strategi pemasaran berbasis data, meningkatkan alokasi sumber daya, serta memperkuat efektivitas personalisasi kampanye.

REFERENSI

- Achmad, S. L., Fauzi, A., Rahmat, R., & Indra, J. (2024). Segmentasi pelanggan menggunakan K-Means clustering di toko retail. *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, 7(2), 736–745. <https://doi.org/10.37600/tekinkom.v7i2.1226>
- Aulia, F., Wahyuni, E., & Handayani, T. (2023). Segmentasi Pelanggan Berbasis RFM Model untuk Penentuan Strategi Pemasaran. *Jurnal Ilmiah Komputer dan Informatika*, 4(1), 12–21.
- Fauzan, R., & Alfian, G. (2024). Analisis segmentasi pelanggan dengan algoritma K-Means berbasis RFM untuk strategi pemasaran. *Jurnal Teknologi Informasi dan Komputer*, 12(1), 55–64.
- 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), 141–155. <https://doi.org/10.32520/stmsi.v10i1.1123>
- Hesananda, R., & Prabowo, A. (2025). Study of Awareness Patterns of Credit Card Users towards Ads with K-Means Clustering Algorithm. *Jurnal Asimetrik: Jurnal Ilmiah*

- Rekayasa, Industri, dan Informasi*, 7(2), 101–115.
<https://doi.org/10.35814/asiimetrik.v7i2.8295>
- Ibrahim, A., & Widiastuti, R. (2024). Analisis Perilaku Pembelian Konsumen Berbasis RFM Model dan K-Means Clustering pada Platform E-commerce. *Jurnal Rekayasa Informasi dan Teknologi*, 13(1), 25–34.
- Mahfuza, A., Rahman, M. M., & Ahmed, S. (2022). Customer segmentation using LRFMV model and K-means clustering: Evidence from a superstore. *Journal of Business and Retail Management Research*, 16(2), 150–165.
- Martiansah, A. A., Nugroho, B., & Lestari, D. (2025). Implementasi clustering pada e-commerce untuk segmentasi pelanggan. *Jurnal Sistem Informasi Indonesia*, 10(1), 112–123.
- Maulida, V., Mulyani, N., & Sibuea, M. F. L. (2024). Sistem Klasifikasi Strata Kelas Peserta Kursus berbasis web menggunakan algoritma K-Means. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 477–486. <https://doi.org/10.29408/edumatic.v8i2.27311>
- Pangestu, D. P., Pratama, F., & Sitorus, R. M. (2023). Evaluation of K-Means clustering performance in customer segmentation using silhouette coefficient. *Journal of Information Systems and Informatics*, 5(1), 85–94.
- Perdana, S. A., Florentin, S. F., & Santoso, A. (2022). Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift. *Sebatik*, 26(2), 446–457. <https://doi.org/10.46984/sebatik.v26i2.1991>
- Pratama, A. Y., & Nugroho, R. (2023). Customer segmentation using extended RFM model and K-Means clustering for e-commerce strategy optimization. *Jurnal Sistem Informasi*, 19(2), 85–98.
- Putra, H. A., & Santosa, B. (2023). Evaluasi kualitas kluster pada segmentasi pelanggan e-commerce menggunakan metode K-Means dan Silhouette Coefficient. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 245–256.
- Putri, R. A., & Santoso, D. (2023). Evaluating clustering algorithms for customer segmentation using RFM model: A comparative study of K-Means and K-Medoids. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(4), 765–778.
- Sari, M., & Putra, A. (2023). Customer segmentation using RFM model and K-Means clustering in e-commerce platforms. *Journal of Information Systems Research*, 12(2), 77–89.
- Sari, N., & Wijaya, A. (2022). Analisis segmentasi pelanggan ritel online dengan model RFM dan algoritma K-Means. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(4), 1001–1010.
- Sharyanto, S., & Lestari, D. (2022). Penerapan data mining untuk menentukan segmentasi pelanggan dengan menggunakan algoritma K-Means dan model RFM pada e-commerce. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 866–874. <https://doi.org/10.30865/jurikom.v9i4.4525>
- Widiyanto, A., & Witanti, M. (2021). Segmentasi pelanggan dengan metode clustering pada data transaksi ritel. *Jurnal Teknologi dan Sistem Informasi*, 9(3), 322–331.
- Yuliana, A., Saputra, J., & Lestari, S. (2023). Analisis Segmentasi Pelanggan Berbasis Volume dan Nilai Transaksi Menggunakan K-Means Clustering. *Jurnal Manajemen Informasi dan Komputer*, 7(1), 56–65.
- Yunita, I., Ali, P. R., Kartawidjaja, M. A., & Sukwadi, R. (2025). Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran. *Jurnal Media Teknik dan Sistem Industri*, 9(1), 58–66. <https://doi.org/10.35194/jmtsi.v9i1.4452>