

# Segmentasi Pelanggan E-Commerce Menggunakan Pendekatan RFM (Recency, Frequency, Monetary) dan Algoritma K-Means Clustering

<sup>1</sup>Bagus Riyadi Fitra, <sup>2</sup>Genta Aprilian, <sup>3</sup>Reza Rifaldy Pratama

<sup>1,2,3</sup>Program Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

Email: <sup>1</sup>[bagus.riyadi555@gmail.com](mailto:bagus.riyadi555@gmail.com), <sup>2</sup>[gntaprln@gmail.com](mailto:gntaprln@gmail.com), <sup>3</sup>[rezarfldy@gmail.com](mailto:rezarfldy@gmail.com)

## Abstract

*Increasingly tight competition in the e-commerce industry requires businesses to gain a deeper understanding of customer behavior so that marketing strategies can be more precisely targeted. A widely used approach for understanding customer value and behavior is RFM (Recency, Frequency, Monetary) analysis, combined with an unsupervised machine learning algorithm, K-Means Clustering. This study aims to segment e-commerce customers based on three main variables: Recency, Frequency, and Monetary. The research stages include data standardization using StandardScaler, determination of the optimal number of clusters using the Elbow Method and Silhouette Score, clustering using the K-Means algorithm, and interpretation of each cluster's characteristics. The results show that 8,635 customers were divided into four clusters with distinct characteristics: champions, loyal customers, occasional customers, and at-risk/lost customers. These segmentation results can serve as a basis for developing more personalized and efficient marketing strategies for each customer group.*

**Keywords:** Customer Segmentation, RFM Analysis, K-Means Clustering, E-Commerce, Customer Relationship Management.

## Abstrak

Persaingan yang semakin ketat di industri e-commerce mengharuskan bisnis untuk mendapatkan pemahaman yang lebih dalam tentang perilaku pelanggan sehingga strategi pemasaran dapat ditargetkan dengan lebih tepat. Pendekatan yang banyak digunakan untuk memahami nilai dan perilaku pelanggan adalah analisis RFM (Recency, Frequency, Monetary), dikombinasikan dengan algoritma pembelajaran mesin tanpa pengawasan, K-Means Clustering. Studi ini bertujuan untuk melakukan segmentasi pelanggan e-commerce berdasarkan tiga variabel utama: Recency, Frequency, dan Monetary. Tahapan penelitian meliputi standarisasi data menggunakan StandardScaler, penentuan jumlah cluster optimal menggunakan Metode Elbow dan Silhouette Score, clustering menggunakan algoritma K-Means, dan interpretasi karakteristik setiap cluster. Hasil menunjukkan bahwa 8.635 pelanggan terbagi menjadi empat cluster dengan karakteristik yang berbeda: champion, pelanggan setia, pelanggan sesekali, dan pelanggan berisiko/hilang. Hasil segmentasi ini dapat menjadi dasar untuk mengembangkan strategi pemasaran yang lebih personal dan efisien untuk setiap kelompok pelanggan.

Kata Kunci: Analisis RFM, K-Means Clustering, E-Commerce, Manajemen Hubungan Pelanggan.

## A. PENDAHULUAN

Pertumbuhan transaksi digital telah mendorong platform e-commerce untuk mengelola data pelanggan dalam jumlah besar setiap harinya. Setiap pelanggan memiliki pola perilaku transaksi yang berbeda-beda, baik dari sisi seberapa baru mereka bertransaksi, seberapa sering, maupun seberapa besar nilai belanjanya. Apabila seluruh pelanggan diperlakukan secara seragam dalam strategi pemasaran, perusahaan berisiko memboroskan anggaran promosi pada pelanggan yang sebenarnya kurang potensial, sekaligus gagal memberikan perhatian khusus pada pelanggan yang benar-benar bernilai tinggi.

Salah satu pendekatan klasik dalam Customer Relationship Management (CRM) untuk mengatasi permasalahan ini adalah analisis RFM (Recency, Frequency, Monetary),

yang pertama kali diperkenalkan oleh Hughes (1994) sebagai metode untuk menilai pelanggan berdasarkan riwayat transaksinya. Konsep RFM kemudian banyak dikombinasikan dengan algoritma pembelajaran mesin tanpa label, salah satunya K-Means Clustering, untuk mengelompokkan pelanggan ke dalam segmen-segmen yang lebih objektif dan dapat ditindaklanjuti secara bisnis (Barus et al., 2023; Robo et al., 2024).

Berdasarkan latar belakang tersebut, penelitian ini melakukan segmentasi pelanggan pada data transaksi e-commerce menggunakan kombinasi analisis RFM dan algoritma K-Means Clustering, dengan penentuan jumlah

kluster optimal melalui Elbow Method dan Silhouette Score.

### 1.1 Rumusan Masalah

Bagaimana karakteristik pelanggan e-commerce jika dikelompokkan berdasarkan variabel Recency, Frequency, dan Monetary?

Berapa jumlah kluster optimal yang dapat dibentuk dari data RFM tersebut menggunakan Elbow Method dan Silhouette Score.

Bagaimana profil dan strategi pemasaran yang sesuai untuk masing-masing kluster pelanggan yang terbentuk?

### 1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk: (1) membentuk variabel RFM dari data transaksi pelanggan e-commerce; (2) menentukan jumlah kluster optimal menggunakan Elbow Method dan Silhouette Score; (3) melakukan segmentasi pelanggan menggunakan algoritma K-Means Clustering; serta (4) menginterpretasikan karakteristik setiap kluster sebagai dasar rekomendasi strategi pemasaran.

### 1.3 Manfaat Penelitian

Secara akademis, penelitian ini diharapkan dapat menjadi referensi penerapan kombinasi RFM Analysis dan K-Means Clustering pada domain e-commerce. Secara praktis, hasil segmentasi diharapkan dapat membantu pelaku usaha e-commerce dalam menyusun strategi retensi, akuisisi, dan reaktivasi pelanggan yang lebih efisien.

## B. TINJUAN PUSTAKA

### 2.1 Customer Relationship Management (CRM)

CRM merupakan strategi bisnis yang berfokus pada pemahaman, pengelolaan, dan peningkatan hubungan dengan pelanggan agar loyalitas dan profitabilitas perusahaan dapat meningkat. Salah satu elemen penting dalam CRM adalah customer profiling, yaitu pengelompokan pelanggan berdasarkan karakteristik atau perilaku transaksi yang serupa, sehingga perusahaan dapat memberikan layanan dan penawaran yang lebih relevan bagi tiap kelompok (Barus et al., 2023).

### 2.2 Analisis RFM (Recency, Frequency, Monetary)

RFM Analysis adalah teknik analitik yang menilai pelanggan berdasarkan tiga dimensi riwayat transaksinya: Recency (selisih waktu sejak transaksi terakhir), Frequency (jumlah transaksi dalam periode tertentu), dan Monetary (total nilai transaksi). Konsep ini pertama kali dipopulerkan oleh Hughes (1994) dan sejak itu menjadi salah satu metode standar dalam database marketing untuk menilai nilai pelanggan (customer value) serta memprediksi kecenderungan perilaku pembelian di masa

depan. Christy et al. (2021) dan Gustriansyah et al. (2020) menunjukkan bahwa RFM dapat dikombinasikan dengan teknik klustering untuk menghasilkan segmentasi yang lebih objektif dibandingkan dengan metode skoring manual.

### 2.3 K-Means Clustering

K-Means merupakan algoritma unsupervised learning yang mengelompokkan data ke dalam K kluster berdasarkan kedekatan jarak (umumnya jarak Euclidean) antara titik data dengan titik pusat kluster (centroid). Algoritma ini bekerja secara iteratif: menetapkan setiap titik data ke centroid terdekat, kemudian memperbarui posisi centroid berdasarkan rata-rata anggota klasternya, hingga konvergen. K-Means dipilih secara luas dalam penelitian segmentasi pelanggan karena relatif sederhana, efisien untuk data berskala besar, dan mudah diinterpretasikan (John et al., 2023; Robo et al., 2024). Implementasi pada penelitian ini menggunakan pustaka scikit-learn (Pedregosa et al., 2011).

### 2.4 Elbow Method dan Silhouette Score

Penentuan jumlah kluster (K) yang optimal pada K-Means umumnya dilakukan melalui dua pendekatan yang saling melengkapi. Elbow Method mengukur Within-Cluster Sum of Square (WCSS), yaitu total jarak kuadrat antara setiap titik data dengan centroid klasternya, lalu memilih nilai K pada titik di mana penurunan WCSS mulai melandai (titik siku/elbow). Silhouette Score mengukur seberapa baik suatu titik data cocok dengan klasternya sendiri dibandingkan dengan kluster lain, dengan rentang nilai -1 hingga 1; semakin mendekati 1, semakin baik pemisahan antar kluster (Shutaywi & Kachouie, 2021). Kedua metode ini umum digunakan bersamaan, seperti pada penelitian Barus et al. (2023) dan Samidi et al. (2023), karena masing-masing memiliki kelebihan dan keterbatasan yang berbeda.

### 2.5 Penelitian Terdahulu

Beberapa penelitian terdahulu telah menerapkan kombinasi RFM dan K-Means pada konteks bisnis yang beragam. Barus et al. (2023) menerapkannya pada data transaksi perusahaan remitansi dan menghasilkan tiga kluster pelanggan (average customer, may-not-lost customer, dan best customer). Robo et al. (2024) menerapkan pendekatan serupa pada industri ritel dan menemukan sepuluh segmen pelanggan, termasuk segmen "Champions" dan "Hibernating". John et al. (2023) membandingkan beberapa algoritma klustering untuk segmentasi pelanggan ritel di Inggris. Penelitian-penelitian tersebut menjadi dasar pijakan metodologis bagi penelitian ini, yang diterapkan secara khusus pada data transaksi e-commerce.

## C. METODOLOGI PENELITIAN

### 3.1 Data Penelitian

Data yang digunakan merupakan data transaksi pelanggan pada sebuah platform e-commerce yang telah diolah menjadi format RFM, dengan atribut *user\_id*, *Recency*, *Frequency*, dan *Monetary*. Berdasarkan hasil akhir klustering (Tabel 2), dataset penelitian ini terdiri dari 8.635 pelanggan unik. Contoh lima baris pertama data RFM ditampilkan pada Tabel 1.

**Table 1.** Contoh data RFM pelanggan

| user_id | Recency | Frequency | Monetary |
|---------|---------|-----------|----------|
| U000001 | 569     | 1         | 1322.96  |
| U000002 | 326     | 4         | 1195.74  |
| U000003 | 25      | 3         | 419.93   |
| U000004 | 66      | 3         | 4503.50  |
| U000005 | 300     | 1         | 12.66    |

### 3.2 Tahapan Penelitian

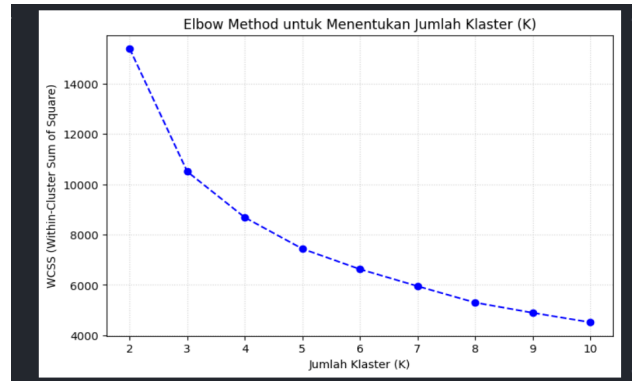
Penelitian ini dilakukan melalui beberapa tahapan berikut: (1) pembentukan variabel RFM dari data transaksi; (2) standarisasi data menggunakan *StandardScaler*; (3) penentuan jumlah kluster optimal (*K*) menggunakan *Elbow Method* dan *Silhouette Score*; (4) proses klustering menggunakan algoritma *K-Means*; dan (5) profiling serta interpretasi karakteristik tiap kluster untuk menghasilkan rekomendasi strategi pemasaran.

### 3.3 Standardisasi Data

Karena variabel *Recency*, *Frequency*, dan *Monetary* memiliki satuan dan rentang nilai yang berbeda jauh — misalnya *Monetary* dapat bernilai ribuan sedangkan *Frequency* hanya berkisar satu digit — maka diperlukan standarisasi data sebelum proses klustering. Tanpa standarisasi, variabel dengan rentang nilai besar seperti *Monetary* akan mendominasi perhitungan jarak Euclidean pada algoritma *K-Means*, sehingga variabel lain menjadi kurang berpengaruh terhadap hasil kluster. Standarisasi dilakukan menggunakan *StandardScaler* dari pustaka *scikit-learn* (Pedregosa et al., 2011), yang mengubah setiap variabel agar memiliki rata-rata 0 dan standar deviasi 1.

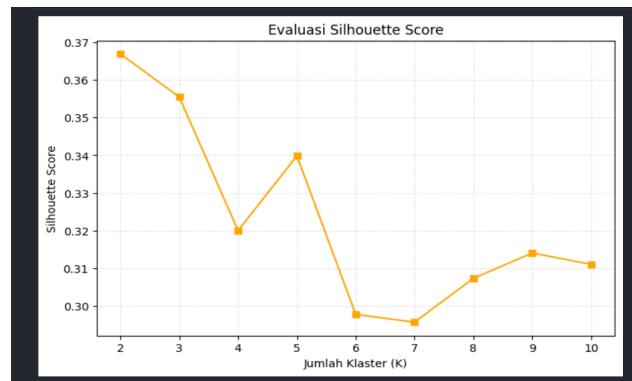
### 3.4 Penentuan Jumlah Kluster Optimal

Penentuan nilai *K* dilakukan dengan mengevaluasi rentang *K* = 2 hingga *K* = 10 menggunakan dua metode, yaitu *Elbow Method* dan *Silhouette Score*.



**Gambar 1.** Elbow Method untuk menentukan jumlah kluster (*K*)

Grafik *Elbow Method* pada Gambar 1 menunjukkan penurunan nilai *WCSS* yang tajam pada *K* = 2 hingga *K* = 4, kemudian penurunannya mulai melandai setelah *K* = 4, membentuk pola siku (*elbow*) di sekitar titik tersebut.



**Gambar 2.** Evaluasi *Silhouette Score* untuk *K* = 2 sampai *K* = 10

Sementara itu, grafik *Silhouette Score* pada Gambar 2 menunjukkan nilai tertinggi pada *K* = 2 (0,367) dan *K* = 3 (0,355), sedikit menurun pada *K* = 4 (0,320), kemudian naik kembali pada *K* = 5 (0,340) sebelum menurun tajam pada *K* = 6 dan *K* = 7.

Berdasarkan kedua grafik tersebut, terdapat sedikit perbedaan indikasi antara *Elbow Method* dan *Silhouette Score* — hal yang lazim terjadi pada data dunia nyata yang tidak memiliki batas kluster yang benar-benar tegas (Samidi et al., 2023). Penelitian ini memilih *K* = 4 dengan pertimbangan: (1) *K* = 4 berada tepat pada titik siku grafik *WCSS*; (2) nilai *Silhouette Score* pada *K* = 4 masih tergolong wajar (0,320) dan tidak jauh berbeda dari nilai tertinggi; serta (3) empat kluster menghasilkan segmentasi yang lebih actionable dan selaras dengan kerangka segmentasi RFM yang umum digunakan dalam praktik pemasaran (misalnya *Champions*, *Loyal*, *Occasional*, dan *At-Risk*), dibandingkan dengan dua atau tiga kluster yang

cenderung terlalu umum untuk menyusun strategi yang berbeda secara nyata.

### 3.5 Algoritma K-Means Clustering

Setelah  $K = 4$  ditetapkan, proses klustering dilakukan menggunakan algoritma K-Means dengan tahapan: (1) inialisasi titik pusat (centroid) awal secara acak; (2) menghitung jarak Euclidean setiap titik data terhadap seluruh centroid dan menetapkan ke klaster dengan jarak terdekat; (3) memperbarui posisi centroid berdasarkan rata-rata anggota klasternya; dan (4) mengulangi langkah (2) dan (3) hingga posisi centroid konvergen atau tidak berubah secara signifikan.

## D. HASIL DAN PEMBAHASAN

### 4.1 Hasil Klustering

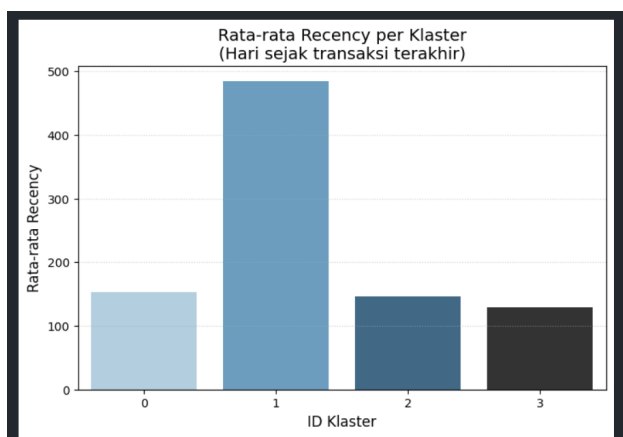
Proses klustering menghasilkan empat klaster pelanggan dengan karakteristik rata-rata Recency, Frequency, dan Monetary yang berbeda-beda, sebagaimana dirangkum pada Tabel 2.

| Klaster | Rata-rata Recency | Rata-rata Frequency | Rata-rata Monetary | Jumlah Pelanggan |
|---------|-------------------|---------------------|--------------------|------------------|
| 0       | 153.21            | 1.54                | 474.92             | 2.686            |
| 1       | 484.49            | 1.40                | 548.68             | 2.349            |
| 2       | 146.84            | 3.15                | 1.158.18           | 2.466            |
| 3       | 129.43            | 4.21                | 2.767.23           | 1.134            |

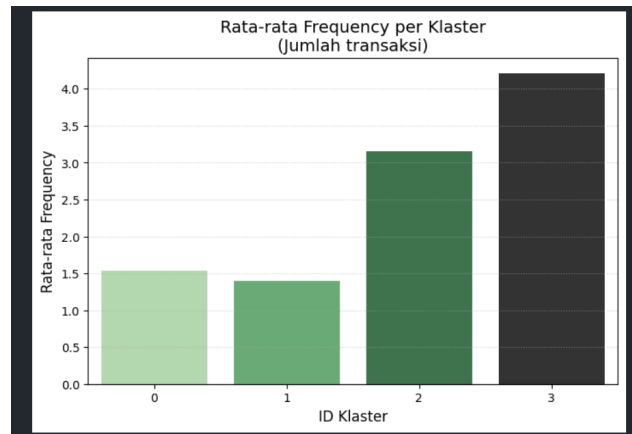
Table 2. Ringkasan karakteristik tiap klaster

### 4.2 Karakteristik Tiap Klaster

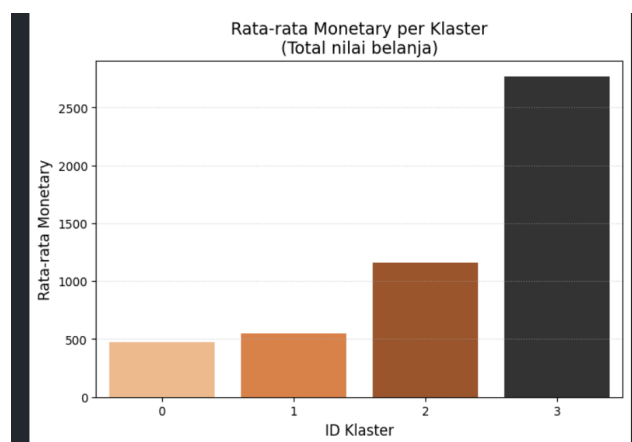
Untuk memperjelas perbedaan antar klaster, ditampilkan visualisasi rata-rata Recency, Frequency, dan Monetary per klaster pada Gambar 3, 4, dan 5.



Gambar 3. Rata-rata Recency per klaster



Gambar 4. Rata-rata Frequency per klaster



Gambar 5. Rata-rata Monetary per klaster

Klaster 1 memiliki nilai Recency rata-rata tertinggi (484,49 hari), jauh di atas klaster lainnya, yang mengindikasikan pelanggan pada klaster ini sudah lama tidak melakukan transaksi. Sebaliknya, Klaster 3 memiliki Frequency (4,21) dan Monetary (2.767,23) rata-rata tertinggi, dengan Recency yang juga rendah (129,43 hari), menunjukkan pelanggan paling aktif dan bernilai tinggi. Klaster 2 berada pada posisi menengah-tinggi, sedangkan Klaster 0 menunjukkan pelanggan dengan frekuensi dan nilai belanja paling rendah namun recency yang masih relatif wajar. Pola ini konsisten dengan temuan Robo et al. (2024) yang juga mengidentifikasi segmen “Champions” dan “Hibernating”/at-risk pada penerapan RFM dan K-Means di industri ritel.

Berdasarkan karakteristik tersebut, setiap klaster diberi nama segmen beserta rekomendasi strategi pemasarannya, sebagaimana dirangkum pada Tabel 3.

| Klaster | Nama Segmen   | Karakteristik   | Rekomendasi Strategi  |
|---------|---|---|---|
| 3       | Champions (Pelanggan Terbaik)                             | Baru bertransaksi, paling sering membeli, nilai belanja tertinggi                 | Prioritaskan program loyalitas/VIP, early access produk baru, upsell/cross-sell |
| 2       | Loyal Customers (Pelanggan Setia)                         | Recency baik, frekuensi & nilai belanja menengah-tinggi                           | Dorong agar naik kelas ke Champions melalui rekomendasi produk personal         |
| 0       | Occasional Customers (Pelanggan Baru/Jarang Bertransaksi) | Recency menengah, frekuensi & nilai belanja rendah                                | Kampanye edukasi produk, insentif transaksi kedua, onboarding                   |
| 1       | At-Risk / Lost Customers (Pelanggan Berisiko Hilang)      | Recency sangat tinggi (lama tidak bertransaksi), frekuensi & nilai belanja rendah | Kampanye reaktivasi (win-back), penawaran diskon khusus, survei alasan churn    |

**Table 3.** Interpretasi segmen dan rekomendasi strategi pemasaran

### 4.3 Pembahasan

Hasil segmentasi menunjukkan bahwa mayoritas pelanggan (Klaster 0 dan Klaster 1, sekitar 58% dari total pelanggan) berada pada kategori dengan nilai transaksi rendah hingga berisiko hilang. Hal ini mengindikasikan adanya peluang besar bagi perusahaan untuk meningkatkan retensi pelanggan melalui kampanye reaktivasi yang tepat sasaran, alih-alih menjalankan promosi yang seragam untuk seluruh basis pelanggan. Di sisi lain, Klaster 3 yang hanya berjumlah 1.134 pelanggan (sekitar 13%) namun memiliki nilai Monetary rata-rata tertinggi merupakan kelompok prioritas yang perlu dijaga loyalitasnya, karena kehilangan pelanggan pada segmen ini berdampak signifikan terhadap pendapatan secara keseluruhan.

Sebagai catatan metodologis, pemilihan  $K = 4$  pada penelitian ini didasarkan pada kombinasi titik siku Elbow Method dan pertimbangan interpretasi bisnis, mengingat Silhouette Score tertinggi sebenarnya berada pada  $K = 2$  dan  $K = 3$ . Pendekatan ini sejalan dengan praktik umum pada penelitian segmentasi pelanggan, di mana jumlah klaster yang dipilih perlu mempertimbangkan baik validitas statistik maupun kebermanfaatan praktis bagi pengambilan keputusan bisnis (Samidi et al., 2023).

## E. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Penelitian ini berhasil melakukan segmentasi terhadap 8.635 pelanggan e-commerce menjadi empat klaster menggunakan kombinasi analisis RFM dan algoritma K-Means Clustering, dengan jumlah klaster ditentukan melalui Elbow Method dan Silhouette Score. Keempat klaster yang terbentuk adalah Champions (Klaster 3), Loyal Customers (Klaster 2), Occasional Customers (Klaster 0), dan At-Risk/Lost Customers (Klaster 1). Hasil segmentasi ini dapat dijadikan dasar penyusunan strategi pemasaran yang lebih personal, seperti program loyalitas

untuk pelanggan bernilai tinggi dan kampanye reaktivasi untuk pelanggan berisiko hilang.

### 5.2 Saran

Penelitian selanjutnya dapat menambahkan variabel lain seperti Customer Lifetime Value (CLV) atau kategori produk untuk memperkaya hasil segmentasi. Perlu dilakukan perbandingan dengan algoritma klustering lain (misalnya Hierarchical/Agglomerative Clustering atau DBSCAN) untuk menguji konsistensi hasil segmentasi, mengingat penelitian lain menunjukkan algoritma lain dapat memberikan Silhouette Score yang lebih tinggi pada data RFM. Sumber data, periode pengambilan data, dan tahap pembersihan data (data cleaning) perlu dicantumkan secara eksplisit pada bagian Metodologi untuk meningkatkan transparansi dan reproducibility penelitian. Hasil segmentasi sebaiknya divalidasi lebih lanjut dengan data perilaku pelanggan pasca-implementasi strategi (misalnya tingkat retensi atau response rate kampanye) untuk mengukur efektivitas segmentasi secara langsung.

## F. DAFTAR PUSTAKA

- (1) Barus, O. P., Nathasya, C., & Pangaribuan, J. J. (2023). The implementation of RFM analysis to customer profiling using K-Means clustering. *Mathematical Modelling of Engineering Problems*, 10(1), 298–303. <https://doi.org/10.18280/mmep.100135>
- (2) Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University – Computer and Information Sciences*, 33(10), 1251–1257. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- (3) Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on K-Means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 470–477. <https://doi.org/10.11591/ijeecs.v18.i1.pp470-477>
- (4) Hughes, A. M. (1994). *Strategic database marketing: The masterplan for starting and managing a profitable, customer-based marketing program*. Probus Publishing.
- (5) John, J. M., Shobayo, O., & Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the UK retail market. *Analytics*, 2(4), 809–823. <https://doi.org/10.3390/analytics2040042>
- (6) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). *Scikit-learn: Machine learning*

- in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- (7) Robo, S., Melani, P. I., Fernatyanan, P., Widiyanto, M. R., & Bariyyah, S. K. (2024). The application of customers segmentation using RFM analysis method and K-Means clustering to improve marketing strategy. *IJISTECH (International Journal of Information System and Technology)*, 8(3), 200–211. <https://doi.org/10.30645/ijistech.v8i3.370>
- (8) Samidi, Suladi, R. Y., & Kusumaningsih, D. (2023). Comparison of the RFM model's actual value and score value for clustering. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(6), 1430–1438. <https://doi.org/10.29207/resti.v7i6.5416>
- (9) Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 759. <https://doi.org/10.3390/e23060759>