

Penerapan Algoritma K-Means dan DBSCAN Untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce

Howard Chen¹, Joni², Devi³

^{1,2,3}Teknik Informatika, STMIK Time, Indonesia

Email: ¹howardchen53@email.com, ²joni_hgw@yahoo.com, ³d3v1wu89@email.com

ABSTRAK

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan pada data transaksi perdagangan elektronik (*e-commerce*) menggunakan algoritma *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*. Data transaksi ditransformasikan menggunakan model *Recency*, *Frequency*, dan *Monetary (RFM)* serta dinormalisasi dengan metode *Min-max scaling*. Evaluasi performa kedua algoritma diukur berdasarkan nilai *silhouette score*, jumlah kelompok (*cluster*) yang terbentuk, dan efisiensi waktu komputasi. Hasil eksperimen menunjukkan bahwa algoritma *K-Means* lebih unggul dan fungsional untuk segmentasi pelanggan pada *dataset* ini karena berhasil membentuk dua kelompok pelanggan yang stabil dengan waktu komputasi instan sebesar 0,004 detik. Sebaliknya, algoritma *DBSCAN* mengalami kegagalan fungsi segmentasi (*failing to cluster*) dengan hanya menghasilkan satu kelompok tunggal, meskipun secara matematis memiliki nilai *silhouette score* sebesar 0,427. Selain proses analisis, penelitian ini juga mengimplementasikan sistem simulasi berbasis web menggunakan kerangka kerja Flask untuk mempermudah visualisasi hasil segmentasi. Penelitian ini menyimpulkan bahwa algoritma *K-Means* merupakan metode yang paling adaptif dan direkomendasikan untuk mendukung pengambilan keputusan bisnis pada platform ini.

Kata Kunci: *clustering, K-Means, DBSCAN, e-commerce, Segmentasi Pelanggan*

ABSTRACT

This study aimed to perform customer segmentation on *e-commerce* transaction data using *K-Means* and *Density-Based Spatial clustering of Applications with noise (DBSCAN)* algorithms. Transaction data were transformed using the *Recency*, *Frequency*, and *Monetary (RFM)* model and normalized using the *Min-max scaling* method. The performance of both algorithms was evaluated based on the *silhouette score*, the number of generated clusters, and computational time efficiency. The experimental results indicated that the *K-Means* algorithm was superior and more functional for customer segmentation on this dataset, successfully forming two stable clusters with an instant processing time of 0.004 seconds. Conversely, the *DBSCAN* algorithm failed to perform proper segmentation by generating only a single cluster, despite producing a mathematically higher *silhouette score* of 0.427. Additionally, this study implemented a web-based simulation system using the Flask framework to facilitate the visualization of segmentation results. This study concluded that the *K-Means* algorithm was the most adaptive and recommended method to support business decision-making.

Keywords: *clustering, K-Means, DBSCAN, e-commerce, Customer Segmentation*

Penulis Korespondensi:

Nama Howard Chen

Email: howardchen53@gmail.com

Article Info

Diterima: 21 Mei 2026

Direvisi: 27 Mei 2026

Disetujui: 28 Mei 2026

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Perkembangan teknologi informasi mendorong pertumbuhan transaksi digital pada sektor perdagangan melalui platform *e-commerce*. Aktivitas transaksi yang semakin tinggi menghasilkan data pelanggan dalam jumlah besar, seperti waktu transaksi, frekuensi pembelian, dan total pengeluaran. Data tersebut memiliki potensi untuk digunakan dalam mendukung pengambilan keputusan bisnis, terutama dalam memahami perilaku pelanggan dan menentukan strategi pemasaran yang lebih tepat sasaran [1].

Salah satu pendekatan yang umum digunakan untuk memahami perilaku pelanggan adalah segmentasi pelanggan. Segmentasi pelanggan dilakukan dengan mengelompokkan pelanggan berdasarkan kemiripan karakteristik atau pola transaksi sehingga perusahaan dapat mengenali kelompok pelanggan bernilai tinggi, pelanggan aktif, maupun pelanggan yang berisiko berhenti melakukan transaksi. Pendekatan ini membantu perusahaan menyusun strategi promosi, program loyalitas, dan pengelolaan hubungan pelanggan secara lebih efektif [11].

Pada bidang *data mining*, teknik *clustering* menjadi metode yang banyak digunakan untuk melakukan segmentasi pelanggan karena mampu membentuk kelompok tanpa memerlukan label data terlebih dahulu [6]. Dua algoritma yang sering digunakan adalah *K-Means* dan *DBSCAN*. *K-Means* bekerja dengan membagi data ke dalam sejumlah *cluster* berdasarkan kedekatan terhadap *centroid* dan dikenal memiliki proses komputasi yang cepat serta sederhana [8]. Namun, metode ini sensitif terhadap *outlier* dan kurang optimal ketika pola data tidak membentuk kelompok yang teratur [2]. Sebaliknya, *DBSCAN* membentuk *cluster* berdasarkan kepadatan data sehingga lebih mampu mendeteksi *noise* dan menangani bentuk *cluster* yang kompleks tanpa menentukan jumlah *cluster* sejak awal [9].

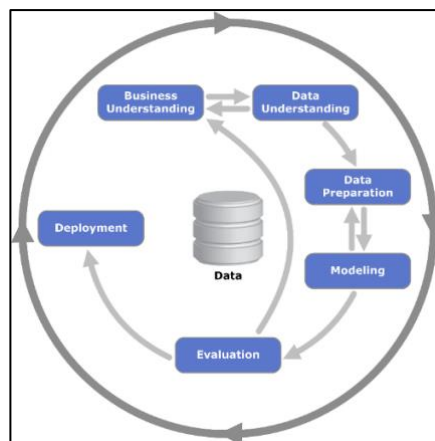
Meskipun kedua algoritma tersebut telah banyak digunakan, pemilihan metode *clustering* yang sesuai masih menjadi tantangan karena kualitas segmentasi sangat dipengaruhi oleh karakteristik data transaksi. Beberapa penelitian sebelumnya menunjukkan hasil yang berbeda. Paramita dan Hariguna [2] menunjukkan bahwa *K-Means* memberikan efisiensi komputasi yang lebih baik, sedangkan *DBSCAN* menghasilkan kualitas *cluster* yang lebih baik pada distribusi data tertentu. Penelitian lain juga menunjukkan bahwa penggunaan atribut perilaku pelanggan seperti *Recency*, *Frequency*, dan *Monetary* (RFM) dapat meningkatkan representasi karakteristik pelanggan dalam proses segmentasi [10], [12].

Berdasarkan kondisi tersebut, penelitian ini menggunakan data transaksi *e-commerce* periode 2023–2024 yang diperoleh dari *Kaggle* dan melakukan transformasi data menggunakan model RFM. Data yang telah diproses kemudian dikelompokkan menggunakan algoritma *K-Means* dan *DBSCAN*. Perbandingan dilakukan menggunakan indikator evaluasi berupa *silhouette score*, jumlah *cluster* yang terbentuk, dan waktu proses untuk memperoleh gambaran performa masing-masing algoritma.

Kontribusi penelitian ini terletak pada perbandingan dua pendekatan *clustering* pada data transaksi *e-commerce* yang telah ditransformasikan menggunakan RFM serta penyediaan sistem berbasis web untuk mempermudah proses *clustering* dan visualisasi hasil segmentasi pelanggan. Hasil penelitian diharapkan dapat menjadi referensi dalam menentukan metode segmentasi pelanggan yang sesuai dengan karakteristik data transaksi dan mendukung pengambilan keputusan pada lingkungan *e-commerce*.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) yang terdiri dari tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [5]. Pendekatan ini dipilih karena mampu memberikan alur penelitian yang sistematis dalam proses pengolahan *data mining* serta sering digunakan pada penelitian *clustering* dan segmentasi pelanggan. Tahapan penelitian dimulai dari pengumpulan *dataset* transaksi pelanggan *e-commerce*, kemudian dilakukan *preprocessing* data, transformasi menggunakan model RFM, proses *clustering* menggunakan algoritma *K-Means* dan *DBSCAN*, evaluasi hasil *clustering*, hingga visualisasi hasil segmentasi pelanggan sebagaimana ditunjukkan pada Gambar 1. di bawah ini .



Gambar 1. Kerangka Kerja *CRISP-DM*

2.1. Dataset Penelitian

Dataset penelitian diperoleh dari platform *Kaggle* dan berisi data transaksi pelanggan *e-commerce* pada periode 2023–2024. *Dataset* dipilih karena memiliki karakteristik transaksi yang beragam sehingga sesuai digunakan untuk proses segmentasi pelanggan berbasis *clustering*. Data transaksi kemudian diolah untuk membentuk atribut *Recency*, *Frequency*, dan *Monetary* (RFM) sebagai dasar pengelompokan pelanggan.

Atribut utama yang digunakan pada penelitian ini meliputi:

1. *Customer ID*
2. *Order Date*
3. *Quantity*
4. *Price*
5. *Total Sales*

Atribut tersebut dipilih karena mampu merepresentasikan perilaku pelanggan dari sisi waktu transaksi terakhir, intensitas transaksi, dan total nilai pembelian yang dilakukan pelanggan. Hasil transformasi RFM kemudian digunakan sebagai masukan pada algoritma *K-Means* dan *DBSCAN*.

2.2. Persiapan Data

Tahap persiapan data (*data preparation*) dilakukan untuk meningkatkan kualitas data sebelum proses *clustering* dilakukan. *dataset* yang digunakan diperoleh dari platform *Kaggle* dan terdiri dari 5.000 data transaksi pelanggan dengan 18 atribut yang merepresentasikan aktivitas transaksi pada platform *e-commerce*.

Pada tahap ini dilakukan proses *preprocessing* untuk memastikan data siap digunakan dalam proses analisis dan pengelompokan. Proses *preprocessing* meliputi pemeriksaan nilai kosong (*missing value*), data duplikat, serta data yang tidak valid. Berdasarkan hasil pemeriksaan, tidak ditemukan *missing value*, data duplikat, maupun data yang tidak sesuai sehingga tidak dilakukan proses penghapusan data.

Selanjutnya dilakukan proses transformasi data untuk membentuk atribut *Recency*, *Frequency*, dan *Monetary* (RFM) yang digunakan sebagai dasar segmentasi pelanggan. Model RFM dipilih karena mampu merepresentasikan perilaku pelanggan berdasarkan waktu transaksi terakhir, frekuensi transaksi, serta total nilai transaksi yang dilakukan pelanggan.

Setelah proses transformasi selesai, dilakukan normalisasi menggunakan metode *Min–Max Scaling* untuk menyamakan rentang nilai setiap atribut agar tidak terjadi dominasi fitur tertentu selama proses perhitungan jarak pada algoritma *clustering*. Normalisasi dilakukan dengan mengubah nilai atribut ke dalam rentang 0–1 sehingga proses segmentasi menggunakan algoritma *K-Means* dan *DBSCAN* dapat menghasilkan pengelompokan yang lebih optimal.

Tabel 1. Hasil *preprocessing* Data

Tahapan	Hasil
Jumlah Data	5.000 data
<i>missing value</i>	Tidak ditemukan (0 data)
Data Duplikat	Tidak ditemukan (0 data)
Data Tidak Valid	Tidak ditemukan (0 data)
Transformasi Data	Menggunakan model RFM
Normalisasi Data	<i>Min–Max Scaling</i> (0–1)

2.3. Transformasi Data Menggunakan RFM

Model *Recency*, *Frequency*, dan *Monetary* (RFM) digunakan untuk merepresentasikan perilaku pelanggan berdasarkan histori transaksi yang dilakukan pada platform *e-commerce*. Pendekatan ini dipilih karena mampu menyederhanakan data transaksi menjadi tiga indikator utama yang lebih representatif untuk proses segmentasi pelanggan.

Komponen RFM yang digunakan meliputi:

1. *Recency* (R) menunjukkan selisih waktu sejak pelanggan melakukan transaksi terakhir. Semakin kecil nilai *Recency*, maka pelanggan dianggap semakin aktif.
2. *Frequency* (F) menunjukkan jumlah transaksi yang dilakukan pelanggan dalam periode pengamatan. Nilai yang tinggi menunjukkan intensitas transaksi yang lebih besar.
3. *Monetary* (M) menunjukkan total nilai transaksi yang dilakukan pelanggan selama periode pengamatan.

Setelah atribut RFM terbentuk, dilakukan normalisasi menggunakan metode *Min–Max Scaling* untuk menyamakan rentang nilai setiap atribut ke skala 0–1 sehingga proses perhitungan jarak pada algoritma *clustering* dapat berjalan lebih optimal.

Tabel 2. Contoh Hasil Transformasi RFM

<i>Customer ID</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
C001	10	15	450.000
C002	35	4	125.000
C003	5	20	700.000
C004	18	10	350.000

2.4. Penerapan Algoritma *K-Means*

K-Means merupakan algoritma *clustering* berbasis *centroid* yang digunakan untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik transaksi hasil transformasi RFM. Algoritma ini bekerja dengan menentukan pusat *cluster* (*centroid*) kemudian menghitung jarak setiap data terhadap *centroid* menggunakan *Euclidean Distance*.

Tahapan *K-Means* pada penelitian ini meliputi:

1. Menentukan jumlah *cluster* (*k*) menggunakan *Elbow Method*.
2. Menentukan *centroid* awal menggunakan *K-Means++*.
3. Menghitung jarak setiap data terhadap *centroid*.
4. Mengelompokkan data ke *cluster* terdekat.
5. Memperbarui *centroid* hingga *konvergen*.

Perhitungan jarak menggunakan:

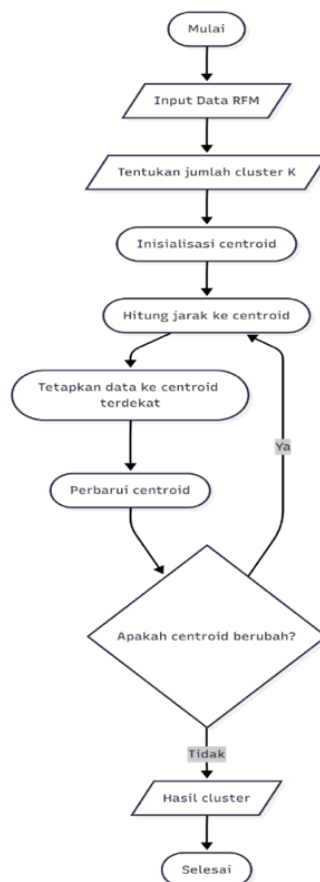
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Parameter penelitian:

Tabel 3. Parameter Algoritma *K-Means*

Parameter	Nilai
Jumlah <i>cluster</i>	2–10
Inisialisasi <i>centroid</i>	<i>K-Means++</i>
Iterasi maksimum	300
Distance metric	<i>Euclidean</i>

Untuk memberikan gambaran yang lebih jelas mengenai alur proses *K-Means* yang digunakan dalam penelitian ini, berikut disajikan sebuah *flowchart* yang merangkum tahapan utama mulai dari pengolahan data awal hingga proses *iteratif* pembaruan *centroid*. Alur proses penerapan algoritma *K-Means* dalam penelitian ini digambarkan dalam bentuk *flowchart* untuk menunjukkan hubungan antar tahapan mulai dari input data RFM hingga terbentuknya *cluster* akhir, sebagaimana ditunjukkan pada Gambar 2. di bawah ini.



Gambar 2. Penerapan Algoritma *K-Means*

2.5. Penerapan Algoritma DBSCAN

DBSCAN (*Density-Based Spatial clustering of Applications with noise*) digunakan untuk mengelompokkan pelanggan berdasarkan tingkat kepadatan data transaksi. Berbeda dengan *K-Means*, *DBSCAN* tidak memerlukan jumlah *cluster* di awal tetapi menggunakan parameter *epsilon* (ϵ) dan *MinPts*.

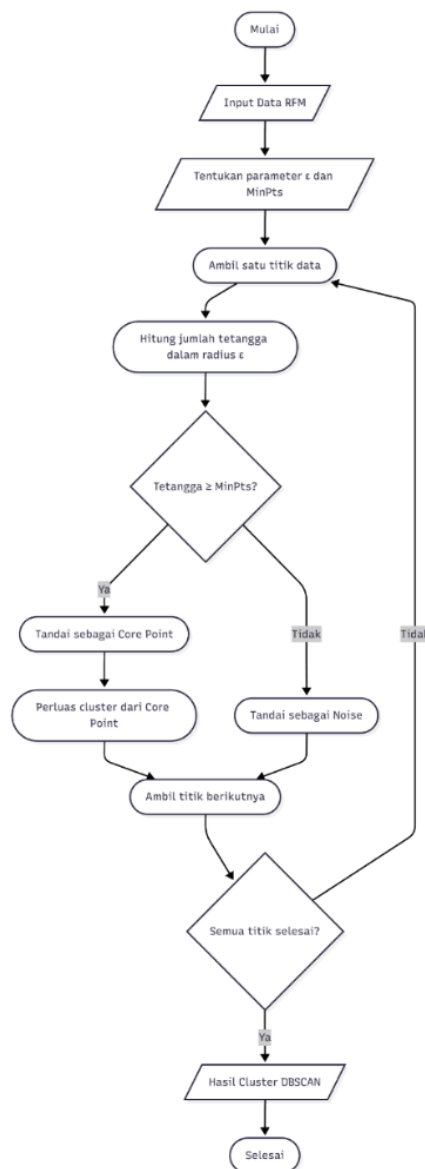
Tahapan *DBSCAN*:

1. Menentukan nilai *epsilon* (ϵ).
2. Menentukan nilai *MinPts*.
3. Mengidentifikasi *core point*.
4. Membentuk *cluster* berdasarkan kepadatan data.
5. Menandai data di luar kepadatan sebagai *noise*.

Tabel 4. Parameter Algoritma *DBSCAN*

Parameter	Nilai
<i>Epsilon</i> (ϵ)	Eksperimen
<i>MinPts</i>	3–6
<i>Distance metric</i>	<i>Euclidean</i>

Untuk memperjelas alur proses *DBSCAN* yang digunakan dalam penelitian ini, berikut ditampilkan diagram *flowchart* yang menggambarkan tahapan lengkap mulai dari input data hingga terbentuknya *cluster* akhir. proses penerapan algoritma *DBSCAN* dalam penelitian ini digambarkan melalui *flowchart* yang menunjukkan tahapan pembentukan *cluster* berdasarkan kepadatan data serta identifikasi *noise*, sebagaimana ditunjukkan pada Gambar 3. di bawah ini.



Gambar 3. Penerapan Algoritma *DBSCAN*

2.6. Evaluasi Hasil *clustering*

Evaluasi dilakukan untuk mengukur kualitas hasil segmentasi pelanggan yang dihasilkan oleh algoritma *K-Means* dan *DBSCAN*. Pengukuran dilakukan menggunakan tiga indikator utama.

1. *Silhouette score*: mengukur kedekatan data dalam *cluster* dan keterpisahan antar *cluster*.
2. Jumlah *Cluster*: menunjukkan pola segmentasi yang terbentuk.
3. Waktu Komputasi: mengukur efisiensi algoritma.

Nilai *silhouette score* berada pada rentang -1 hingga 1 . Semakin mendekati nilai 1 menunjukkan kualitas *cluster* yang semakin baik.

Selain evaluasi numerik, dilakukan visualisasi menggunakan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data sehingga hasil segmentasi dapat diamati secara visual dalam ruang dua dimensi.

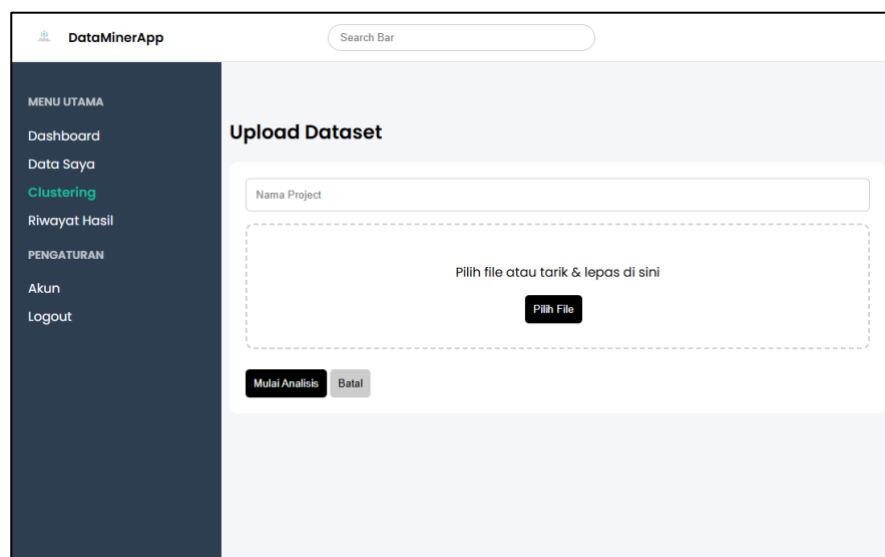
3. HASIL DAN PEMBAHASAN

3.1. Hasil Implementasi Sistem

Hasil penelitian ini tidak hanya menghasilkan proses segmentasi pelanggan menggunakan algoritma *K-Means* dan *DBSCAN*, tetapi juga menghasilkan aplikasi berbasis web yang digunakan sebagai media simulasi proses *clustering* serta visualisasi hasil segmentasi pelanggan. Sistem dikembangkan menggunakan bahasa pemrograman *Python* dengan *framework* *Flask* sehingga pengguna dapat melakukan proses analisis melalui antarmuka yang lebih mudah digunakan.

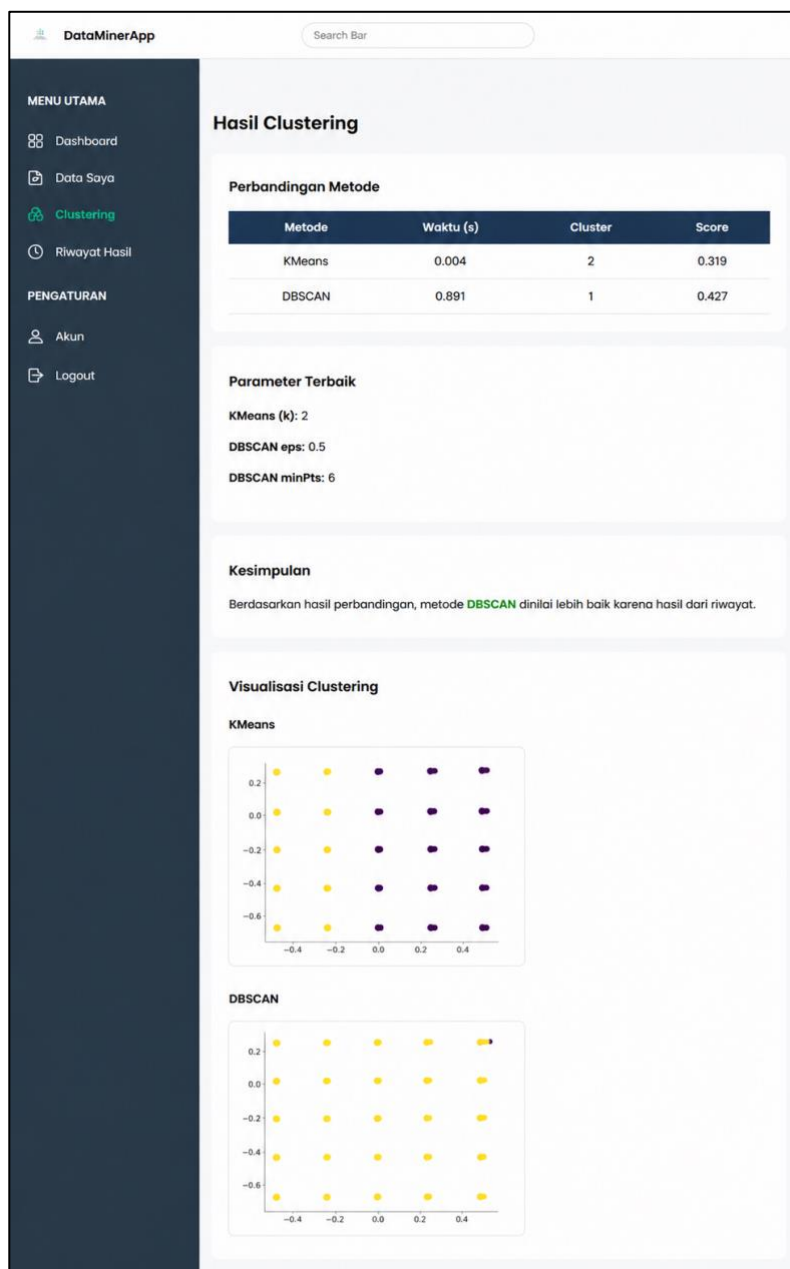
Sistem yang dibangun mendukung beberapa proses utama, mulai dari pengunggahan *dataset*, proses *clustering*, hingga penyajian hasil evaluasi dan visualisasi segmentasi pelanggan. Dalam paper ini hanya ditampilkan halaman utama yang secara langsung berkaitan dengan proses penelitian.

Halaman ini digunakan untuk melakukan analisis *clustering*. Pengguna dapat mengunggah *dataset* dan memberikan nama *project*, kemudian sistem akan memproses data menggunakan metode *K-Means* dan *DBSCAN* untuk menghasilkan pengelompokan data. Tampilan antarmuka dari halaman *clustering* dapat dilihat seperti pada Gambar 4. di bawah ini.



Gambar 4. Halaman *clustering*

Halaman ini menampilkan hasil perbandingan antara algoritma *K-Means* dan *DBSCAN* dalam proses *clustering* data. Evaluasi dilakukan berdasarkan nilai *silhouette score*, jumlah *cluster* yang terbentuk, serta waktu proses. Berdasarkan hasil tersebut, sistem menentukan metode terbaik yang paling sesuai dengan karakteristik data. Tampilan antarmuka dari halaman Hasil Perbandingan dapat dilihat seperti pada Gambar 5. di bawah ini.



Gambar 5. Halaman Hasil Perbandingan

3.2. Hasil Penerapan Algoritma *K-Means*

Penerapan algoritma *K-Means* dilakukan menggunakan data hasil transformasi *Recency*, *Frequency*, dan *Monetary* (RFM) yang telah dinormalisasi menggunakan metode *Min-Max Scaling*. Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method* dengan variasi nilai $k = 2-10$ untuk memperoleh jumlah *cluster* yang paling sesuai dengan karakteristik data pelanggan.

Hasil pengujian menunjukkan bahwa algoritma *K-Means* mampu membentuk *cluster* yang relatif stabil dengan waktu komputasi yang lebih cepat dibandingkan metode *DBSCAN*. Pengelompokan pelanggan dilakukan berdasarkan pola transaksi yang direpresentasikan melalui atribut RFM.

Tabel 5. Hasil Penerapan Algoritma *K-Means*

Parameter	Hasil
Metode	<i>K-Means</i>
Jumlah Cluster	2
<i>Silhouette score</i>	0,319
Waktu Komputasi	0,004 detik

Berdasarkan hasil evaluasi pada Tabel 5, algoritma *K-Means* menghasilkan 2 *cluster* dengan nilai *silhouette score* sebesar 0,319 dan waktu komputasi sebesar 0,004 detik. Meskipun nilai *silhouette score* berada pada kategori *fair*, struktur

pengelompokan yang dihasilkan terbukti valid dan aplikatif secara fungsional bisnis karena mampu memisahkan basis data pelanggan secara tegas menjadi kelompok yang fungsional, didukung oleh efisiensi waktu pemrosesan yang sangat instan.

3.3. Hasil Penerapan Algoritma DBSCAN

Algoritma DBSCAN diterapkan menggunakan data hasil transformasi RFM yang telah dinormalisasi. Parameter *epsilon* (ϵ) dan *MinPts* ditentukan melalui proses eksperimen untuk memperoleh hasil *clustering* yang optimal.

Hasil pengujian menunjukkan bahwa DBSCAN mampu membentuk *cluster* berdasarkan kepadatan data serta mendeteksi sebagian data yang berada di luar pola utama sebagai *noise*. Kemampuan tersebut menjadi keunggulan dibandingkan metode berbasis *centroid* karena tidak seluruh data dipaksa masuk ke dalam *cluster* tertentu.

Tabel 6. Hasil Penerapan Algoritma DBSCAN

Parameter	Hasil
Metode	DBSCAN
Jumlah Cluster	1
Silhouette score	0,427
Waktu Komputasi	0,891 detik

Berdasarkan hasil evaluasi pada Tabel 6, algoritma DBSCAN menghasilkan 1 *cluster* dengan nilai *silhouette score* sebesar 0,427 dan waktu komputasi sebesar 0,891 detik. Meskipun secara matematis nilai *silhouette score* DBSCAN lebih tinggi dibandingkan *K-Means*, secara fungsional *data mining* metode ini mengalami kegagalan segmentasi (*failing to cluster*). Fenomena *single-clustering* (hanya membentuk 1 *cluster* besar) mengindikasikan bahwa algoritma menyapu bersih seluruh data pelanggan ke dalam satu keranjang raksasa, sehingga tidak menghasilkan variasi kelompok perilaku yang dapat digunakan untuk strategi operasional bisnis. Hal ini dipicu oleh nilai parameter *epsilon* ($\epsilon = 0.5$) yang kurang sensitif terhadap kerapatan data RFM yang homogen setelah dinormalisasi.

3.4. Perbandingan Hasil clustering

Setelah dilakukan proses *clustering* menggunakan algoritma *K-Means* dan DBSCAN, dilakukan evaluasi untuk membandingkan performa masing-masing algoritma berdasarkan kualitas *cluster*, jumlah *cluster*, serta efisiensi waktu komputasi.

Tabel 7. Perbandingan Hasil clustering

Metode	Silhouette score	Jumlah Cluster	Waktu
K-Means	0,319	2	0,004 detik
DBSCAN	0,427	1	0,891 detik

Berdasarkan hasil evaluasi pada Tabel 7, algoritma DBSCAN memperoleh nilai *silhouette score* sebesar 0,427, sedangkan algoritma *K-Means* memperoleh nilai sebesar 0,319. Pada antarmuka sistem (Gambar 5), kesimpulan otomatis menyatakan bahwa metode DBSCAN dinilai lebih baik karena sistem diprogram untuk mendeteksi nilai *Silhouette score* tertinggi.

Namun, secara fungsional *data mining*, rekomendasi otomatis tersebut kurang tepat untuk diimplementasikan pada strategi bisnis. Hal ini dikarenakan algoritma DBSCAN mengalami kegagalan segmentasi (*single-clustering*) dengan hanya menghasilkan 1 *cluster* besar. Oleh karena itu, algoritma *K-Means* memberikan hasil segmentasi yang jauh lebih aplikatif bagi pihak manajemen *e-commerce* dengan membagi pelanggan ke dalam 2 *cluster* fungsional.

Interpretasi terhadap karakteristik rata-rata (*mean*) dari nilai *Recency*, *Frequency*, dan *Monetary* pada masing-masing *cluster K-Means* dijabarkan secara mendalam sebagai berikut:

1. **Cluster 0 – Pelanggan Potensial Utama (Core / Potential Loyalist Customers):** Kelompok ini didominasi oleh pelanggan yang memiliki nilai *Recency* yang rendah (baru saja melakukan transaksi), dikombinasikan dengan rata-rata skor *Frequency* (intensitas pembelian) yang tinggi serta nilai *Monetary* (total nominal belanja) yang besar. Strategi yang direkomendasikan adalah program retensi intensif seperti pemberian akses eksklusif ke produk baru atau program poin loyalitas.
2. **Cluster 1 – Pelanggan Pasif / Berisiko Hilang (Hibernating / At-risk customers):** Kelompok ini dicirikan oleh nilai *Recency* yang tinggi (sudah lama tidak aktif bertransaksi) serta didominasi oleh nilai *Frequency* dan *Monetary* yang rendah. Strategi pemasaran yang tepat untuk kelompok ini adalah kampanye re-aktivasi (*Re-engagement marketing*) berupa pengiriman voucher belanja khusus atau diskon besar untuk memicu transaksi kembali.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma *K-Means* merupakan metode yang paling tepat, aplikatif, dan efisien untuk melakukan segmentasi pelanggan pada data transaksi *e-commerce* ini dibandingkan dengan DBSCAN. Berdasarkan hasil evaluasi, *K-Means* sukses membentuk 2 *cluster* pelanggan yang fungsional dan representatif untuk kebutuhan perumusan strategi pemasaran dengan waktu pemrosesan yang sangat instan sebesar 0,004 detik. Sebaliknya, algoritma DBSCAN terbukti tidak efektif dan mengalami kegagalan fungsi segmentasi karena hanya menghasilkan 1 *cluster* besar (terjadi *single-clustering*),

meskipun membuahkan *silhouette score* tinggi sebesar 0,427 akibat bias evaluasi pada satu kelompok homogen. Implementasi sistem simulasi berbasis web menggunakan kerangka kerja Flask telah berhasil memfasilitasi visualisasi hasil ini secara riil. Untuk penelitian selanjutnya, disarankan melakukan optimasi pencarian parameter *epsilon* pada *DBSCAN* menggunakan metode *K-Distance Graph* agar algoritma dapat memetakan kepadatan data secara lebih sensitif.

REFERENSI

- [1] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using *K-Means*, *K-Medoids*, and *DBSCAN* Methods," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 11, no. 1, p. 32, Apr. 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
- [2] A. S. Paramita and T. Hariguna, "Comparison of *K-Means* and *DBSCAN* Algorithms for Customer Segmentation in *e-commerce*," *Journal of Digital Market and Digital Currency*, vol. 1, no. 1, pp. 43–62, Jun. 2024, doi: 10.47738/jdmcdc.v1i1.3.
- [3] C. Poonsin and P. Boonma, "EFFICIENT SEGMENTATION OF CUSTOMERS BASED ON RFM ANALYSIS", Data Science and Engineering (DSE) RECORD, March. 2025, [Online]. Tersedia: <https://datascience.cmu.ac.th/storage/articles/74.pdf> [Diakses: 05 Juli 2025].
- [4] J. M. John, O. Shobayo, and B. Ogunleye, "An Exploration of *clustering* Algorithms for Customer Segmentation in the UK Retail Market," *Analytics*, vol. 2, no. 4, pp. 809–823, Oct. 2023, doi: 10.3390/analytics2040042.
- [5] A. Aglasia and I. Agus, "Unveiling Hidden Customer Segments in *e-commerce* Using *DBSCAN clustering* on Demographic and Behavioral Insights," *International Journal for Applied Information Management*, vol. 4, no. 3, pp. 128–140, 2024, doi: 10.47738/ijaim.v4i3.85.
- [6] H. Qing, J. Zhang, and D. Fu, "Data Mining Technology in Business Data Analysis," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1852/2/022045.
- [7] H. Ramadhan, M. Rizal Abdan Kamaludin, M. Alfian Nasrullah, and D. Rolliawati, "Comparison of Hierarchical, *K-Means* and *DBSCAN clustering* Methods for Credit Card Customer Segmentation Analysis Based on Expenditure Level", *Journal of Applied Informatics and Computing (JAIC)*, December 2023, [Online]. Tersedia: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/5790/2143> [Diakses: 12 September 2025]
- [8] H. Kilari, S. Edara, G. Ratna Sai Yarra, and D. Varma Gadhiraju, "Customer Segmentation using *K-Means clustering*", *International Journal of Engineering Research & Technology (IJERT)*, March 2022, [Online]. Tersedia: <https://www.ijert.org/research/customer-segmentation-using-k-means-clustering-IJERTV11IS030152.pdf> [Diakses: 12 September 2025].
- [9] H. Saragih and J. Manurung, "Customer segmentation analysis using *DBSCAN* method in marketing research of retail company."
- [10] R. Martiansah, S. Monalisa, F. Muttakin, and M. Fronita, "Customer Segmentation Analysis Through RFM-D Model and *K-Means* Algorithm," 2025.
- [11] D. A. Awaliyah, Budi Prasetyo, R. Muzayanah, and A. D. Lestari, "Optimizing Customer Segmentation in *Online retail* Transactions through the Implementation of the *K-Means clustering* Algorithm," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 539–548, Jun. 2024, doi: 10.15294/sji.v11i2.6137.
- [12] A. L. M. Tampubolon, T. M. E. Y. Butar Butar, and S. Rochimah, "Segmentasi Pelanggan Majalah pada Situs Web *e-commerce* dengan *K-Means++* dan Metode RFM," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1243–1252, Dec. 2024, doi: 10.25126/jtiik.2024118208.
- [13] C. Hafidz Ardana *et al.*, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode *K-Means clustering*," 2024.
- [14] J. Homepage, M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Customer Segmentation with *clustering* Algorithm Based on *Recency*, *Frequency*, and *Monetary* (RFM) Attributes Segmentasi Pelanggan dengan Algoritma *clustering* Berdasarkan Atribut *Recency*, *Frequency* dan *Monetary* (RFM)," vol. 5, no. 1, pp. 48–56, 2025, doi: 10.57152/malcom.v5v1.1491.
- [15] R. Fitriyanto, "Optimalisasi Pengelompokan Konsumen dengan Multi Internal Metric Validation dan Boxplot Analysis," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 8, no. 1, pp. 16–21, 2025.