Jurnal Komputer dan Informatika
Vol 19 No 1, April 2024: hlm 43-51

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

EVALUASI EFEKTIFITAS ALGORTIMA *K-MEANS* DAN DBSCAN DALAM *CLUSTERING* DATA RUMAH SAKIT UNTUK OPTIMALISASI LAYANAN KESEHATAN

Andri Susilo
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: andri.535210074@stu.untar.ac.id,

ABSTRAK

Penelitian ini mengevaluasi efektivitas algoritma *K-Means* dan DBSCAN dalam *clustering* data rumah sakit untuk mendukung optimalisasi layanan kesehatan. Dataset yang digunakan berasal dari situs data.gov.au, mencakup atribut seperti jumlah tempat tidur, lokasi geografis, dan jenis pendanaan. Evaluasi dilakukan menggunakan metrik *Silhouette Score* untuk menilai kualitas *clustering*. Hasil menunjukkan bahwa *K-Means* menghasilkan hasil yang konsisten dengan Silhouette Score sebesar 0.7879 pada jumlah cluster kecil (C=2), sedangkan DBSCAN menunjukkan performa yang sangat baik dengan *Silhouette Score* mencapai 0.9986 pada konfigurasi optimal (ε=0.1, min_samples=2). Penelitian ini menyimpulkan bahwa *K-Means* lebih efektif untuk cluster yang sederhana, sementara DBSCAN unggul dalam menangani distribusi data yang kompleks dan *noise*.

Kata kunci DBSCAN, *K-Means, Silhouette Score*

ABSTRACT

This study evaluates the effectiveness of the K-Means and DBSCAN algorithms in clustering hospital data to support healthcare optimization. The dataset used comes from the data.gov.au site, including attributes such as number of beds, geographic location, and funding type. The evaluation was carried out using the Silhouette Score metric to assess the quality of clustering. The results show that K-Means produces consistent results with a Silhouette Score of 0.7879 at a small number of clusters (C=2), while DBSCAN shows very good performance with a Silhouette Score reaching 0.9986 at the optimal configuration (ε=0.1, min_samples=2). This study concludes that K-Means is more effective for simple clusters, while DBSCAN excels in handling complex data distributions and noise.

Keywords *DBSCAN, K-Means, Silhouette Score*

1. PENDAHULUAN

Analisis data dalam sektor kesehatan telah menjadi perhatian utama dalam upaya meningkatkan kualitas layanan medis, efisiensi operasional rumah sakit, dan pengambilan keputusan berbasis data [1]. Seiring dengan kemajuan teknologi, institusi kesehatan menghasilkan volume data yang semakin besar, mencakup berbagai informasi mulai dari demografi pasien, kapasitas rumah sakit, hingga indikator kesehatan masyarakat [2]. Tantangan utama yang dihadapi adalah bagaimana memanfaatkan data tersebut secara efektif untuk mendukung perencanaan strategis, optimalisasi sumber daya, serta peningkatan pelayanan kesehatan [3]. Dalam konteks ini, analisis berbasis machine learning, khususnya teknik clustering, menjadi solusi potensial yang semakin banyak digunakan [4].

Penelitian ini menggunakan dataset publik dari situs *data.gov.au*, yang berisi informasi mengenai rumah sakit umum di Australia. Dataset ini mencakup berbagai atribut penting, seperti nama rumah sakit, lokasi geografis, jumlah tempat tidur, dan jenis layanan yang tersedia. Data ini menawarkan peluang untuk menerapkan algoritma clustering guna mengidentifikasi pola yang relevan untuk stratifikasi pasien, pengelompokan fasilitas kesehatan, dan pengelolaan sumber daya

secara efisien. Penggunaan dataset ini relevan karena mencerminkan variasi karakteristik rumah sakit dalam konteks geografis yang beragam, sehingga memungkinkan eksplorasi kemampuan algoritma dalam menangani data yang kompleks.

Teknik clustering, yang mengelompokkan data berdasarkan kesamaan karakteristik, dapat membantu institusi kesehatan dalam memahami struktur data yang kompleks [5]. Misalnya, stratifikasi pasien berbasis karakteristik medis dapat mendukung penerapan intervensi yang lebih terarah, sementara pengelompokan data rumah sakit memungkinkan perencanaan sumber daya yang lebih efisien [6]. Urgensi penelitian ini terletak pada kebutuhan untuk mengatasi berbagai keterbatasan yang ada dalam analisis data tradisional, seperti ketidakmampuan mengidentifikasi pola yang tidak linier atau ketidakefisienan dalam menangani dataset yang besar dan tidak terstruktur [7] .

Algoritma clustering seperti K-Means dan DBSCAN telah terbukti efektif dalam analisis data kesehatan. K-Means, sebagai metode partisi, membagi data ke dalam sejumlah cluster yang ditentukan sebelumnya berdasarkan jarak Euclidean [8]. Algoritma ini unggul dalam kesederhanaan dan efisiensinya, sehingga cocok untuk dataset dengan struktur cluster yang jelas dan berbentuk bulat [9]. Namun, sensitivitas K-Means terhadap outlier dan ketergantungannya pada jumlah cluster yang harus ditentukan sebelumnya menjadi kelemahan signifikan [10]. Sebaliknya, DBSCAN, yang merupakan algoritma berbasis kepadatan, mampu mengidentifikasi cluster dengan bentuk yang kompleks dan menangani noise secara efektif [11].

Hal ini menjadikannya pilihan yang lebih fleksibel untuk dataset dengan distribusi yang tidak teratur [12]. Meskipun kedua algoritma tersebut telah banyak diterapkan secara terpisah, masih terdapat kesenjangan penelitian dalam mengintegrasikan atau membandingkan keduanya untuk memahami keunggulan dan kelemahannya dalam konteks analisis data kesehatan [13]. Beberapa penelitian sebelumnya menunjukkan efektivitas algoritma ini dalam pengelompokan pasien berdasarkan fenotipe atau analisis data genetik [14]. tetapi belum banyak yang membahas aplikasinya secara khusus dalam pengelompokan data rumah sakit untuk optimalisasi operasional.

Penelitian ini bertujuan untuk mengevaluasi efektivitas K-Means dan DBSCAN dalam clustering data rumah sakit, dengan fokus pada stratifikasi pasien dan pengelolaan sumber daya. Dengan menggabungkan pendekatan kedua algoritma, penelitian ini diharapkan dapat memberikan wawasan baru bagi institusi kesehatan dalam memilih metode clustering yang sesuai dengan karakteristik dataset mereka. Analisis yang dilakukan tidak hanya akan mencakup evaluasi algoritma dari segi akurasi, tetapi juga mempertimbangkan aspek efisiensi komputasi dan kemampuan menangani noise.

Dalam konteks yang lebih luas, penelitian ini berkontribusi pada pengembangan sistem kesehatan yang lebih responsif dan adaptif terhadap kebutuhan pasien. Dengan memanfaatkan teknik clustering yang tepat, institusi kesehatan dapat merancang strategi yang lebih efektif dalam pencegahan dan penanganan penyakit, serta optimalisasi sumber daya yang terbatas. Dengan pendekatan metodologis yang rigor dan dukungan literatur terkini, penelitian ini diharapkan memberikan kontribusi yang signifikan terhadap pengembangan kebijakan dan praktik di sektor kesehatan.

2. METODE PENELITIAN

2.1 Dataset

Dataset yang digunakan dalam penelitian ini diambil dari situs *data.gov.au*, yang berisi informasi tentang rumah sakit umum di Australia. Dataset ini terdiri dari 696 entri dengan 19 atribut, yang mencakup berbagai aspek penting terkait rumah sakit, mulai dari lokasi, kapasitas, hingga

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

pelaporan data ke sistem kesehatan nasional. Atribut-atribut dalam dataset ini mencakup informasi geografi, operasional, dan status pelaporan rumah sakit yang berguna untuk tujuan analisis clustering.

Beberapa fitur yang terdapat dalam dataset adalah informasi geografis seperti *State* (negara bagian rumah sakit berada), nama rumah sakit (*Hospital name*), alamat lengkap (*Address Line 1* dan *Address Line 2*), serta *Remoteness area*, yang menggambarkan tingkat keterpencilan rumah sakit. Data operasional mencakup jumlah tempat tidur yang tersedia (*Number of available beds*) dan *Local Hospital Network*, yang menunjukkan jaringan rumah sakit yang mengelola rumah sakit tersebut.

Selain itu, dataset ini juga mencakup atribut yang berhubungan dengan sistem pelaporan rumah sakit ke berbagai sistem kesehatan nasional, seperti *provided data for NPHED*, *NHMD*, *NAPEDCD*, dan sebagainya. Informasi tentang klasifikasi rumah sakit berdasarkan kelompok peer dan tipe pendanaan rumah sakit juga termasuk dalam dataset, yang sangat berguna untuk memahami karakteristik rumah sakit di Australia. Deskripsi dataset dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi dataset.

Atribut	Deskripsi	Tipe Data	Jumlah nilai uik	Nilai *Missing*
State	Negara bagian tempat rumah sakit berada	Kategorikal	8	0
Hospital name	Nama rumah sakit	Kategorikal	693	3
Establishment ID	ID unik untuk setiap rumah sakit	Kategorikal	696	0
Medicare Provider No.	Nomor penyedia layanan Medicare	Kategorikal	445	251
Local Hospital Network identifier	ID untuk jaringan rumah sakit lokal	Numerik	667	0
Local Hospital Network	Nama jaringan rumah sakit lokal	Kategorikal	667	29
Address Line 1	Alamat rumah sakit (baris pertama)	Kategorikal	696	0
Address Line 2	Alamat rumah sakit (baris kedua)	Kategorikal	696	0
asgc_ra	Kode pengelompokan geografis rumah sakit	Numerik	696	0
Remoteness area	Tingkat keterpencilan rumah sakit (misal, *Major Cities*)	Kategorikal	5	0
Number of available beds	Jumlah tempat tidur yang tersedia	Numerik	696	0
2012-13 Peer Group code	Kode kelompok rumah sakit berdasarkan kinerja dan layanan	Kategorikal	5	0
Peer Group Name	Nama kelompok rumah sakit berdasarkan kode peer group	Kategorikal	5	0
Provided data for NPHED	Status pelaporan data ke NPHED (Yes/No)	Kategorikal	2	0
Provided data for NHMD	Status pelaporan data ke NHMD (Yes/No)	Kategorikal	2	0

Atribut	Deskripsi	Tipe Data	Jumlah nilai uik	Nilai *Missing*
Provided data for NAPEDCD	Status pelaporan data ke NAPEDCD (Yes/No)	Kategorikal	2	0
Provided data for ESWT	Status pelaporan data ke ESWT (Yes/No)	Kategorikal Kategorikal	2	0
Provided data for NNAPCD	Status pelaporan data ke NNAPCD (Yes/No)		2	0
IHPA funding designation	Tipe pendanaan rumah sakit berdasarkan IHPA	Kategorikal	3	0

Dataset tersebut memberikan informasi yang sangat berguna untuk analisis clustering, terutama dalam konteks optimasi layanan kesehatan. Data yang terkandung di dalamnya dapat digunakan untuk mengeksplorasi pengelompokan rumah sakit berdasarkan karakteristik tertentu, seperti kapasitas, jenis layanan, dan keterpencilan geografis. Dengan menggunakan teknik clustering, rumah sakit dapat dikelompokkan berdasarkan kesamaan, sehingga dapat membantu dalam pengambilan keputusan terkait alokasi sumber daya, perencanaan perawatan, serta evaluasi kualitas layanan.

2.1.1 Pra-pemrosesan data

Proses preprocessing data dilakukan untuk mempersiapkan dataset dalam format yang sesuai dengan algoritma clustering. Langkah pertama adalah pemilihan fitur, di mana enam atribut utama dipilih berdasarkan relevansi terhadap analisis. Fitur-fitur tersebut meliputi jumlah tempat tidur, area keterpencilan, jaringan rumah sakit lokal, jenis pendanaan, kode grup rekanan, dan status penyediaan data untuk NPHED. Selanjutnya, data yang memiliki nilai kosong atau hilang dihapus untuk memastikan dataset bersih dan konsisten, sehingga algoritma clustering dapat berjalan dengan optimal.

Langkah berikutnya adalah pengkodean fitur kategorikal. Atribut yang memiliki nilai berupa kategori, seperti *remoteness area* dan *local hospital network*, dikonversi menjadi bentuk numerik menggunakan metode *label encoding*. Selain itu, atribut biner seperti *provided data for NPHED*, yang memiliki nilai "Yes" dan "No", diubah menjadi nilai numerik 1 dan 0. Transformasi ini penting untuk memastikan semua atribut dapat diproses oleh algoritma clustering yang hanya bekerja dengan data numerik.

Tahap akhir preprocessing adalah normalisasi data. Proses ini dilakukan untuk memastikan bahwa semua fitur memiliki skala yang seragam. Normalisasi mengubah data ke skala standar dengan rata-rata 0 dan standar deviasi 1, sehingga setiap atribut memiliki bobot yang sama dalam analisis clustering. Normalisasi juga mencegah fitur dengan nilai rentang besar mendominasi hasil clustering.

2.2 Metode

Penelitian ini menggunakan dua algoritma clustering utama, yaitu K-Means dan DBSCAN
.

2.2.1 K-Means

K-Means adalah salah satu algoritma clustering paling populer yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik [15]. Algoritma ini termasuk dalam kategori clustering partisi, di mana data dibagi menjadi sejumlah *k* cluster yang telah ditentukan sebelumnya. Tujuan utama K-Means adalah meminimalkan total jarak dalam cluster, sehingga titik data dalam cluster yang sama lebih mirip satu sama lain dibandingkan dengan data di cluster lain

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[16]. Algoritma ini bekerja secara iteratif dengan mengoptimalkan posisi centroid, yang merupakan titik rata-rata dari semua data dalam cluster tertentu.

Proses K-Means dimulai dengan inisialisasi sejumlah *k* centroid secara acak. Selanjutnya, setiap titik data ditugaskan ke cluster dengan centroid terdekat berdasarkan jarak Euclidean. Setelah penugasan selesai, centroid diperbarui dengan menghitung rata-rata koordinat dari seluruh data dalam cluster tersebut. Proses ini berulang hingga tidak ada perubahan signifikan pada posisi centroid atau hingga batas iterasi maksimum tercapai [17].

K-Means memiliki beberapa keunggulan, seperti efisiensi komputasi yang tinggi untuk dataset berukuran besar, implementasi yang sederhana, serta kemampuan memberikan hasil clustering yang jelas pada data dengan distribusi yang terstruktur [18]. Namun, algoritma ini juga memiliki keterbatasan, seperti sensitivitas terhadap inisialisasi centroid awal, ketergantungan pada jumlah cluster yang harus ditentukan sebelumnya, dan kurangnya kemampuan untuk menangani cluster dengan bentuk yang tidak reguler atau data dengan outlier yang signifikan [19].

2.2.2 DBSCAN

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah algoritma clustering berbasis kepadatan yang dirancang untuk mengelompokkan data menjadi cluster berdasarkan densitas lokal di sekitar titik data [20]. Algoritma ini efektif dalam menangani dataset yang mengandung noise dan cluster dengan bentuk yang tidak teratur. Berbeda dengan algoritma partisi seperti K-Means, DBSCAN tidak memerlukan penentuan jumlah cluster secara eksplisit, melainkan bergantung pada dua parameter utama, yaitu epsilon (ε) dan min_*samples* [21]. Parameter ε menentukan radius maksimum di sekitar titik data yang digunakan untuk mendefinisikan tetangga, sedangkan min_samples adalah jumlah minimum titik dalam radius ε agar suatu titik dapat dianggap sebagai titik inti (*core point*) [22].

Proses DBSCAN dimulai dengan memilih titik data acak, kemudian menghitung jumlah tetangganya dalam radius ε. Jika jumlah tetangga memenuhi syarat min_*samples*, titik tersebut menjadi titik inti, dan cluster mulai dibentuk dengan memasukkan semua tetangga dalam radius tersebut [23]. Proses ini berlanjut dengan memeriksa tetangga dari setiap titik inti baru hingga tidak ada lagi titik yang dapat ditambahkan ke dalam cluster. Data yang tidak masuk dalam radius ε dari titik inti mana pun dianggap sebagai noise [24].

Keunggulan DBSCAN terletak pada kemampuannya untuk mengidentifikasi cluster dengan bentuk yang arbitrer, menangani data dengan noise, dan tidak memerlukan penentuan jumlah cluster di awal. Namun, algoritma ini juga memiliki kelemahan, seperti sensitivitas terhadap parameter ε dan *min_samples*, sehingga pemilihan parameter yang tidak tepat dapat mengurangi kualitas clustering [25]. Selain itu, DBSCAN kurang efektif pada dataset dengan kepadatan yang bervariasi, karena algoritma mengasumsikan kepadatan yang relatif seragam di seluruh dataset. Meskipun demikian, DBSCAN telah digunakan secara luas dalam berbagai bidang, termasuk analisis data geografis, deteksi anomali, dan pengelompokan data kesehatan, terutama pada dataset yang memiliki distribusi kompleks dan banyak outlier [26].

2.2.3 *Sihouutte Score*

Silhouette Score adalah metrik evaluasi yang digunakan untuk menilai kualitas clustering dengan mengukur seberapa baik objek-objek data berada dalam cluster yang benar dan seberapa jauh mereka dari cluster lain [27]. Metode ini mempertimbangkan dua aspek utama, jarak rata-rata dari suatu titik data ke titik-titik lain dalam cluster yang sama (*a*) dan jarak rata-rata dari titik data tersebut ke titik-titik dalam cluster terdekat (*b*). Nilai Silhouette Score untuk setiap titik data dihitung dengan rumus:

$$s = \frac{b - a}{max(a, b)} \quad (1)$$

Di mana:

- A: Rata-rata jarak suatu titik ke semua titik lain dalam cluster yang sama (koherensi intra-cluster).
- b: Rata-rata jarak suatu titik ke semua titik dalam cluster terdekat yang berbeda (separasi inter-cluster).
- Nilai Silhouette Score untuk seluruh dataset dihitung sebagai rata-rata nilai ss untuk semua titik data. Nilai ini berada dalam rentang [-1,1], di mana:
- Nilai mendekati 1 menunjukkan bahwa data dikelompokkan dengan baik ke dalam cluster yang benar, dan jarak antar cluster jauh.
- Nilai mendekati 0 menunjukkan bahwa data berada di perbatasan antara dua cluster, sehingga pemisahan antar cluster kurang jelas.
- Nilai negatif (< 0) menunjukkan bahwa data lebih dekat ke cluster lain daripada cluster tempat data tersebut ditugaskan, yang menunjukkan clustering yang buruk.

Silhouette Score sangat berguna dalam mengevaluasi hasil clustering tanpa memerlukan informasi label atau kebenaran ground-truth, sehingga sering digunakan dalam clustering tak terawasi (*unsupervised clustering*). Metode ini memberikan pandangan holistik tentang bagaimana setiap data sesuai dengan cluster masing-masing, dan apakah jumlah cluster yang dipilih optimal.

Namun, meskipun *Silhouette Score* merupakan metrik yang kuat, terdapat beberapa batasan. Salah satu kelemahannya adalah sensitivitas terhadap jarak antar cluster yang tidak merata, sehingga pada data dengan distribusi yang kompleks, interpretasi *Silhouette Score* perlu dilakukan dengan hati-hati. Selain itu, metrik ini juga bergantung pada metrik jarak yang digunakan, seperti jarak Euclidean, Manhattan, atau lainnya, yang dapat memengaruhi hasil evaluasi [28].

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, algoritma K-Means dan DBSCAN diuji dengan 10 kali eksperimen yang berbeda untuk mengevaluasi efektivitas keduanya dalam mengelompokkan data rumah sakit. Setiap eksperimen dilakukan dengan mengubah parameter-parameter tertentu, seperti jumlah cluster dan nilai dari *random state* untuk K-Means dan eps serta *min_samples* untuk DBSCAN. Hasil evaluasi dilakukan dengan menggunakan Silhouette Score, yang mengukur seberapa baik pemisahan antara cluster dan kualitas internal cluster tersebut. Hasil eksperimen K-Means dan DBSCAN dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Hasil Eksperimen K-Means

Eksperimen Ke-	Parameter		Silhouette Score
	Nilai Cluster	Nilai Random State	
1	2	1	0.787
2	2	10	0.787
3	2	42	0.787
4	2	50	0.787
5	2	100	0.787
6	3	1	0.379
7	3	10	0.674
8	3	42	0.674
9	3	50	0.699
10	3	100	0.674
Rata-rata Silhouette Score			

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Eksperimen menggunakan algoritma K-Means menunjukkan bahwa dengan *C=2*, Silhouette Score tetap konsisten tinggi pada nilai 0.7879, terlepas dari variasi random state (1, 10, 42, 50, dan 100). Hasil ini menunjukkan bahwa dua cluster yang terbentuk memiliki kualitas yang baik, dengan jarak antar cluster yang jelas dan data dalam cluster yang saling berdekatan. Namun, ketika jumlah cluster dinaikkan menjadi *C=3*, kualitas clustering menurun secara signifikan, dengan Silhouette Score bervariasi antara 0.3795 hingga 0.6997. Hal ini mengindikasikan bahwa struktur cluster yang lebih kompleks tidak sesuai dengan distribusi data, sehingga menurunkan efektivitas pemisahan antar cluster.

Eksperimen Ke-	**Parameter**		**Silhouette Score**
	eps	**Min Samples**	
1	0.1	1	0.070
2	0.3	1	0.074
3	0.5	1	0.078
4	0.7	1	0.078
5	0.9	1	0.075
6	0.1	18	0.998
7	0.3	2	0.955
8	0.5	2	0.902
9	0.7	2	0.902
10	0.9	2	0.766
Rata-rata Silhouette Score			

Eksperimen menggunakan algoritma DBSCAN menunjukkan bahwa hasil clustering sangat bergantung pada pengaturan parameter eps dan min*_samples*. Pada pengaturan awal dengan *min_samples=1*, DBSCAN menghasilkan jumlah cluster yang sangat banyak, yaitu antara 622 hingga 638 cluster, dengan Silhouette Score yang rendah, berkisar antara 0.0703 hingga 0.0788. Hal ini mencerminkan bahwa parameter tersebut menyebabkan DBSCAN cenderung mengidentifikasi banyak cluster kecil dengan kualitas pemisahan yang buruk. Sebaliknya, ketika min_samples ditingkatkan menjadi 2, hasil clustering membaik secara signifikan. Pada pengaturan *eps=0.1*, DBSCAN menghasilkan 18 cluster dengan Silhouette Score mencapai 0.9986, yang menunjukkan kualitas clustering yang hampir sempurna. Eksperimen tambahan dengan eps yang lebih besar, seperti 0.3, 0.5, dan 0.7, tetap menghasilkan kualitas clustering yang baik, dengan Silhouette Score berkisar antara 0.7665 hingga 0.9553. Hasil ini menegaskan bahwa DBSCAN dapat menghasilkan cluster yang sangat terpisah dan terstruktur dengan baik ketika parameter disesuaikan secara optimal.

Berdasarkan hasil eksperimen, kedua algoritma menunjukkan performa yang berbeda tergantung pada karakteristik data dan parameter yang digunakan. K-Means lebih stabil dengan jumlah cluster yang lebih sedikit, dan menghasilkan Silhouette Score yang lebih tinggi ketika *C=2*. Namun, algoritma ini tidak dapat menangani distribusi data yang lebih kompleks atau jumlah cluster yang lebih banyak dengan baik. Sebaliknya, DBSCAN menunjukkan kemampuan yang lebih baik dalam menangani cluster dengan distribusi yang tidak teratur dan noise, terutama ketika parameter min_samples disesuaikan dengan benar.

Meskipun DBSCAN cenderung menghasilkan lebih banyak cluster pada pengaturan awal, kemampuan untuk menghasilkan cluster yang lebih besar dan lebih terpisah dengan *min_samples=2* dan *eps=0.1* menunjukkan bahwa DBSCAN lebih unggul dalam menangani kompleksitas data yang lebih tinggi dan menghasilkan kualitas clustering yang lebih baik.

4. KESIMPULAN

Hasil eksperimen menunjukkan bahwa K-Means lebih efektif pada dataset dengan dua cluster yang jelas, sedangkan DBSCAN lebih unggul pada data dengan distribusi yang lebih kompleks dan terdapat noise. Hasil Silhouette Score yang tinggi untuk DBSCAN pada pengaturan *eps=0.1* dan *min_samples=2* menunjukkan bahwa algoritma ini lebih efektif dalam menghasilkan cluster yang sangat terpisah dan terstruktur dengan baik. Oleh karena itu, pemilihan algoritma clustering harus mempertimbangkan karakteristik dan kompleksitas data yang digunakan, serta pemilihan parameter yang tepat untuk mencapai hasil clustering yang optimal.

DAFTAR PUSTAKA

[1] K. a. K. M. a. P. F. Moniung, "Kualitas Pelayanan Publik Poli Anak Di Rumah Sakit Umum Daerah Noongan Kabupaten Minahasa Provinsi Sulawesi Utara," *Jurnal Eksekutif,* vol. 2, 2020.

[2] D. K. a. E. E. a. M. F. a. H. S. Yudityawati, "Analisis Kepuasan Terhadap Mutu Pelayanan Menggunakan Metode Importance Performance Analysis Pada Pasien Rawat Jalan Rs "X"," dalam *UMMagelang Conference Series*, 2022, pp. 685-697.

[3] I. M. V.-B. H. D. M. E. d. K. P. K. d. P. K. Medan, "Lubis, Syafirda Alifah and Agustina, Dewi and Hafidzah, Fidiana and Barus, Maharani Br and Lubis, Putri Ananda and Nasution, Yulia Adinda," *Jurnal Kesehatan Tambusa,* vol. 5, pp. 5760-5769, 2024.

[4] T. L. a. K. R. R. Wiemken, "Machine learning in epidemiology and health outcomes research," *Annu Rev Public Health,* vol. 41, pp. 21-36, 2020.

[5] R. a. W. A. Ramadhan, "IMPLEMENTASI DATA MINING KLASTERISASI DENGAN ALGORITMA K-MEANS UNTUK PENGELOMPOKKAN KELURAHAN DI PROVINSI DKI JAKARTA BERDASARKAN DISTRIBUSI BANTUAN PANGAN BULOG DI PT YASA ARTHA TRIMANUNGGAL," dalam *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2024, pp. 390-397.

[6] S. A. Kustiyanti, "Smart Hospital: Konsep, Implementasi, dan Tantangan," dalam *Transformasi Rumah Sakit Indonesia Menuju Era Masyarakat*, 2023, p. 161.

[7] M. a. W. A. a. F. L. a. B. S. a. S. J. a. L. K. B. a. P. N. a. C. L. Subramanian, "Precision medicine in the era of artificial intelligence: implications in chronic disease management," *Journal of translational medicine,* vol. 18, pp. 1-12, 2020.

[8] P. a. M. P. Bhattacharjee, "A survey of density based clustering algorithms," *Frontiers of Computer Science,* vol. 15, pp. 1-27, 2021.

[9] D. a. N. F. a. W. S. a. A. D. D. Hastari, "Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan: Application of K-Means and K-Medoids Algorithms for Grouping Country Data Based on Socio-Economic and Health Factors," dalam *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 2023, pp. 274-281.

[10] H. a. L. J. a. Z. X. a. F. M. Hu, "An effective and adaptable K-means algorithm for big data cluster analysis," *Pattern Recognition,* vol. 139, 2023.

[11] H. V. a. G. A. a. D. S. Singh, "A Literature survey based on DBSCAN algorithms," dalam *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2022, pp. 751-758.

[12] Z. a. W. J. a. H. K. Cai, "Adaptive density-based spatial clustering for massive data analysis," *IEEE Access,* vol. 8, 2020.

[13] M. I. a. L. R. Y. a. A. M. A. K. a. H. M. S. a. C. M. K. H. a. K. B. Pramanik, "Healthcare informatics and analytics in big data," *Expert Systems with Applications,* vol. 152, 2020.

[14] A. a. N. S. a. R. I. Rehman, "Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities," *Multimedia Systems,* vol. 28, 2022.

[15] W. Gunawan, "Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot," *Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi,* vol. 6, pp. 91-98, 2021.

[16] F. a. L. Z. a. W. R. a. L. X. Nie, "An effective and efficient algorithm for K-means clustering with new formulation," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, 2022.

[17] A. a. o. Zakir, "Implementasi Algoritma K-Means Untuk Clustering Judul Skripsi Universitas Harapan Medan," *Jurnal Media Informatika,* vol. 4, pp. 40-47, 2022.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[18] S. a. N. B. I. a. A. Z. Setyaningtyas, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang,* vol. 10, pp. 52-61, 2022.
[19] R. J. a. B. S. a. A. S. Kasim, "Implementasi Metode K-Means Untuk Clustering Data Penduduk Miskin Dengan Systematic Random Sampling," *Prosiding SISFOTEK,* vol. 5, pp. 95-101, 2021.
[20] W. Gunawan, "Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot," *Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi,* vol. 6, pp. 91-98, 2021.
[21] B. a. R. T. a. J. A. Biantara, "Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat," *Scientific Student Journal for Information, Technology and Science,* vol. 4, pp. 88-94, 2023.
[22] D. Deng, "DBSCAN clustering algorithm based on density," dalam *2020 7th international forum on electrical engineering and automation (IFEEA)*, 949-953, IEEE, 2020.
[23] R. a. P. H. a. D. Y. a. W. J. a. S. Q. a. L. Y. a. Z. J. a. Y. P. S. Zhang, "Automating DBSCAN via deep reinforcement learning," dalam *Proceedings of the 31st acm international conference on information \& knowledge management*, 2022, pp. 2620-2630.
[24] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan," *Jurnal Informatika dan Teknik Elektro Terapan,* vol. 12, 2024.
[25] G. Mahardika, "CLUSTERING DATA SENSOR IOT DENGAN ALGORITMA DBSCAN," *Jurnal Dunia Data,* vol. 1, 2024.
[26] F. M. a. W. S. H. a. S. N. Y. Pranata, "Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 8, 2024.
[27] W. A. a. H. K. Silamantha, "Analisis RFM dan K-Means Clustering untuk Segmentasi Pelanggan pada PT. Sanutama Bumi Arto," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen),* vol. 5, 2024.
[28] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan," *Jurnal Informatika dan Teknik Elektro Terapan,* vol. 12, 2024.
[29] T. D. a. V. A. a. F. R. Harjanto, "Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode dbscan clustering (studi kasus data dinas kesehatan kabupaten lebong)," *Rekursif: Jurnal Informatika,* vol. 9, 2021.
[30] T. D. a. V. A. a. F. R. Harjanto, "Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode dbscan clustering (studi kasus data dinas kesehatan kabupaten lebong)," *Rekursif: Jurnal Informatika,* vol. 9, 2021.
[31] B. a. R. T. a. J. A. Biantara, "Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat," *Scientific Student Journal for Information, Technology and Science,* vol. 4, pp. 88-94, 2023.