

ANALISIS PERBANDINGAN PERFORMA METODE K-MEANS DAN DBSCAN UNTUK CLUSTERING PELANGGAN

Muhammad Ali Bazmul, Dwi Hartanti, Aprilisa Arum Sari

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55, Tipe, Kec. Serengan, Kota Surakarta, Jawa Tengah, Indonesia

220103268@mhs.udb.ac.id

ABSTRAK

Segmentasi pelanggan merupakan strategi penting untuk memahami karakteristik konsumen dan mendukung pengambilan keputusan pemasaran yang lebih efektif. Namun, pemilihan algoritma *clustering* yang kurang sesuai dengan karakteristik data dapat menghasilkan kualitas segmentasi yang kurang optimal. Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan DBSCAN dalam melakukan segmentasi pelanggan pada *Shopping Mall Customer Segmentation Dataset* yang terdiri dari 15.079 data pelanggan. Metode penelitian meliputi *preprocessing data*, normalisasi menggunakan *StandardScaler*, penerapan algoritma K-Means dan DBSCAN, serta evaluasi hasil clustering menggunakan *Silhouette Score* dan *Davies-Bouldin Index (DBI)*. Penentuan jumlah *cluster* pada K-Means dilakukan menggunakan *Elbow Method*, sedangkan parameter DBSCAN ditentukan melalui pencarian parameter (*grid search*). Hasil penelitian menunjukkan bahwa K-Means menghasilkan 5 *cluster* dengan nilai *Silhouette Score* sebesar 0,2669 dan DBI sebesar 1,0650, sedangkan DBSCAN menghasilkan 2 *cluster* dengan nilai *Silhouette Score* sebesar 0,1074 dan DBI sebesar 0,7681. Berdasarkan hasil evaluasi dan visualisasi *cluster*, K-Means menunjukkan performa yang lebih baik untuk segmentasi pelanggan pada dataset yang digunakan.

Kata kunci : *clustering, K-Means, DBSCAN, segmentasi pelanggan, evaluasi cluster*

1. PENDAHULUAN

Perkembangan teknologi informasi mendorong perusahaan untuk mengumpulkan data pelanggan dalam jumlah besar dari berbagai aktivitas transaksi digital. Data tersebut memiliki nilai strategis karena dapat dimanfaatkan untuk memahami perilaku konsumen, meningkatkan personalisasi layanan, dan mendukung pengambilan keputusan bisnis yang lebih tepat sasaran. Namun, tanpa proses analisis yang tepat, data hanya menjadi kumpulan informasi yang belum memberikan manfaat optimal [1], [2].

Salah satu teknik yang banyak digunakan dalam mengolah data pelanggan adalah *clustering*, yaitu metode pengelompokan data berdasarkan kemiripan karakteristik antar objek tanpa memerlukan label kelas. Dalam konteks bisnis, *clustering* berperan penting dalam segmentasi pelanggan untuk mengidentifikasi kelompok pelanggan potensial, pelanggan loyal, maupun pelanggan berisiko rendah sehingga strategi pemasaran dapat disusun secara lebih efektif dan efisien [3], [4].

Berbagai algoritma *clustering* dapat digunakan untuk segmentasi pelanggan, di antaranya K-Means dan DBSCAN. K-Means merupakan algoritma berbasis *centroid* yang populer karena sederhana, cepat, dan efektif pada data dengan pola *cluster* yang jelas. Namun, metode ini sensitif terhadap *outlier* dan mengharuskan jumlah *cluster* ditentukan di awal. Sebaliknya, DBSCAN menggunakan pendekatan kepadatan data sehingga mampu mendeteksi *noise* serta membentuk *cluster* dengan pola tidak beraturan, meskipun hasilnya sangat dipengaruhi oleh pemilihan parameter epsilon dan minPts [5], [6].

Perbedaan karakteristik tersebut menunjukkan bahwa tidak ada satu metode *clustering* yang selalu unggul pada semua jenis data. Oleh karena itu, diperlukan analisis perbandingan performa agar pemilihan algoritma dapat disesuaikan dengan struktur data yang digunakan.

Untuk menyelesaikan permasalahan tersebut, penelitian ini melakukan perbandingan antara K-Means dan DBSCAN menggunakan *Shopping Mall Customer Segmentation Dataset* yang memiliki atribut numerik berupa usia, pendapatan tahunan, dan *spending score*. Proses penelitian dilakukan melalui tahap *preprocessing data*, normalisasi menggunakan *StandardScaler*, penerapan algoritma *clustering*, serta evaluasi hasil menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index*. *Silhouette Score* digunakan untuk mengukur kedekatan data dalam *cluster* yang sama serta pemisahannya terhadap *cluster* lain, sedangkan *Davies-Bouldin Index* mengevaluasi rasio kedekatan intra-*cluster* dan jarak antar *cluster*. Nilai *Silhouette Score* yang lebih tinggi dan nilai *Davies-Bouldin Index* yang lebih rendah menunjukkan kualitas *clustering* yang lebih baik [7].

Berdasarkan rangkaian proses tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma K-Means dan DBSCAN pada segmentasi pelanggan sehingga dapat memberikan rekomendasi metode yang paling sesuai berdasarkan karakteristik data yang dianalisis.

2. TINJAUAN PUSTAKA

Penelitian mengenai segmentasi pelanggan menggunakan teknik *clustering* telah banyak dilakukan karena mampu membantu perusahaan

memahami pola perilaku konsumen dari data yang tersedia. Salah satu algoritma yang sering digunakan adalah K-Means karena memiliki proses komputasi yang sederhana dan efisien untuk data numerik berskala besar.

Penelitian oleh Ardana et al. menunjukkan bahwa metode K-Means efektif dalam mengelompokkan pelanggan penjualan online berdasarkan pola transaksi. Hasil segmentasi yang diperoleh dapat digunakan sebagai dasar penyusunan strategi pemasaran yang lebih tepat sasaran[8].

Penelitian lain oleh Baiq Nikum Yulisasih et al. juga membuktikan bahwa K-Means mampu mengidentifikasi kelompok pelanggan potensial berdasarkan riwayat pembelian, sehingga mendukung pengambilan keputusan bisnis berbasis data [9].

Selain K-Means, algoritma DBSCAN juga banyak digunakan terutama pada data yang mengandung *noise* dan memiliki bentuk *cluster* tidak beraturan. Penelitian Susilo menyatakan bahwa DBSCAN memiliki performa baik dalam mendeteksi pola data kompleks serta memisahkan *outlier* dari data utama. Namun, K-Means tetap lebih stabil ketika diterapkan pada data dengan struktur sederhana dan terpisah jelas [10].

Tabel 1. Perbandingan karakteristik k-means dan dbscan

Aspek	K-Means	DBSCAN
Pendekatan	<i>Centroid-based</i>	<i>Density-based</i>
Jumlah <i>cluster</i>	Ditentukan diawal (K)	Ditentukan otomatis berdasarkan kepadatan
Bentuk <i>cluster</i>	Cocok untuk <i>cluster</i> bulat	Dapat menangani bentuk <i>cluster</i> tidak beraturan
Deteksi <i>noise</i>	Tidak tersedia	Mampu mendeteksi <i>noise/outlier</i>
Sensitif terhadap <i>outlier</i>	Tinggi	Rendah
Parameter utama	Jumlah <i>cluster</i> (K)	Epsilon dan minPts
Kompleksitas	Relatif cepat	Lebih tinggi dibanding K-Means

Berdasarkan Tabel 1, algoritma K-Means dan DBSCAN memiliki karakteristik yang berbeda dalam proses pembentukan *cluster*. K-Means lebih unggul dalam efisiensi komputasi dan cocok digunakan pada data yang memiliki pola *cluster* yang jelas, sedangkan DBSCAN lebih efektif dalam menangani data yang mengandung *noise* serta memiliki bentuk *cluster* yang tidak beraturan. Oleh karena itu, pemilihan algoritma *clustering* perlu disesuaikan dengan karakteristik data yang akan dianalisis.

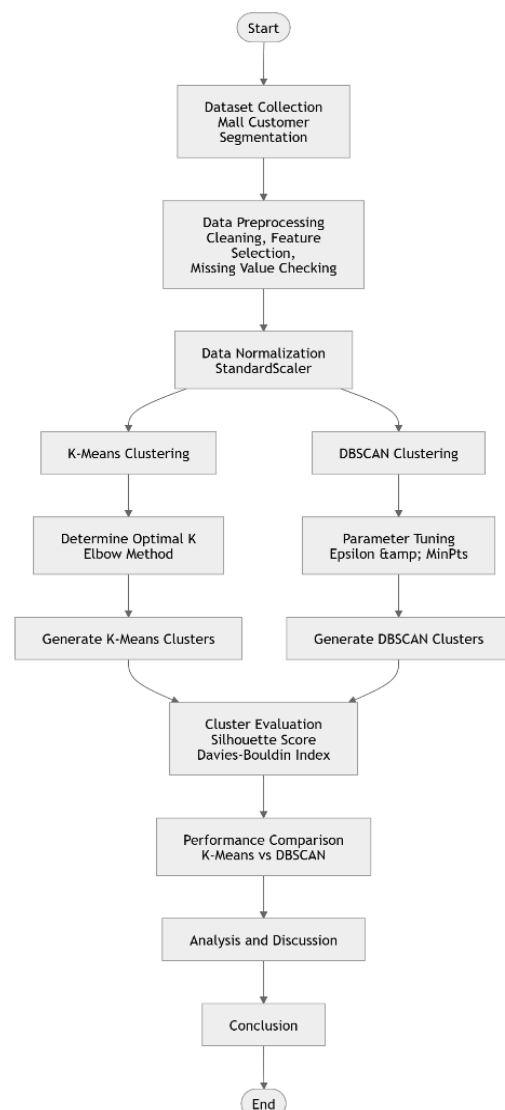
Penelitian komparatif antara K-Means dan DBSCAN dilakukan oleh Saputra dan Yusuf pada data pelanggan transportasi publik. Hasil penelitian menunjukkan DBSCAN lebih efektif pada distribusi data yang kompleks, sedang K-Means lebih optimal pada pola *cluster* yang jelas dan terstruktur [11].

Temuan serupa juga dilaporkan oleh Hammam dan Utomo, di mana K-Means memperoleh nilai *Silhouette Score* lebih tinggi, sedangkan DBSCAN unggul dalam mendeteksi *noise* dan pola tidak beraturan [12].

Berdasarkan beberapa penelitian terdahulu, dapat disimpulkan bahwa performa algoritma *clustering* sangat dipengaruhi oleh karakteristik data yang digunakan. Oleh karena itu, penelitian ini berfokus pada analisis perbandingan performa K-Means dan DBSCAN pada dataset *Shopping Mall Customer Segmentation* menggunakan metrik evaluasi internal berupa *Silhouette Score* dan *Davies-Bouldin Index* untuk menentukan metode yang paling sesuai.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa algoritma K-Means dan DBSCAN pada proses segmentasi pelanggan. Fokus penelitian adalah mengevaluasi kualitas hasil *clustering* berdasarkan karakteristik data menggunakan metrik evaluasi internal.



Gambar 1. Flowchart penelitian

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan dataset *Shopping Mall Customer Segmentation* yang kemudian melalui tahap *preprocessing* berupa pengecekan data, seleksi atribut, dan normalisasi menggunakan *StandardScaler*. Selanjutnya data diproses menggunakan algoritma K-Means dan DBSCAN secara terpisah. Hasil *clustering* dari kedua algoritma kemudian dievaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* untuk mengetahui kualitas *cluster* yang terbentuk. Tahap akhir penelitian adalah analisis perbandingan performa kedua algoritma dan penarikan kesimpulan.

Dataset yang digunakan dalam penelitian ini adalah *Shopping Mall Customer Segmentation* Dataset yang diperoleh dari platform Kaggle. Dataset terdiri dari 15.079 data pelanggan dengan lima atribut yaitu *Customer ID*, *Gender*, *Age*, *Annual Income*, dan *Spending Score*. Pada penelitian ini hanya digunakan atribut numerik yang relevan yaitu *Age*, *Annual Income*, dan *Spending Score*. Atribut *Age* dipilih karena merepresentasikan karakteristik demografis pelanggan. *Annual Income* digunakan untuk menggambarkan kemampuan ekonomi pelanggan, sedangkan *Spending Score* digunakan untuk merepresentasikan perilaku konsumsi pelanggan. Kombinasi ketiga atribut tersebut dianggap mampu menggambarkan karakteristik pelanggan secara lebih komprehensif sehingga cocok digunakan dalam proses segmentasi. Sebelum dilakukan proses *clustering*, data diperiksa untuk memastikan tidak terdapat *missing value* maupun data duplikat yang dapat mempengaruhi hasil analisis.

Tahap awal penelitian adalah *preprocessing data* yang meliputi pengecekan data kosong, seleksi atribut, serta normalisasi data menggunakan *StandardScaler*. Normalisasi dilakukan agar seluruh atribut memiliki skala yang sebanding sehingga tidak memengaruhi proses perhitungan jarak pada algoritma *clustering*.

Setelah tahap *preprocessing*, dilakukan penerapan algoritma K-Means dan DBSCAN. Pada K-Means, penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow Method*. Selanjutnya model dijalankan dengan jumlah *cluster* terbaik untuk menghasilkan kelompok pelanggan.

Pada algoritma DBSCAN, pencarian parameter dilakukan menggunakan pencarian parameter (*grid search*) dengan kombinasi nilai epsilon (ϵ) sebesar 0,1; 0,15; dan 0,2 serta nilai minimum poin (*minPts*) sebesar 3, 5, 7, 10, dan 15. Setiap kombinasi parameter dievaluasi menggunakan *Silhouette Score* untuk menentukan parameter terbaik yang menghasilkan kualitas *clustering* paling optimal. Parameter terbaik selanjutnya digunakan pada tahap *clustering* dan evaluasi akhir menggunakan *Silhouette Score* dan *Davies-Bouldin Index*.

Hasil *clustering* dari kedua algoritma kemudian dievaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index*. Nilai *Silhouette Score* yang lebih tinggi menunjukkan kualitas *cluster* yang lebih baik,

sedangkan nilai *Davies-Bouldin Index* yang lebih rendah menandakan pemisahan *cluster* yang lebih optimal.

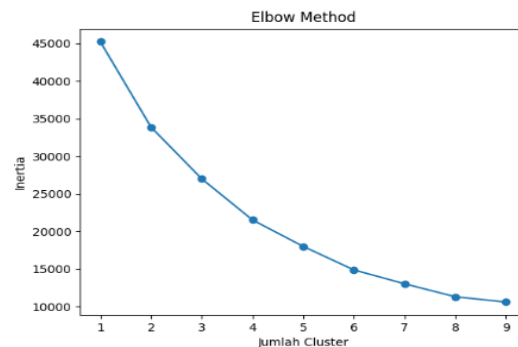
Tahapan penelitian terdiri dari: (1) pengumpulan dataset, (2) *preprocessing data*, (3) penerapan K-Means, (4) penerapan DBSCAN, (5) evaluasi hasil, dan (6) analisis perbandingan performa kedua algoritma untuk menentukan metode yang paling sesuai terhadap karakteristik data pelanggan.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Eksperimen

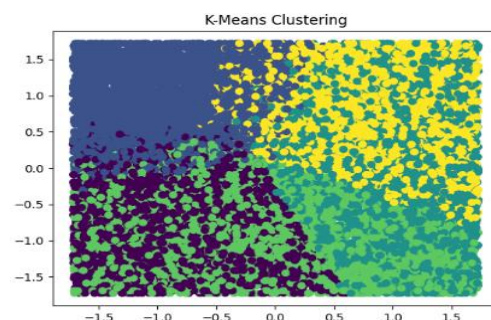
Pada penelitian ini dilakukan pengujian terhadap algoritma K-Means dan DBSCAN menggunakan dataset *Shopping Mall Customer Segmentation* dengan atribut *Age*, *Annual Income*, dan *Spending Score*. Sebelum proses *clustering*, data terlebih dahulu melalui tahap normalisasi menggunakan *StandardScaler* untuk menyamakan skala antar atribut.

Penentuan jumlah *cluster* optimal pada algoritma K-Means dilakukan menggunakan metode *Elbow Method*. Metode ini digunakan untuk melihat titik penurunan signifikan (*elbow*) pada nilai *inertia* terhadap jumlah *cluster*.



Gambar 2. Elbow method

Berdasarkan grafik dari Gambar 2, terlihat bahwa penurunan nilai *inertia* mulai melandai pada nilai $k = 5$. Hal ini menunjukkan jika jumlah *cluster* optimal yang digunakan pada penelitian ini sebanyak 5 *cluster*. Selanjutnya, algoritma K-Means diterapkan pada data yang telah dinormalisasi. Hasil *clustering* divisualisasikan menggunakan *scatter plot* berdasarkan atribut *Annual Income* dan *Spending Score*.



Gambar 3. Hasil clustering k-means

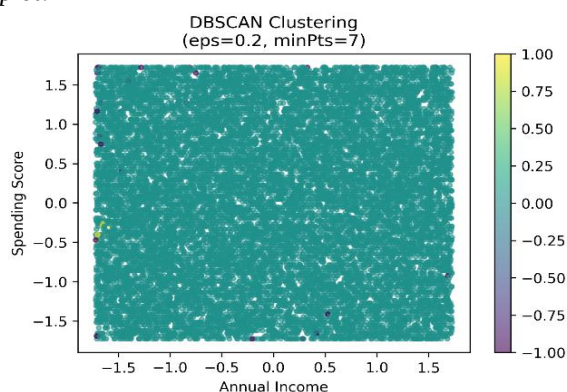
Hasil visualisasi pada Gambar 3 menunjukkan bahwa data pelanggan berhasil dikelompokkan ke dalam beberapa *cluster* dengan pemisahan yang cukup jelas. Setiap *cluster* merepresentasikan kelompok pelanggan dengan karakteristik tertentu, seperti pelanggan dengan tingkat pendapatan dan pengeluaran yang berbeda. *Cluster* yang terbentuk menunjukkan adanya perbedaan karakteristik pelanggan berdasarkan tingkat pendapatan dan pengeluaran. Beberapa *cluster* menggambarkan pelanggan dengan pendapatan tinggi dan pengeluaran tinggi yang berpotensi menjadi pelanggan prioritas, sementara *cluster* lainnya menunjukkan pelanggan dengan tingkat pengeluaran yang lebih rendah. Hasil ini menunjukkan bahwa K-Means mampu menghasilkan segmentasi pelanggan yang lebih mudah diinterpretasikan.

Berdasarkan hasil pengujian dengan menggunakan pencarian parameter (*grid search*), kombinasi parameter terbaik diperoleh pada nilai $\epsilon = 0,2$ dan $\text{minPts} = 7$ dengan nilai *Silhouette Score* sebesar 0,1074. Parameter tersebut kemudian digunakan pada proses *clustering* DBSCAN untuk menghasilkan segmentasi pelanggan yang akan dibandingkan dengan hasil K-Means.

Tabel 2. Tabel parameter terbaik dan hasil evaluasi dbscan

Parameter	Nilai
Eps	0,2
minPts	7
Jumlah cluster	2
Silhouette score	0.1074
DBI	0.7681

Pengujian juga dilakukan menggunakan algoritma DBSCAN dengan beberapa kombinasi parameter epsilon (ϵ) dan minimum points (minPts). Hasil *clustering* divisualisasikan dalam bentuk *scatter plot*.



Gambar 4. Hasil clustering dbscan

Berdasarkan hasil visualisasi di Gambar 4, hasil *clustering* menggunakan DBSCAN dengan parameter terbaik $\epsilon = 0,2$ dan $\text{minPts} = 7$ menghasilkan dua *cluster* pelanggan. Berdasarkan visualisasi hasil *clustering*, sebagian besar data masih terkonsentrasi pada satu kelompok besar, sedangkan kelompok lainnya memiliki jumlah anggota yang relatif sedikit.

Nilai *Silhouette Score* sebesar 0,1074 menunjukkan bahwa pemisahan antar *cluster* yang dihasilkan masih kurang optimal dibandingkan K-Means. Kondisi ini mengindikasikan bahwa struktur data pelanggan pada dataset yang digunakan lebih sesuai untuk pendekatan berbasis *centroid* dibandingkan pendekatan berbasis kepadatan yang digunakan oleh DBSCAN.

4.2. Analisis dan Pembahasan

Evaluasi dilakukan menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index* untuk mengukur kualitas hasil *clustering*.

Tabel 3. Tabel hasil evaluasi

Algoritma	Jumlah Cluster	Silhouette Score	DBI
K-Means	5	0.2669	1.0650
DBSCAN	2	0.1074	0.7681

Berdasarkan hasil evaluasi pada Tabel 3, algoritma K-Means menghasilkan nilai *Silhouette Score* sebesar 0,2669 dan *Davies-Bouldin Index* (DBI) sebesar 1,0650, sedangkan DBSCAN menghasilkan nilai *Silhouette Score* sebesar 0,1074 dan DBI sebesar 0,7681. Secara teoretis, nilai DBI yang lebih rendah pada DBSCAN mengindikasikan rasio pemisahan antar *cluster* yang baik. Namun, hasil visualisasi pada Gambar 4 menunjukkan bahwa rendahnya nilai DBI tersebut disebabkan oleh penumpukan mayoritas data pada satu kelompok besar tunggal, sehingga variasi internal *cluster* menjadi sangat kecil secara umum.

Sebaliknya, K-Means menunjukkan performa yang lebih objektif dalam konteks segmentasi pelanggan dengan nilai *Silhouette Score* yang jauh lebih tinggi, menandakan kedekatan data dalam *cluster* yang lebih solid. Hal ini membuktikan bahwa karakteristik *Shopping Mall Customer Segmentation* Dataset memiliki pola distribusi yang relatif terstruktur dan berbasis jarak, sehingga lebih cocok diproses menggunakan pendekatan *centroid* seperti K-Means dibandingkan pendekatan berbasis kepadatan seperti DBSCAN yang kesulitan membentuk segmentasi optimal akibat tidak adanya area kepadatan yang kontras pada data tersebut.

Dari sudut pandang implementasi bisnis, penemuan ini menegaskan bahwa segmentasi pelanggan yang dihasilkan oleh K-Means jauh lebih aplikatif untuk menyusun strategi pemasaran. Keberhasilan K-Means dalam membagi pelanggan ke dalam lima segmen yang proporsional memungkinkan perusahaan ritel untuk mengidentifikasi kelompok konsumen secara spesifik, mulai dari kelompok pelanggan bernilai tinggi (*high-value*), pelanggan potensial, hingga pelanggan dengan tingkat pengeluaran rendah. Sebaliknya, hasil klusterisasi DBSCAN yang cenderung mengelompokkan mayoritas data ke dalam satu *cluster* dominan menyebabkan karakteristik antar pelanggan menjadi kurang terpisahkan secara jelas. Hasil analisis komparatif ini sejalan dengan temuan Saputra dan

Yusuf [11] serta Hammam dan Utomo [12] yang menunjukkan bahwa meskipun algoritma berbasis kepadatan memiliki keunggulan tersendiri, K-Means tetap memberikan stabilitas dan interpretasi yang lebih optimal pada struktur data pelanggan yang terdistribusi secara teratur.

Secara metodologis, eksperimen ini menegaskan bahwa keterbatasan teoretis dari masing-masing algoritma sangat dipengaruhi oleh bagaimana parameter operasionalnya dikonfigurasi terhadap struktur data. Meskipun K-Means memiliki kelemahan bawaan berupa keharusan menentukan jumlah kelompok di awal dan sensitif terhadap *outlier*, penerapan metode *Elbow* dalam penelitian ini terbukti efektif untuk memitigasi kendala tersebut dengan menemukan titik kelandaian nilai *inertia* yang optimal pada $k=5$. Sebaliknya, keunggulan DBSCAN yang mampu menentukan jumlah *cluster* secara otomatis dan mendeteksi *noise* menjadi kurang optimal ketika dihadapkan pada dataset pelanggan yang tidak memiliki batas kepadatan yang kontras. Dengan demikian, keberhasilan penemuan parameter terbaik melalui pencarian parameter (*grid search*) sekalipun tetap sangat bergantung pada kesesuaian antara mekanisme kerja algoritma dengan karakteristik internal dari distribusi data yang dianalisis.

4.3. Implikasi Hasil Penelitian

Hasil penelitian menunjukkan bahwa algoritma K-Means lebih sesuai digunakan pada dataset *Shopping Mall Customer Segmentation* karena mampu menghasilkan *cluster* yang lebih jelas berdasarkan nilai evaluasi yang diperoleh. Temuan ini dapat menjadi referensi bagi perusahaan ritel dalam memilih metode segmentasi pelanggan yang sesuai dengan karakteristik data yang dimiliki. Selain itu, hasil penelitian juga menunjukkan pentingnya pemilihan algoritma *clustering* yang tepat karena setiap algoritma memiliki kelebihan dan keterbatasan yang berbeda dalam menangani struktur data pelanggan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil eksperimen dan analisis perbandingan performa antara algoritma K-Means dan DBSCAN pada *Shopping Mall Customer Segmentation Dataset*, dapat disimpulkan bahwa K-Means menunjukkan performa yang lebih optimal dan relevan untuk karakteristik data yang diteliti. K-Means berhasil membentuk 5 *cluster* pelanggan yang terstruktur dengan nilai *Silhouette Score* sebesar 0,2669 dan *Davies-Bouldin Index* (DBI) sebesar 1,0650. Di sisi lain, DBSCAN menghasilkan 2 *cluster* dengan nilai *Silhouette Score* sebesar 0,1074 dan DBI sebesar 0,7681. Meskipun secara teoretis DBSCAN memperoleh nilai DBI yang lebih rendah (lebih baik), namun hasil visualisasi menunjukkan terjadinya sentralisasi data di mana mayoritas objek menumpuk hanya pada satu *cluster* besar. Sebaliknya, K-Means menghasilkan pembagian kelompok yang jauh lebih proporsional, pemisahan antar *cluster* yang lebih jelas,

serta segmentasi yang lebih mudah diinterpretasikan untuk kebutuhan strategi bisnis. Dengan demikian, untuk dataset pelanggan yang memiliki pola distribusi relatif terstruktur tanpa *noise* yang dominan, pendekatan berbasis *centroid* seperti K-Means jauh lebih disarankan dibandingkan pendekatan berbasis kepadatan (*density-based*) seperti DBSCAN.

Penelitian ini memiliki beberapa keterbatasan, di antaranya adalah penggunaan atribut yang baru berfokus pada variabel numerik tertentu, yaitu *Age*, *Annual Income*, dan *Spending Score*. Selain itu, proses optimasi parameter pada algoritma DBSCAN masih terbatas pada kombinasi rentang nilai epsilon dan minPts tertentu. Pengujian kualitas kelompok juga baru mengandalkan dua metrik evaluasi internal, yaitu *Silhouette Score* dan *Davies-Bouldin Index*, sehingga belum meninjau performa algoritma dari sudut pandang metrik klasterisasi lainnya. Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk menggunakan dataset dengan karakteristik yang lebih beragam serta menambahkan atribut pelanggan yang lebih komprehensif, seperti data kategorikal atau riwayat transaksi menggunakan pendekatan RFM (*Recency, Frequency, Monetary*). Eksplorasi parameter DBSCAN juga dapat diperluas dengan rentang pencarian yang lebih dinamis untuk membuka peluang ditemukannya konfigurasi kepadatan *cluster* yang lebih optimal. Terakhir, disarankan pula untuk memperluas komparasi algoritma dengan melibatkan metode klasterisasi lain, seperti *Hierarchical Clustering* atau *Gaussian Mixture Model* (GMM), disertai metrik evaluasi tambahan agar menghasilkan analisis perbandingan yang lebih mendalam dan valid.

DAFTAR PUSTAKA

- [1] B. N. Yuliasih, H. Herman, S. Sunardi, and H. Yuliansyah, "Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation," *Ilk. J. Ilm.*, vol. 16, no. 3, pp. 330–342, 2024.
- [2] 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, 2023.
- [3] W. G. Mokodaser, F. Soewignyo, and T. I. Soewignyo, "Unveiling Banking Customer Pattern Through K-Means and DBSCAN Cluster Evaluation," *CogITo Smart J.*, vol. 11, no. 1, pp. 182–192, 2025.
- [4] T. Setyani, H. Nopijani Heidy, K. Sari, Y. Indriani, and H. Sulistiani, "Perbandingan Kinerja Algoritma K-Means dan DBSCAN dalam Segmentasi Nasabah Berdasarkan Data Pemasaran Bank," *Pros. Semin. Nas. Teknol. Komput. dan Sains*, vol. 3, no. 1, pp. 376–382, 2025.
- [5] N. A. Yolandari, L. E. Butarbutar, M. F. Zulfi, Arnita, and F. Ramadhani, "Analisis

- Perbandingan K-Means Dan Dbscan Dalam Pengelompokan Data Travel Review Ratings Menggunakan Evaluasi Silhouette Index Dan Davies-Bouldin Index,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025.
- [6] P. Nadya, P. Astina, I. W. Supriana, and I. M. Satria, “Perbandingan Metode Clustering K-Means , GMM , dan DBSCAN Berbasis Fitur RFM,” vol. 4, pp. 295–306, 2026.
- [7] G. Vardakas, I. Papakostas, and A. Likas, “Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters,” 2024.
- [8] C. H. Ardana, A. A. A. A. A. Khoyum, and M. Faisal, “Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024.
- [9] B. N. Yulisasih, H. Herman, and S. Sunardi, “K-Means Clustering Method For Customer Segmentation Based On Potential Purchases,” *J. ELTIKOM*, vol. 8, no. 1, pp. 83–90, 2024.
- [10] A. Susilo, “Evaluasi Efektifitas Algoritma K-Means Dan Dbscan Dalam Clustering Data Rumah Sakit Untuk Optimalisasi Layanan Kesehatan,” *J. Komput. dan Inform.*, vol. 19, no. 1, pp. 43–51, 2024.
- [11] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM: Comparison of the DBSCAN and K-MEANS Algorithms in Segmenting Customers Using Public Transportation of Transjakarta Usin,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024.
- [12] Mochamad Hammam and Tegar Utomo, “Clustering data siswa putus sekolah dengan algoritma k-means dan dbscan,” *J. Komput. dan Inform.*, vol. 18, no. 2, pp. 150–159, 2023.