



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 965–976

Analisis Segmentasi Pelanggan Berbasis Data Transaksi Menggunakan Metode K-Means

Nadhif Arif Shidqi, Hasbi Firmansyah, Wahyu Asriyani,
Rizki Prasetyo Tulodo

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4017>

How to Cite this Article

APA : Arif Shidqi, N., Firmansyah, H. ., Asriyani, W., & Prasetyo Tulodo, R. (2025). Analisis Segmentasi Pelanggan Berbasis Data Transaksi Menggunakan Metode K-Means . Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 965 – 976.
<https://doi.org/10.32672/mister.v3i1.4017>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Analisis Segmentasi Pelanggan Berbasis Data Transaksi Menggunakan Metode K-Means

Nadhif Arif Shidqi¹, Hasbi Firmansyah², Wahyu Asriyani³, Rizki Prasetyo Tulodo⁴

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Kota Tegal,
Tegal, Indonesia^{1,2,4}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Pendidikan, Universitas
Pancasakti Kota Tegal, Tegal, Indonesia³

*Email: nadifarif023@gmail.com¹, hasbifirmansyah@upstegal.ac.id², asriyani1409@gmail.com³,
rizki.prasetyo.tulodo@gmail.com⁴

Diterima: 10-12-2025

| Disetujui: 20-12-2025

| Diterbitkan: 22-12-2025

ABSTRACT

This study aims to analyze customer segmentation based on transaction data using the K-Means Clustering method as an effort to improve the effectiveness of data-driven marketing strategies. The main problem identified is the suboptimal utilization of customer transaction data in forming accurate and representative segments, causing difficulties for companies in prioritizing customers and designing more targeted marketing interventions. The research uses a customer transaction dataset that includes behavioral attributes such as purchase frequency, transaction value, balance, credit limit, and payment activities. The research process consists of data preprocessing, attribute transformation, handling of missing values, normalization, K-Means modeling, model application, and cluster quality evaluation using the Davies–Bouldin Index. The results show that the K-Means algorithm successfully grouped the customers into five clusters with distinct behavioral characteristics. The evaluation produced a Davies–Bouldin Index score of 3.012, indicating that the separation quality between clusters can still be improved; however, the clustering structure provides an initial overview of customer segments. These findings emphasize that a clustering approach can yield relevant insights into customer transaction patterns and serve as a foundation for developing more targeted marketing strategies, service personalization, and customer retention programs.

Keywords: K-Means Clustering, Customer Segmentation, Transaction Data, Data Preprocessing, Davies–Bouldin Index, Marketing Strategy, Customer Behavior.

ABSTRAK

Penelitian ini bertujuan untuk menganalisis segmentasi pelanggan berdasarkan data transaksi menggunakan metode K-Means Clustering sebagai upaya meningkatkan efektivitas strategi pemasaran berbasis data. Permasalahan utama yang diidentifikasi adalah belum optimalnya pemanfaatan data transaksi pelanggan dalam membentuk segmen yang akurat dan representatif, sehingga perusahaan sering mengalami kesulitan dalam menentukan prioritas pelanggan serta merancang intervensi pemasaran yang lebih tepat sasaran. Penelitian ini menggunakan dataset transaksi pelanggan yang mencakup atribut perilaku seperti frekuensi pembelian, nilai transaksi, saldo, limit kredit, serta aktivitas pembayaran. Proses penelitian meliputi tahapan praproses data, transformasi atribut, penanganan missing values, normalisasi, pemodelan K-Means, penerapan model, serta evaluasi kualitas kluster menggunakan Davies–Bouldin Index. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data pelanggan ke dalam

lima klaster dengan karakteristik perilaku yang berbeda. Evaluasi menggunakan Davies–Bouldin Index menghasilkan nilai 3.012 yang mengindikasikan bahwa kualitas pemisahan antar klaster masih dapat ditingkatkan, namun struktur klaster tetap dapat memberikan gambaran awal mengenai segmen pelanggan. Temuan ini menegaskan bahwa pendekatan clustering mampu memberikan wawasan yang relevan mengenai pola transaksi pelanggan dan dapat dijadikan dasar dalam penyusunan strategi pemasaran yang lebih terarah, personalisasi layanan, serta pengembangan program retensi pelanggan.

Katakunci: K-Means Clustering, Segmentasi Pelanggan, Data Transaksi, Praproses Data, Indeks Davies–Bouldin, Strategi Pemasaran, Perilaku Pelanggan

PENDAHULUAN

Perkembangan pesat transaksi online dan digitalisasi dalam likuiditas data pelanggan membuat perusahaan semakin dihadapkan pada kebutuhan untuk memahami perilaku konsumen dengan lebih mendalam. Segmentasi pelanggan, yaitu proses mengelompokkan pelanggan ke dalam kelompok-kelompok homogen berdasarkan pola transaksi atau karakteristik lain, kini menjadi fondasi penting dalam strategi pemasaran, retensi, dan personalisasi layanan perusahaan. Dengan segmentasi yang tepat, perusahaan dapat mengenali kelompok pelanggan bernilai tinggi (high-value customers), pelanggan loyal, serta pelanggan dengan perilaku potensial churn, sehingga memudahkan dalam merancang strategi pemasaran yang lebih efisien dan tertarget.

Metode clustering, khususnya algoritma K-Means Clustering, telah banyak digunakan dalam literatur sebagai teknik segmentasi pelanggan yang efisien dan mudah diterapkan. K-Means menawarkan keseimbangan antara kompleksitas komputasi dan interpretabilitas hasil, sehingga cocok untuk dataset transaksi dalam skala besar. Dalam sebuah penelitian oleh *Customer Segmentation using K-Means Clustering* (2022), K-Means berhasil mengelompokkan pelanggan berdasarkan data transaksi menjadi segmen yang berbeda, yang kemudian dapat digunakan perusahaan untuk strategi pemasaran yang lebih terfokus (Awaliyah et al., 2024). Penelitian ini tersedia dalam versi PDF dan menguraikan secara sistematis bagaimana clustering diaplikasikan pada data pelanggan nyata.

Lebih jauh, dalam studi *Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm* (2024), K-Means diterapkan pada dataset transaksi e-commerce dengan menggunakan metode RFM (Recency, Frequency, Monetary) untuk membentuk segmen pelanggan. Hasil menunjukkan bahwa pembentukan tiga klaster memberikan segmentasi yang bermakna dan memberikan wawasan tentang perilaku pelanggan berdasarkan nilai dan frekuensi transaksi (Ardana et al., 2024). Sementara itu, penelitian lokal seperti *Analisis Segmentasi Pelanggan Menggunakan Metode K-Means Clustering* (2022) menerapkan K-Means pada data pelanggan supermarket dan menemukan dua klaster utama: konsumen prioritas dan konsumen biasa, yang kemudian digunakan sebagai dasar strategi promosi dan loyalitas pelanggan (Nugraha et al., 2025).

Selain itu, penelitian lain menunjukkan bahwa penerapan K-Means pada data ritel skala besar mampu meningkatkan efektivitas analisis perilaku pelanggan melalui pemilihan klaster optimal (Winaryanti & Hadi, 2024), sementara studi lain menyoroti bahwa kombinasi metode K-Means dengan fitur RFM dapat menghasilkan segmentasi dengan akurasi lebih tinggi (Putri et al., 2024). Penelitian mengenai segmentasi pelanggan berbasis machine learning juga menegaskan bahwa K-Means merupakan algoritma yang paling efisien untuk dataset non-hirarkis berukuran besar (Tabianan & Velu, 2022). Studi lain menambahkan bahwa K-Means mampu memberikan hasil segmentasi yang stabil selama proses pra-proses data diterapkan dengan benar, khususnya pada normalisasi dan transformasi fitur (Atmoko et al., 2025). Penelitian selanjutnya pada data e-commerce Indonesia mengonfirmasi bahwa metode ini sangat efektif untuk mengidentifikasi perilaku belanja konsumen berdasarkan nilai transaksi dan frekuensinya (Yusak et al., 2024). Selain itu, analisis komparatif dengan algoritma clustering lainnya menunjukkan bahwa K-Means masih unggul dalam kesederhanaan dan kecepatan komputasi (No et al., 2025), serta memperoleh skor evaluasi yang lebih baik ketika diuji menggunakan Silhouette Coefficient (Arifianto et al., 2024).

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menerapkan metode K-Means pada dataset transaksi pelanggan aktual untuk mengidentifikasi segmen pelanggan dengan karakteristik berbeda. Melalui segmentasi ini diharapkan dapat diperoleh wawasan yang mendalam mengenai perilaku pelanggan, nilai pelanggan, dan potensi intervensi strategis seperti personalisasi promosi, retensi, atau program loyalitas yang dapat meningkatkan efisiensi pemasaran dan profitabilitas perusahaan.

METODE PENELITIAN

Analisa Permasalahan

Dalam era digital saat ini, perusahaan menghadapi tantangan besar dalam mengolah dan memanfaatkan data transaksi pelanggan yang jumlahnya terus meningkat secara signifikan. Data tersebut pada dasarnya menyimpan berbagai pola perilaku konsumen, namun sering kali tidak dimanfaatkan secara optimal karena kurangnya proses analitik yang memadai. Banyak perusahaan masih menggunakan segmentasi statis, bersifat subjektif, dan tidak berbasis data, sehingga menghasilkan strategi pemasaran yang tidak terarah dan tidak sesuai dengan kebutuhan pelanggan (Model, 2021). Ketidakmampuan mengidentifikasi kelompok pelanggan berdasarkan pola pembelian aktual menyebabkan perusahaan kesulitan dalam menentukan prioritas pelanggan bernilai tinggi, pelanggan dengan risiko churn, dan pelanggan loyal yang memerlukan program retensi khusus (Elektro, 2024).

Selain itu, kompleksitas data transaksi seperti variasi frekuensi belanja, nilai pembelian, dan waktu transaksi membuat analisis manual menjadi tidak efektif. Penelitian menunjukkan bahwa perusahaan yang tidak melakukan segmentasi berbasis data cenderung mengalami inefisiensi dalam alokasi anggaran pemasaran serta tidak mampu meningkatkan personalisasi layanan kepada pelanggan (Robo et al., 2024). Oleh karena itu, dibutuhkan metode analisis yang mampu menangani data besar secara sistematis dan menghasilkan pengelompokan pelanggan secara akurat.

Algoritma *K-Means Clustering* telah diakui sebagai salah satu metode paling efektif untuk segmentasi pelanggan karena mampu mengelompokkan data berdasarkan kemiripan karakteristik dengan cepat, skalabel, dan mudah diinterpretasikan (Pradina et al., 2024). Berbagai penelitian sebelumnya menunjukkan bahwa K-Means mampu menghasilkan segmentasi yang lebih representatif dalam konteks ritel, e-commerce, maupun industri jasa (Aslantas et al., 2023). Permasalahan inti yang ingin dijawab dalam penelitian ini meliputi bagaimana mengolah data transaksi pelanggan menjadi informasi yang bermakna, bagaimana membentuk segmen pelanggan yang homogen menggunakan K-Means, serta bagaimana hasil segmentasi tersebut dapat digunakan untuk mendukung strategi pemasaran yang lebih efektif dan terarah.

Rancangan Metode

1. Pengumpulan Dataset

Tabel 1. Dataset Pelanggan

CU ST_ ID	BALA NCE	BALANCE _FREQUE NCY	PUR CHA SES	ONEOFF_ PURCHAS ES	INSTALLME NTS_PURCH ASES	CASH_ ADVANCE	PURCHASE S_FREQUE NCY	ONEOFF_PURC HASES_FREQU ENCY
C10 001	40.90 0.749	0.818182	95.4	0	95.4	0	0.166667	0
C10 002	3.202. 467.4 16	0.909091	0	0	0	6.442. 945.48 3	0	0
C10 003	2.495. 148.8 62	1	773. 17	773.17	0	0	1	1
C10 004	1.666. 670.5 42	0.636364	1499	1499	0	205.78 8.017	0.083333	0.083333
C10 005	817.7 14.33 5	1	16	16	0	0	0.083333	0.083333

C10 006	1.809. 828.7 51	1	1333 .28	0	1333.28	0	0.666667	0
C10 007	627.2 60.80 6	1	7091 .01	6402.63	688.38	0	1	1
C10 008	1.823. 652.7 43	1	436. 2	0	436.2	0	1	0
C10 009	1.014. 926.4 73	1	861. 49	661.49	200	0	0.333333	0.083333
C10 010	152.2 25.97 5	0.545455	1281 .5	1281.6	0	0	0.166667	0.166667
C10 011	1.293. 124.9 39	1	920. 12	0	920.12	0	1	0
C10 012	630.7 94.74 4	0.818182	1492 .18	1492.18	0	0	0.25	0.25
C10 013	151.6 92.86 2	1	3217 .99	2500.23	717.76	0	1	0.25

PURCHASES_INST ALLMENTS_FREQ UENCY	CASH_ADVA NCE_FREQU ENCY	CASH_A DVANCE _TRX	PURCH ASES_T RX	CREDI T_LI MIT	PAYM ENTS	MINIMU M_PAYM ENTS	PRC_FUL L_PAYM ENT	TE NU RE
0.083333	0	0	2	1000	201.8 02.08 4	139.509.7 87	0	12
0	0.25	4	0	7000	4.103 .032. 597	1.072.340 .217	0.22222 2	12
0	0	0	12	7500	622.0 66.74 2	627.284.7 87	0	12
0	0.083333	1	1	7500	0		0	12
0	0	0	1	1200	678.3 34.76 3	244.791.2 37	0	12

0.583333	0	0	8	1800	140.0 05.77 7	2.407.246 .035	0	12
1	0	0	64	1350 0	6.354 .314. 328	198.065.8 94	1	12
1	0	0	12	2300	679.0 65.08 2	53.203.39 9	0	12
0.25	0	0	5	7000	688.2 78.56 8	311.963.4 09	0	12
0	0	0	3	1100 0	1.164 .770. 591	100.302.2 62	0	12
1	0	0	12	1200	1.083 .301. 007	2.172.697 .765	0	12
0	0	0	6	2000	705.6 18.62 7	155.549.0 69	0	12
0.916667	0	0	26	3000	608.2 63.68 9	490.207.0 13	0.25	12

Data penelitian diperoleh dari dataset transaksi pelanggan, atribut-atribut tersebut mencerminkan perilaku pelanggan dalam melakukan transaksi, seperti jumlah pembelian, saldo, limit kredit, penggunaan cash advance, dan nilai pembayaran. Dataset diimpor ke RapidMiner menggunakan operator.

Praproses Data

Tahap praproses data dilakukan untuk meningkatkan kualitas data sebelum masuk ke proses clustering. Tahapan praproses dapat dilihat pada gambar 1.



Gambar 1. Tahapan Proses

a. Retrieve (Retrieve customerkuu)

Operator Retrieve digunakan untuk mengambil dataset yang tersimpan di dalam *Local Repository* RapidMiner. Dataset yang digunakan merupakan data transaksi pelanggan yang telah disimpan sebelumnya dalam format tabel. Operator ini menjadi titik awal proses karena menyediakan data mentah yang akan dianalisis pada tahap selanjutnya.

b. Transformasi Atribut Nominal

Atribut yang bersifat kategorikal diubah menjadi data numerik menggunakan operator Nominal to Numerical. Transformasi ini diperlukan karena algoritma K-Means melakukan pengelompokan berdasarkan perhitungan jarak, sehingga seluruh atribut harus berada dalam bentuk numerik.

c. Penanganan Missing Values

Beberapa atribut pada dataset memiliki nilai kosong yang ditandai dengan simbol "?". Operator Replace Missing Values digunakan untuk mengganti nilai hilang dengan nilai rata-rata dari atribut terkait. Penanganan ini dilakukan untuk menghindari distorsi hasil clustering akibat data yang tidak lengkap.

d. Normalisasi Data

Normalisasi dilakukan untuk menyeragamkan skala antar atribut sehingga tidak ada atribut tertentu yang mendominasi perhitungan jarak. Proses normalisasi diterapkan secara otomatis melalui konfigurasi operator K-Means.

Pemodelan K-Means Clustering

Penerapan algoritma K-Means dilakukan untuk mengelompokkan pelanggan berdasarkan kemiripan perilaku transaksi. Parameter yang digunakan dalam model K-Means pada penelitian ini meliputi:

- Jumlah klaster (k): 5
- Maximum iterations: 100
- Distance measure: Euclidean distance
- Random seed: 2001

Model K-Means kemudian dijalankan untuk membentuk lima klaster pelanggan yang masing-masing merepresentasikan pola transaksi yang berbeda.

Penerapan Model pada Dataset

Setelah model berhasil dibangun, operator Apply Model digunakan untuk menerapkan model tersebut pada seluruh dataset. Proses ini menghasilkan label klaster untuk masing-masing pelanggan. Setiap klaster dianalisis berdasarkan kecenderungan atribut dominannya, seperti intensitas pembelian, besar saldo, atau rata-rata pembayaran.

Evaluasi Kualitas Klaster

Evaluasi klaster dilakukan dengan menggunakan operator Performance (Clustering) untuk mengukur kualitas dan validitas hasil pengelompokan. Evaluasi dilakukan melalui beberapa indikator seperti:

- Nilai indeks Davies-Bouldin
- Jarak antar pusat klaster
- Sebaran anggota klaster
- Variasi internal dalam setiap klaster

Tahap evaluasi ini memastikan bahwa klaster yang dihasilkan memiliki keterpisahan yang baik dan struktur yang jelas sehingga dapat digunakan untuk analisis segmentasi pelanggan secara akurat.

Tahapan Algoritma K-Means dan Rumus Matematikanya

Algoritma K-Means merupakan metode *unsupervised learning* yang bertujuan membagi data ke dalam kkk klaster berdasarkan kedekatan jarak. Tahapan-tahapan algoritma K-Means dapat dijelaskan sebagai berikut:

1. Inisialisasi Jumlah Klaster

Tentukan jumlah klaster yang ingin dibentuk, yaitu k.
Selanjutnya pilih pusat awal klaster (*centroid*) secara acak:

$$\mathcal{C} = \{c_1, c_2, \dots, c_k\}$$

2. Menghitung Jarak Data ke Setiap Centroid

Jarak antara titik data $\mathcal{X}_i$ dengan centroid $\mathcal{C}_j$ dihitung menggunakan rumus Euclidean Distance:

$$d(\mathcal{X}_i, \mathcal{C}_j) = \sqrt{\sum_{m=0}^n (\mathcal{X}_{im} - \mathcal{C}_{jm})^2}$$

Keterangan:

- $\mathcal{X}_i$ = data ke-i
- $\mathcal{C}_j$ = centroid klaster ke-j
- m = atribut ke-m
- n = jumlah atribut

3. Pengelompokan Data ke Klaster Terdekat

Setiap data $\mathcal{X}_i$ dimasukkan ke klaster yang memiliki jarak minimum terhadap centroid:

$$\text{Cluster}(\mathcal{X}_i) = \arg \min d(\mathcal{X}_i, \mathcal{C}_j)$$

Artinya, data dimasukkan ke klaster dengan nilai jarak terkecil.

4. Pembaruan Nilai Centroid

Setelah semua data masuk ke klaster masing-masing, centroid diperbarui dengan mengambil nilai rata-rata dari semua titik data dalam klaster tersebut.

Centroid baru dihitung dengan rumus:

$$\mathcal{C}_j = \frac{1}{|K_j|} \sum_{\mathcal{X}_i \in K_j} \mathcal{X}_i$$

Keterangan:

- K_j = kumpulan data dalam kluster j
- $|K_j|$ = jumlah data pada kluster j

5. Pengecekan Konvergensi

Proses iterasi akan berhenti jika:

1. Centroid tidak berubah lagi
atau
2. Perubahan posisi centroid sangat kecil
atau
3. Jumlah iterasi sudah mencapai batas maksimum

Secara matematis:

$$C_j^{(t)} = C_j^{(t-1)}$$

Jika kondisi terpenuhi → algoritma berhenti.

Jika tidak → kembali ke langkah 2.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pemodelan algoritma K-Means menggunakan RapidMiner berdasarkan data transaksi pelanggan. Proses analisis difokuskan pada dua keluaran utama, yaitu pembentukan kluster dan evaluasi kualitas kluster menggunakan Davies–Bouldin Index (DBI). Seluruh hasil ditampilkan dalam bentuk narasi sesuai standar penulisan ilmiah.

1. Hasil Pembentukan Cluster

Algoritma K-Means berhasil mengelompokkan 160 data transaksi ke dalam 5 kluster. Distribusi jumlah anggota setiap kluster ditunjukkan pada table sebagai berikut:

Tabel 2. Hasil Pembentukan Cluster

Cluster 0	59 items
Cluster 1	64 items
Cluster 2	11 items
Cluster 3	15 items
Cluster 4	11 items
Total	160 items

Distribusi tersebut menunjukkan bahwa Cluster 0 dan Cluster 1 merupakan kluster dengan anggota paling banyak, sedangkan Cluster 2 dan Cluster 4 memiliki anggota paling sedikit. Perbedaan jumlah

anggota ini mengindikasikan adanya variasi karakteristik pelanggan yang cukup signifikan dalam data transaksi.

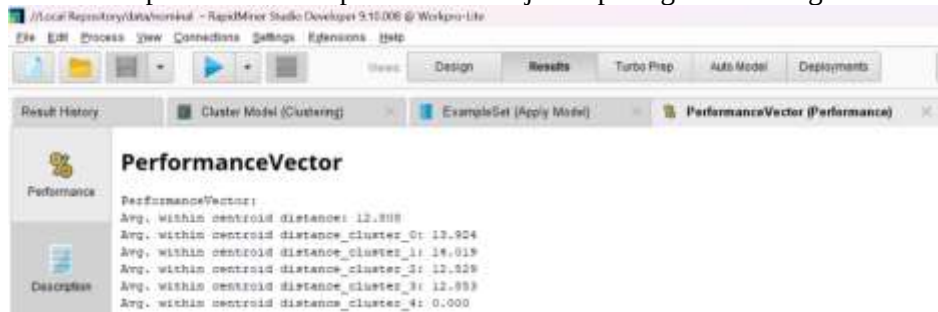
Setiap kluster dibentuk berdasarkan kemiripan nilai atribut yang digunakan pada proses clustering. Semakin kecil jarak antara data dengan centroid, semakin homogen suatu kluster.

2. Evaluasi Kualitas Kluster

Evaluasi model dilakukan menggunakan metrik Average Within-Centroid Distance dan Davies–Bouldin Index yang merupakan ukuran umum untuk menilai kualitas pemisahan kluster.

2.1 Average Within-Centroid Distance

Nilai jarak rata-rata setiap kluster terhadap centroid ditunjukkan pada gambar sebagai berikut:



Gambar 2. Hasil Evaluasi Cluster Average Within-Centroid Distance

Nilai-nilai tersebut menunjukkan tingkat kekompakan (compactness) setiap kluster. Kluster dengan nilai yang lebih kecil memiliki anggota yang lebih mirip satu sama lain. Nilai 0.000 pada Cluster 4 menunjukkan adanya kemungkinan bahwa kluster tersebut hanya berisi data dengan nilai atribut yang sangat serupa atau centroid berhimpit dengan salah satu titik data.

2.2 Davies–Bouldin Index



Gambar 3. Hasil Evaluasi Davies–Bouldin Index

Nilai DBI dari model adalah:

Davies–Bouldin Index (DBI) = 3.012

Nilai DBI menunjukkan tingkat pemisahan dan kekompakan antar kluster. Semakin kecil nilai DBI, semakin baik kualitas kluster. Nilai 3.012 mengindikasikan bahwa pemisahan antar kluster belum optimal,

namun struktur klaster tetap dapat digunakan sebagai dasar analisis segmentasi pelanggan. Nilai DBI yang masih cukup besar dapat disebabkan oleh:

1. Variansi data yang tinggi pada beberapa atribut.
2. Centroid antar klaster yang relatif berdekatan.
3. Perbedaan jumlah anggota klaster yang cukup ekstrem.

3. Pembahasan

Hasil pengelompokan menunjukkan bahwa pelanggan terbagi ke dalam lima segmen dengan karakteristik yang berbeda. Dua klaster besar (Cluster 0 dan 1) menggambarkan kelompok pelanggan dengan pola transaksi yang umum atau mendominasi dalam dataset. Sementara klaster yang lebih kecil (Cluster 2, 3, dan 4) mengindikasikan adanya kelompok pelanggan dengan pola transaksi yang lebih spesifik atau berbeda dari mayoritas.

Nilai DBI sebesar 3.012 menunjukkan bahwa walaupun algoritma telah membentuk klaster, tingkat pemisahan antar klaster masih dapat ditingkatkan. Hal ini dapat diperbaiki dengan melakukan:

- normalisasi atau standardisasi ulang data,
- penyesuaian jumlah klaster (misalnya mencoba $k = 4$ atau $k = 6$),
- atau pemilihan atribut yang lebih relevan untuk representasi perilaku pelanggan.

Secara keseluruhan, model K-Means yang dihasilkan dapat memberikan gambaran awal mengenai segmentasi pelanggan berdasarkan data transaksi. Segmentasi ini penting untuk mendukung strategi yang lebih tepat sasaran, seperti rekomendasi produk, promosi yang dipersonalisasi, atau analisis profitabilitas per segmen.

KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma K-Means mampu melakukan segmentasi pelanggan berdasarkan data transaksi dengan membentuk lima klaster yang merepresentasikan variasi perilaku pelanggan. Tahapan praproses data seperti transformasi atribut, penanganan missing value, dan normalisasi terbukti penting untuk memastikan kualitas data sebelum pemodelan, sehingga proses pengelompokan dapat berjalan optimal. Hasil clustering menunjukkan bahwa dua klaster memiliki jumlah anggota paling besar, mencerminkan kelompok pelanggan dengan pola transaksi dominan, sementara klaster lainnya menggambarkan segmen pelanggan dengan karakteristik transaksi yang lebih khusus. Evaluasi menggunakan Davies-Bouldin Index menghasilkan nilai 3.012, yang mengindikasikan bahwa pemisahan klaster belum ideal akibat variansi data yang tinggi dan kedekatan antar centroid, meskipun struktur klaster yang terbentuk tetap dapat memberikan gambaran awal yang berguna dalam memahami segmentasi pelanggan. Secara keseluruhan, metode K-Means dapat menjadi alat analisis yang efektif untuk mendukung pengambilan keputusan strategis, seperti perancangan promosi, personalisasi layanan, dan identifikasi nilai pelanggan, dengan potensi peningkatan kualitas clustering melalui optimasi parameter dan seleksi atribut yang lebih tepat.

DAFTAR PUSTAKA

- Ardana, C. H., Aldita, A., Aisyah, A., Khoyum, A., & Faisal, M. (2024). *Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering*. 9(1), 1–9.
<https://doi.org/10.14421/jiska.2024.9.1.1-9>
- Arifianto, F., Hasudungan, J., Muzaky, A., & Achsan, H. T. Y. (2024). *Segmentasi Pelanggan Berdasarkan Recency, Frequency, dan Monetary dengan K-Means Clustering : Studi Kasus Toko Pakaian Almost*

- Famous komponen penting dari strategi bisnis modern . Dalam era komputer dan internet saat ini , untuk mengumpulkan informasi b. 10(1), 122–135. <https://doi.org/10.37012/jtik.v10i1.2096>
- Aslantaş, G., Gençgöl, M., Rumelli, M., & Özşaraç, M. (2023). Customer Segmentation Using K-Means Clustering Algorithm and RFM Model K- Means Kümeleme Algoritması ve RFM Modeli Kullanarak Müşteri Segmentasyonu. 25(74), 491–503. <https://doi.org/10.21205/deufmd.2023257418>
- Atmoko, F. D., Studi, P., Informatika, T., & Ulama, U. N. (2025). Penerapan K-Means Clustering untuk Segmentasi Pelanggan Fathoni Dwi Atmoko Perusahaan di berbagai industri telah beralih ke Customer Relationship Management (CRM) yang canggih , di mana segmentasi pelanggan adalah fondasi utamanya . Dengan Segmentasi Pe. <https://doi.org/10.61132/uranus.v3i2.1214>
- Awaliyah, D. A., Budi Prasetyo, Muzayanah, R., & Lestari, A. D. (2024). Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm. *Scientific Journal of Informatics*, 11(2), 539–548. <https://doi.org/10.15294/sji.v11i2.6137>
- Elektro, J. T. (2024). K-means clustering method for customer segmentation based on potential purchases. 8(1), 83–90. <https://doi.org/10.31961/eltikom.v8i1.1137>
- Model, M. (2021). *JITE (Journal of Informatics and Telecommunication Engineering)*. 5(July). <https://doi.org/10.31289/jite.v5i1.5182>
- No, V., Artiarno, A. M., Setiaji, P., & Nugraha, F. (2025). *Edumatic : Jurnal Pendidikan Informatika K-Means Clustering untuk Segmentasi Pelanggan : Mengungkap Pola Pembelian Strategi Pemasaran pada Sektor Ritel*. 9(2), 442–451. <https://doi.org/10.29408/edumatic.v9i2.30336>
- Nugraha, R. D., Adelia, D. D., & Rivaldi, D. (2025). Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma K- Means Clustering. 5(2), 141–148. <https://doi.org/10.47709/digitech.v5i2.6340>
- Pradina, D. A., Kurniawati, Y., Afwan, A. S., & Heikal, J. (2024). RFM Segmentation and K – Means Clustering of Skincare Product (Case study Scarlett). 6(2), 213–216. <https://doi.org/10.55338/saintek.v6i2.3644>
- Putri, Y., Aldo, D., Ilham, W., Padang, U. T., & Cendekia, C. I. (2024). Retail Marketing Strategy Optimization : Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering. 8(4), 2155–2163. <https://doi.org/10.33395/sinkron.v8i4.14000>
- Robo, S., Melani, P. I., Fernatyanan, P., & Riandi, M. (2024). The Application of Customers Segmentation Using RFM Analysis Method and K-Means Clustering to Improve Marketing Strategy. 8(158), 200–211. <https://doi.org/10.30645>
- Tabianan, K., & Velu, S. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. 1–15. <https://doi.org/10.3390/su14127243>
- Winaryanti, H. S., & Hadi, H. P. (2024). Customer Segmentation Using K-Means Clustering with RFM Method (Case Study : PT . Dewangga Travindo Semarang). 9(1), 147–160. <https://doi.org/10.62411/jais.v9i1.10440>
- Yusak, A., Rumapea, N., Pratiwi, D., & Sari, S. (2024). Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency , Frequency , Monetary (RFM). 6(3), 292–299. <https://doi.org/10.55338/saintek.v6i3.4607>