

COMPARISON OF DBSCAN AND K-MEANS CLUSTER ANALYSIS WITH PATH-ANOVA IN CLUSTERING WASTE MANAGEMENT BEHAVIOUR PATTERNS

Muhammad Rizal Zuhdi¹, Hafizh Syihabuddin Al Jauhar^{*2}, Adji Achmad Rinaldo Fernandes³,
Ni Wayan Surya Wardhani⁴

^{1,2,3}Research field of Computational Statistics and Data Science, Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Brawijaya

Email: ¹muha_rizalz@student.ub.ac.id, ²aljauhar.hafizh@student.ub.ac.id, ³fernandes@ub.ac.id

(Article received: December 4, 2024; Revision: January 7, 2025; published: February 20, 2025)

Abstract

This study aims to compare the effectiveness of DBSCAN and K-Means cluster analysis methods in clustering waste management behaviour patterns in Batu City. The data used is secondary data from previous research with a total of 395 respondents taken using the quota sampling method. DBSCAN classifies data based on density with the main parameters epsilon and MinPts, while K-Means uses the average centroid to determine the cluster. The analysis results show that DBSCAN produces a silhouette index of 0.664, which is higher than K-Means with a value of 0.574. DBSCAN also successfully identified noise as much as 10 data that did not belong to any cluster, while K-Means did not have a similar mechanism. The results of Path-ANOVA show that DBSCAN is the most optimal clustering with a more significant partition difference value. Further tests were conducted to strengthen the validation of clustering results using Path-ANOVA. Both methods produced two main clusters, with the second cluster showing higher quality in terms of maintenance, quality, and ease of use of environmental hygiene facilities. This research emphasises the importance of choosing an appropriate clustering method to ensure optimal clustering results, especially in data with complex characteristics.

Keywords: cluster analysis, DBSCAN, K-Means, Path-ANOVA, waste management

PERBANDINGAN ANALISIS CLUSTER DBSCAN DAN K-MEANS DENGAN PATH-ANOVA DALAM PENGELOMPOKAN POLA PERILAKU PENGELOLAAN SAMPAH

Abstrak

Penelitian ini bertujuan untuk membandingkan efektivitas metode analisis cluster DBSCAN dan K-Means dalam mengelompokkan pola perilaku pengelolaan sampah di Kota Batu. Data yang digunakan merupakan data sekunder dari penelitian sebelumnya dengan total 395 responden yang diambil menggunakan metode quota sampling. DBSCAN mengelompokkan data berdasarkan kepadatan dengan parameter utama epsilon dan MinPts, sedangkan K-Means menggunakan rata-rata centroid untuk menentukan cluster. Hasil analisis menunjukkan bahwa DBSCAN menghasilkan indeks *silhouette* sebesar 0,664, yang lebih tinggi dibandingkan K-Means dengan nilai 0,574. DBSCAN juga berhasil mengidentifikasi noise sebanyak 10 data yang tidak masuk dalam cluster manapun, sementara K-Means tidak memiliki mekanisme serupa. Uji lanjut dilakukan untuk memperkuat validasi hasil *clustering* menggunakan *Path-ANOVA*. Diperoleh bahwa DBSCAN merupakan *Clustering* paling optimal dengan nilai perbedaan partisi yang lebih signifikan. Kedua metode menghasilkan dua cluster utama, dengan cluster kedua menunjukkan kualitas lebih tinggi dalam hal perawatan, kualitas, dan kemudahan penggunaan sarana kebersihan lingkungan. Penelitian ini menekankan pentingnya memilih metode clustering yang sesuai untuk memastikan hasil pengelompokan yang optimal, terutama dalam data dengan karakteristik yang kompleks.

Kata kunci: analisis cluster, DBSCAN, K-Means, Path-ANOVA, pengelolaan sampah

1. PENDAHULUAN

Analisis *cluster* adalah sekelompok teknik multivariat yang tujuan utamanya adalah mengelompokkan objek berdasarkan karakteristik

yang mereka miliki [1]. Dalam kelompok yang sama, objek-objek ini memiliki jarak relatif yang serupa satu sama lain. Dua metode utama digunakan dalam analisis *cluster*: metode hierarki dan metode

nonhierarki. Perbedaan antara keduanya terletak pada cara penentuan jumlah kelompok. Metode hierarki tidak memerlukan jumlah kelompok yang telah ditentukan sebelumnya, sementara metode nonhierarki menentukan jumlah kelompok di awal. Metode hierarki melibatkan beberapa pendekatan seperti *single linkage*, *average linkage*, *complete linkage*, dan *ward linkage*. Di sisi lain, metode nonhierarki mencakup metode *K-Means* dan metode *Fuzzy K-Means* (FCM).

Menurut [2] analisis *cluster K-Means* adalah algoritma *unsupervised machine learning* yang paling umum digunakan untuk mempartisi kumpulan data yang diberikan ke dalam satu set k kelompok, di mana k mewakili jumlah kelompok yang telah ditentukan sebelumnya. Algoritma ini mengklasifikasikan objek dalam beberapa kelompok, sehingga objek-objek dalam kelompok yang sama memiliki kemiripan yang tinggi, sedangkan objek-objek dari kelompok yang berbeda memiliki kemiripan yang rendah. Dalam analisis *cluster K-Means*, setiap *cluster* diwakili oleh pusatnya atau *centroid* yang sesuai dengan rata-rata titik yang ditetapkan ke *cluster*. Namun, pada analisis *cluster K-Means* terdapat masalah yaitu tidak semua objek pada penelitian memiliki karakter yang dapat digolongkan dalam suatu *cluster* tertentu. Beberapa objek dengan karakter unik yang jauh berbeda dengan objek lain. Objek ini merupakan *noise* yang jika masuk dalam suatu *cluster* dapat mempengaruhi identifikasi karakter dan mengurangi akuratnya hasil *cluster* yang dibentuk [3].

Objek yang teridentifikasi sebagai *noise* dapat memberikan informasi yang tidak bisa diberikan oleh objek lain pada data, misalnya karena *noise* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki [4]. Beberapa algoritma dalam teknik *clustering* yang mampu mengatasi permasalahan *noise* (*outlier*) adalah *Density-Based Spatial Clustering of Applications with Noise*, *K-Medoids*, *Self-Organizing Map*, dan *Fuzzy C-Means*. [3] menyatakan bahwa algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) mampu mendeteksi adanya *noise* pada data yang efektif untuk mengelompokkan data berdasarkan kepadatan data. Sistem algoritma ini mengelompokkan data dengan mengidentifikasi titik yang tidak termasuk di dalam *cluster* sebagai *noise* dan mengelompokkan titik yang memiliki kemiripan yang tinggi.

DBSCAN diperkenalkan oleh Ester, algoritma ini merupakan teknik *clustering* non-parametrik [5]. Algoritma ini mendapatkan penghargaan *the test of the time award* pada tahun 2014 oleh ACM (*Association for Computing Machinery*) saat konferensi data mining. DBSCAN adalah jenis *cluster* berbasis partisi yang dimana daerah yang lebih padat dianggap sebagai *cluster* dan daerah yang kepadatannya rendah disebut *noise* [6]. Dalam

DBSCAN tidak perlu memberikan nilai input untuk jumlah *cluster*. Oleh karena itu, *cluster* terbentuk berdasarkan kepadatan, bukan dari bentuk asumsi apapun [7]. Keunggulan utama dari algoritma ini tidak akan memasukkan data yang dianggap *noise* kedalam *cluster* manapun. Tentunya ini menjadi keunggulan sendiri mengingat data *noise* bisa jadi memiliki karakteristik yang jauh berbeda dengan kumpulan data pada umumnya.

Penelitian ini bertujuan untuk membandingkan performa kedua metode *clustering* yaitu DBSCAN dan *K-Means* dengan pengembangan uji lanjut *Path-ANOVA*. Penggunaan *Path-ANOVA* difungsikan secara khusus untuk mengevaluasi perbedaan signifikan antara hasil *clustering* dengan DBSCAN dan *K-Means*, terhadap variabel-variabel yang memengaruhi *outcome* utama. Teknik ini lebih luas karena tidak hanya memungkinkan analisis pengaruh langsung antar variabel, tetapi juga mengidentifikasi efek mediasi yang berkontribusi terhadap perbedaan performa kedua metode. Dengan mengintegrasikan pendekatan ANOVA [8] [9] dalam jalur [10], *Path-ANOVA* diharapkan mampu memberikan gambaran yang lebih akurat mengenai faktor-faktor yang berkontribusi secara signifikan terhadap *cluster* yang dihasilkan, sehingga memberikan pemahaman yang lebih kompleks untuk kebutuhan pengambilan keputusan berbasis data.

2. METODE PENELITIAN

2.1. Data

Penelitian ini menggunakan data sekunder yang diambil dari penelitian yang dilakukan oleh [11]. Populasi dalam penelitian tersebut mencakup masyarakat Kota Batu, yang berjumlah 217.871 keluarga. Sampel penelitian ini terdiri dari masyarakat yang tinggal di Kecamatan Batu dan Kecamatan Bumiaji, dengan total sampel yang ditetapkan menggunakan *quota sampling* yaitu sebanyak 395 responden.

2.2. Analisis Cluster DBSCAN

DBSCAN merupakan salah satu algoritma *clustering* yang mengelompokkan objek ke dalam *cluster* dimana kategori atau class data ujinya belum diketahui (*unsupervised learning*). DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. Konsep kepadatan yang dimaksud dalam DBSCAN adalah jumlah data yang berada dalam radius ϵ dari setiap data. Jika jumlah data dalam radius ϵ lebih dari atau sama dengan $MinPts$ (jumlah minimal data dalam radius ϵ), data tersebut masuk dalam kategori kepadatan yang diinginkan, jumlah data dalam radius tersebut termasuk data itu sendiri [12].

Algoritma DBSCAN membutuhkan dua parameter penting, yaitu ϵ (*Eps*) dan jumlah data tetangga minimal untuk membentuk kelompok ($MinPts$). Algoritma dari DBSCAN sebagai berikut [13]:

1. Pilih titik awal p secara *random*.
2. Inisialisasi parameter input *MinPts* dan *Eps*.
3. Hitung *Eps* atau semua jarak densitas terjangkau terhadap p menggunakan jarak *euclidean*

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \quad (0.1)$$

Dimana x_{ia} adalah variabel ke- a dari objek i ($i = 1, \dots, n; a = 1, \dots, p$) dan d_{ij} adalah nilai *Euclidean Distance*.

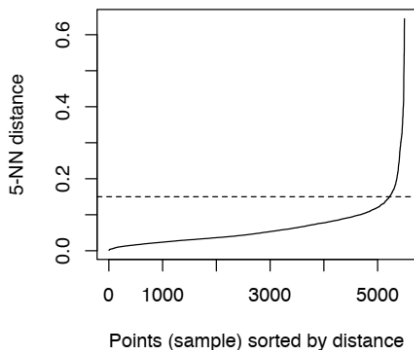
4. Jika jumlah titik yang memenuhi *Eps* lebih dari *MinPts* maka titik p adalah titik pusat (*core*) dan *cluster* terbentuk.
5. Ulangi langkah 3 – 4 hingga semua titik diproses. Jika p adalah titik border dan tidak ada titik yang densitas terjangkau terhadap p , maka proses dilanjutkan ke titik yang lain.

Menurut [11] terdapat beberapa keuntungan yang dimiliki oleh algoritma DBSCAN, antara lain:

1. Mengenali *cluster* non-konveks.
2. Partisi dengan jumlah *cluster* yang paling tepat diperoleh secara otomatis.
3. Tidak perlu menggunakan indeks untuk menentukan jumlah *cluster* yang sesuai dalam sebuah partisi.

2.3. Algoritma K-Nearest Neighbor

Terdapat dua parameter utama dalam pengelompokan data menggunakan DBSCAN yaitu *epsilon* dan *MinPts* [15]. Penentuan *epsilon* dan *MinPts* dapat dilakukan dengan bantuan algoritma *K-Nearest Neighbour* (KNN). Algoritma KNN merupakan suatu algoritma untuk melakukan klasifikasi pada objek terhadap data pembelajaran yang jaraknya paling dekat dengan objek tersebut [16]. Untuk menentukan *epsilon* dan *MinPts* dapat menggunakan bantuan grafik *k-dist*.



Gambar 1. Contoh Grafik *k-dist*
Sumber: Kassambara (2017)

Ilustrasi di atas merupakan grafik *k-dist* dengan nilai k sebesar 5. Perubahan tajam (berbentuk siku) yang terjadi pada grafik akan dianggap sebagai *epsilon* dan nilai k digunakan sebagai *MinPts*. Garis vertikal pada grafik merupakan ambang batas yang

ditentukan berdasarkan kenaikan grafik yang signifikan pada keseluruhan grafik *k-dist*. Garis ambang batas bertujuan untuk mencegah kesalahan dalam penentuan nilai *epsilon* di mana jika nilai *eps* terlalu besar maka mengakibatkan pencilaan tergabung dalam kelompok. Langkah-langkah dalam menentukan *epsilon* dan *MinPts* sebagai berikut.

1. Hitung nilai *k-dist* untuk semua titik pada setiap nilai k . Nilai *k-dist* diperoleh dari nilai rata-rata jarak *euclidean* suatu objek dengan k titik terdekat.
2. Urutkan nilai *k-dist* secara meningkat dari rata-rata paling kecil hingga paling besar.
3. Amati titik dengan perubahan paling tajam pada setiap grafik *k-dist*, lalu tandai sebagai ambang batas.
4. Perubahan pada nilai *k-dist* atau titik yang membentuk siku akan digunakan sebagai nilai *epsilon* dan nilai k sebagai *MinPts* yang sesuai.

2.4. Analisis Cluster K-Means

Menurut [17], analisis *K-Means* merupakan salah satu pendekatan dalam analisis *cluster* yang tidak bersifat hierarkis. Metode ini bertujuan untuk membagi objek-objek yang ada ke dalam sejumlah klaster atau kelompok, berdasarkan ciri-ciri atau karakteristik yang dimiliki oleh masing-masing objek. Dengan cara ini, objek-objek yang memiliki ciri-ciri serupa akan dikelompokkan bersama dalam satu klaster, sementara objek-objek yang memiliki ciri-ciri yang berbeda akan ditempatkan dalam klaster yang berbeda pula. Tujuan utama dari pengelompokan ini adalah untuk mengurangi variasi di dalam satu klaster sebanyak mungkin, sekaligus meningkatkan variasi antar klaster.

Langkah-langkah algoritma *K-Means* adalah sebagai berikut [12].

1. Menentukan k sebagai jumlah *cluster* yang ingin dibentuk
2. Mengalokasikan objek ke dalam *cluster* secara acak
3. Menentukan pusat *cluster* (*centroid*) dari data yang ada pada masing-masing *cluster* dengan persamaan (0.2).

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (0.2)$$

Keterangan

C_{kj} : pusat-pusat *cluster* ke- k pada variabel ke- j ($j = 1, 2, \dots, p$)
 n : banyak data pada *cluster* ke- k

4. Menentukan jarak setiap objek dengan setiap *centroid* dengan menggunakan ukuran jarak *Euclidean* pada persamaan (0.3).

$$d_{kj} = \sqrt{\sum_{j=1}^n (x_{ij} - v_{kj})^2} \quad (0.3)$$

Keterangan:

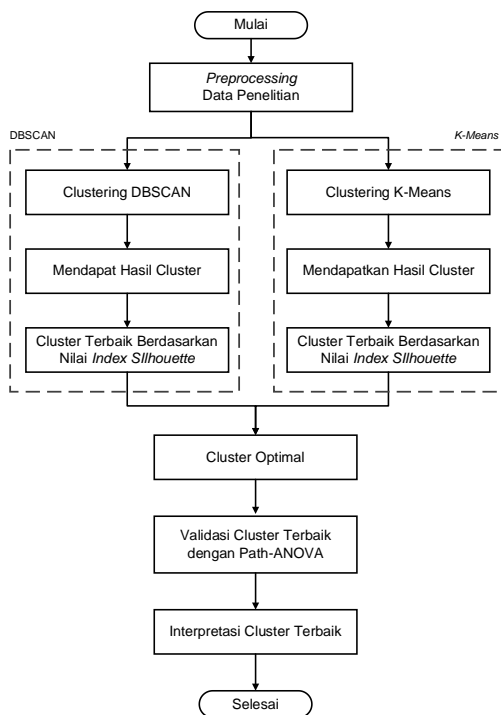
d_{kj} : *Euclidean distance* ke- k untuk variabel ke- j

- k : Indeks *cluster*
- j : Indeks dari variabel
- x_{ij} : Nilai objek ke- i yang ada di dalam *cluster* tersebut untuk variabel ke- j
- v_{kj} : Centroid *cluster* ke- k untuk variabel ke- j

5. Mengelompokkan objek berdasarkan jarak ke *centroid* terdekat.
6. Mengulangi kembali langkah 3-5 sampai tidak ada lagi perpindahan objek atau tidak ada perubahan pada objektifnya

2.5. Diagram Alir Penelitian

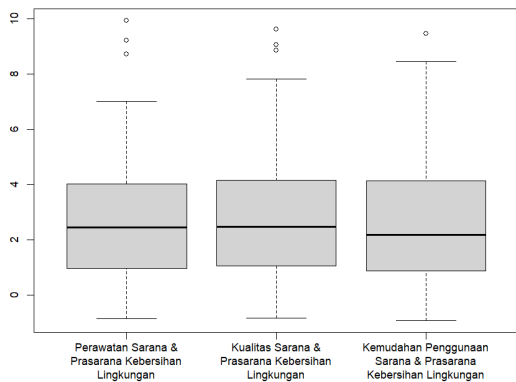
Alur penelitian secara runtut dan rinci tertuang pada Gambar 2. berikut.



Gambar 2. Diagram Alir Penelitian

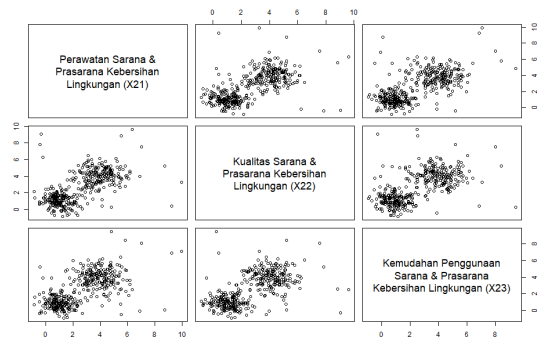
3. HASIL DAN PEMBAHASAN

3.1. Visualisasi Data



Gambar 3. Visualisasi Boxplot Indikator Kualitas Sarana dan Prasarana

Gambar 3 menunjukkan adanya *outlier* pada setiap indikator, yang diwakili oleh titik-titik data yang berada di luar garis whisker pada boxplot. Pada indikator perawatan sarana dan prasarana kebersihan lingkungan terdapat tiga *outlier*, demikian pula pada indikator kualitas sarana dan prasarana kebersihan lingkungan, sementara indikator kemudahan penggunaan sarana dan prasarana kebersihan lingkungan hanya menunjukkan satu *outlier*. Keberadaan *outlier* ini menunjukkan adanya data yang berada jauh dari nilai rata-rata kelompok, yang dapat disebabkan oleh faktor tertentu atau variasi dalam penilaian responden. Situasi ini menekankan pentingnya penerapan metode analisis yang *robust* terhadap *outlier* agar hasil pengelompokan tidak terdistorsi oleh nilai-nilai ekstrem tersebut. Selain mendeteksi *outlier*, pada visualisasi data ini juga akan diidentifikasi pola pengelompokan yang mungkin terbentuk dalam data. Visualisasi scatter plot dapat dilihat pada Gambar 4.



Gambar 4. Visualisasi Scatter Plot

Gambar 4 memperlihatkan pola pengelompokan data yang cukup jelas, dengan dua *cluster* yang terpisah pada setiap hubungan antar variabel. Pola ini mengindikasikan adanya pembagian alami dalam data menjadi dua kelompok, sehingga diperlukan analisis *cluster* untuk mengeksplorasi lebih lanjut [18]. Selain itu, Gambar 3 juga menunjukkan bahwa salah satu kelompok memiliki penyebaran data yang lebih padat, sementara beberapa titik berada agak jauh dari pusat kelompok utama. Fenomena ini mungkin mengindikasikan adanya data *outlier* atau *noise* yang memerlukan analisis lebih lanjut guna memahami karakteristik setiap *cluster* dengan lebih mendalam.

3.2. Uji Statistik Hopkins

Langkah awal sebelum melakukan analisis *cluster* adalah mendapatkan nilai statistik *Hopkins*. Nilai statistik *Hopkins* berkisar antara 0 sampai 1. Statistik *Hopkins* bertujuan untuk mengetahui kecenderungan sebaran data atau terbentuknya kelompok dari suatu set data dengan mengukur jarak antara objek sampel acak yang sudah dibuat sebelumnya dengan objek-objek asli pada dataset. Semakin kecil nilai perbandingan jarak antar objek, semakin kecil kemungkinan data memiliki struktur yang tidak acak.

Berdasarkan uji statistik *Hopkins* didapatkan hasil dari persamaan **Error! Reference source not found.** yaitu sebesar 0,796 di mana nilai statistik *hopkins* lebih dari 0,5 dan mendekati 1 yang berarti data pada penelitian ini memiliki kemungkinan struktur yang tidak acak atau terdapat kecenderungan terbentuknya kelompok antar objek.

3.3. Hasil Cluster menggunakan DBSCAN

Penelitian ini menggunakan metode *clustering* DBSCAN. Pada analisis *cluster* DBSCAN, pengelompokan dilakukan berdasarkan kepadatan (*density*) dengan memanfaatkan dua parameter utama, yaitu *MinPts* dan *epsilon*. Nilai indeks *silhouette* yang diperoleh dari berbagai kombinasi *MinPts* dan *epsilon* ditampilkan pada Tabel 1.

Tabel 1. Nilai Indeks *Silhouette* dari Berbagai Kombinasi *MinPts* dan *Epsilon*

MinPts	ϵ	IS	MinPts	ϵ	IS
3	0,2	0,143	7	0,2	0,241
	0,3	0,419		0,3	0,587
	0,4	0,488		0,4	0,654
	0,5	0,488		0,5	0,488
	0,6	0,505		0,6	0,505
	0,7	0,523		0,7	0,523
	0,8	0,523		0,8	0,523
	0,9	0,523		0,9	0,523
	1	0,523		1	0,523
	0,2	0,118		0,2	0,263
4	0,3	0,416	8	0,3	0,555
	0,4	0,488		0,4	0,641
	0,5	0,488		0,5	0,664
	0,6	0,505		0,6	0,505
	0,7	0,523		0,7	0,523
	0,8	0,523		0,8	0,523
	0,9	0,523		0,9	0,523
	1	0,523		1	0,523
	0,2	0,372		0,2	0,293
	5	0,3		0,603	9
0,4		0,660	0,4	0,630	
0,5		0,488	0,5	0,664	
0,6		0,505	0,6	0,505	
0,7		0,523	0,7	0,523	
0,8		0,523	0,8	0,523	
0,9		0,523	0,9	0,523	
1		0,523	1	0,523	
0,2		0,277	0,2	0,290	
6		0,3	0,592	10	
	0,4	0,660	0,4		0,625
	0,5	0,488	0,5		0,664
	0,6	0,505	0,6		0,505
	0,7	0,523	0,7		0,523
	0,8	0,523	0,8		0,523
	0,9	0,523	0,9		0,523
	1	0,523	1		0,523

Tabel 1 menunjukkan bahwa nilai indeks *silhouette* tertinggi dicapai saat parameter *epsilon* diatur sebesar 0,5, dengan *MinPts* bernilai 8, 9, atau 10. Nilai indeks *silhouette* ini menunjukkan bahwa konfigurasi parameter tersebut menghasilkan kualitas pengelompokan yang optimal, di mana jarak antar-*cluster* cukup besar dan distribusi data dalam setiap *cluster* cukup padat [19]. Jumlah anggota dalam setiap *cluster* yang dihasilkan menggunakan metode DBSCAN dapat dilihat secara rinci pada Tabel 2.

Tabel 2. Banyaknya Anggota Cluster Metode DBSCAN

MinPts	ϵ	Cluster	Banyak Anggota	IS
8, 9, 10	0,5	Noise	10	0,664
		Cluster 1	189	
		Cluster 2	196	

Berdasarkan Tabel 2, metode DBSCAN dengan nilai *epsilon* sebesar 0,5 dan *MinPts* bernilai 8, 9, serta 10 menghasilkan dua *cluster*. *Cluster* pertama terdiri dari 189 responden, sementara *cluster* kedua mencakup 196 responden. Selain itu, terdapat 10 responden yang dikelompokkan sebagai *noise*. *Noise* ini merujuk pada data yang tidak mengikuti pola utama dalam *cluster* dan oleh karena itu dikeluarkan untuk menjaga kualitas pengelompokan. Kehadiran *noise* menunjukkan adanya variasi yang signifikan dalam distribusi data, yang mungkin diakibatkan oleh faktor *outlier* atau ketidaksesuaian pola [20]. Hasil pengelompokan DBSCAN menunjukkan indeks *silhouette* sebesar 0,664, yang mengindikasikan kualitas pengelompokan dengan tingkat kekompakan dan pemisahan yang baik.

3.4. Hasil Cluster menggunakan K-Means

Penelitian ini menggunakan metode *clustering* K-Means. Pada analisis *cluster* K-Means, pengelompokan dilakukan berdasarkan rata-rata *centroid*. Hasil analisis *cluster* metode K-Means di tampilkan pada Tabel 3.

Tabel 3. Banyaknya Anggota Cluster Metode K-Means

Cluster	Banyak Anggota	Silhouette Index
Cluster 1	190	0,574
Cluster 2	205	

Metode *K-Means* menghasilkan dua *cluster*, di mana *cluster* pertama terdiri dari 190 responden, sementara *cluster* kedua mencakup 205 responden. Perbedaan jumlah anggota di antara kedua *cluster* ini menunjukkan adanya variasi karakteristik di antara responden, yang dikelompokkan berdasarkan rata-rata *centroid* [21]. Metode *K-Means* bekerja dengan mengelompokkan data berdasarkan jarak terdekat responden ke pusat *cluster* (*centroid*), sehingga perbedaan ukuran *cluster* juga dapat mengindikasikan distribusi data yang tidak merata dalam dataset [22]. Variasi ini memberikan gambaran awal mengenai keragaman pola di antara kelompok responden yang terbentuk [23]. Pada metode *K-Means*, diperoleh nilai indeks *silhouette* sebesar 0,574, yang mengindikasikan kualitas pengelompokan yang kurang baik.

3.5. Hasil Path-ANOVA

Hasil Analisis dari DBSCAN dan *K-Means* selanjutnya akan divalidasi kembali menggunakan *Path-ANOVA*. Hasil analisis perbedaan ini dapat dilihat pada Tabel 4. berikut.

Tabel 4. Perbandingan Perbedaan Cluster dengan *Path-ANOVA*

Metode	Cluster	Nilai-P	Keputusan
DBSCAN	Cluster 1	< 0,001	Signifikan
	Cluster 2		
K-Means	Cluster 1	0,012	Signifikan
	Cluster 2		

Tabel 4 menunjukkan bahwa pada Cluster 1 dan Cluster 2 diperoleh nilai-p < 0,001, artinya terdapat perbedaan yang bermakna pada hasil *Clustering* DBSCAN. Pada *K-Means*, diperoleh nilai-p 0,012, yang berarti perbedaan cluster ini juga signifikan tetapi tidak lebih baik dari hasil cluster menggunakan DBSCAN. Hasil ini menunjukkan bahwa metode DBSCAN dan *K-Means* berhasil mempartisi data menjadi cluster dengan perbedaan yang signifikan melalui pendekatan *Path-ANOVA*. DBSCAN memiliki signifikansi yang lebih kuat (nilai-p lebih kecil), tetapi *K-Means* juga menghasilkan perbedaan yang signifikan. Hal ini sejalan dengan hasil penelitian [24] namun terdapat perbedaan analisis dan model data yang digunakan.

4. DISKUSI

Setelah ditentukan penentuan banyak *cluster*, langkah selanjutnya adalah melihat karakteristik setiap *cluster* dengan melihat nilai rata-rata yang diperoleh dari setiap *cluster*. Menurut [19] penentuan karakteristik pada setiap *cluster* yang terbentuk menggunakan rata-rata skor dengan interpretasi berdasarkan Tabel 5.

Tabel 5. Kriteria Rata-rata Skor

Rata-Rata Skor	Keterangan
1,00 – 1,50	Sangat buruk
1,51 – 2,50	Buruk
2,51 – 3,50	Sedang/cukup
3,51 – 4,50	Baik
4,51 – 5,00	Sangat baik

Berdasarkan kriteria interpretasi dalam Tabel 5, karakteristik setiap *cluster* yang dihasilkan melalui berbagai metode pengelompokan dapat dilihat pada Tabel 6.

Tabel 6. Karakteristik Setiap Cluster Berbagai Metode

Metode	Cluster	X21	X22	X23
DBSCAN	Cluster 1	3,605	3,540	3,525
	Cluster 2	4,624	4,590	4,569
K-Means	Cluster 1	3,605	3,539	3,525
	Cluster 2	4,629	4,609	4,584

Berdasarkan Tabel 6, Penjelasan nilai rata-rata masing-masing *cluster* dapat diketahui bahwa:

1) Cluster 1

Cluster 1, yang terbentuk melalui metode DBSCAN dan *K-Means*, dikategorikan sebagai *cluster* baik karena nilai rata-rata indikatornya berkisar antara 3,525 hingga 3,605, yang berada dalam kategori baik berdasarkan skala interpretasi. Responden dalam *Cluster 1* menunjukkan perhatian yang positif terhadap aspek-aspek kebersihan

lingkungan, seperti perawatan sarana dan prasarana kebersihan lingkungan, kualitas sarana dan prasarana kebersihan, serta kemudahan penggunaan fasilitas kebersihan.

Karakteristik responden dalam *cluster* ini mencerminkan kesadaran dan kepedulian tinggi terhadap pentingnya menjaga kualitas sarana kebersihan serta efisiensi dalam penggunaannya. Kondisi ini menunjukkan bahwa pada *cluster* ini, responden cenderung memiliki tingkat kepuasan yang lebih tinggi terhadap fasilitas yang ada, serta komitmen dalam menjaga kebersihan yang dapat mendukung tercapainya lingkungan yang lebih baik.

2) Cluster 2

Cluster 2, yang terbentuk melalui analisis menggunakan metode DBSCAN dan *K-Means*, dikategorikan sebagai *cluster* sangat baik berdasarkan skala interpretasi, dengan nilai rata-rata indikator yang berkisar antara 4,584 hingga 4,629. Nilai ini menunjukkan bahwa responden dalam *Cluster 2* mempersepsikan perawatan, kualitas, dan kemudahan penggunaan sarana kebersihan lingkungan pada tingkat yang sangat baik.

Responden dalam *Cluster 2* menunjukkan kepuasan yang sangat tinggi terhadap perawatan sarana dan prasarana kebersihan, menilai kualitas sarana kebersihan secara keseluruhan sebagai sangat baik, serta merasakan kemudahan penggunaan fasilitas kebersihan. Tingkat apresiasi yang tinggi terhadap fasilitas ini mengindikasikan bahwa kelompok ini cenderung memiliki kesadaran dan praktik yang sangat positif terhadap kebersihan lingkungan, sehingga dapat berkontribusi signifikan terhadap terciptanya lingkungan yang lebih bersih dan tertib.

5. KESIMPULAN

Hasil Penelitian menunjukkan bahwa metode DBSCAN memiliki keunggulan dalam mendeteksi noise dan menghasilkan pengelompokan dengan indeks *silhouette* yang lebih tinggi dibandingkan *K-Means*, yaitu sebesar 0,664. Selain itu, DBSCAN mampu mengidentifikasi dua cluster utama yang lebih berbeda signifikan dengan distribusi yang lebih kompak serta 10 data noise yang tidak termasuk dalam cluster manapun. Sebaliknya, metode *K-Means* menghasilkan indeks *silhouette* sebesar 0,574, dengan distribusi yang cenderung lebih merata tetapi kurang efisien dalam menangani *outlier* [25]. Berdasarkan hasil uji menggunakan *Path-ANOVA*, tampak bahwa kedua metode ini mengindikasikan adanya karakteristik perilaku yang signifikan dalam pengelolaan sampah di masing-masing cluster, dengan cluster kedua secara konsisten menunjukkan kualitas yang lebih tinggi berdasarkan rata-rata skor indikator. Artinya Pemerintah Kota Batu diharapkan untuk lebih fokus menggalakkan pemberdayaan pengelolaan sampah di area *cluster 1*. Dengan demikian pemerataan

kualitas pengelolaan sampah di Kota Batu dapat tercapai.

DAFTAR PUSTAKA

- [1] Y. S. Afrianti, U. S. Pasaribu, F. H. Sulaiman, G. Angelia, and H. J. Wattimanela, "K-Means And Agglomerative Hierarchy Clustering Analysis On The Stainless Steel Corrosion Problem," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, pp. 0589–0602, Mar. 2024, doi: 10.30598/barekengvol18iss1pp0589-0602.
- [2] A. Kassambara, *Practical guide to cluster analysis in R: Unsupervised machine learning*, vol. 1. Sthda, 2017.
- [3] I. D. Id, A. Astrid, and E. Mahdiyah, "Modifikasi DBSCAN (Density-Based Spatial Clustering With Noise) pada Objek 3 Dimensi," *Jurnal Komputer Terapan*, vol. 3, no. 1, pp. 41–52, 2017, [Online]. Available: <http://jurnal.pcr.ac.id>
- [4] E. T. K. Dewi, A. Agoestanto, and S. Sunarmi, "Metode Least Trimmed Square (LTS) Dan Mm-Estimation Untuk Mengestimasi Parameter Regresi Ketika Terdapat Outlier," *UNNES Journal of Mathematics*, vol. 5, no. 1, pp. 47–54, 2016.
- [5] T. N. Tran, K. Drab, and M. Daszykowski, "Revised DBSCAN algorithm to cluster data with dense adjacent clusters," *Chemometrics and Intelligent Laboratory Systems*, vol. 120, pp. 92–96, Jan. 2013, doi: 10.1016/j.chemolab.2012.11.006.
- [6] P. B. Nagpal and P. A. Mann, "Comparative Study of Density based Clustering Algorithms," *Int J Comput Appl*, vol. 27, no. 11, pp. 44–77, Dec. 2011, doi: 10.5120/3341-4600.
- [7] S. U. Rehman, S. Asghar, S. Fong, and S. Sarasvady, "DBSCAN: Past, present and future," in *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, IEEE, Feb. 2014, pp. 232–238. doi: 10.1109/ICADIWT.2014.6814687.
- [8] D. S. Moore and G. P. McCabe, *Introduction to the practice of statistics eight edition*. WH Freeman/Times Books/Henry Holt & Co., 2014.
- [9] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons., 2021.
- [10] J. F. Hair, W. C. Black, and R. E. Anderson, *Multivariate Data Analysis (8th ed.)*. Pearson Education, 2020.
- [11] S. Solimun, A. A. R. Fernandes, I. Rahmawati, R. Isaskar, L. Muflikhah, and F. L. N. Rasyidah, "Cluster Integration Path Analysis to Model PT Pelindo II's Market Mapping," *International Journal of Circuits, Systems and Signal Processing*, vol. 15, pp. 1833–1841, Jan. 2022, doi: 10.46300/9106.2021.15.198.
- [12] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. 2014.
- [13] A. S. Devi, I. K. G. D. Putra, and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," *Lontar Komput. J. Ilm. Teknol. Inf.*, p. 185, Dec. 2015, doi: 10.24843/LKJITI.2015.v06.i03.p05.
- [14] S. Scitovski, "A density-based clustering algorithm for earthquake zoning," *Comput Geosci*, vol. 110, pp. 90–95, Jan. 2018, doi: 10.1016/j.cageo.2017.08.014.
- [15] P. N. Tan, M. Steinbach, and V. Kumar, *Data mining cluster analysis: basic concepts and algorithms*. 2013.
- [16] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," vol. 9, no. 1, pp. 57–68, 2012.
- [17] M. N. M. Ediyanto and N. Satyahadewi, "Pengklasifikasian karakteristik dengan metode K-Means cluster analysis.," *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 2, no. 2, 2013.
- [18] A. Aljohani, "Optimizing Patient Stratification in Healthcare: A Comparative Analysis of Clustering Algorithms for EHR Data," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, p. 173, Jul. 2024, doi: 10.1007/s44196-024-00568-8.
- [19] D. P. Isnarwaty and I. Irhamah, "Text Clustering pada Akun TWITTER Layanan Ekspedisi JNE, J&T, dan Pos Indonesia Menggunakan Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dan K-Means," *J. Sains Dan Seni ITS*, vol. 8, no. 2, pp. D137–D144, Feb. 2020, doi: 10.12962/j23373520.v8i2.49094.
- [20] M. Hahsler, M. Piekenbrock, and D. Doran, "Fast Density-Based Clustering with R," *J. Stat. Softw.*, vol. 91, no. 1, 2019, doi: 10.18637/jss.v091.i01.
- [21] A. J. Alifah, S. Saepudin, and C. Irawan, "Implementation of the K-Means Clustering Algorithm in Analyzing Public Satisfaction Regarding Public Services (studi Case: Balai Pengujian Standar Instrumen Tanaman Industri Dan Penyegar)".
- [22] Z. Mustofa and I. S. Suasana, "Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi," *J. Teknol. Inf. DAN Komun.*, vol. 9, no. 1,

- pp. 1–10, Apr. 2020, doi: 10.51903/jtikp.v9i1.162.
- [23] D. Adillah, N. Manurung, and A. Dermawan, “Implementation of K-Means Clustering Analysis to Determine Barriers to Online Learning Case Study: Swasta Yappendak Tinjowan Junior High School”.
- [24] F. Nurulhikmah and D. N. E. Abdi, “Classification of Foods Based on Nutritional Content Using K-Means and DBSCAN Clustering Methods,” *Teknika*, vol. 13, no. 3, pp. 481–486, Oct. 2024, doi: 10.34148/teknika.v13i3.1067.
- [25] P. A. Escudero, M. C. López González, and J. L. García Valdecabres, “Optimising Floor Plan Extraction: Applying DBSCAN and K-Means in Point Cloud Analysis of Valencia Cathedral,” *Heritage*, vol. 7, no. 10, pp. 5787–5799, Oct. 2024, doi: 10.3390/heritage7100272.