

SEGMENTASI PELANGGAN MENGGUNAKAN MODEL RFM DAN ALGORITMA CLUSTERING K- MEANS PADA TOKO RITEL

Ahmad Rifqi Noor Alam¹, Nana Suarna², Agus Bahtiar³, Riri Narasati⁴.

Program Studi Teknik Informatika¹²⁴
Program Studi Sistem Informasi³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
ahmadrifqinoor@gmail.com

(*) Corresponding Author : ahmadrifqinoor@gmail.com
Published : 30 Juni 2026

Abstract—This study examines the application of the RFM (Recency, Frequency, Monetary) model combined with the K-Means clustering algorithm to perform customer segmentation in the retail sector. The primary objective of this research is to design a systematic data analysis workflow, determine the optimal number of clusters, and interpret the characteristics of each cluster to support data-driven marketing strategies. The research stages include cleaning transactional data, calculating RFM scores, normalizing features, and conducting clustering using the Elbow method along with cluster quality evaluation through the Davies–Bouldin Index (DBI). The results show that despite the relatively low transaction frequency, transformation and normalization of RFM features successfully form well-identified customer clusters. DBI evaluation indicates that $k = 3$ is the optimal number of clusters with the lowest DBI value of 0.7259. The three clusters represent high-value customers, moderately active customers, and passive customers at risk of churn. Interpretation based on the average RFM values of each cluster generates strategic recommendations such as loyalty programs for high-value customers and reactivation campaigns for at-risk customers. This study demonstrates that the combination of RFM and K-Means remains effective for customer segmentation in low-frequency retail environments and encourages future research involving alternative clustering methods and integration of additional behavioral variables.

Keywords: Segmentasi pelanggan; RFM; K-Means; Davies–Bouldin Index; Data mining

Abstrak—Penelitian ini mengkaji penerapan model RFM (Recency, Frequency, Monetary) yang dikombinasikan dengan algoritma klasterisasi K-Means untuk melakukan segmentasi pelanggan pada sektor ritel. Tujuan utama penelitian ini adalah merancang alur analisis data yang sistematis, menentukan jumlah klaster yang optimal, serta menginterpretasikan karakteristik setiap klaster guna mendukung strategi pemasaran berbasis data. Tahapan penelitian meliputi pembersihan data transaksi, perhitungan skor RFM, normalisasi fitur, serta proses klasterisasi menggunakan metode Elbow dan evaluasi kualitas klaster melalui Davies–Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa meskipun frekuensi transaksi relatif rendah, transformasi dan normalisasi fitur RFM mampu membentuk klaster pelanggan yang teridentifikasi dengan baik. Evaluasi DBI menunjukkan bahwa $k = 3$ merupakan jumlah klaster optimal dengan nilai DBI terendah sebesar 0.7259. Ketiga klaster yang terbentuk menggambarkan pelanggan bernilai tinggi, pelanggan aktif dengan nilai sedang, serta pelanggan pasif yang berisiko churn. Interpretasi berdasarkan nilai rata-rata RFM tiap klaster menghasilkan rekomendasi strategis, seperti program loyalitas untuk pelanggan bernilai tinggi dan kampanye reaktivasi bagi pelanggan berisiko. Penelitian ini membuktikan bahwa kombinasi RFM dan K-Means tetap efektif dalam segmentasi pelanggan pada lingkungan ritel dengan intensitas transaksi rendah, serta mendorong penelitian lanjutan menggunakan metode klasterisasi alternatif dan integrasi variabel perilaku tambahan.

Kata Kunci : Segmentasi pelanggan, RFM, K-Means, Davies–Bouldin Index, Data mining.

INTRODUCTION

Pelaku ritel menghadapi tantangan yang semakin kompleks dalam memahami perilaku konsumen, terutama akibat perubahan struktur

pasar dan preferensi pelanggan yang berubah secara cepat. Pasca pandemi, konsumen mengalami pergeseran signifikan dalam pemilihan kanal belanja, tingkat sensitivitas terhadap keamanan, serta ekspektasi pelayanan, sehingga pola perilaku menjadi tidak stabil dan sulit diprediksi [1]. Pada saat yang sama, tingginya tingkat *information overload* di platform e-commerce mendorong perilaku pembelian impulsif serta peningkatan tingkat pengembalian produk, yang melemahkan akurasi prediksi berbasis perilaku transaksi historis [2].

Perilaku pembelian juga dipengaruhi oleh faktor sosial-ekonomi yang beragam seperti perbedaan dampak ekonomi, perubahan pola substitusi, dan variasi daya beli yang meningkatkan heterogenitas kelompok konsumen [3]. Selain itu, dorongan emosional dan impulsif telah terbukti berperan besar dalam pengambilan keputusan konsumen sehingga mengurangi stabilitas segmentasi tradisional [4].

Preferensi akan keberlanjutan lingkungan juga semakin memengaruhi kriteria pembelian, mengubah motivasi serta komposisi segmen konsumen [5]. Kompleksitas faktor-faktor tersebut menurunkan kekuatan prediktif indikator RFM historis, meningkatkan *noise* dalam data transaksi, dan menuntut pendekatan segmentasi yang lebih kuat dengan menggabungkan sinyal perilaku, konteks, serta aspek psikologis konsumen.

Segmentasi berbasis data muncul sebagai solusi karena mampu menghasilkan pengelompokan konsumen yang lebih akurat dan berbasis perilaku, sehingga mendukung strategi pemasaran yang lebih tepat sasaran. [6] menunjukkan bahwa kemajuan teknologi seperti *machine learning*, *big data analytics*, dan sistem *business intelligence* telah mentransformasi praktik pemasaran menjadi lebih presisi melalui personalisasi kampanye, optimalisasi program layanan pelanggan, serta peningkatan kualitas pengambilan keputusan, meskipun terdapat tantangan terkait kualitas data, privasi, dan kesiapan organisasi.

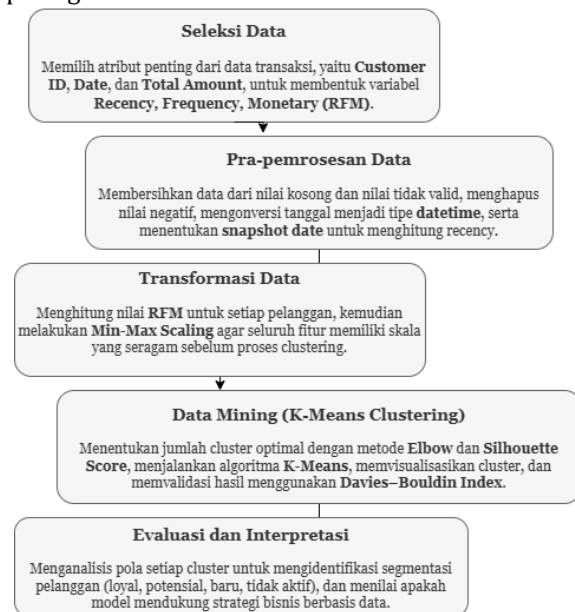
Dalam konteks e-commerce, representasi data seperti RFM yang dipadukan dengan algoritma clustering, terutama k-means, terbukti efektif dalam menghasilkan segmentasi yang terstruktur dan dapat ditindaklanjuti [7].

Pendekatan ini membantu pemasar dalam mengidentifikasi kelompok pelanggan bernilai tinggi, memprediksi tren perilaku, serta merancang strategi personalisasi. Studi empiris oleh [8] memperkuat temuan tersebut dengan

menunjukkan bahwa penerapan k-means dan *fuzzy c-means* pada data digital native mampu mengungkap segmen laten yang memberikan wawasan strategis bagi pemasaran digital. Secara keseluruhan, penelitian-penelitian ini menegaskan bahwa segmentasi berbasis data mampu meningkatkan ketepatan pemasaran, memperkuat keterlibatan pelanggan, dan mengoptimalkan alokasi sumber daya dalam industri ritel modern.

MATERIALS AND METHODS

Eksperimen dalam penelitian ini dirancang mengikuti tahapan metodologi *Knowledge Discovery in Databases* (KDD) yang meliputi: Seleksi data, Pra-pemrosesan, Transformasi, Data Mining, dan Evaluasi/Interpretasi. Pendekatan ini memungkinkan proses eksplorasi dan analisis data dilakukan secara sistematis dan terstruktur dalam rangka segmentasi pelanggan berbasis model RFM dan algoritma K-Means. Tahapan penelitian seperti pada gambar 1



Gambar 1 Tahapan Penelitian

Seleksi Data

Dataset yang digunakan berasal dari sistem transaksi penjualan ritel, dengan atribut utama berupa *'Customer ID'*, *'Date'*, dan *'Total Amount'*. Tahap ini menyaring informasi penting yang relevan untuk membentuk indikator *Recency*, *Frequency*, dan *Monetary*, mengacu pada struktur umum data RFM dalam studi segmentasi pelanggan [9]

Pra-pemrosesan Data

Data yang telah diseleksi selanjutnya diproses untuk mengatasi nilai kosong, mendeteksi dan

menghapus nilai negatif, serta memastikan bahwa format tanggal telah dikonversi ke *tipe datetime*. Langkah ini juga mencakup penentuan *'snapshot date'* untuk menghitung *recency* secara akurat.

Transformasi Data Tahap ini mencakup perhitungan nilai RFM untuk setiap pelanggan: *Recency*: Selisih hari antara transaksi terakhir pelanggan dengan *snapshot date*. *Frequency*: Jumlah transaksi unik oleh pelanggan. *Monetary*: Total nilai belanja kumulatif pelanggan.

Nilai-nilai RFM kemudian dinormalisasi menggunakan teknik *Min-Max Scaling* untuk menghindari dominasi fitur tertentu dalam proses klusterisasi [10].

Proses klusterisasi dilakukan dengan algoritma K-Means. Sebelum diterapkan, nilai optimal 'k' ditentukan menggunakan metode *Elbow* dan *Silhouette Score*. Klaster yang terbentuk divisualisasikan untuk membantu memahami struktur segmen pelanggan. Model validasi dilakukan dengan perhitungan *Davies Bouldin Index* untuk menilai pemisahan antar klaster (Wilbert et al., 2023).

Hasil klaster dianalisis untuk mengidentifikasi profil masing-masing segmen. Profil ini mencakup pelanggan potensial, pelanggan *loyal*, pelanggan baru, atau pelanggan berisiko, berdasarkan pola nilai RFM dalam masing-masing *klaster*. Evaluasi akhir memastikan bahwa model segmentasi yang dibangun dapat mendukung strategi pemasaran berbasis data.

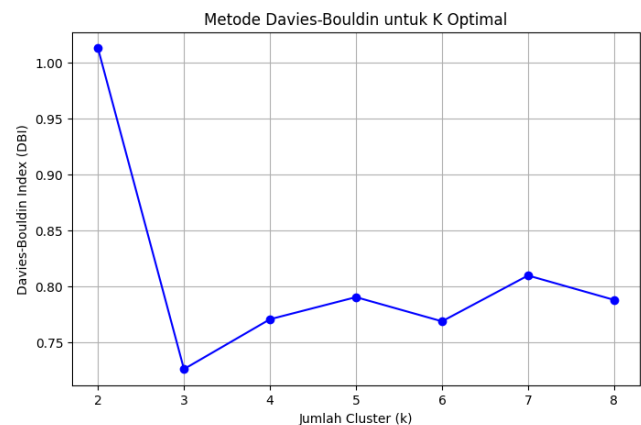
RESULTS AND DISCUSSION

Pada tahap ini, dilakukan proses klusterisasi pelanggan menggunakan algoritma K-Means untuk menemukan pola segmentasi berdasarkan nilai RFM yang telah dinormalisasi. Salah satu langkah paling krusial dalam proses data mining adalah penentuan jumlah klaster (nilai k) yang optimal agar segmentasi yang dihasilkan memiliki validitas tinggi dan interpretabilitas yang baik. Proses Klusterisasi Algoritma K-Means seperti pada gambar 2.

--- Menghitung Skor DBI ---	
K=2,	DBI Score = 1.0129
K=3,	DBI Score = 0.7259
K=4,	DBI Score = 0.7705
K=5,	DBI Score = 0.7903
K=6,	DBI Score = 0.7686
K=7,	DBI Score = 0.8096
K=8,	DBI Score = 0.7878

Gambar 2 Proses Klusterisasi Algoritma K-Means

Gambar keenam menampilkan hasil perhitungan *Davies-Bouldin Index* (DBI) untuk berbagai nilai k dari 2 hingga 8. DBI adalah metrik internal yang digunakan untuk mengevaluasi kualitas klaster, di mana nilai yang lebih rendah mengindikasikan klaster yang lebih kompak dan terpisah dengan baik. Dari hasil tersebut, terlihat bahwa nilai DBI terendah diperoleh pada k 3, dengan skor sebesar 0.7259, yang berarti pemilihan 3 klaster memberikan segmentasi paling optimal secara matematis. Hasil Perhitungan *Davies-Bouldin Index* (DBI) seperti pada gambar 3



Gambar 3 Hasil Perhitungan Davies-Bouldin Index (DBI)

Gambar ketujuh memperkuat temuan tersebut melalui visualisasi grafik garis yang menunjukkan hubungan antara jumlah klaster dan nilai DBI. Grafik ini memperlihatkan penurunan tajam dari k=2 ke k=3, kemudian nilai DBI meningkat kembali dan relatif stabil di k=4 hingga k=8. Hal ini menunjukkan bahwa menambah jumlah klaster setelah k=3 tidak menghasilkan peningkatan signifikan dalam kualitas pemisahan klaster.

Berdasarkan hasil ini, proses klusterisasi selanjutnya dilanjutkan dengan k=3, karena nilai tersebut memberikan keseimbangan terbaik antara

kompleksitas model dan kualitas segmentasi. Strategi ini sejalan dengan pendekatan best practice dalam literatur ilmiah, di mana kombinasi metrik validasi internal dan visualisasi digunakan untuk mengidentifikasi nilai k yang optimal (Fitriyanto, 2025; Wilbert et al., 2023).

Tahap data mining ini menjadi inti dari proses segmentasi pelanggan, karena secara langsung menghasilkan pembagian populasi pelanggan ke dalam kelompok-kelompok homogen yang dapat digunakan untuk strategi pemasaran yang lebih terfokus dan berbasis data.

Tahap ini bertujuan untuk memahami lebih dalam karakteristik dari masing-masing kluster yang terbentuk setelah proses klusterisasi K-Means dilakukan terhadap data RFM (*Recency, Frequency, Monetary*). Evaluasi dilakukan baik secara numerik maupun visual agar segmentasi yang dihasilkan memiliki makna yang jelas dalam konteks perilaku pelanggan. Karakteristik Dari Masing-masing kluster Seperti pada gambar 3

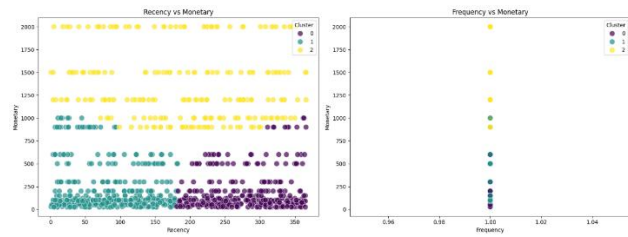
--- Analisis Karakteristik Cluster ---				
	Recency	Frequency	Monetary	Count
Cluster				
2	192.76	1.00	1357.02	228
1	87.76	1.00	211.24	392
0	275.35	1.00	167.88	380

Gambar 4 Karakteristik Dari Masing-masing kluster

Berdasarkan hasil analisis, kluster 2 menunjukkan rata-rata nilai **monetary** tertinggi yaitu 1357.02, dengan *recency* 192.76 dan *frequency* 1.00. Ini mengindikasikan bahwa pelanggan dalam kluster ini meskipun tidak terlalu sering bertransaksi dan tidak terlalu baru, memiliki kontribusi nilai belanja yang sangat tinggi. Kluster ini dapat diidentifikasi sebagai pelanggan bernilai tinggi.

Kluster 1 memiliki *recency* yang rendah (87.76) namun *monetary* yang sedang (211.24), menunjukkan bahwa pelanggan dalam kluster ini adalah pelanggan yang masih aktif dalam bertransaksi namun dengan nilai pembelian yang relatif moderat.

Sebaliknya, kluster 0 memiliki *recency* tertinggi (275.35), menunjukkan bahwa pelanggan pada kluster ini telah cukup lama tidak melakukan pembelian, dengan nilai belanja (*monetary*) yang paling rendah, yaitu 167.88. Kluster ini bisa dianggap sebagai pelanggan yang berisiko untuk churn (berhenti). kluster 0 memiliki *recency* tertinggi (275.35) seperti pada gambar5



Gambar 5 kluster 0 memiliki *recency* tertinggi (275.35)

Gambar pertama menampilkan scatter plot hubungan antara *Recency* dan *Monetary*. Terlihat bahwa kluster 2 (kuning) terkonsentrasi pada nilai *monetary* tinggi, sementara kluster 1 (hijau) berada di tengah dan kluster 0 (ungu) lebih mendominasi area dengan *recency* tinggi dan *monetary* rendah. Hal ini mendukung interpretasi bahwa kluster 2 adalah pelanggan bernilai tinggi, kluster 1 adalah pelanggan aktif dengan nilai sedang, dan kluster 0 adalah pelanggan pasif.

Pada *scatter plot* kedua antara *Frequency* dan *Monetary*, semua kluster memiliki nilai *frequency* yang sama (1.00), sehingga distribusi kluster hanya terlihat pada dimensi *monetary*. Kluster 2 kembali menempati area dengan *monetary* paling tinggi.

CONCLUSION

Penelitian ini berhasil menjawab pertanyaan mengenai bagaimana model RFM dan algoritma K-Means dapat digunakan untuk melakukan segmentasi pelanggan pada data transaksi ritel. Proses seleksi, pra-pemrosesan, dan transformasi data menghasilkan fitur RFM yang valid untuk analisis. Evaluasi menggunakan Davies-Bouldin Index menunjukkan bahwa tiga kluster merupakan struktur paling optimal. Hasil interpretasi memperlihatkan tiga segmen pelanggan utama: pelanggan bernilai tinggi, pelanggan aktif bernilai sedang, dan pelanggan pasif berisiko churn. Temuan ini menegaskan bahwa pendekatan RFM + K-Means mampu menggambarkan perilaku pelanggan secara efektif meskipun frekuensi transaksi rendah dan seragam.

REFERENCE

- [1] D.-C. Dabija, V. Câmpian, B. Philipp, and D. B. Grant, "How did consumers retail purchasing expectations and behaviour switch due to the COVID-19 pandemic?," *Journal of Marketing Analytics*, 2024, doi: 10.1057/s41270-024-00344-9.

- [2] J. Lv and X. Liu, "The impact of information overload of e-commerce platform on consumer return intention," *Int. J. Environ. Res. Public Health*, 2022, doi: 10.3390/ijerph19138060.
- [3] D. Das, A. Sarkar, and A. Debroy, "Impact of COVID-19 on changing consumer behaviour: Lessons from an emerging economy," *Int. J. Consum. Stud.*, 2022, doi: 10.1111/ijcs.12786.
- [4] R. I. Rodrigues, P. Lopes, and M. Varela, "Factors affecting impulse buying behavior of consumers," *Front. Psychol.*, 2021, doi: 10.3389/fpsyg.2021.697080.
- [5] Q. Islam and S. M. F. Ali Khan, "Assessing consumer behavior in sustainable product markets," *Sustainability*, 2024, doi: 10.3390/su16083400.
- [6] A. T. Rosário and J. C. Dias, "How has data-driven marketing evolved," *International Journal of Information Management Data Insights*, 2023, doi: 10.1016/j.jjime.2023.100203.
- [7] M. A. Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, 2023, doi: 10.1007/s10257-023-00640-4.
- [8] M. A. Uddin *et al.*, "Data-driven strategies for digital native market segmentation using clustering," *International Journal of Cognitive Computing in Engineering*, 2024, doi: 10.1016/j.ijcce.2024.04.002.
- [9] F. ; P. Alzami A.; Wibowo S., "Interactive RFM-based customer segmentation dashboard using Streamlit," *International Journal of Data Science*, vol. 7, no. 2, pp. 112–124, 2023.
- [10] P. Wongoutong, "Untitled Study," 2024.
- [11] J. Wilbert, L. Hansen, and G. Pellegrino, "Data-driven segmentation using RFM, clustering, and behavioral modeling in omnichannel retail," *Journal of Business Analytics*, vol. 6, no. 2, 2023, doi: 10.1080/2573234X.2023.2181123.
- [12] A. et al. Fitriyanto, "Untitled Study," 2025.