

# Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan

Yasir Hasan<sup>1</sup>

<sup>1</sup> Universitas Budi Darma, Jl. Sisingamangaraja No.338, Kota Medan, 20219, Sumatera Utara

## ARTICLE INFORMATION

Received: Februari, 2024

Revised: Maret, 2024

Available online: April 2024

## KEYWORDS

K-Means Clustering, DBSCAN Clustering, Silhouette Score, Davies-Bouldin\_Index, employee performance

## CORRESPONDENCE

Phone: 081375557689

E-mail: yasirhasan.kom@gmail.com

## ABSTRACT

Evaluation of Clustering results is a critical step in unsupervised data analysis. Clustering algorithms such as K-Means and DBSCAN are often used, but choosing the right algorithm for a dataset can be challenging. Application of Silhouette Score and Davies-Bouldin Index as internal evaluation metrics to evaluate Clustering results from K-Means and DBSCAN. The K-Means and DBSCAN methods were chosen because of their popularity and ability to handle various types of data. The Silhouette Score provides a measure of how well each data point is placed in its own Cluster compared to other Clusters, while the Davies-Bouldin Index evaluates how far away the Cluster is from the others. This research was carried out by implementing both algorithms on an experimental dataset, namely the employee performance dataset and comparing the evaluation results using these two metrics. The experimental results of both evaluation metrics provide useful insights in evaluating the quality of employee performance Clustering from K-Means and DBSCAN. Thus, the use of the Silhouette Score and Davies-Bouldin Index can be an effective guide in choosing an appropriate Clustering algorithm for a dataset without the need for ground truth labels.

## PENDAHULUAN

Evaluasi hasil *Clustering* adalah tahap penting dalam analisis data tanpa supervisi[1], di mana algoritma *Clustering* seperti K-Means dan DBSCAN sering digunakan untuk mengelompokkan data ke dalam kelompok yang memiliki kesamaan tertentu[2], [3]. Namun, memilih algoritma yang sesuai untuk suatu dataset dapat menjadi tantangan karena setiap algoritma memiliki kelebihan dan kekurangannya sendiri. Oleh karena itu, penting untuk memiliki metrik evaluasi yang dapat memberikan pemahaman yang mendalam tentang kualitas hasil *Clustering*[4].

Metode evaluasi internal seperti *Silhouette Score* dan *Davies-Bouldin Index* telah menjadi fokus penelitian dalam evaluasi *Clustering*. *Silhouette Score* memberikan ukuran seberapa baik setiap titik data ditempatkan dalam *Cluster* mereka sendiri dibandingkan dengan *Cluster* lain, sementara *Davies-Bouldin Index* mengevaluasi seberapa jauh *Cluster* tersebut berada dari yang lain[5], [6]. Kedua metrik ini memberikan informasi yang berharga tentang kualitas dan kehomogenan *Cluster* yang dihasilkan oleh algoritma *Clustering*[7].

Beberapa penelitian sebelumnya telah mengkaji penggunaan *Silhouette Score* dan *Davies-Bouldin Index* dalam evaluasi *Clustering*. Contohnya, penelitian oleh Peneliti Kholiq Amrulloh menunjukkan bahwa *Silhouette Score* dapat digunakan untuk mengevaluasi kualitas pembagian data menjadi *Cluster* transaksi penjualan toko retail[7], sementara penelitian oleh Peneliti Firman Ashari dkk menyoroti keefektifan *Davies-Bouldin Index* dan metode-metode lainnya dalam mengevaluasi seberapa jauh *Cluster* tersebut berada dari *Cluster* lainnya dalam klasifikasi wilayah terdampak banjir di Jakarta. Namun penggunaan kedua metrik ini secara khusus dalam konteks algoritma K-Means dan DBSCAN masih perlu dieksplorasi lebih lanjut [4], [8], [9].

Dalam konteks pemilihan algoritma *Clustering* yang tepat, karakteristik data juga memainkan peran penting. Algoritma seperti K-Means dan DBSCAN dipilih karena popularitas dan kemampuannya dalam menangani berbagai jenis data. Namun, untuk memahami bagaimana metrik evaluasi internal seperti *Silhouette Score* dan *Davies-Bouldin Index* berkinerja pada kedua algoritma ini, dilakukan penelitian terhadap dataset yang digunakan sebagai data uji coba adalah dataset kinerja karyawan.

Oleh karena itu, penelitian ini bertujuan untuk mengisi celah pengetahuan tersebut dengan mengkaji penggunaan *Silhouette Score* dan *Davies-Bouldin Index* sebagai metrik evaluasi internal untuk mengevaluasi hasil *Clustering* dari algoritma K-Means dan DBSCAN. Dataset yang digunakan adalah dataset kinerja karyawan untuk alternatif *Clustering* sebagai data percobaan. Dengan membandingkan hasil evaluasi menggunakan kedua metrik tersebut, diharapkan penelitian ini dapat memberikan wawasan yang lebih baik dalam memilih algoritma *Clustering* yang sesuai untuk suatu dataset tanpa kebutuhan akan label *ground truth*.

Tujuan Khusus Penelitian: (1). Menganalisis secara mendalam kualitas hasil *Clustering* yang dihasilkan oleh algoritma K-Means dan DBSCAN pada dataset kinerja karyawan jenis dataset tanpa supervisi. (2). Implementasikan *Silhouette Score* dan *Davies-Bouldin Index* sebagai metrik evaluasi internal untuk mengevaluasi hasil *Clustering* dari K-Means dan DBSCAN. (3). Membandingkan kinerja metrik evaluasi internal (*Silhouette Score* dan *Davies-Bouldin Index*) dalam mengevaluasi kualitas *Clustering* antara algoritma K-Means dan DBSCAN. (4). Menyelidiki hubungan antara nilai *Silhouette Score* dan *Davies-Bouldin Index* dengan karakteristik struktural dari hasil *Clustering* dari K-Means dan DBSCAN. (5). Menyajikan panduan yang efektif dalam memilih algoritma *Clustering* yang paling sesuai untuk suatu dataset tanpa memerlukan label *ground truth* berdasarkan hasil evaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index*. (6). Menyediakan pemahaman yang lebih mendalam tentang keunggulan dan kelemahan dari kedua algoritma *Clustering* (K-Means dan DBSCAN) dalam konteks evaluasi *Clustering* tanpa supervisi menggunakan metode evaluasi internal yang dipilih.

Penelitian ini memiliki urgensi yang tinggi dalam konteks analisis data tanpa supervisi dan penggunaan algoritma *Clustering* seperti K-Means dan DBSCAN. Evaluasi hasil *Clustering* merupakan langkah kritis dalam proses analisis data, karena dapat memberikan wawasan yang mendalam tentang struktur data yang tersembunyi. Dalam lingkungan yang semakin kompleks dan beragam jenis dataset, pemilihan algoritma *Clustering* yang tepat menjadi krusial untuk memastikan hasil *Clustering* yang optimal[4], [10], [11].

Algoritma *Clustering* seperti K-Means dan DBSCAN sering menjadi pilihan utama dalam berbagai aplikasi karena popularitas dan kemampuannya dalam menangani berbagai jenis data. Namun, tantangan muncul dalam memilih algoritma yang paling sesuai untuk dataset yang spesifik. Oleh karena itu, penelitian ini memberikan kontribusi signifikan dengan mengusulkan penerapan *Silhouette Score* dan *Davies-Bouldin Index* sebagai metrik evaluasi internal untuk mengevaluasi hasil *Clustering* dari K-Means dan DBSCAN.

### METODE PENELITIAN

Metode penelitian Pengukuran *Silhouette Score* dan *Davies-Bouldin Index* untuk Evaluasi Hasil *Cluster* dari K-Means dan DBSCAN terdiri dari beberapa tahapan sebagai berikut:

1. Pemilihan Dataset Percobaan:  
Memilih dataset percobaan yaitu kinerja karyawan yang representatif dan beragam untuk mengevaluasi kualitas *Clustering* dari algoritma K-Means dan DBSCAN.
2. Implementasi Algoritma K-Means dan DBSCAN:
  - a. Mengimplementasikan algoritma K-Means dan DBSCAN menggunakan bahasa pemrograman atau perangkat lunak analisis data yang tersedia.
  - b. Memastikan bahwa implementasi algoritma sesuai dengan definisi dan prinsip operasional yang telah ditetapkan.
3. Pengukuran *Silhouette Score* dan *Davies-Bouldin Index*:
  - a. Menghitung *Silhouette Score* dan *Davies-Bouldin Index* untuk setiap *Cluster* yang dihasilkan oleh algoritma K-Means dan DBSCAN.
  - b. Melakukan perhitungan menggunakan rumus yang telah dijelaskan sebelumnya, dengan memperhatikan perhitungan jarak antara titik data dan *Centroid Cluster*.
4. Analisis dan Evaluasi Hasil:
  - a. Menganalisis hasil pengukuran *Silhouette Score* dan *Davies-Bouldin Index* untuk mengevaluasi kualitas *Clustering* dari K-Means dan DBSCAN.
  - b. Membandingkan performa keduanya dengan memperhatikan perbedaan dalam nilai metrik evaluasi, serta mengidentifikasi kelebihan dan kelemahan masing-masing algoritma.
5. Interpretasi dan Kesimpulan:
  - a. Menginterpretasi hasil evaluasi metrik *Silhouette Score* dan *Davies-Bouldin Index* untuk mendapatkan pemahaman yang lebih baik tentang kualitas *Clustering* dari K-Means dan DBSCAN.
  - b. Menyimpulkan temuan dan implikasi dari hasil evaluasi, serta memberikan rekomendasi atau panduan dalam memilih algoritma *Clustering* yang sesuai untuk suatu dataset tanpa kebutuhan akan label *ground truth*.

### HASIL DAN PEMBAHASAN

Evaluasi hasil *Clustering* dari algoritma K-Means dan DBSCAN yang berbeda tentunya memberikan performa yang berbeda. Evaluasi dilakukan dengan menggunakan dua metrik utama, yaitu *Silhouette Score* dan *Davies-Bouldin Index*. Untuk memberikan pemahaman yang jelas, dataset yang digunakan berisi 20 karyawan dengan lima atribut berbeda: Disiplin (at1), Performa(at2), Tanggung Jawab (at3), Komunikasi (at4), dan Kerjasama (at5).

Tabel 1. Dataset alternatif kinerja karyawan

No.	Nama	Disiplin (at1)	Performa(at2)	Tanggung Jawab (at3)	Komunikasi (at4)	Kerjasama (at5)
1	Karyawan1	6	7	5	7	7
2	Karyawan2	7	8	7	8	7
3	Karyawan3	8	7	6	7	7
4	Karyawan4	6	6	7	7	6
5	Karyawan5	8	8	8	8	7
6	Karyawan6	8	5	6	7	7
7	Karyawan7	7	7	8	7	7
8	Karyawan8	6	7	7	8	7
9	Karyawan9	7	8	7	8	8
10	Karyawan10	6	5	6	6	8
11	Karyawan11	7	8	6	5	6
12	Karyawan12	6	6	6	7	7
13	Karyawan13	7	8	7	7	6
14	Karyawan14	5	7	7	8	7
15	Karyawan15	8	6	7	7	7
16	Karyawan16	6	8	6	7	8

No.	Nama	Disiplin (at1)	Performa(at2)	Tanggung Jawab (at3)	Komunikasi (at4)	Kerjasama (at5)
17	Karyawan17	7	8	7	8	8
18	Karyawan18	6	7	5	7	6
19	Karyawan19	7	7	7	6	7
20	Karyawan20	7	8	8	7	7

Penjelasan nilai data dan variabel-variabel input sangat penting sebelum melakukan penerapan algoritma. Penjelasan nilai dari tabel di atas sangat sederhana, yaitu sangat baik = 8, baik = 7, cukup baik = 6, kurang = 5. Berikutnya untuk keterangan *Cluster 1(C1) = Excellent*, *Cluster 2(C2) = Profesional*, dan *Cluster 3(C3) = Need motivation*. Untuk data pada tabel 2 tidak dilakukan normalisasi dan langsung digunakan penerapan K-Means maupun DBSCAN. Selain itu juga beberapa variabel yang digunakan dalam beberapa metode dan algoritma didefinisikan sebagai berikut:

K-Means: Jumlah  $k = 3$ , yaitu C1, C2, C3

Maksimal iterasi = 50 kali

DBSCAN: Jumlah  $k = 3$ , yaitu C1, C2, C3

epsilon ( $\epsilon$ ) = 2.5

Jumlah min sampel minS = 5

### 1. Clustering K-Means

Penerapan  $k = 3$  untuk yaitu hasil *Cluster 1*, *Cluster 2*, dan *Cluster 3*. Namun sebelumnya harus menentukan *Centroid* awal dari dataset yang digunakan. Teknik penentuan *Centroid* awal digunakan teknik random. *Centroid* awal yang dipilih adalah c1 dari data karyawan1, c2 dari data karyawan16, dan c3 dari data karyawan17.

Tabel 2. *Centroid* awal

<i>Centroid</i> Awal	Cat1	Cat2	Cat3	Cat4	Cat5
c1	6	7	5	7	7
c2	6	8	6	7	8
c3	7	8	7	8	8

1. Tentukan jarak *Euclidean* dari setiap titik ke *Centroid*. Jarak *Euclidean* didapatkan dari akar ( $\sqrt{\phantom{x}}$ ) pengurangan nilai dari setiap titik dataset (p) ke setiap nilai *Centroid* awal (q) yang diberikan pangkat sesuai dengan rumusnya, maka ditulis:

$$Distance = \sqrt{(at1 - Cat1)^2 + (at2 - Cat2)^2 + (at3 - Cat3)^2 + (at4 - Cat4)^2 + (at5 - Cat5)^2}$$

$$\text{Jarak Karyawan1 ke c1} = \sqrt{(6 - 6)^2 + (7 - 7)^2 + (5 - 5)^2 + (7 - 7)^2 + (7 - 7)^2} = 0$$

$$\text{Jarak Karyawan1 ke c2} = \sqrt{(6 - 6)^2 + (7 - 8)^2 + (5 - 6)^2 + (7 - 7)^2 + (7 - 8)^2} = 1.73$$

$$\text{Jarak Karyawan1 ke c3} = \sqrt{(6 - 7)^2 + (7 - 8)^2 + (5 - 7)^2 + (7 - 8)^2 + (7 - 8)^2} = 2.83$$

Contoh perhitungan jarak *Euclidean* titik Karyawan1 ke *Centroid* awal di atas, juga berlaku untuk titik Karyawan2 hingga Karyawan20. Penentuan *Cluster* dari penentuan awal dilakukan dengan cara membandingkan nilai ketiga jarak tersebut. Hasil ke c1 = 0, hasil ke c2 = 1.73, dan hasil ke c3 = 2.83, dari ketiga nilai tersebut Hasil ke c1 = 0 adalah nilai terkecil dipilih sebagai *Cluster*. Penentuan *Cluster* dengan membandingkan nilai terkecil atau dari *Centroid* terdekat dapat dilihat pada tabel berikut.

Tabel 3. Penentuan Jarak *Euclidean* dan *Cluster*

Nama	Hasil ke c1	Hasil ke c2	Hasil ke c3	<i>Cluster</i>
Karyawan1	0.00	1.73	2.83	C1
Karyawan2	2.65	2.00	1.00	C3
Karyawan3	2.24	2.45	2.24	C1
Karyawan4	2.45	3.00	3.16	C1
Karyawan5	3.87	3.16	1.73	C3
Karyawan6	3.00	3.74	3.61	C1
Karyawan7	3.16	2.65	2.00	C3
Karyawan8	2.24	2.00	1.73	C3
Karyawan9	2.83	1.73	0.00	C3
Karyawan10	2.65	3.16	3.87	C1
Karyawan11	2.83	3.00	3.74	C1

Nama	Hasil ke c1	Hasil ke c2	Hasil ke c3	Cluster
Karyawan12	1.41	2.24	2.83	C1
Karyawan13	2.65	2.45	2.24	C3
Karyawan14	2.45	2.24	2.45	C2
Karyawan15	3.00	3.16	2.65	C3
Karyawan16	1.73	0.00	1.73	C2
Karyawan17	2.83	1.73	0.00	C3
Karyawan18	1.00	2.45	3.32	C1
Karyawan19	2.45	2.24	2.45	C2
Karyawan20	3.32	2.45	1.73	C3

2. Ekstraksi hasil *Cluster* awal untuk penentuan *Centroid* baru

Kelompok ketiga *Cluster* awal dipisahkan sehingga menampilkan anggota masing-masing.

*Cluster*1: Karyawan1, Karyawan3, Karyawan4, Karyawan6, Karyawan10, Karyawan11, Karyawan12, Karyawan18

*Cluster*2: Karyawan14, Karyawan16, Karyawan19

*Cluster*3: Karyawan2, Karyawan5, Karyawan7, Karyawan8, Karyawan9, Karyawan13, Karyawan15, Karyawan17, Karyawan20

Dari ketiga kelompok di atas, setiap nilai pada *Cluster* pada setiap titik karyawan dijumlahkan dan dibagi dengan rata-rata jumlah untuk membuat tiga *Centroid* baru.

*Cluster*1: 8 anggota karyawan

$at1 = (6+8+6+8+6+7+6+6) / 8 = 6.625$ ,  $at2 = (7+7+6+5+5+8+6+7) / 8 = 6.375$ ,  $at3 = (5+6+7+6+6+6+6+5) / 8 = 5.875$   
 $at4 = (7+7+7+7+6+5+7+7) / 8 = 6.625$ ,  $at5 = (7+7+6+7+8+6+7+6) / 8 = 6.75$ ,

*Cluster*2: 3 anggota karyawan

$at1 = (5+6+7) / 3 = 6$ ,  $at2 = (7+8+7) / 3 = 7.3$ ,  $at3 = (7+6+7) / 3 = 6.67$ ,  $at4 = (8+7+6) / 3 = 7$ ,  $at5 = (7+8+7) / 3 = 7.33$

*Cluster*3: 9 anggota karyawan

$at1 = (7+8+7+6+7+7+8+7+7) / 9 = 7.11$ ,  $at2 = (8+8+7+7+8+8+6+8+8) / 9 = 7.55$ ,  $at3 = (7+8+8+7+7+7+7+7+8) / 9 = 7.33$ ,  
 $at4 = (8+8+7+8+8+7+7+8+7) / 9 = 7.55$ ,  $at5 = (7+7+7+7+8+6+7+8+7) / 9 = 7.11$

Hasil semua dari *Cluster* awal tersebut diunakan untuk *Centroid* baru yang dilanjutkan perhitungan algoritmanya.

Tabel 4. *Centroid* baru untuk iterasi 1

<i>Centroid</i> baru	Cat1	Cat2	Cat3	Cat4	Cat5
c1	6.62	6.37	5.87	6.62	6.75
c2	6	7.3	6.67	7	7.33
c3	7.11	7.55	7.33	7.55	7.11

3. Iterasi dan Konvergensi

Iterasi merupakan pengerjaan ulang algoritma K-Means seperti dari awal perhitungan namun menggunakan *Centroid* baru. Sedangkan konvergensi menyatakan iterasi perhitungan berhenti. Penandaan konvergensi terjadi jika dua iterasi terakhir memiliki jumlah anggota *Cluster* sama. Jika tidak ada kesamaan maka iterasi dilakukan berulang hingga maksimal iterasi sebanyak 50 iterasi. Selain itu konvergensi dihentikan jika jumlah iterasi tercukupi namun jumlah anggota *Cluster* belum sama. Pada perhitungan algoritma ini konvergensi terjadi pada jumlah *Cluster* pada iterasi 2 sama dengan jumlah *Cluster* pada iterasi 1.

Tabel 5. Konvergensi dari iterasi 1 ke iterasi 2

Klaster A	C1	C3	C1	C1	C3	C1	C3	C3	C3	C1	C1	C1	C3	C2	C3	C2	C3	C1	C2	C3
Klaster iterasi1	C1	C3	C1	C1	C3	C1	C3	C2	C3	C1	C1	C1	C3	C2	C1	C2	C3	C1	C1	C3
Klaster iterasi2	C1	C3	C1	C1	C3	C1	C3	C2	C3	C1	C1	C1	C3	C2	C1	C2	C3	C1	C1	C3

4. Hasil *Cluster* K-means

Hasil *Cluster* k-means diterapkan

Tabel 6. Hasil *Cluster* K-Means

No	Nama	Disiplin (at1)	Performa (at2)	Tanggung Jawab (at3)	Komunikasi (at4)	Kerjasama (at5)	Cluster	Keterangan Cluster
1	Karyawan1	6	7	5	7	7	C1	Excellent
2	Karyawan2	7	8	7	8	7	C3	Need motivation

No	Nama	Disiplin (at1)	Performa (at2)	Tanggung Jawab (at3)	Komunikasi (at4)	Kerjasama (at5)	Cluster	Keterangan Cluster
3	Karyawan3	8	7	6	7	7	C1	Excellent
4	Karyawan4	6	6	7	7	6	C1	Excellent
5	Karyawan5	8	8	8	8	7	C3	Need motivation
6	Karyawan6	8	5	6	7	7	C1	Excellent
7	Karyawan7	7	7	8	7	7	C3	Need motivation
8	Karyawan8	6	7	7	8	7	C2	Professional
9	Karyawan9	7	8	7	8	8	C3	Need motivation
10	Karyawan10	6	5	6	6	8	C1	Excellent
11	Karyawan11	7	8	6	5	6	C1	Excellent
12	Karyawan12	6	6	6	7	7	C1	Excellent
13	Karyawan13	7	8	7	7	6	C3	Need motivation
14	Karyawan14	5	7	7	8	7	C2	Professional
15	Karyawan15	8	6	7	7	7	C1	Excellent
16	Karyawan16	6	8	6	7	8	C2	Professional
17	Karyawan17	7	8	7	8	8	C3	Need motivation
18	Karyawan18	6	7	5	7	6	C1	Excellent
19	Karyawan19	7	7	7	6	7	C1	Excellent
20	Karyawan20	7	8	8	7	7	C3	Need motivation

## 1.2 Clustering DBSCAN

Clustering dengan DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) mampu mengidentifikasi Cluster dan noise dalam data. Data dengan predikat *noise* dieliminasi dianggap tidak layak dimasukkan ke dalam Cluster. DBSCAN mengidentifikasi wilayah dengan kepadatan tinggi dalam ruang data. Langkah-langkah DBSCAN:

### 1. Membuat tabel matriks jarak

Tabel matriks jarak dibuat untuk melihat kepadatan data dalam wilayah radius menggunakan jarak *Euclidean*. Untuk setiap p adalah titik data karyawan yang akan dikurangi dengan q yaitu 20 titik data karyawan yang ada. Cara ini berlaku untuk setiap p titik karyawan. Berikut perhitungan awal jarak *Euclidean* DBSCAN.

$$\text{Jarak pKaryawan1 dengan qKaryawan1} = \sqrt{(6-6)^2 + (7-7)^2 + (5-5)^2} = 0$$

$$\text{Jarak pKaryawan1 dengan qKaryawan2} = \sqrt{(6-7)^2 + (7-8)^2 + (5-7)^2} = 2.65$$

$$\text{Jarak pKaryawan1 dengan qKaryawan3} = \sqrt{(6-8)^2 + (7-7)^2 + (5-6)^2} = 2.24$$

$$\text{Jarak pKaryawan1 dengan qKaryawan4} = \sqrt{(6-6)^2 + (7-6)^2 + (5-7)^2} = 2.45$$

$$\text{Jarak pKaryawan1 dengan qKaryawan5} = \sqrt{(6-8)^2 + (7-8)^2 + (5-8)^2} = 3.87$$

$$\text{Jarak pKaryawan1 dengan qKaryawan6} = \sqrt{(6-8)^2 + (7-5)^2 + (5-6)^2} = 3$$

$$\text{Jarak pKaryawan1 dengan qKaryawan7} = \sqrt{(6-7)^2 + (7-7)^2 + (5-8)^2} = 3.16$$

$$\text{Jarak pKaryawan1 dengan qKaryawan8} = \sqrt{(6-6)^2 + (7-7)^2 + (5-7)^2} = 2.24$$

$$\text{Jarak pKaryawan1 dengan qKaryawan9} = \sqrt{(6-7)^2 + (7-8)^2 + (5-7)^2} = 2.83$$

$$\text{Jarak pKaryawan1 dengan qKaryawan10} = \sqrt{(6-6)^2 + (7-5)^2 + (5-6)^2} = 2.65$$

$$\text{Jarak pKaryawan1 dengan qKaryawan11} = \sqrt{(6-7)^2 + (7-8)^2 + (5-6)^2} = 2.83$$

$$\text{Jarak pKaryawan1 dengan qKaryawan12} = \sqrt{(6-6)^2 + (7-6)^2 + (5-6)^2} = 1.14$$

$$\text{Jarak pKaryawan1 dengan qKaryawan13} = \sqrt{(6-7)^2 + (7-8)^2 + (5-7)^2} = 2.65$$

$$\text{Jarak pKaryawan1 dengan qKaryawan14} = \sqrt{(6-5)^2 + (7-7)^2 + (5-7)^2} = 2.45$$

$$\text{Jarak pKaryawan1 dengan qKaryawan15} = \sqrt{(6-8)^2 + (7-6)^2 + (5-7)^2} = 3$$

$$\text{Jarak pKaryawan1 dengan qKaryawan16} = \sqrt{(6-6)^2 + (7-8)^2 + (5-6)^2} = 1.73$$

$$\text{Jarak pKaryawan1 dengan qKaryawan17} = \sqrt{(6-7)^2 + (7-8)^2 + (5-7)^2} = 2.83$$

$$\text{Jarak pKaryawan1 dengan qKaryawan18} = \sqrt{(6-6)^2 + (7-7)^2 + (5-5)^2} = 1$$

$$\text{Jarak pKaryawan1 dengan qKaryawan19} = \sqrt{(6-7)^2 + (7-7)^2 + (5-7)^2} = 2.45$$

$$\text{Jarak pKaryawan1 dengan qKaryawan20} = \sqrt{(6-7)^2 + (7-8)^2 + (5-8)^2} = 3.32$$

Perhitungan untuk jarak *Euclidean* titik data pkaryawan2 ke setiap qkaryawan dan berlaku juga ke semua titik data pkaryawan. Berikut adalah matriks jarak:

Tabel 7. Matriks jarak *Euclidean*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	2.65	2.24	2.45	3.87	3	3.16	2.24	2.83	2.65	2.83	1.41	2.65	2.45	3	1.73	2.83	1	2.45	3.32
2	2.65	0	2	2.65	1.41	3.46	1.73	1.41	1	4	3.32	2.65	1.41	2.24	2.45	2	1	2.83	2.24	1.41
3	2.24	2	0	2.65	2.45	2	2.24	2.45	2.24	3.16	2.65	2.24	2	3.32	1.41	2.45	2.24	2.45	1.73	2.45
4	2.45	2.65	2.65	0	3.32	2.65	2	1.73	3.16	2.65	3.16	1.41	2.24	2	2.24	3	3.16	2.24	2	2.65
5	3.87	1.41	2.45	3.32	0	3.74	1.73	2.45	1.73	4.69	3.87	3.61	2	3.32	2.45	3.16	1.73	4	2.65	1.41
6	3	3.46	2	2.65	3.74	0	3	3.16	3.61	2.45	3.87	2.24	3.46	3.87	1.41	3.74	3.61	3.16	2.65	3.74
7	3.16	1.73	2.24	2	1.73	3	0	1.73	2	3.32	3.16	2.45	1.73	2.45	1.73	2.65	2	3.32	1.41	1
8	2.24	1.41	2.45	1.73	2.45	3.16	1.73	0	1.73	3.16	3.61	1.73	2	1	2.45	2	1.73	2.45	2.24	2
9	2.83	1	2.24	3.16	1.73	3.61	2	1.73	0	3.87	3.74	2.83	2.24	2.45	2.65	1.73	0	3.32	2.45	1.73
10	2.65	4	3.16	2.65	4.69	2.45	3.32	3.16	3.87	0	3.87	1.73	4	3.32	2.83	3.16	3.87	3.16	2.65	4
11	2.83	3.32	2.65	3.16	3.87	3.87	3.16	3.61	3.74	3.87	0	3.16	2.24	4	3.32	3	3.74	2.65	2	3
12	1.41	2.65	2.24	1.41	3.61	2.24	2.45	1.73	2.83	1.73	3.16	0	2.65	2	2.24	2.24	2.83	1.73	2	3
13	2.65	1.41	2	2.24	2	3.46	1.73	2	2.24	4	2.24	2.65	0	2.65	2.45	2.45	2.24	2.45	1.73	1.41
14	2.45	2.24	3.32	2	3.32	3.87	2.45	1	2.45	3.32	4	2	2.65	0	3.32	2.24	2.45	2.65	2.83	2.65
15	3	2.45	1.41	2.24	2.45	1.41	1.73	2.45	2.65	2.83	3.32	2.24	2.45	3.32	0	3.16	2.65	3.16	1.73	2.45
16	1.73	2	2.45	3	3.16	3.74	2.65	2	1.73	3.16	3	2.24	2.45	2.24	3.16	0	1.73	2.45	2.24	2.45
17	2.83	1	2.24	3.16	1.73	3.61	2	1.73	0	3.87	3.74	2.83	2.24	2.45	2.65	1.73	0	3.32	2.45	1.73
18	1	2.83	2.45	2.24	4	3.16	3.32	2.45	3.32	3.16	2.65	1.73	2.45	2.65	3.16	2.45	3.32	0	2.65	3.46
19	2.45	2.24	1.73	2	2.65	2.65	1.41	2.24	2.45	2.65	2	2	1.73	2.83	1.73	2.24	2.45	2.65	0	1.73
20	3.32	1.41	2.45	2.65	1.41	3.74	1	2	1.73	4	3	3	1.41	2.65	2.45	2.45	1.73	3.46	1.73	0

## 2. Parameter Input

Menerapkan Parameter input  $\epsilon$  (*epsilon*) = 2.5 berguna memilih nilai jarak maksimal  $\geq 2.5$  dari matriks jarak *Euclidean*. pada tabel 7 titik 1 terdapat 8 nilai dalam radius  $\epsilon = 2.5$  dan nilai tersebut diurutkan. Lihat tabel 8 dan 9, minS = 5 berguna untuk menentukan titik inti dari data dengan jumlah minimum sebanyak 5, pada tabel tersebut ditandai dengan batas minS berwarna biru muda.

Tabel 8. Pemilihan  $\geq 2.5$  titik 1 sampai 10

No.	titik1	No.	titik2	No.	titik3	No.	titik4	No.	titik5	No.	titik6	No.	titik7	No.	titik8	No.	titik9	No.	titik10
18	1.00	9	1.00	15	1.41	12	1.41	2	1.41	15	1.41	20	1.00	14	1.00	17	0.00	12	1.73
12	1.41	17	1.00	19	1.73	8	1.73	20	1.41	3	2.00	19	1.41	2	1.41	2	1.00	6	2.45
16	1.73	5	1.41	2	2.00	7	2.00	7	1.73	12	2.24	2	1.73	4	1.73	5	1.73		
3	2.24	8	1.41	6	2.00	14	2.00	9	1.73	10	2.45	5	1.73	7	1.73	8	1.73		
8	2.24	13	1.41	13	2.00	19	2.00	17	1.73			8	1.73	9	1.73	16	1.73		
4	2.45	20	1.41	1	2.24	13	2.24	13	2.00			13	1.73	12	1.73	20	1.73		
14	2.45	7	1.73	7	2.24	15	2.24	3	2.45			15	1.73	17	1.73	7	2.00		
19	2.45	3	2.00	9	2.24	18	2.24	8	2.45			4	2.00	13	2.00	3	2.24		
		16	2.00	12	2.24	1	2.45	15	2.45			9	2.00	16	2.00	13	2.24		
		14	2.24	17	2.24	2						17	2.00	20	2.00	14	2.45		
		19	2.24	5	2.45	3						3	2.24	1	2.24	19	2.45		
		15	2.45	8	2.45	4						12	2.45	19	2.24				
		10	4.00	16	2.45	5						14	2.45	3	2.45				
				18	2.45	6								5	2.45				
				20	2.45	9								15	2.45				
						10								18	2.45				
						11													
						16													
						17													
						20													

Tabel 9. Pemilihan nilai yang  $\geq 2.5$  titik lanjutan 11 sampai 20

[illegible]

### 3. Identifikasi Titik Inti

Identifikasi titik inti, titik tepi, dan titik noise menggunakan  $\epsilon = 2.5$  dan  $\text{minS} = 5$

Titik Inti: Titik yang memiliki setidaknya 5 tetangga dalam radius 2.5.

Jika titik memiliki 5 atau lebih tetangga dalam radius 2.5, maka titik tersebut adalah titik inti.

Titik Tepian: Titik yang berada dalam radius 2.5 namun berjumlah kurang dari 5 yaitu 4. Pada tabel 9 tidak ada titik tepi.

Titik Noise: Titik yang tidak masuk dalam dua kategori di atas berjumlah  $<$  dari titik tepian yaitu  $< 4$  atau  $< \min S - 1$ .

Tabel 10. Penentuan titik inti

No	Nama	Tetangga dalam Radius 2.5	Kategori
1	Karyawan1	18, 12, 16, 3, 8, 4, 14, 19	Titik Inti
2	Karyawan2	9, 17, 5, 8, 13, 20, 7, 3, 16, 14, 19, 15, 10	Titik Inti
3	Karyawan3	15, 19, 2, 6, 13, 1, 7, 9, 12, 17, 5, 8, 16, 18, 20	Titik Inti
4	Karyawan4	12, 8, 7, 14, 19, 13, 15, 18, 1	Titik Inti
5	Karyawan5	2, 20, 7, 9, 17, 13, 3, 8, 15, 18	Titik Inti
6	Karyawan6	15, 3, 12, 10	Titik Noise
7	Karyawan7	20, 19, 2, 5, 8, 13, 15, 4, 9, 17, 3, 12, 14	Titik Inti
8	Karyawan8	14, 2, 4, 7, 9, 12, 17, 13, 16, 20, 1, 19, 3, 5, 15, 18	Titik Inti
9	Karyawan9	17, 2, 5, 8, 16, 20, 7, 3, 13, 14, 19	Titik Inti
10	Karyawan10	12, 6	Titik Noise
11	Karyawan11	19, 13	Titik Noise
12	Karyawan12	1, 4, 8, 10, 18, 14, 19, 3, 6, 15, 16, 7	Titik Inti
13	Karyawan13	2, 20, 7, 19, 3, 5, 8, 4, 9, 11, 17, 15, 16, 18	Titik Inti
14	Karyawan14	8, 4, 12, 2, 16, 1, 7, 9, 17	Titik Inti
15	Karyawan15	3, 6, 7, 19, 4, 12, 2, 5, 8, 13, 20, 17	Titik Inti
16	Karyawan16	1, 9, 17, 2, 8, 12, 14, 19, 3, 13, 18, 20	Titik Inti
17	Karyawan17	9, 2, 5, 8, 16, 20, 7, 3, 13, 14, 19	Titik Inti
18	Karyawan18	1, 12, 4, 3, 8, 13, 16, 5	Titik Inti
19	Karyawan19	7, 3, 13, 15, 20, 4, 11, 12, 2, 8, 16, 1, 9, 17	Titik Inti



No	Nama	Tetangga dalam Radius 2.5	Kategori
20	Karyawan20	7, 2, 5, 13, 9, 17, 19, 8, 3, 15, 16	Titik Inti

#### 4. Membentuk Cluster

Cluster dibentuk berdasarkan hubungan antar titik inti dan tetangganya dalam radius 2.5. Penyatuan cluster terjadi karena banyak titik inti dan tetangga yang saling beririsan. Gunakan informasi titik inti untuk membentuk cluster. Setiap titik inti akan membentuk cluster dengan tetangganya.

Karyawan1 memiliki 8 tetangga dalam radius 2.5 (18, 12, 16, 3, 8, 4, 14, 19). Jika dua titik inti memiliki tetangga yang sama atau salah satu titik inti adalah tetangga dari titik inti lainnya, maka cluster mereka akan disatukan.

Proses Penyatuan Cluster, Cluster 1 dibentuk dari titik inti yang saling terhubung dan tetangganya.

Cluster 1 dimulai dari Karyawan1:

Karyawan1 : 18, 12, 16, 3, 8, 4, 14, 19.

Karyawan2 : 9, 17, 5, 8, 13, 20, 7, 3, 16, 14, 19, 15, 10.

Karyawan3 : 15, 19, 2, 6, 13, 1, 7, 9, 12, 17, 5, 8, 16, 18, 20.

Karyawan1 dan Karyawan2 berbagi beberapa tetangga: 3, 8, 16, 19, 14 namun tidak cukup kuat dalam hubungan rantai untuk membentuk satu cluster

Karyawan1 dan Karyawan3 berbagi beberapa tetangga: 16, 3, 8, 19, 12, 18.

Karyawan3 berhubungan dengan Karyawan1 melalui tetangganya.

Berikutnya Karyawan1 diperiksa ke Karyawan lainnya Sehingga didapat untuk Cluster 1:

Tabel 11. Cluster 1

No	Nama	Tetangga dalam Radius 2.5
1	Karyawan1	18, 12, 16, 3, 8, 4, 14, 19
3	Karyawan3	15, 19, 2, 6, 13, 1, 7, 9, 12, 17, 5, 8, 16, 18, 20
4	Karyawan4	12, 8, 7, 14, 19, 13, 15, 18, 1
12	Karyawan12	1, 4, 8, 10, 18, 14, 19, 3, 6, 15, 16, 7
14	Karyawan14	8, 4, 12, 2, 16, 1, 7, 9, 17
18	Karyawan18	1, 12, 4, 3, 8, 13, 16, 5
19	Karyawan19	7, 3, 13, 15, 20, 4, 11, 12, 2, 8, 16, 1, 9, 17

Cluster 2 dibentuk dari titik inti lainnya yang memiliki cukup banyak tetangga dalam radius 2.5.

Cluster 2 dimulai dari Karyawan2:

Karyawan2: 9, 17, 5, 8, 13, 20, 7, 3, 16, 14, 19, 15, 10.

Karyawan5: 2, 20, 7, 9, 17, 13, 3, 8, 15, 18.

Karyawan5 berhubungan dengan Karyawan2 melalui tetangganya

Tabel 12. Cluster 2

No	Nama	Tetangga dalam Radius 2.5
2	Karyawan2	9, 17, 5, 8, 13, 20, 7, 3, 16, 14, 19, 15, 10
5	Karyawan5	2, 20, 7, 9, 17, 13, 3, 8, 15, 18
7	Karyawan7	20, 19, 2, 5, 8, 13, 15, 4, 9, 17, 3, 12, 14
8	Karyawan8	14, 2, 4, 7, 9, 12, 17, 13, 16, 20, 1, 19, 3, 5, 15, 18
9	Karyawan9	17, 2, 5, 8, 16, 20, 7, 3, 13, 14, 19
13	Karyawan13	2, 20, 7, 19, 3, 5, 8, 4, 9, 11, 17, 15, 16, 18
17	Karyawan17	9, 2, 5, 8, 16, 20, 7, 3, 13, 14, 19
20	Karyawan20	7, 2, 5, 13, 9, 17, 19, 8, 3, 15, 16

Cluster 3 dibentuk dari titik inti yang tersisa dengan tetangganya.

Cluster 3 dari Karyawan15 dan Karyawan16:

Karyawan15: 3, 6, 7, 19, 4, 12, 2, 5, 8, 13, 20, 17.

Karyawan16: 1, 9, 17, 2, 8, 12, 14, 19, 3, 13, 18, 20

Tabel 13. Cluster C3

No	Nama	Tetangga dalam Radius 2.5
15	Karyawan15	3, 6, 7, 19, 4, 12, 2, 5, 8, 13, 20, 17
16	Karyawan16	1, 9, 17, 2, 8, 12, 14, 19, 3, 13, 18, 20

Titik *noise* merupakan data Karyawan yang tidak digunakan sebagai pembentuk *Cluster*, dikarenakan jumlah data tidak mencukupi batas jumlah minimum sampel data. Titik Karyawan6 = 4, Karyawan10 = 2, dan Karyawan11 = 2. Ketiga titik data karyawan ini disebut titik *noise*.

Tabel 14. Titik *noise*

No	Nama	Tetangga dalam Radius 2.5
6	Karyawan6	15, 3, 12, 10, 7
10	Karyawan10	12, 6



11	Karyawan11	19, 13
----	------------	--------

5. Hasil *Cluster* DBSCAN  
Hasil *Cluster* k-means diterapkan

Tabel 15. Hasil *Cluster* DBSCAN

No	Nama	Disiplin (at1)	Performa (at2)	Tanggung Jawab (at3)	Komunikasi (at4)	Kerjasama (at5)	Cluster	Keterangan Cluster
1	Karyawan1	6	7	5	7	7	C1	Excellent
2	Karyawan2	7	8	7	8	7	C2	Professional
3	Karyawan3	8	7	6	7	7	C1	Excellent
4	Karyawan4	6	6	7	7	6	C1	Excellent
5	Karyawan5	8	8	8	8	7	C2	Professional
6	Karyawan6	8	5	6	7	7	Noise	---
7	Karyawan7	7	7	8	7	7	C2	Professional
8	Karyawan8	6	7	7	8	7	C2	Professional
9	Karyawan9	7	8	7	8	8	C2	Professional
10	Karyawan10	6	5	6	6	8	Noise	---
11	Karyawan11	7	8	6	5	6	Noise	---
12	Karyawan12	6	6	6	7	7	C1	Excellent
13	Karyawan13	7	8	7	7	6	C2	Professional
14	Karyawan14	5	7	7	8	7	C1	Excellent
15	Karyawan15	8	6	7	7	7	C3	Need motivation
16	Karyawan16	6	8	6	7	8	C3	Need motivation
17	Karyawan17	7	8	7	8	8	C2	Professional
18	Karyawan18	6	7	5	7	6	C1	Excellent
19	Karyawan19	7	7	7	6	7	C1	Excellent
20	Karyawan20	7	8	8	7	7	C2	Professional

### 1.3 Evaluasi *Silhouette Score*

Evaluasi ini dilakukan terhadap hasil *Cluster* dari algoritma K-Means dan DBSCAN. Sebelum memulai pengukuran. Gunakan jarak *Euclidean* antar setiap titik data yang sudah didapatkan sebelumnya.

- a. Mengukur Hasil *Cluster* K-Means

1. a(i): Rata-rata jarak antara titik data i dengan semua titik data lain dalam *Clusternya* sendiri.

$$a(1) = \frac{\sum \text{Jarak Euclidean ke anggota lain di C1}}{\text{Jumlah anggota lain di C1}}$$

Rata-rata pada *Cluster*:

#### **Cluster 1**

Karyawan1 ke titik lainnya di *Cluster* 1

Karyawan3 = 2.24, Karyawan4 = 2.45, Karyawan6 = 3, Karyawan10 = 2.65, Karyawan11 = 2.83, Karyawan12 = 1.41, Karyawan15 = 3, Karyawan18 = 1, Karyawan19 = 2.45

$$a(1) = (2.24 + 2.45 + 3 + 2.65 + 2.83 + 1.41 + 3 + 1 + 2.45) / 10 = 21.03 / 10 = 2.1$$

Karyawan3 ke titik lainnya di *Cluster* 1

Karyawan1 = 2.24, Karyawan4 = 2.65, Karyawan6 = 2, Karyawan10 = 3.16, Karyawan11 = 2.65, Karyawan12 = 2.24, Karyawan15 = 1.41, Karyawan18 = 2.45, Karyawan19 = 1.73

$$a(2) = (2.24 + 2.65 + 2 + 3.16 + 2.65 + 2.24 + 1.41 + 2.45 + 1.73) / 10 = 20.53 / 10 = 2.05$$

perhitungan a yang sama dilakukan untuk setiap titik

#### **Cluster 1**

berikut hasil a(1) dan yang lainnya di *Cluster* 1

Karyawan1 a(1) = 2.1, Karyawan3 a(3) = 2.05, Karyawan4 a(4) = 2.14, Karyawan6 a(6) = 2.34, Karyawan10 a(10) = 2.51, Karyawan11 a(11) = 2.75, Karyawan12 a(12) = 1.82, Karyawan15 a(15) = 2.13, Karyawan18 a(18) = 2.22, Karyawan19 a(19) = 1.99

#### **Cluster 2**

Karyawan8 a(8) = 1, Karyawan14 a(14) = 1.08, Karyawan16 a(16) = 1.41,

#### **Cluster 3**

Karyawan2 a(2) = 1.14, Karyawan5 a(5) = 1.43, Karyawan7 a(7) = 1.46, Karyawan9 a(9) = 1.24, Karyawan13 a(13) = 1.58, Karyawan17 a(17) = 1.24, Karyawan20 a(20) = 1.24

2. b(i): Rata-rata jarak antara titik data i dengan semua titik data dari *Cluster* terdekat yang bukan *Cluster* yang sama.  
 Karyawan1 ke setiap titik di *Cluster* 2  
 Karyawan1 ke Karyawan 8 = 2.24, Karyawan1 ke Karyawan14 = 2.45, Karyawan1 ke Karyawan 16 = 1.73  
 $b(1) = (2.24 + 2.45 + 1.73) / 3 = 6.42 / 3 = 2.14$   
 lakukan cara yang sama untuk setiap karyawan anggota *Cluster* 1 ke *Cluster* 2, berikut hasilnya:  
**b(i) dari titik *Cluster* 1 ke *Cluster* 2.**  
 Karyawan1 b(1) = 2.14, Karyawan3 b(3) = 2.74, Karyawan4 b(4) = 2.24, Karyawan6 b(6) = 3.59, Karyawan10 b(10) = 3.21, Karyawan11 b(11) = 3.54, Karyawan12 b(12) = 1.99, Karyawan15 b(15) = 2.98, Karyawan18 b(18) = 2.52, Karyawan19 b(19) = 2.44  
**b(i) untuk *Cluster* 2**  
 Jarak ke *Cluster* terdekat untuk titik dalam *Cluster* 1  
 Karyawan8 b(8) = 2.52, Karyawan14 b(14) = 2.98, Karyawan16 b(16) = 2.72  
**b(i) untuk *Cluster* 3**  
 Jarak ke *Cluster* terdekat untuk titik dalam *Cluster* 2  
 Karyawan2 b(2) = 1.88, Karyawan5 b(5) = 2.98, Karyawan7 b(7) = 2.28, Karyawan9 b(9) = 1.97, Karyawan13 b(13) = 2.37, Karyawan17 b(17) = 1.97, Karyawan20 b(20) = 2.37
3. s(i): *Silhouette Score* untuk titik data i, penghitungan dimulai dari karyawan1 s(i)  

$$\text{Karyawan1 } S(1) = \frac{b_1 - a_1}{\max(a_1, b_1)} \quad S(1) = \frac{2.14 - 2.1}{\max(2.1, 2.14)} \quad \text{Karyawan1 } S(1) = \frac{0.04}{2.14} \quad S(1) = 0.02$$
  
 Selanjutnya dengan menggunakan nilai a(i) dan b(i) didapatkan semua s(i) untuk semua titik data hasil *Cluster* K-Means.

Tabel 16. *Silhouette Score Cluster* K-Means

Kelas	Nama	a(i)	b(i)	s(i)
Cluster 1	Karyawan1	2.10	2.14	0.02
	Karyawan3	2.05	2.74	0.25
	Karyawan4	2.14	2.24	0.04
	Karyawan6	2.34	3.59	0.35
	Karyawan10	2.51	3.21	0.22
	Karyawan11	2.75	3.54	0.22
	Karyawan12	1.82	1.99	0.09
	Karyawan15	2.13	2.98	0.29
	Karyawan18	2.22	2.52	0.12
	Karyawan19	1.99	2.44	0.18
Cluster 2	Karyawan8	1.00	2.52	0.60
	Karyawan14	1.08	2.98	0.64
	Karyawan16	1.41	2.72	0.48
Cluster 3	Karyawan2	1.14	1.88	0.39
	Karyawan5	1.43	2.98	0.52
	Karyawan7	1.46	2.28	0.36
	Karyawan9	1.24	1.97	0.37
	Karyawan13	1.58	2.37	0.33
	Karyawan17	1.24	1.97	0.37
	Karyawan20	1.24	2.37	0.48

4. Rata-rata *Silhouette Score* s(i)
- Cluster 1**  

$$= \frac{0.02 + 0.25 + 0.04 + 0.35 + 0.22 + 0.22 + 0.09 + 0.29 + 0.12 + 0.18}{10} = 0.178$$
  
 Rata-rata *Silhouette Score* untuk *Cluster* 1 adalah 0.178, yang menunjukkan bahwa karyawan dalam *Cluster* ini memiliki tingkat kualitas *Clusterisasi* yang rendah. Beberapa karyawan, seperti Karyawan6 dan Karyawan15, memiliki *Silhouette Score* yang lebih tinggi, tetapi sebagian besar memiliki nilai yang rendah.
- Cluster 2**  

$$= \frac{0.6 + 0.64 + 0.48}{3} = 0.57$$
  
 Rata-rata *Silhouette Score* untuk *Cluster* 2 adalah 0.57, yang menunjukkan bahwa karyawan dalam *Cluster* ini memiliki kualitas yang lebih tinggi terhadap *Clusterisasi* mereka dibandingkan dengan *Cluster* 1. Skor yang lebih tinggi ini menunjukkan bahwa mereka lebih dekat dengan pusat *Cluster* mereka dan jauh dari *Cluster* lain.
- Cluster 3**  

$$= \frac{0.39 + 0.52 + 0.36 + 0.37 + 0.33 + 0.37 + 0.48}{7} = 0.403$$
  
 Rata-rata *Silhouette Score* untuk *Cluster* 3 adalah 0.403. Ini menunjukkan bahwa karyawan dalam *Cluster* ini memiliki kualitas *Clusterisasi* yang lebih baik dibandingkan dengan *Cluster* 1 tetapi sedikit lebih rendah dibandingkan dengan *Cluster* 2. Karyawan5 dan Karyawan20 memiliki *Silhouette Score* yang lebih tinggi, menunjukkan bahwa lebih baik dengan *Clusterisasi* mereka.

b. Mengukur Hasil *Cluster* DBSCAN

1. a(i): Rata-rata jarak antara titik data i dengan semua titik data lain dalam *Clusternya* sendiri. Teknik yang digunakan sama dengan perhitungan a(i) hasil *Cluster* K-Means sebelumnya dan berikut hasilnya

**Cluster 1**

Karyawan1 a(1) = 2.14, Karyawan3 a(3) = 2.52, Karyawan4 a(4) = 2.25, Karyawan12 a(12) = 1.97, Karyawan14 a(14) = 2.61, Karyawan18 a(18) = 2.25, Karyawan19 a(19) = 2.38

**Cluster 2**

Karyawan2 a(2) = 1.17, Karyawan5 a(5) = 1.59, Karyawan7 a(7) = 1.49, Karyawan8 a(8) = 1.63, Karyawan9 a(9) = 1.3, Karyawan13 a(13) = 1.63, Karyawan17 a(17) = 1.3, Karyawan20 a(20) = 1.34

**Cluster 3**

Karyawan15 a(15) = 1.58, Karyawan16 a(16) = 1.58,

2. b(i): Rata-rata jarak antara titik data i dengan semua titik data dari *Cluster* terdekat yang bukan *Clusternya*. Teknik yang digunakan sama dengan perhitungan b(i) hasil *Cluster* K-Means sebelumnya dan berikut hasilnya.

**Cluster 1** Karyawan1 b(1) = 2.94, Karyawan3 b(3) = 2.76, Karyawan4 b(4) = 2.61, Karyawan12 b(12) = 2.72, Karyawan14 b(14) = 2.4, Karyawan18 b(18) = 3.14, Karyawan19 b(19) = 2.11

**Cluster 2** Karyawan2 b(2) = 2.47, Karyawan5 b(5) = 3.32, Karyawan7 b(7) = 2.43, Karyawan8 b(8) = 1.98, Karyawan9 b(9) = 2.75, Karyawan13 b(13) = 2.34, Karyawan17 b(17) = 2.75, Karyawan20 b(20) = 2.75

**Cluster 3** Karyawan15 b(15) = 2.41, Karyawan16 b(16) = 2.47

3. s(i): *Silhouette Score* untuk titik data i, penghitungan dimulai dari karyawan1 s(i), dengan cara yang sama dengan sebelumnya, berikut hasil *Silhouette Score* untuk hasil *Cluster* DBSCAN.

Tabel 17. *Silhouette Score Cluster* DBSCAN

Kelas	Nama	a(i)	b(i)	s(i)
Cluster 1	Karyawan1	2.14	2.94	0.27
	Karyawan3	2.52	2.76	0.09
	Karyawan4	2.25	2.61	0.14
	Karyawan12	1.97	2.72	0.28
	Karyawan14	2.61	2.40	-0.08
	Karyawan18	2.25	3.14	0.28
	Karyawan19	2.38	2.11	-0.11
Cluster 2	Karyawan2	1.17	2.47	0.53
	Karyawan5	1.59	3.32	0.52
	Karyawan7	1.49	2.43	0.39
	Karyawan8	1.63	1.98	0.18
	Karyawan9	1.30	2.75	0.53
	Karyawan13	1.63	2.34	0.30
	Karyawan17	1.30	2.75	0.53
	Karyawan20	1.34	2.75	0.51
Cluster 3	Karyawan15	1.58	2.41	0.34
	Karyawan16	1.58	2.47	0.36
Noise	Karyawan6	-	-	-
	Karyawan10	-	-	-
	Karyawan11	-	-	-

4. Rata-rata *Silhouette Score* s(i)

$$\text{Cluster 1} = \frac{0.27+0.09+0.14+0.28+0.08+0.28+0.11}{7} = 0.12$$

Rata-rata *Silhouette Score* untuk *Cluster* 1 adalah 0.12, yang menunjukkan bahwa karyawan dalam *Cluster* ini memiliki tingkat kualitas *Clusterisasi* yang rendah. Karyawan14 dan Karyawan19 memiliki nilai negatif, menunjukkan bahwa lebih dekat dengan *Cluster* lain daripada *Cluster* mereka sendiri.

$$\text{Cluster 2} = \frac{0.53+0.52+0.39+0.18+0.53+0.3+0.53+0.51}{8} = 0.44$$

Rata-rata *Silhouette Score* untuk *Cluster* 2 adalah 0.44, yang menunjukkan bahwa karyawan dalam *Cluster* ini memiliki kualitas yang lebih tinggi terhadap *Clusterisasi* mereka dibandingkan dengan *Cluster* 1. Skor yang lebih tinggi ini menunjukkan bahwa mereka lebih dekat dengan pusat *Cluster* mereka dan jauh dari *Cluster* lain. Karyawan2, Karyawan9,

dan Karyawan17 memiliki nilai  $s(i)$  tertinggi (0.53), menunjukkan bahwa mereka sangat dekat dengan anggota *Cluster* mereka dan jauh dari *Cluster* lain.

$$\text{Cluster 3} = \frac{0.34+0.36}{2} = 0.52$$

Rata-rata *Silhouette Score* untuk *Cluster 3* adalah 0.52. Ini menunjukkan bahwa karyawan dalam *Cluster* ini memiliki kualitas *Clusterisasi* yang lebih baik dibandingkan dengan *Cluster 1* tetapi sedikit lebih rendah dibandingkan dengan *Cluster 2*. Meskipun hanya terdiri dari dua karyawan, nilai  $s(i)$  yang tinggi menunjukkan bahwa mereka sangat dekat satu sama lain dan jauh dari *Cluster* lain.

#### 1.4 Evaluasi Davies-Bouldin Index (DBI)

Evaluasi ini dilakukan terhadap hasil *Cluster* dari algoritma K-Means dan DBSCAN. Sebelum memulai pengukuran.

a. Mengukur DBI Hasil *Cluster* K-Means

1. Hitung *Centroid* untuk setiap *Cluster*: *Centroid* adalah rata-rata dari setiap atribut di *Cluster* tersebut.

Terdapat tiga *Cluster*: C1, C2, dan C3.

**Centroid C1:**

$$C1 = \left( \frac{6+8+6+8+6+7+6+8+6+7}{10}, \frac{7+7+6+5+5+8+6+6+7+7}{10}, \frac{5+6+7+6+6+6+6+7+5+7}{10}, \frac{7+7+7+7+6+5+7+7+7+6}{10}, \frac{7+7+6+7+8+6+7+7+6+7}{10} \right)$$

$$C1 = (6.8, 6.4, 6.1, 6.6, 6.8)$$

**Centroid C2:**

$$C2 = \left( \frac{6+5+6}{3}, \frac{7+7+8}{3}, \frac{7+7+6}{3}, \frac{8+8+7}{3}, \frac{7+7+8}{3} \right)$$

$$C2 = (5.67, 7.33, 6.67, 7.67, 7.33)$$

**Centroid C3:**

$$C3 = \left( \frac{7+8+7+7+7+7+7}{7}, \frac{8+8+7+8+8+8+8}{7}, \frac{7+8+8+7+7+7+8}{7}, \frac{8+8+7+8+7+8+7}{7}, \frac{7+7+7+8+6+8+7}{7} \right)$$

$$C3 = (7.14, 7.86, 7.43, 7.57, 7.14)$$

2. Hitung jarak rata-rata (*dispersion*) dalam *Cluster* jarak *Euclidean*. Jarak rata-rata dari setiap anggota *Cluster* ke *Centroid* *Cluster* tersebut.

Jarak Karyawan1 ke c1

$$= \sqrt{(6-6.8)^2 + (7-6.4)^2 + (5-6.1)^2 + (7-6.6)^2 + (7-6.8)^2} = 1.55$$

Berikut hasilnya:

**Penghitungan jarak-jarak semua Karyawan Cluster 1 ke c1**

Karyawan1 = 1.55, Karyawan3 = 1.42, Karyawan4 = 1.55, Karyawan6 = 1.90, Karyawan10 = 2.10, Karyawan11 = 2.41, Karyawan12 = 1.00, Karyawan15 = 1.62, Karyawan18 = 1.73, Karyawan19 = 1.27

**Jarak ke c2**

Karyawan8 = 0.74, Karyawan14 = 0.94, Karyawan16 = 1.38

**Jarak ke c3**

Karyawan2 = 0.75, Karyawan5 = 1.18, Karyawan7 = 1.36, Karyawan9 = 0.84, Karyawan13 = 1.60, Karyawan17 = 0.48, Karyawan20 = 1.06

Selanjutnya hitung semua jarak c1, c2, dan c3

$$\text{Rata-rata c1} = \frac{1.55+1.42+1.55+1.90+2.10+2.41+1+1.62+1.73+1.27}{10} = 1.66$$

$$\text{Rata-rata c2} = \frac{0.74+0.94+1.38}{3} = 1.02$$

$$\text{Rata-rata c3} = \frac{0.63+1.13+1.19+1.36+1.41+1.36+0.84}{3} = 1.13$$

Jarak Rata-rata dalam *Cluster* C1: SC1=1.66

Jarak Rata-rata dalam *Cluster* C2: SC2=1.02

Jarak Rata-rata dalam *Cluster* C3: SC3=1.13

3. Hitung jarak antar *Centroid*: Ini adalah jarak antara *Centroid* dari dua *Cluster* yang berbeda.

*Centroid* C1 = (6.8, 6.4, 6.1, 6.6, 6.8)

*Centroid* C2 = (5.67, 7.33, 6.67, 7.67, 7.33)

*Centroid* C3 = (7.14, 7.86, 7.43, 7.57, 7.14)

Jarak C1 ke C2

$$= \sqrt{(6.8-5.67)^2 + (6.4-7.33)^2 + (6.1-6.67)^2 + (6.6-7.67)^2 + (6.8-7.33)^2} = 1.97$$

Jarak C1 ke C3

$$= \sqrt{(6.8-7.14)^2 + (6.4-7.86)^2 + (6.1-7.43)^2 + (6.6-7.57)^2 + (6.8-7.14)^2} = 2.25$$

Jarak C2 ke C3

$$= \sqrt{(5.7 - 7.14)^2 + (7.33 - 7.86)^2 + (6.67 - 7.43)^2 + (7.67 - 7.57)^2 + (7.33 - 7.14)^2} = 1.75$$

Jarak antara *Centroid* C1 dan *Centroid* C2:  $d(C1, C2) = 1.97$

Jarak antara *Centroid* C1 dan *Centroid* C3:  $d(C1, C3) = 2.25$

Jarak antara *Centroid* C2 dan *Centroid* C3:  $d(C2, C3) = 1.75$

4. Hitung *Davies-Bouldin Index*.

Terlebih dahulu hitung ratio  $R_{ij}$

$$R_{12} = \frac{SC1 + SC2}{d(C1, C2)} = \frac{1.66 + 1.02}{1.97} = 1.36$$

$$R_{13} = \frac{SC1 + SC3}{d(C1, C3)} = \frac{1.66 + 1.13}{1.97} = 1.27$$

$$R_{23} = \frac{SC2 + SC3}{d(C2, C3)} = \frac{1.02 + 1.13}{1.75} = 1.23$$

$$DBI = \frac{DC1 + DC2 + DC3}{3} = \frac{1.36 + 1.36 + 1.24}{3} = 1.32$$

Nilai DBI sebesar 1.32 menunjukkan bahwa *Clustering* yang dilakukan cukup baik dalam memisahkan data ke dalam *Cluster* yang terdefinisi dengan baik. Angka ini relatif rendah, yang berarti bahwa *Cluster* yang dihasilkan memiliki kesamaan tinggi antar anggota dalam *Cluster* dan perbedaan yang jelas antar *Cluster*.

b. Mengukur DBI Hasil *Cluster* DBSCAN

1. Hitung *Centroid* untuk setiap *Cluster*: *Centroid* adalah rata-rata dari setiap atribut di *Cluster* tersebut.

Terdapat tiga *Cluster*: C1, C2, dan C3.

***Centroid* C1:**

$$C1 = \left( \frac{6 + 8 + 6 + 6 + 5 + 6 + 7}{7}, \frac{7 + 7 + 6 + 6 + 7 + 7 + 7}{7}, \frac{5 + 6 + 7 + 6 + 7 + 5 + 7}{7}, \right. \\ \left. \frac{7 + 7 + 7 + 7 + 8 + 7 + 6}{7}, \frac{7 + 7 + 6 + 7 + 7 + 6 + 7}{7} \right)$$

$$C1 = (6.42, 6.71, 6.14, 7, 6.71)$$

***Centroid* C2:**

$$C2 = \left( \frac{7 + 8 + 7 + 6 + 7 + 7 + 7 + 7}{8}, \frac{8 + 8 + 7 + 7 + 8 + 8 + 8 + 8}{8}, \frac{7 + 8 + 8 + 7 + 7 + 7 + 7 + 8}{8}, \right. \\ \left. \frac{8 + 8 + 7 + 8 + 8 + 7 + 8 + 7}{8}, \frac{7 + 7 + 7 + 7 + 8 + 6 + 8 + 7}{8} \right)$$

$$C2 = (7, 7.75, 7.38, 7.63, 7.13)$$

***Centroid* C3:**

$$C3 = \left( \frac{8 + 6}{2}, \frac{6 + 8}{2}, \frac{7 + 6}{2}, \frac{7 + 7}{2}, \frac{7 + 8}{2} \right)$$

$$C3 = (7, 7, 6.5, 7, 7.5)$$

2. Hitung jarak rata-rata (*dispersion*) dalam *Cluster* jarak *Euclidean*. Jarak rata-rata dari setiap anggota *Cluster* ke *Centroid* *Cluster* tersebut.

Jarak Karyawan1 ke c1

$$= \sqrt{(6 - 6.42)^2 + (7 - 6.71)^2 + (5 - 6.14)^2 + (7 - 7)^2 + (7 - 6.71)^2} = 1.28$$

Berikut hasil penghitungan

**jarak-jarak semua Karyawan *Cluster* 1 ke c1**

Karyawan1 = 1.28, Karyawan3 = 1.64, Karyawan4 = 1.39, Karyawan12 = 0.89, Karyawan14 = 1.98, Karyawan18 = 1.44, Karyawan19 = 1.50

**Jarak ke c2**

Karyawan2 = 0.60, Karyawan5 = 1.27, Karyawan7 = 1.17, Karyawan8 = 1.36, Karyawan9 = 1.05, Karyawan13 = 1.37, Karyawan17 = 1.05, Karyawan20 = 0.93

**Jarak ke c3**

Karyawan17 = 1.58, Karyawan20 = 1.58

Selanjutnya hitung semua jarak c1, c2, dan c3

**Rata-rata c1**

$$= \frac{1.28 + 1.64 + 1.39 + 0.89 + 1.98 + 1.44 + 1.50}{7} = 1.44$$

**Rata-rata c2**

$$= \frac{0.60 + 1.27 + 1.17 + 1.36 + 1.05 + 1.37 + 1.05 + 0.93}{8} = 1.10$$

**Rata-rata c3**

$$= \frac{1.58 + 1.58}{2} = 1.58$$

Jarak Rata-rata dalam *Cluster* C1: SC1=1.44

Jarak Rata-rata dalam *Cluster* C2: SC2=1.10

Jarak Rata-rata dalam *Cluster* C3: SC3=1.58

3. Hitung jarak antar *Centroid*: Ini adalah jarak antara *Centroid* dari dua *Cluster* yang berbeda.

*Centroid* C1 = (6.42, 6.71, 6.14, 7, 6.71)

*Centroid* C2 = (7, 7.75, 7.38, 7.63, 7.13)

*Centroid* C3 = (7, 7, 6.5, 7, 7.5)

Jarak C1 ke C2

$$= \sqrt{(7 - 6.42)^2 + (7.75 - 6.71)^2 + (7.38 - 6.14)^2 + (7.63 - 7)^2 + (7.13 - 6.71)^2} = 1.88$$

Jarak C1 ke C3

$$= \sqrt{(7 - 6.42)^2 + (7 - 6.71)^2 + (6.5 - 6.14)^2 + (7 - 7)^2 + (7.5 - 6.71)^2} = 1.08$$

Jarak C2 ke C3

$$= \sqrt{(7 - 7)^2 + (7.75 - 7)^2 + (7.38 - 6.5)^2 + (7.63 - 7)^2 + (7.13 - 7.5)^2} = 1.37$$

Jarak antara *Centroid* C1 dan *Centroid* C2:  $d(C1, C2) \approx 1.88$

Jarak antara *Centroid* C1 dan *Centroid* C3:  $d(C1, C3) \approx 1.08$

Jarak antara *Centroid* C2 dan *Centroid* C3:  $d(C2, C3) \approx 1.37$

4. Hitung *Davies-Bouldin Index*.

Terlebih dahulu hitung ratio  $R_{ij}$

$$R_{12} = \frac{SC1 + SC2}{d(C1, C2)} = \frac{1.44 + 1.10}{1.88} \approx 1.35$$

$$R_{13} = \frac{SC1 + SC3}{d(C1, C3)} = \frac{1.44 + 1.58}{1.08} \approx 2.78$$

$$R_{23} = \frac{SC2 + SC3}{d(C2, C3)} = \frac{1.10 + 1.58}{1.37} \approx 1.96$$

$$DBI = \frac{DC1 + DC2 + DC3}{3} = \frac{2.78 + 1.96 + 2.78}{3} \approx 2.51$$

Nilai DBI sebesar 2.51 yaitu  $> 2.5$  ini menunjukkan bahwa *Clustering* yang dilakukan tidak baik, tidak optimal dan saling tumpang tindih atau kurang terpisah dengan baik.

## 1.5 Hasil Evaluasi

Setelah selesai menghitung hasil *Silhouette Score* dan *Davies-Bouldin Index* dari hasil *Clustering* dengan algoritma K-Means dan DBSCAN.

1. *Silhouette Score* dari hasil *Cluster* K-Means

Pada hasil K-Means, *Cluster* 2 menunjukkan kualitas *Cluster* terbaik dengan rata-rata *Silhouette Score* sebesar 0.57, diikuti oleh *Cluster* 3 (0.40) dan *Cluster* 1 (0.18).

2. *Silhouette Score* dari hasil *Cluster* DBSCAN

Pada hasil DBSCAN, *Cluster* 3 menunjukkan kualitas *Cluster* terbaik dengan rata-rata *Silhouette Score* sebesar 0.52, diikuti oleh *Cluster* 2 (0.44) dan *Cluster* 1 (0.12).

3. *Davies-Bouldin Index*

Nilai DBI untuk K-Means (1.32) lebih rendah dibandingkan DBSCAN (2.51), menunjukkan bahwa *Cluster* yang dihasilkan oleh K-Means lebih baik dalam hal separasi dan kohesi dibandingkan DBSCAN.

## KESIMPULAN

Evaluasi ini membantu memahami kekuatan dan kelemahan kedua algoritma dalam konteks data yang digunakan. Dengan demikian dapat disimpulkan sabagi berikut:

1. Algoritma K-Means menghasilkan *Cluster* dengan nilai *Davies-Bouldin Index* (DBI) lebih rendah (1.32) dibandingkan DBSCAN (2.51), menunjukkan bahwa K-Means lebih baik dalam hal separasi dan kohesi *Cluster*.
2. Rata-rata *Silhouette Score* pada K-Means menunjukkan bahwa *Cluster* 2 memiliki kualitas *Cluster* terbaik (0.57), diikuti oleh *Cluster* 3 (0.40) dan *Cluster* 1 (0.18). Sedangkan pada DBSCAN, *Cluster* 3 memiliki kualitas *Cluster* terbaik (0.52), diikuti oleh *Cluster* 2 (0.44) dan *Cluster* 1 (0.12).

Secara keseluruhan, K-Means lebih efektif dalam menghasilkan *Cluster* yang lebih kohesif dan terpisah dengan baik dibandingkan DBSCAN berdasarkan nilai DBI dan *Silhouette Score*.

## DAFTAR PUSTAKA

- [1] F. E. Ozturk, N. Demirel, and M. Bilgisi, "Comparison of the Methods to Determine Optimal Number of Cluster," 2023. [Online]. Available: [www.dergipark.gov.tr/veri](http://www.dergipark.gov.tr/veri)
- [2] A. Karami and R. Johansson, "Choosing DBSCAN Parameters Automatically using Differential Evolution," Apr. 2014.
- [3] A. Alalousi, R. Razif, M. AbuAlhaj, M. Anbar, and S. Nizam, "A Preliminary Performance Evaluation of K-means, KNN and EM Unsupervised Machine Learning Methods for Network Flow Classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 2, p. 778, Apr. 2016, doi: 10.11591/ijece.v6i2.8909.
- [4] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, Dec. 2023, doi: 10.59653/ijmars.v2i01.504.

- [5] A. A. A and L. R. Nair Associate Professor, "A NOVEL STUDY OF SILHOUETTE METHOD TO SOLVE THE ISSUES OF OUTLIER AND IMPROVE THE QUALITY OF CLUSTER," *Journal of Data Acquisition and Processing*, vol. 38, no. 2, pp. 3099–3118, 2023, doi: 10.5281/zenodo.777129.
- [6] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [7] K. Amrulloh, T. Hendro Pudjiantoro, P. Nurul Sabrina, and A. Id Hadiana, "Comparison Between Davies-Bouldin Index and Silhouette Coefficient Evaluation Methods in Retail Store Sales Transaction Data Clusterization Using K-Medoids Algorithm," Cimahi, Jul. 2022. [Online]. Available: [www.kaggle.com](http://www.kaggle.com),
- [8] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J (Basel)*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.
- [9] M. Hahsler, M. Piekenbrock, and D. Doran, "DbSCAN: Fast density-based clustering with R," *J Stat Softw*, vol. 91, 2019, doi: 10.18637/jss.v091.i01.
- [10] J. Yang, Z. Sun, and Y. Chen, "Fault detection using the clustering-kNN rule for gas sensor arrays," *Sensors (Switzerland)*, vol. 16, no. 12, Dec. 2016, doi: 10.3390/s16122069.
- [11] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *ILKOM Jurnal Ilmiah*, vol. 13, no. 1, pp. 31–35, Apr. 2021, doi: 10.33096/ilkom.v13i1.763.31-35.