PROSEMNASPROIT : Prosiding Seminar Nasional Ilmu Teknik
Volume. 2 Nomor. 2, Desember 2025
E-ISSN : 3063-5713; P-ISSN : 3063-4709; Hal. 375-384
DOI: https://doi.org/10.61132/prosemnasproit.v2i2.84
Tersedia: https://prosiding.aritekin.or.id/index.php/PROSEMNASPROIT

Segmentasi Pelanggan Ritel Global dan Inggris Menggunakan RFM dan K-Means *Clustering*

Prayitno[1], Irawan[2*], Marrylinteri Istoningtyas[3]
[1-3]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia
Email: prayitno4704@gmail.com[1] , irend.irawan.irend@gmail.com[2] , marrylinteri@unama.ac.id[3]*

Alamat : Jalan Jendral Sudirman Thehok – Jambi
**Penulis Korespondensi: irend.irawan.irend@gmail.com* *

Abstract: *Transaction logs in online retail provide opportunities for data-driven customer segmentation. This study segments customers at two scopes global (all countries) and United Kingdom (UK) using Recency, Frequency, and Monetary (RFM) features derived from the Online Retail transaction dataset. After cleaning cancellations and invalid records, RFM variables are computed per customer and normalized. K-Means clustering is applied separately for global and UK data, while the number of clusters is selected via the elbow criterion and validated using internal indices. The best configuration for both scopes yields five clusters, with moderate separation quality based on the silhouette score. Cluster profiling indicates distinct groups ranging from low-frequency low-spending customers to highly frequent high-spending customers. The comparison between global and UK segmentation shows similar structural patterns, yet different proportions across segments, supporting targeted retention and value-driven marketing actions.*

Keywords: *customer segmentation; RFM; K-Means; clustering; online retail; UK market*

Abstrak: Log transaksi pada ritel daring membuka peluang segmentasi pelanggan berbasis data. Penelitian ini melakukan segmentasi pada dua cakupan global (seluruh negara) dan Inggris (UK) menggunakan fitur Recency, Frequency, dan Monetary (RFM) yang dihitung dari dataset transaksi Online Retail. Setelah pembersihan data (pembatalan dan record tidak valid), variabel RFM dihitung per pelanggan dan dinormalisasi. Algoritma K-Means diterapkan secara terpisah untuk data global dan UK, sedangkan jumlah cluster dipilih dengan kriteria elbow dan dievaluasi menggunakan indeks internal. Konfigurasi terbaik pada kedua cakupan menghasilkan lima cluster, dengan kualitas pemisahan sedang berdasarkan silhouette score. Hasil profiling menunjukkan perbedaan segmen mulai dari pelanggan bernilai rendah hingga pelanggan bernilai tinggi. Perbandingan global vs UK memperlihatkan pola struktur segmen yang serupa namun proporsi pelanggan yang berbeda, sehingga dapat mendukung tindakan pemasaran yang lebih terarah.

Kata kunci: segmentasi pelanggan; RFM; K-Means; clustering; online retail; pasar UK

1. LATAR BELAKANG

Pertumbuhan ritel daring menghasilkan data transaksi berukuran besar yang dapat dimanfaatkan untuk membangun segmentasi pelanggan berbasis perilaku dan mendukung keputusan pemasaran yang lebih presisi (Wong et al., 2024). Teknik segmentasi membantu perusahaan mengelompokkan pelanggan yang heterogen sehingga pesan, kanal, dan penawaran dapat disesuaikan dengan karakteristik masing-masing kelompok (Zhou et al., 2021). Penentuan nilai pelanggan melalui variabel *Recency, Frequency, dan Monetary (RFM)* telah digunakan untuk mengukur kontribusi dan potensi pelanggan dalam konteks manajemen hubungan pelanggan (Safari et al., 2016). Pemeringkatan berbasis RFM terbukti efektif untuk

Naskah Masuk: 27 November 2025; Revisi: 28 November 2025; Diterima: 28 Desember 2025;
Terbit: 31 Desember 2025

mengidentifikasi pelanggan bernilai tinggi serta pelanggan yang mulai menurun aktivitasnya sehingga mendukung prioritisasi retensi (Christy et al., 2021).

Penerapan K-Means pada fitur RFM memungkinkan pembentukan *cluster* pelanggan yang mudah diprofilkan berdasarkan kedekatan perilaku pembelian (Anitha & Patil, 2022). Pengayaan analisis RFM dengan teknik data mining lain seperti *association rule mining* dapat memberikan wawasan tambahan terkait pola nilai pelanggan dan strategi ritel (Pai et al., 2025). Pemilihan jumlah cluster yang tepat dan evaluasi kualitas *cluster* sangat penting, dan *silhouette analysis* merupakan salah satu cara populer untuk menilai pemisahan cluster secara internal (Shutaywi & Kachouie, 2021). Kajian yang meninjau pasar ritel Inggris menunjukkan bahwa variasi algoritma *clustering* dapat menghasilkan struktur segmen yang berbeda sehingga perlu dilakukan pengujian dan interpretasi yang konsisten (John et al., 2023).

2. KAJIAN TEORITIS

Pendekatan hierarchical berbasis *Formal Concept Analysis* pada data RFM menunjukkan bahwa struktur relasi atribut dapat digunakan untuk membangun segmentasi yang lebih informatif (Rungruang et al., 2024). Tinjauan sistematis pada kasus *e-commerce* menekankan bahwa pemilihan fitur, metode clustering, dan skema evaluasi adalah komponen inti untuk mencapai segmentasi yang berguna bagi personalisasi (Alves Gomes & Meisen, 2023). Integrasi *customer lifetime value* dengan model migrasi pelanggan menambah dimensi temporal pada segmentasi sehingga perubahan perilaku pelanggan dapat dipahami lebih baik (Kanchanapoom & Chongwatpol, 2023). Studi empiris memperlihatkan bahwa kombinasi RFM dan K-Means dapat memetakan perbedaan pola pembelian pelanggan secara nyata dalam data transaksi ritel (Wu et al., 2022).

Model segmentasi *multi-view* berlapis menunjukkan bahwa K-Means dapat diperluas untuk mengakomodasi lebih dari satu representasi data pelanggan dalam proses pengelompokan (Handoko & Wibowo, 2020). Optimasi proses clustering pada analisis RFM dilaporkan dapat meningkatkan kualitas segmen melalui perbaikan pemilihan parameter dan pengurangan variasi *intra-cluster* (Gustriansyah et al., 2020). Karena perilaku belanja dapat berubah dari waktu ke waktu, segmentasi pelanggan sering dipandang sebagai masalah dinamis yang membutuhkan metode pengelompokan yang adaptif (Sivaguru & Punniyamoorthy, 2020). Pendekatan segmentasi modern juga memanfaatkan *deep learning* dan *swarm intelligence* untuk menangkap pola yang lebih kompleks pada data ritel (Talaat et al., 2023).

3. METODE PENELITIAN

A. Dataset dan *Preprocessing*

Dataset yang digunakan berupa log transaksi ritel daring yang memuat informasi faktur, kode produk, deskripsi, kuantitas, tanggal transaksi, harga, identitas pelanggan, dan negara. Tahap praproses dilakukan untuk memperoleh data yang representatif: menghapus transaksi pembatalan/retur, menghapus kuantitas atau harga yang tidak valid, serta menangani nilai hilang pada identitas pelanggan. Setelah pembersihan, kolom pendapatan (*revenue*) dihitung sebagai hasil kali kuantitas dan harga per baris transaksi, kemudian digunakan sebagai dasar perhitungan nilai belanja pelanggan.

B. Pembentukan Fitur RFM

Fitur perilaku pelanggan dibentuk menggunakan pendekatan RFM pada tingkat pelanggan. *Recency* dihitung sebagai selisih hari antara tanggal referensi (tanggal transaksi terakhir pada dataset) dan tanggal pembelian terakhir pelanggan. *Frequency* dihitung sebagai jumlah faktur (*invoice*) unik yang dimiliki pelanggan. *Monetary* dihitung sebagai total pendapatan pelanggan selama periode pengamatan. Pada penelitian ini segmentasi dilakukan pada dua cakupan berupa global, yaitu seluruh transaksi lintas negara, dan UK, yaitu subset transaksi dengan negara United Kingdom.

C. Normalisasi dan K-Means Clustering

Karena skala *Recency, Frequency,* dan *Monetary* berbeda jauh, data RFM dinormalisasi sebelum proses *clustering*. Algoritma K-Means diterapkan pada data RFM yang telah dinormalisasi untuk membentuk sejumlah K *cluster*. Pemilihan K dilakukan dengan menguji beberapa nilai K dan mengevaluasi perubahan nilai inertia (*elbow*), kemudian dipilih K terbaik yang memberikan keseimbangan antara kompleksitas dan kualitas pemisahan.

D. Evaluasi dan Interpretasi Cluster

Kualitas cluster dievaluasi menggunakan metrik internal, yaitu *silhouette score, Calinski-Harabasz score*, dan *Davies-Bouldin score*. Setelah jumlah cluster ditetapkan, profil cluster dianalisis menggunakan rata-rata *Recency, Frequency*, dan *Monetary* serta proporsi anggota pada setiap *cluster*. Sebagai tambahan interpretasi visual, hasil clustering divisualisasikan dalam proyeksi 2D menggunakan PCA untuk melihat kecenderungan pemisahan antar cluster.

4. HASIL DAN PEMBAHASAN

A. Praproses Data

Tabel 1. Persentase nilai hilang per kolom pada data awal

Kolom	missing_pct
CustomerID	24.93
Description	0.27
StockCode	0.00
InvoiceNo	0.00
Quantity	0.00
InvoiceDate	0.00
UnitPrice	0.00
Country	0.00

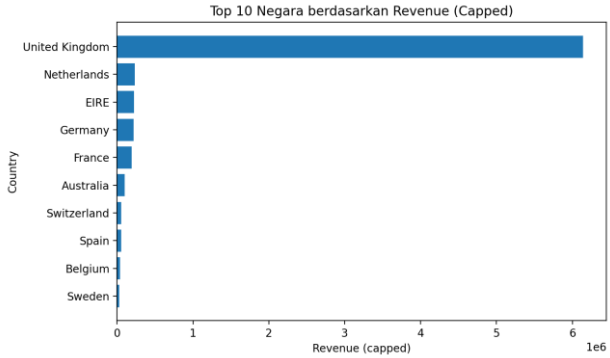
Tabel 1 menunjukkan bahwa nilai hilang terutama terdapat pada *CustomerID* dan sebagian kecil pada *Description*, sehingga pembersihan difokuskan pada *record* tanpa identitas pelanggan. Setelah pembersihan, data global yang digunakan berukuran 397.884 baris (10 kolom) dan subset UK berukuran 354.321 baris (10 kolom).

B. Distribusi Negara

Tabel 2. Ringkasan pelanggan, invoice, dan revenue per negara (Top 10)

Country	customers	invoices	revenue	revenue_capped
United Kingdom	3920	16646	7,308,391.55	6,140,218.02
Netherlands	9	94	285,446.34	233,115.39
EIRE	3	260	265,545.90	225,325.27
Germany	94	457	228,867.14	220,958.82
France	87	389	209,024.05	193,789.56
Australia	9	57	138,521.31	101,527.31
Spain	30	90	61,577.11	53,269.24
Switzerland	21	51	56,443.95	55,643.15
Belgium	25	98	41,196.34	41,196.34
Sweden	8	36	38,378.33	31,238.71

Tabel 2 menunjukkan dominasi transaksi dari *United Kingdom*, baik dari jumlah pelanggan maupun *invoice*, sehingga analisis khusus UK relevan untuk melihat karakter segmen pada pasar utama.



Gambar 1. Top 10 negara berdasarkan revenue (capped)

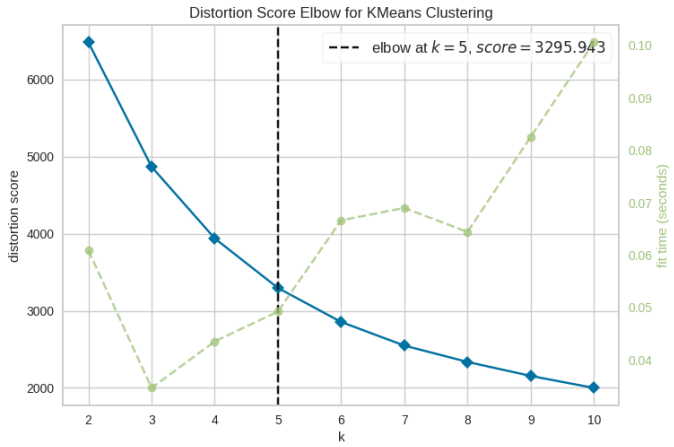
Gambar 1 memperlihatkan perbandingan revenue antar negara, di mana UK memiliki kontribusi paling besar dibanding negara lain; temuan ini menjadi dasar pemilihan konteks global dan UK sebagai fokus penelitian.

C. Hasil Clustering Global

Tabel 3. Statistik deskriptif RFM pada cakupan global (per pelanggan)

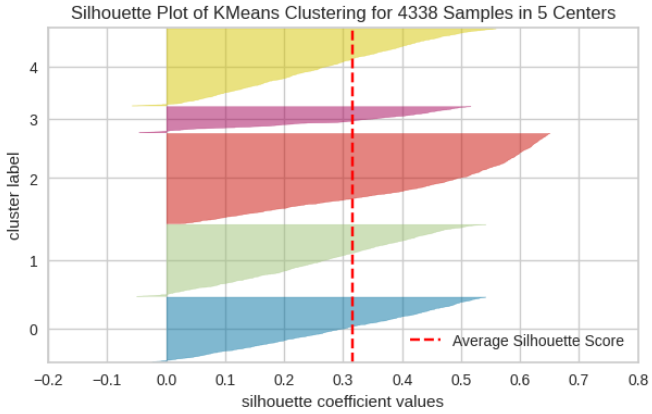
Variabel	count	mean	std	min	25%	50%	75%	max
Recency	4,338.00	92.54	100.01	1.00	18.00	51.00	142.00	374.00
Frequency	4,338.00	4.27	7.70	1.00	1.00	2.00	5.00	209.00
Monetary	4,338.00	2,054.27	8,989.23	3.75	307.42	674.49	1,661.74	280,206.02

Tabel 3 menunjukkan sebaran RFM pada pelanggan global (n=4.338) dengan variasi Monetary yang tinggi, sehingga normalisasi diperlukan sebelum clustering.



Gambar 2. Pemilihan jumlah cluster (elbow) untuk data global

Gambar 2 menunjukkan titik belokan yang mengindikasikan K=5 sebagai pilihan yang seimbang antara penurunan inertia dan kompleksitas model.



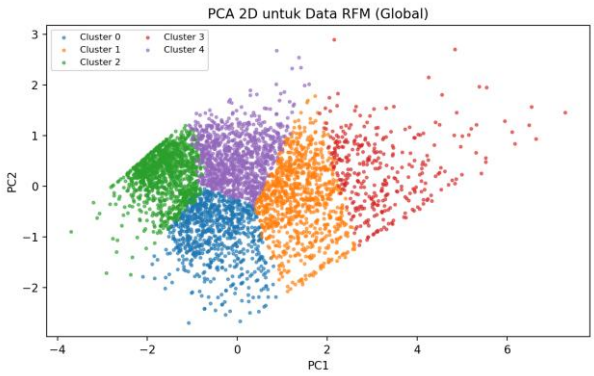
Gambar 3. Visualisasi evaluasi *silhouette* untuk clustering global

Nilai *silhouette* global sebesar 0,3161 mengindikasikan pemisahan cluster berada pada tingkat sedang, namun masih cukup untuk memetakan perbedaan perilaku belanja pelanggan.

Tabel 4. Profil rata-rata RFM per cluster (Global)

Cluster	customers	avg_recency	avg_frequency	avg_monetary	customers_pct
0	851	27.99	1.61	389.04	19.62
1	944	18.55	5.80	2,168.93	21.76
2	1185	211.15	1.21	277.20	27.32
3	343	12.19	20.63	13,780.08	7.91
4	1015	104.14	3.13	1,455.97	23.40

Tabel 4 memperlihatkan bahwa cluster 3 memiliki recency paling rendah, frequency paling tinggi, dan monetary paling besar (segmen bernilai sangat tinggi), sedangkan cluster 2 memiliki recency tinggi dan monetary rendah (segmen bernilai rendah/dorman).



Gambar 4. Proyeksi 2D menggunakan PCA untuk clustering global

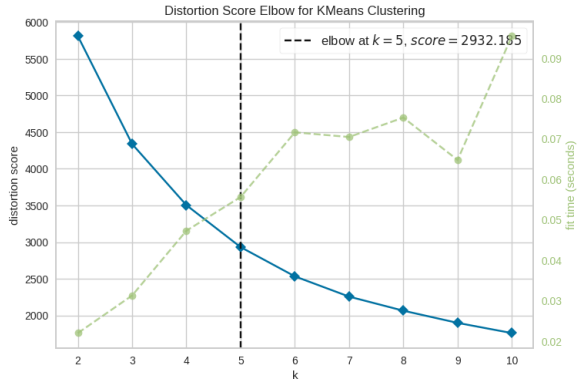
Gambar 4 menampilkan proyeksi PCA 2D; meskipun terjadi tumpang tindih pada beberapa area, cluster bernilai tinggi cenderung lebih terpisah dibanding cluster bernilai rendah.

D. Hasil Clustering UK

Tabel 5. Statistik deskriptif RFM pada cakupan UK (per pelanggan)

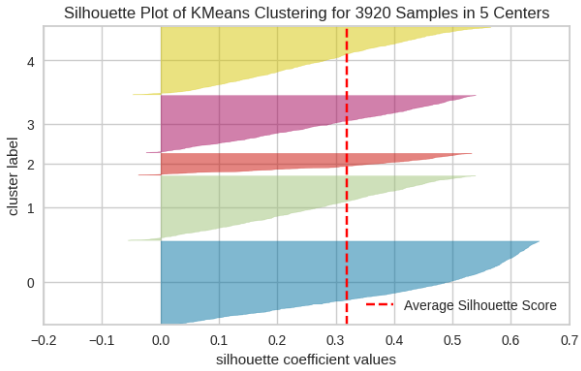
Variabel	count	mean	std	min	25%	50%	75%	max
Recency	3,920.00	92.21	99.53	1.00	18.00	51.00	143.00	374.00
Frequency	3,920.00	4.25	7.20	1.00	1.00	2.00	5.00	209.00
Monetary	3,920.00	1,864.39	7,482.82	3.75	300.28	652.28	1,576.59	259,657.30

Tabel 5 menunjukkan pola sebaran RFM pada UK (n=3.920) yang mirip dengan global, namun tetap terdapat pelanggan dengan Monetary yang sangat tinggi pada pasar utama.



Gambar 5. Pemilihan jumlah *cluster* (*elbow*) untuk data UK

Gambar 5 menunjukkan pola elbow yang konsisten, sehingga K=5 dipilih untuk subset UK.



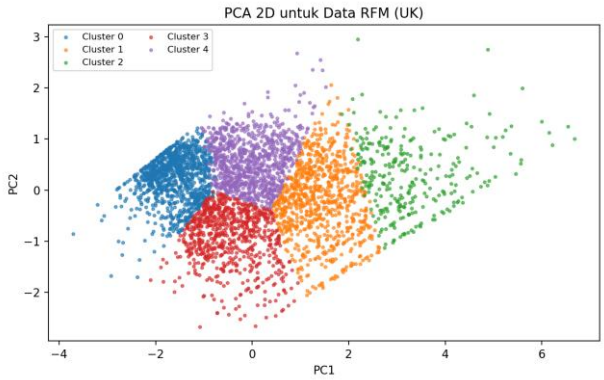
Gambar 6. Visualisasi evaluasi silhouette untuk clustering UK

Nilai *silhouette* UK sebesar 0,3189 menunjukkan kualitas pemisahan yang sedikit lebih baik dibanding global, meskipun masih berada pada kategori sedang.

Tabel 6. Profil rata-rata RFM per cluster (UK)

Cluster	customers	avg_recency	avg_frequency	avg_monetary	customers_pct
0	1106	206.57	1.21	269.51	28.21
1	859	19.02	5.93	2,120.19	21.91
2	291	11.32	20.93	12,558.09	7.42
3	761	26.90	1.65	381.12	19.41
4	903	102.86	3.18	1,378.34	23.04

Tabel 6 memperlihatkan *cluster* 2 sebagai segmen terbaik pada UK (*recency* rendah, *frequency* dan *monetary* tinggi), sedangkan cluster 0 menunjukkan karakter pelanggan yang cenderung *dorman* (*recency* tinggi, *frequency* rendah).



Gambar 7. Proyeksi 2D menggunakan PCA untuk clustering UK

Gambar 7 memperlihatkan proyeksi PCA 2D untuk subset UK dan membantu melihat pola penyebaran cluster pada dua komponen utama.

E. Perbandingan Evaluasi Global vs UK

Tabel 7. Ringkasan metrik evaluasi internal untuk clustering (Global vs UK)

Cakupan	Optimal_K	Silhouette	Calinski_Harabasz	Davies_Bouldin
Global	5	0.32	3,193.95	0.99
UK	5	0.32	2,946.69	0.98

Tabel 7 menunjukkan bahwa kedua cakupan menghasilkan K=5 dengan silhouette yang mirip; perbedaan skor Calinski-Harabasz dan Davies-Bouldin menunjukkan variasi kepadatan dan pemisahan cluster antara global dan UK.

5. KESIMPULAN DAN SARAN

Penelitian ini menyusun segmentasi pelanggan ritel daring pada cakupan global dan UK menggunakan fitur RFM dan K-Means. Hasil evaluasi menunjukkan bahwa K=5 merupakan jumlah cluster yang stabil untuk kedua cakupan dengan kualitas pemisahan sedang berdasarkan *silhouette score*. Profil cluster berhasil membedakan segmen bernilai sangat tinggi, segmen bernilai menengah, dan segmen pelanggan yang cenderung dorman/berisiko.

Secara operasional, segmen bernilai tinggi dapat diprioritaskan untuk program loyalitas, sedangkan segmen dorman/berisiko dapat ditargetkan dengan kampanye reaktivasi. Penelitian lanjutan dapat menambahkan fitur perilaku lain dan menguji algoritma *clustering* alternatif untuk meningkatkan ketajaman segmen.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada penyelenggara PROSEMNASPROIT serta pihak-pihak yang telah memberikan dukungan selama proses penyusunan penelitian ini.

DAFTAR REFERENSI

Alves Gomes, D., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. Information Systems and e-Business Management, 21, 527–570. https://doi.org/10.1007/s10257-023-00640-4

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University – Computer and Information Sciences, 34(5), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

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

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

Handoko, A. F., & Wibowo, A. (2020). Three-layer data clustering model for multi-view customer segmentation using K-Means. *International Journal of Recent Technology and Engineering*, 8(6), 1840–1845. https://doi.org/10.35940/ijrte.F7962.038620

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

Kanchanapoom, K., & Chongwatpol, J. (2023). Integrated customer lifetime value and customer migration model to improve customer segmentation and marketing strategies. *Journal of Marketing Analytics*, 11, 172–185. https://doi.org/10.1057/s41270-022-00158-7

Pai, P.-Y., Lin, S.-W., & Lu, W.-M. (2025). Integration of association rule mining and RFM analysis with machine learning for e-commerce customer value segmentation: A sustainable retail perspective. Quality & Quantity. Advance online publication. https://doi.org/10.1007/s11135-025-02259-8

Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K., & Muangprathub, J. (2024). RFM model customer segmentation based on hierarchical approach using formal concept analysis. Expert Systems with Applications, 241, 122605. https://doi.org/10.1016/j.eswa.2023.121449

Safari, F., Safari, N., & Montazer, G. A. (2016). Customer lifetime value determination based on RFM model. Marketing Intelligence & Planning, 34(4), 446–461. https://doi.org/10.1108/MIP-03-2015-0060

Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications. Entropy, 23(6), 759. https://doi.org/10.3390/e23060759

Sivaguru, R., & Punniyamoorthy, M. (2020). Modified dynamic fuzzy c-means clustering algorithm—Application in dynamic customer segmentation. Applied Intelligence, 50, 1922–1942. https://doi.org/10.1007/s10489-019-01626-x

Wong, C.-G., Tong, T.-C., & Haw, D.-S. (2024). Exploring customer segmentation in e-commerce using RFM analysis with hierarchical clustering and K-Means clustering. *Journal of Telecommunications and the Digital Economy*, 12(3), 97–125. https://doi.org/10.18080/jtde.v12n3.978

Wu, F.-X., Shi, B.-S., Lin, C.-J., Tsai, L.-C., Li, R.-H., Yang, C.-Z., & Xu, J. (2022). Research on segmenting e-commerce customer through an improved K-medoids and RFM model. Computational Intelligence and Neuroscience, 2022, Article 9930613. https://doi.org/10.1155/2022/9930613

Talaat, F. M., Aljadani, A., Alharthi, B., Farsi, M. A., Badawy, M., & Elhosseini, M. (2023). A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing. *Mathematics*, *11*(18), 3930. https://doi.org/10.3390/math11183930

Zhou, F., Wei, Q., & Xu, M. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*, 61, 102588. https://doi.org/10.1016/j.jretconser.2021.102588