



# Analisis Metode Elbow SSE, Silhouette Score, dan Jaccard Stability dalam Pemilihan Jumlah Kluster Data yang Optimal

Budi Hartono\*, Veronica Lusiana

Fakultas Teknologi Informasi dan Industri, Program Studi Teknik Informatika, Universitas Stikubank, Semarang, Indonesia

Email: <sup>1\*</sup>budihartono@edu.unisbank.ac.id, <sup>2</sup>vero@edu.unisbank.ac.id

Email Penulis Korespondensi: budihartono@edu.unisbank.ac.id

**Abstrak**—Penelitian ini membahas pemilihan jumlah kluster (K) yang optimal pada algoritma K-Means dengan memanfaatkan kombinasi metode Elbow dengan metrik SSE (*Sum of Squared Errors*) dan Silhouette Score. Permasalahan utama adalah nilai K yang optimal tidak diketahui. Pemilihan K yang terlalu kecil berpotensi menggabungkan pola berbeda dan K yang terlalu besar berpotensi memecah pola yang sama menjadi beberapa kluster. Percobaan menggunakan data uji dua dimensi dengan variasi jumlah data 20, 30, 40, 50, dan 60 data. K-Means dijalankan pada rentang K=2 hingga K=8, kemudian dihitung nilai SSE untuk membentuk kurva Elbow serta nilai Silhouette rata-rata untuk mengevaluasi kualitas kluster. Penelitian ini menambahkan uji kestabilan kluster menggunakan nilai Jaccard Stability. Nilai Silhouette tertinggi yaitu 0.4619 diperoleh dari jumlah data 20 untuk K=2. Nilai Jaccard stability tertinggi yaitu 0.9507 diperoleh dari jumlah data 60 untuk K=2. Hasil percobaan menunjukkan metode Elbow, nilai Silhouette, dan Jaccard stability dapat digunakan secara komplementer dalam menentukan K yang optimal. Pada beberapa data uji, kedua metrik menghasilkan rekomendasi K yang konsisten, sedangkan pada data uji tertentu Elbow dapat memberikan beberapa kandidat sehingga diperlukan validasi menggunakan nilai Silhouette untuk memilih K yang optimal.

**Kata Kunci:** Pengelompokan Data; K-Means Kluster; Elbow SSE; Silhouette Score; Jaccard Stability

**Abstract**—This study discusses the selection of the optimal number of clusters (K) in the K-Means algorithm by utilizing a combination of the Elbow method with the SSE (Sum of Squared Errors) and Silhouette Score metrics. The main problem is that the optimal K value is unknown. Choosing K that is too small can combine different patterns (under-clustering), and choosing K that is too large can break the same pattern into several clusters (over-clustering). The experiment used two-dimensional test data with variations in the number of data 20, 30, 40, 50, and 60. K-Means was run in the range of K = 2 to K = 8, then the SSE value was calculated to form the Elbow curve and the average Silhouette value to evaluate the quality of the cluster. This study added a cluster stability test using the Jaccard Stability value. The highest Silhouette value of 0.4619 was obtained from the data 20 for K = 2. The highest Jaccard stability value of 0.9507 was obtained from 60 data sets for K = 2. The experimental results show that the Elbow method, Silhouette value, and Jaccard stability can be used complementarily in determining the optimal K. In some test data, both metrics produce consistent K recommendations, while in certain test data, Elbow can provide several candidates, so that validation using the Silhouette value is needed to select the optimal K.

**Keywords:** Data Cluster; K-Means Cluster; Elbow SSE; Silhouette Score; Jaccard Stability

## 1. PENDAHULUAN

Penentuan jumlah kluster data (K) yang optimal merupakan persoalan kunci dalam analisis pengelompokan data tanpa label (*unsupervised learning*). Pemilihan K yang tidak tepat berakibat pada struktur kluster yang tidak representatif terhadap data. Nilai K yang terlalu kecil menyebabkan objek yang seharusnya berbeda dipaksa bergabung atau masuk dalam kluster yang sama. Nilai K terlalu besar akan memecah atau memisahkan objek data yang homogen yang seharusnya dalam kluster yang sama. Algoritma K-Means adalah metode pengelompokan data yang banyak dipakai, dengan K ditentukan secara intuitif pada saat awal penyusunan kluster.

Pada penelitian ini permasalahan yang dikaji adalah bagaimana memilih K yang optimal pada kluster K-Means. Dua metode yang digunakan yaitu Elbow dengan metrik SSE (*Sum of Squared Errors*) dan Silhouette Score untuk memilih jumlah kluster yang optimal. Implementasi komputasi pembentukan kluster menggunakan Python. Hasil pengelompokan data adalah jumlah dan daftar anggota setiap kluster.

Berbagai penelitian menunjukkan bahwa persoalan krusial pada kluster data adalah menentukan jumlah kluster yang tepat. Di bidang kesehatan, kebencanaan, dan lingkungan, pemilihan K yang optimal dipakai untuk mendukung pemetaan prioritas dan analisis spasial. Penelitian oleh Diantika mengelompokkan provinsi terkait faktor penyebab stunting serta membandingkan Elbow, Silhouette, dan gap statistics untuk menentukan jumlah kluster terbaik (Diantika et al., 2025). Penelitian oleh Syahkur memakai K-Means++ untuk data gizi balita dan menilai K dengan Elbow-Silhouette serta memvalidasi dengan DBI (Syahkur et al., 2024). Ashari mengklasifikasikan wilayah terdampak banjir di Jakarta menggunakan K-Means dan mengevaluasi dengan Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, dan Rand Index (Ashari et al., 2023). Pada kasus pneumonia balita, Chrisamudra & Abdurrahman menggunakan Silhouette untuk menentukan kluster optimal dan DBI untuk evaluasi hasil, menekankan bahwa klustering dapat menjadi alat bantu membuat kebijakan kesehatan berbasis data (Chrisamudra & Abdurrahman, 2025). Sementara itu, Guntara & Lutfi mengelompokkan faktor penyebab mahasiswa *drop out* sehingga institusi dapat memetakan risiko dan intervensi secara lebih terarah (Guntara & Lutfi, 2023). Penelitian oleh Kurniawan mengelompokkan prioritas negara yang membutuhkan bantuan ke dalam level C1 - C5, sehingga hasil kluster dapat dipakai sebagai dasar penetapan prioritas bantuan (Kurniawan et al., 2023). Pemilihan K yang tepat dapat berdampak langsung pada interpretasi kelompok dan pengambilan keputusan berbasis data.

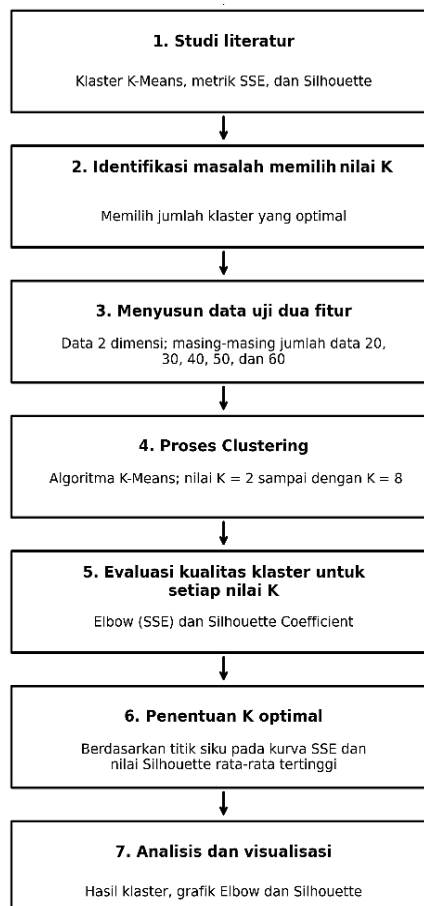
Pada domain ulasan oleh pengguna, Sholeh dan Aeni mengevaluasi kualitas kluster pada data ulasan destinasi wisata dengan membandingkan beberapa metrik yaitu Davies–Bouldin, Elbow, dan Silhouette pada rentang K yang luas (Sholeh & Aeni, 2023). Penelitian oleh Wicaksono menerapkan K-Means pada data penjualan tiket film dan menentukan K terbaik, sehingga pola penjualan dapat dikelompokkan untuk kebutuhan analisis bisnis (Wicaksono et al., 2024). Pada konteks ulasan obat, Juanita dan Cahyono mengelompokkan ulasan pengguna situs drugs.com menggunakan K-Means dan menemukan jumlah kluster, data berbasis opini atau ulasan membutuhkan strategi pemilihan K yang hati-hati (Juanita & Cahyono, 2024). Ritonga dan Hasibuan membandingkan Silhouette dan Elbow pada K-Means dan DBSCAN (Ritonga & Hasibuan, 2025). Rahmawati menyoroti perbandingan Elbow dan Silhouette pada K-Medoids untuk pengelompokan provinsi berdasarkan indikator Indeks Pembangunan Manusia (Rahmawati et al., 2024), sedangkan Kuswardana membandingkan Elbow–Silhouette pada K-Prototype untuk data transaksi pelanggan (Kuswardana et al., 2025).

Rangkaian penelitian terapan memperlihatkan adanya ruang kajian yang lebih sistematis tentang Elbow dengan metrik SSE (*Sum of Squared Errors*), Silhouette Score digunakan untuk menentukan nilai K yang optimal, dan mengukur kestabilan hasil kluster menggunakan Jaccard Stability. Data percobaan menggunakan data dua fitur dengan variasi jumlah data, agar diperoleh pemahaman yang lebih baik mengenai perbedaan rekomendasi nilai K. Data berukuran dua dimensi (x, y) sehingga melalui *plotting* data dapat diamati struktur kluster yang terbentuk.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian ini diawali dengan studi literatur untuk menentukan pendekatan pemilihan jumlah kluster, khususnya melalui metrik Elbow dengan nilai SSE dan Silhouette Score. Selanjutnya dilakukan identifikasi masalah yang berfokus pada penentuan nilai K yang optimal. Data uji kemudian disusun dalam bentuk dua fitur atau dua dimensi dengan variasi jumlah data sebanyak 20, 30, 40, 50, dan 60. Proses pengelompokan dilakukan menggunakan algoritma K-Means pada rentang nilai K antara 2 hingga 8. Kualitas kluster dievaluasi untuk setiap nilai K menggunakan beberapa indikator. Penentuan K optimal didasarkan pada kurva Elbow, nilai Silhouette rata-rata tertinggi, serta stabilitas Jaccard, yang kemudian dianalisis dan divisualisasikan.



**Gambar 1.** Diagram Alir Tahapan Penelitian



Tahapan penelitian dapat dilihat pada Gambar 1, diawali dengan studi literatur untuk membangun landasan teori. Pada tahap ini dibahas konsep klustering K-Means, serta dua metrik evaluasi yang digunakan dalam pemilihan jumlah kluster, yaitu Elbow (SSE) dan dan Silhouette Score untuk mengukur kualitas pemisahan kluster. Selanjutnya dilakukan identifikasi masalah utama penelitian, yaitu bagaimana menentukan nilai K yang optimal sehingga hasil kluster tidak terlalu sedikit (*under-clustering*) maupun terlalu banyak (*over-clustering*).

Penelitian dilanjutkan dengan penyusunan data uji dua fitur. Dataset yang digunakan adalah data dua dimensi (2D) dengan dua atribut atau fitur (x, y) dalam beberapa variasi jumlah data yaitu 20, 30, 40, 50, dan 60. Dataset dua dimensi yang berukuran kecil memudahkan untuk mengamati struktur sebaran data secara visual. Jumlah data yang kecil maka dapat dilihat apakah kandidat K yang diusulkan telah mewakili pola sebaran data. Proses kluster data menggunakan algoritma K-Means. Percobaan dilakukan dengan rentang jumlah K=2 sampai dengan K=8. Dengan pendekatan ini, setiap dataset dapat menghasilkan beberapa kluster yang berbeda (Lusiana et al., 2024).

Pada tahap evaluasi kualitas kluster, setiap nilai K yang diuji dihitung metriknya menggunakan dua pendekatan yaitu Elbow (SSE) dan dan Silhouette Score. Kurva SSE digunakan untuk melihat kecenderungan penurunan *error* dan menemukan titik siku sebagai indikasi K yang efisien, sedangkan Silhouette rata-rata digunakan untuk memilih K dengan kualitas pemisahan kluster terbaik. Berdasarkan dua metrik ini dilakukan penentuan K optimal, yaitu memilih K yang direkomendasikan oleh titik siku SSE dan didukung nilai Silhouette rata-rata tertinggi. Tahap selanjutnya Adalah analisis dan visualisasi hasil kluster, grafik Elbow, nilai Silhouette, dan nilai Jaccard Stability.

## 2.2 Elbow (Sum of Squared Errors)

Algoritma K-Means menjadi metode pengelompokan yang banyak dipakai, dengan K telah ditentukan sejak awal. Untuk menentukan K yang optimal dapat mengadopsi kriteria berbasis *goodness of fit* kluster, diantaranya menggunakan metode Elbow dengan SSE (*Sum of Squared Errors*) dan Silhouette Score. Metode Elbow SSE pada K-Means kluster adalah menentukan jumlah kluster optimal (K) dengan mencari siku (*elbow*) pada grafik, di mana penurunan nilai SSE yaitu ukuran jarak data ke pusat kluster mulai melandai atau melambat drastis (Maori & Evanita, 2023). Hal ini menandakan penambahan kluster sudah tidak memberikan peningkatan yang signifikan. Nilai K yang membentuk siku adalah pilihan terbaik. Dalam konteks K-Means kluster, terdapat istilah *Within-Cluster Sum of Squares* (WCSS) yang merujuk pada konsep yang sama dengan SSE, yaitu menghitung jumlah kuadrat jarak data terhadap pusat kluster (*centroid*) masing-masing. SSE yang merepresentasikan total kuadrat jarak setiap data terhadap centroid kluster tempat data tersebut berada (Shi et al., 2021) (Hartono et al., 2023). Nilai SSE/WCSS akan menurun seiring bertambahnya K, sehingga perubahan laju penurunan yang membentuk siku sebagai indikasi bahwa penambahan kluster berikutnya tidak lagi memberikan perbaikan yang signifikan. SSE untuk suatu K didefinisikan melalui persamaan (1). Metode Elbow menggunakan nilai SSE yang dihitung untuk setiap kandidat jumlah kluster K.

$$SSE(K) = \sum_{k=1}^K \sum_{x_i \in C_k} |x_i - \mu_k|^2 \quad (1)$$

Penentuan jumlah kluster pada algoritma K-Means umumnya dilakukan melalui evaluasi berbasis metrik. Metode Elbow dan Silhouette Score sebagai indikator yang saling melengkapi dalam memilih K terbaik. K-Means dijalankan pada rentang nilai K tertentu. Elbow digunakan untuk mengidentifikasi titik ketika penurunan *error* mulai melambat.

## 2.3 Silhouette Score

Metrik evaluasi Silhouette score atau Silhouette coefficient untuk menilai kualitas kluster dengan mengukur seberapa mirip suatu objek dengan klusternya sendiri dibandingkan dengan kluster tetangga, menghasilkan skor antara -1 hingga 1. Nilai mendekati 1 menunjukkan kluster yang baik dan nilai mendekati -1 menunjukkan kluster yang buruk. Skor ini membantu menentukan jumlah kluster K optimal dengan memilih nilai K yang menghasilkan rata-rata Silhouette score tertinggi, yaitu menunjukkan pengelompokan yang paling efektif (Tan et al., 2019) (Vania & Sari, 2023). Menghitung Silhouette score S(i) sebuah data menggunakan persamaan (2). Rata-rata (*mean*) skor S(i) menjadi indikator kualitas keseluruhan kluster.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Keterangan, a(i) adalah rata-rata jarak titik data ke semua titik lain di kluster yang sama (*intra-cluster distance*), b(i) adalah rata-rata jarak minimum dari titik data ke titik di kluster tetangga terdekat (*inter-cluster distance*). Rentang nilai S(i) adalah dari -1 sampai dengan 1, memiliki interpretasi:

~1 (positif) : objek sangat cocok dengan kluster sendiri dan jauh dari kluster lain.

~ 0 (netral): objek berada di dekat batas antara dua kluster.

~-1 (negatif): objek seharusnya berada di kluster tetangga.

Nilai rata-rata digunakan untuk menilai kualitas seluruh kluster.

Silhouette digunakan untuk menilai kekompakan kluster dan keterpisahan antar kluster (Hasan, 2024) (Shutaywi & Kachouie, 2021). Apabila metrik Elbow dan Silhouette mengarah pada nilai K yang sama maka K optimal menjadi lebih kuat, sebaliknya jika berbeda maka pemilihan K ditetapkan dengan mempertimbangkan K yang memberikan kualitas pemisahan tinggi.



### 2.4 Jaccard Stability

Metrik Jaccard stability digunakan untuk mengukur kestabilan hasil kluster terhadap perubahan inisialisasi K-Means, karena K-Means sensitif terhadap centroid awal. Mengukur kestabilan hasil kluster ketika proses penyusunan kluster diulang sebanyak 10 kali (R=10) untuk setiap nilai K pada data uji yang sama. Pada rentang nilai K antara 2 sampai dengan 8, maka jumlah mengulang (*re-run*) adalah sebanyak 70 kali. Menghitung Jaccard Stability menggunakan persamaan (3). Memiliki rentang nilai 0 sampai dengan 1. Nilai mendekati 1 berarti hasil kluster relatif konsisten atau stabil, nilai mendekati 0 berarti hasil kluster mudah berubah atau tidak stabil (Han et al., 2023) (Tan et al., 2019).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

Keterangan,  $|A \cap B|$  adalah jumlah anggota yang sama pada kedua kluster,  $|A \cup B|$  adalah jumlah anggota gabungan kedua kluster. Nilai  $J(A, B)$  memiliki rentang nilai 0 sampai dengan 1. Metrik Jaccard stability untuk menilai K yang dipilih adalah baik dan konsisten.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Uji

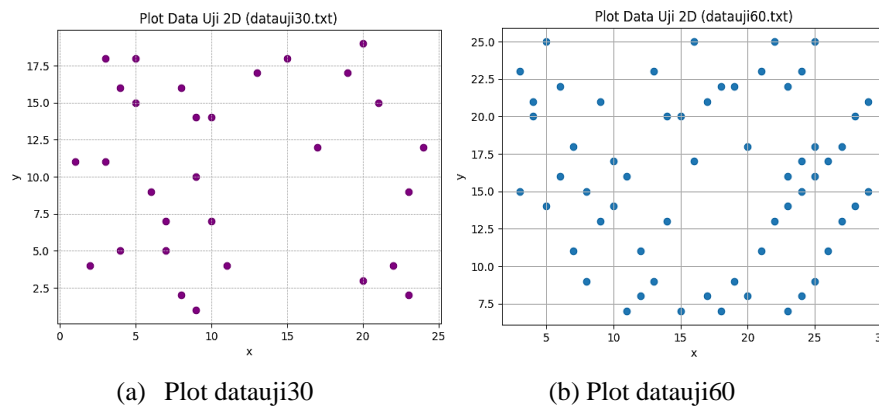
Data uji yang digunakan adalah data dua dimensi yang berjumlah 20, 30, 40, 50, dan 60. Data dapat ditampilkan pada grafik dua dimensi menggunakan koordinat (x,y). Dataset ini disimpan di dalam 5 file text yaitu datauji20, datauji30, datauji40, datauji50, dan datauji60. Bentuk data uji dapat dilihat pada Tabel 1 yang mencatat isi datauji20, datauji30, dan datauji40.

**Tabel 1.** Fitur Data Uji

datauji20.txt		datauji30.txt		datauji40.txt	
No.	Fitur (x, y)	No.	Fitur (x, y)	No.	Fitur (x, y)
1	8, 2	1	8, 2	1	8, 2
2	3, 4	2	3, 18	2	3, 18
3	1, 6	3	1, 11	3	1, 11
4	14, 8	4	17, 12	4	14, 8
5	11, 10	5	10, 7	5	11, 10
6	15, 13	6	6, 9	6	15, 13
7	12, 11	7	9, 14	7	12, 11
8	17, 12	8	7, 5	8	17, 12
9	10, 7	9	9, 1	9	10, 7
10	16, 9	10	4, 5	10	6, 9
11	9, 14	11	5, 15	11	9, 14
12	7, 5	12	8, 16	12	7, 5
13	9, 1	13	9, 10	13	9, 1
14	4, 2	14	7, 7	14	4, 5
15	5, 15	15	10, 14	15	5, 15
16	13, 8	16	22, 4	16	13, 8
17	8, 16	17	20, 19	17	8, 16
18	9, 10	18	23, 9	18	9, 10
19	7, 7	19	24, 12	19	7, 7
20	10, 14	20	23, 2	20	10, 14
		21	20, 3	21	22, 4
		22	21, 15	22	20, 19
		23	15, 18	23	23, 9
		24	19, 17	24	24, 12
		25	2, 4	25	23, 2
		26	3, 11	26	20, 3
		27	13, 17	27	21, 15
		28	4, 16	28	15, 18
		29	5, 18	29	16, 9
		30	11, 4	30	19, 17
				31	2, 4
				32	3, 11
				33	13, 17
				34	14, 6
				35	16, 6

datauji20.txt		datauji30.txt		datauji40.txt	
No.	Fitur (x, y)	No.	Fitur (x, y)	No.	Fitur (x, y)
				36	13, 13
				37	4, 16
				38	14, 5
				39	5, 18
				40	11, 4

Tampilan data (*plotting* data) contoh untuk datauji30 dan datauji60 dapat dilihat pada Gambar 2. Sebaran data terlihat cukup merata tidak terkonsentrasi pada daerah tertentu, sehingga data uji cukup representatif dipakai untuk penelitian ini. Pada datauji30 memiliki jangkauan nilai fitur x antara 1-24, nilai fitur y antara 1-19 yaitu dapat dilihat pada Gambar 2(a). Pada datauji60 memiliki jangkauan nilai fitur x antara 2-28, nilai fitur y antara 7-25 seperti dapat dilihat pada Gambar 2(b).



**Gambar 2.** Plot Data Uji

### 3.2 Hasil Pengujian

Proses memperoleh jumlah kluster data (K) yang optimal menggunakan K-Means cluster dengan mencoba nilai K=1 sampai dengan K=8. Disediakan lima buah dataset, yaitu: datauji20, datauji30, datauji40, datauji50, dan datauji60. Pada metode Elbow (SSE) dan Silhouette score memberi informasi seberapa baik kluster terbentuk, tetapi keduanya tidak selalu menjamin kluster yang baik selalu stabil apabila algoritma diulang. Disini metrik Jaccard stability digunakan untuk menilai K yang dipilih adalah baik dan konsisten.

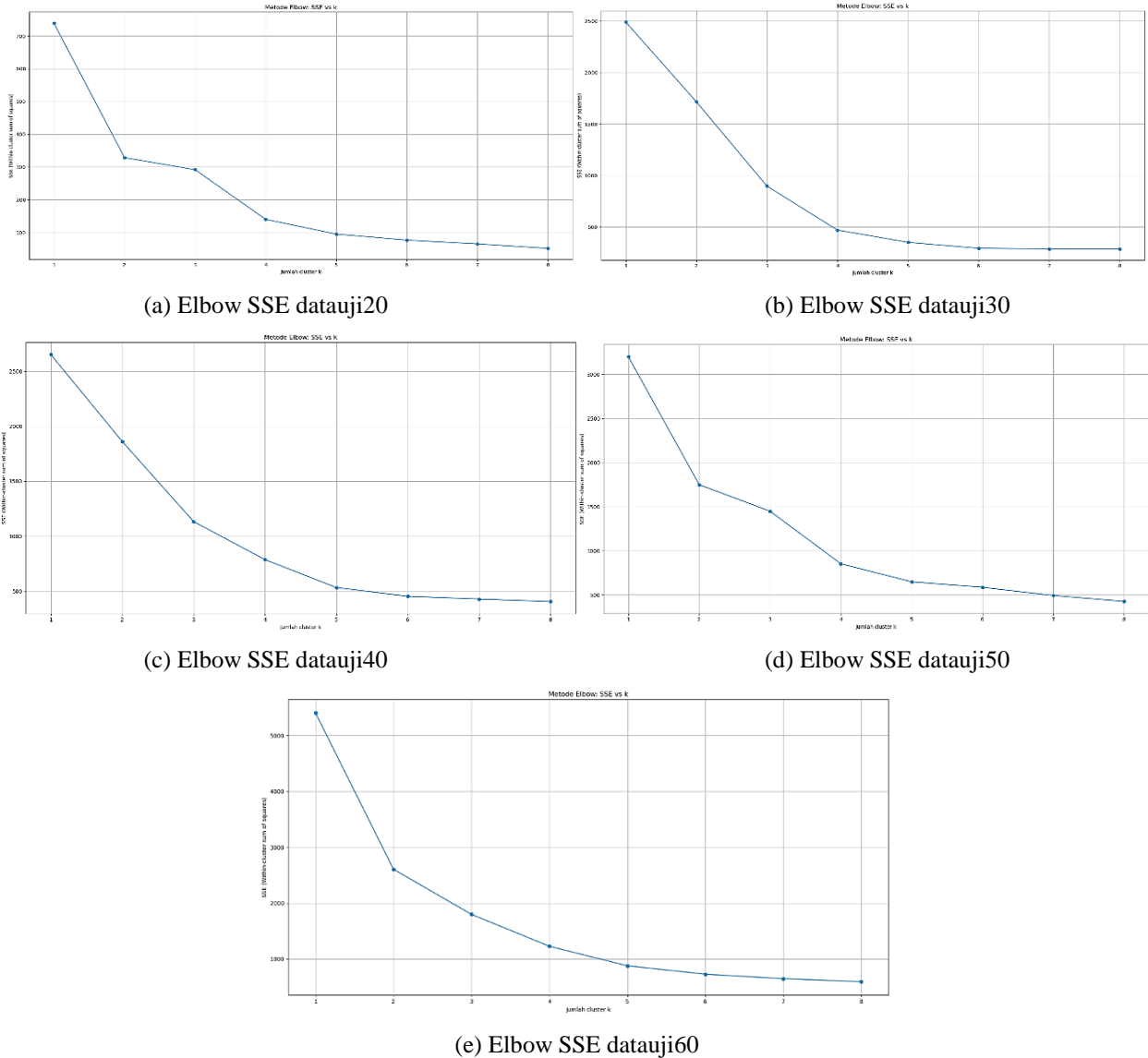
#### 3.2.1 Hasil Metode Elbow (SSE)

Hasil metode Elbow (SSE) dari percobaan lima buah data uji dicatat pada Tabel 2. Tampilan grafis hasil metode ini untuk seluruh data uji dapat dilihat pada Gambar 3.

**Tabel 2.** Hasil Metode Elbow (SSE)

K	datauji20	datauji30	datauji40	datauji50	datauji60
	SSE				
1	739	2489.37	2653.2	3198.8	5405.33
2	328.57	1715.29	1861.94	1747.96	2606.61
3	291.81	897.15	1133.13	1447.59	1799.02
4	140.17	470.01	786.60	851.49	1227.78
5	95.08	351.54	532.54	646.55	878.26
6	76.75	293.64	452.65	584.55	727.84
7	65.17	287.14	427.92	491.53	648.11
8	51.5	286.14	405.74	424.60	593.94

Nilai K yang dipilih adalah yang SSE-nya mulai melandai (*elbow*) seperti dapat diamati melalui Gambar 3. Pada Gambar 3(b) adalah tampilan grafis untuk datauji30, SSE mulai turun tajam dari K=2 ke K=3 lalu penurunannya lebih kecil untuk K=4, K=5 dan seterusnya. Pola pelandaian mulai terlihat di K=4, sehingga K=4 sebagai kandidat terbaik. Pada Gambar 3(c) yaitu tampilan grafis untuk datauji40, SSE mulai turun bertahap dan lebih landai di K=5 dan K=6, lalu penurunannya lebih kecil untuk K=7 dan K=8. Sehingga K=5 atau K=6 menjadi kandidat K yang terbaik. Pada Gambar 3(e) menampilkan grafis datauji60, SSE mulai turun bertahap dan lebih landai di K=2, K=3, dan seterusnya. Sehingga K=2 atau K=3 menjadi kandidat K yang terbaik.



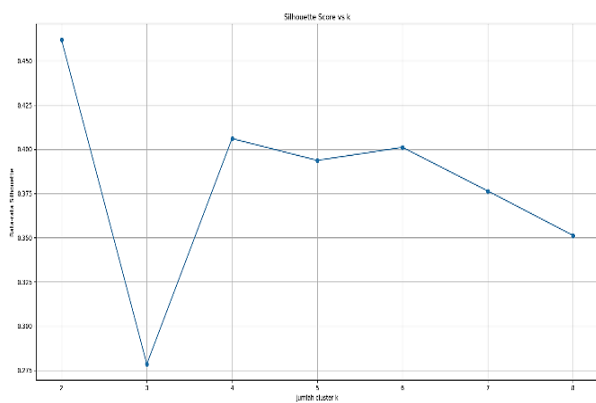
**Gambar 3.** Hasil Metode Elbow SSE

### 3.2.2 Hasil Metode Silhouette score

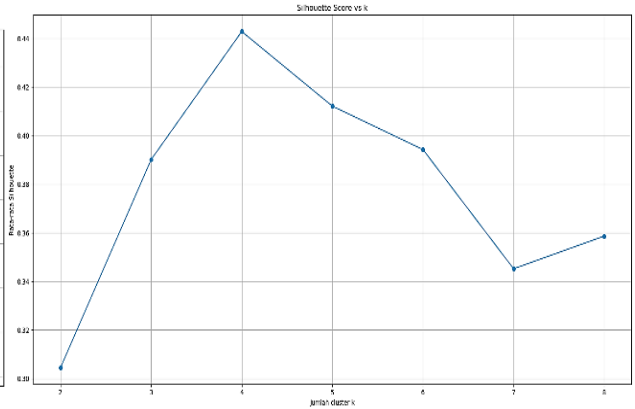
Hasil metode Silhouette score dari percobaan lima buah data uji dicatat pada Tabel 3. Tampilan grafis hasil metode ini untuk seluruh data uji dapat dilihat pada Gambar 4. Dalam pemilihan K, dipilih K dengan Silhouette rata-rata tertinggi yaitu kluster yang paling jelas terpisah.

**Tabel 3.** Hasil Metode Silhouette Score

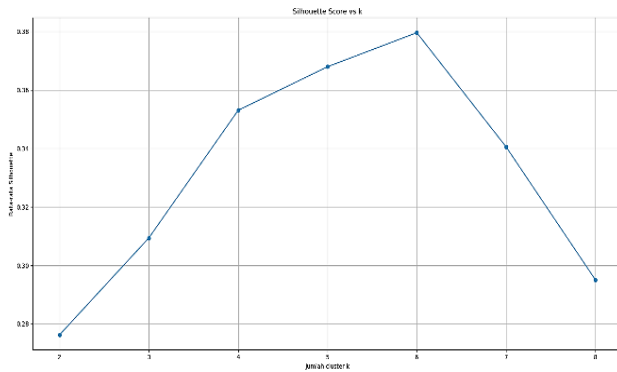
K	datauji20	datauji30	datauji40	datauji50	datauji60
Silhouette score (mean)					
1	-	-	-	-	-
2	0.4619	0.3045	0.2762	0.4014	0.4374
3	0.2785	0.3901	0.3094	0.3399	0.3881
4	0.4061	0.4429	0.3530	0.3604	0.4072
5	0.3937	0.4121	0.3679	0.3749	0.3890
6	0.4011	0.3943	0.3796	0.3693	0.3748
7	0.3764	0.3452	0.3405	0.3536	0.3251
8	0.3513	0.3587	0.2950	0.3427	0.3341



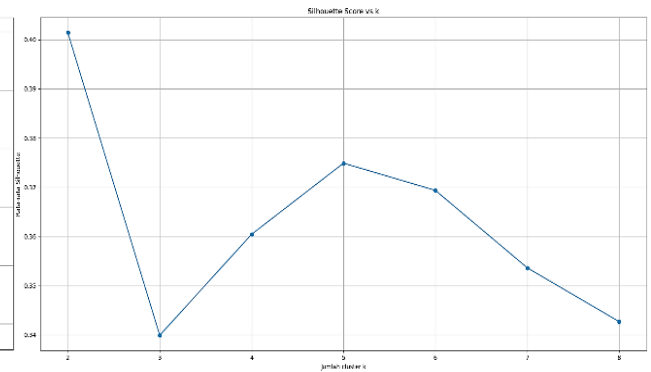
(a) Silhouette Score datauji20



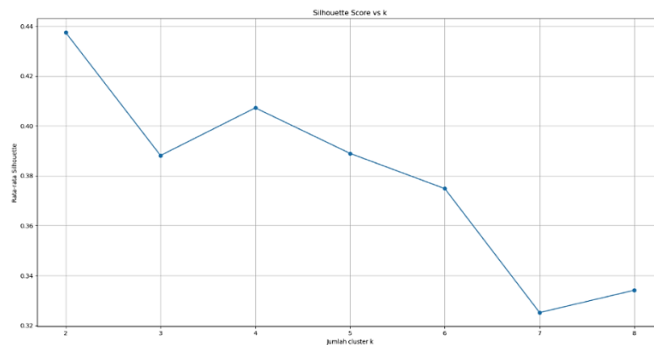
(b) Silhouette Score datauji30



(c) Silhouette Score datauji40



(d) Silhouette Score datauji50



(e) Silhouette Score datauji60

**Gambar 4.** Hasil Metode Silhouette Score

Pada datauji20, nilai *mean* Silhouette tertinggi yaitu 0.4619 untuk K=2, sehingga ini menjadi kandidat K terbaik, dapat dilihat pada Gambar 4(a). Pada datauji30, nilai *mean* Silhouette tertinggi yaitu 0.4429 untuk K=4, ini menjadi kandidat K terbaik, dapat dilihat pada Gambar 4(b). Sedangkan pada datauji40, nilai *mean* Silhouette tertinggi yaitu 0.3796 untuk K=6, ini menjadi kandidat K terbaik, dapat dilihat pada Gambar 4(c).

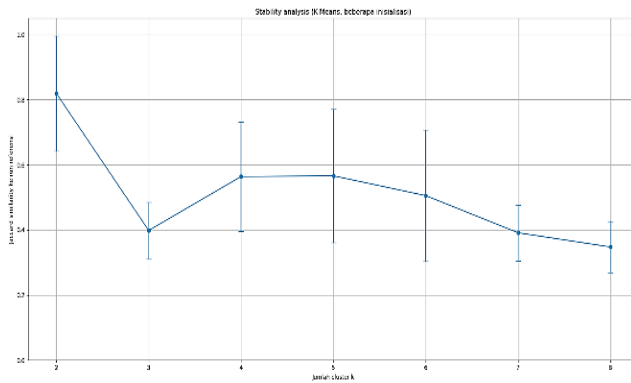
### 3.2.3 Hasil Metode Jaccard Stability

Hasil metode Jaccard stability dari percobaan lima buah data uji dicatat pada Tabel 4. Tampilan grafis hasil metode ini untuk seluruh data uji dapat dilihat pada Gambar 5. Dalam pemilihan K, dipilih K dengan rata-rata Jaccard stability yang tertinggi yang berarti relatif paling stabil.

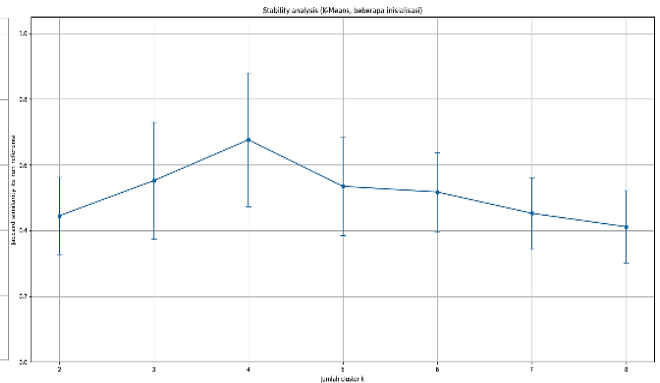
**Tabel 4.** Hasil Metode Jaccard Stability

K	datauji20	datauji30	datauji40	datauji50	datauji60
	Jaccard stability (mean)				
1	-	-	-	-	-
2	0.8190	0.4452	0.4541	0.8332	0.9507
3	0.3983	0.5521	0.4455	0.4539	0.6129
4	0.5637	0.6761	0.4509	0.5497	0.6915

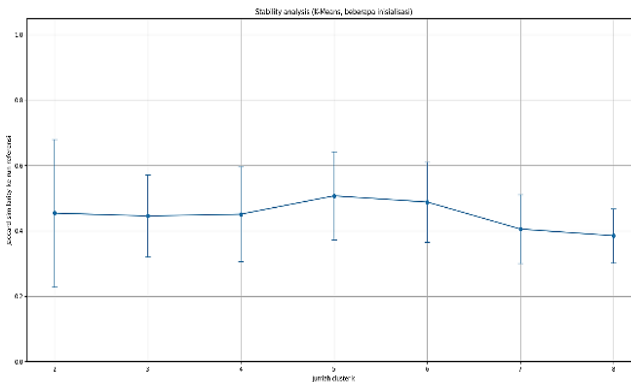
K	datauji20	datauji30	datauji40	datauji50	datauji60
	Jaccard stability (mean)				
5	0.5659	0.5348	0.5072	0.4856	0.5879
6	0.5051	0.5173	0.4881	0.4143	0.5527
7	0.3911	0.4526	0.4056	0.4141	0.4158
8	0.3475	0.4115	0.3851	0.4590	0.3918



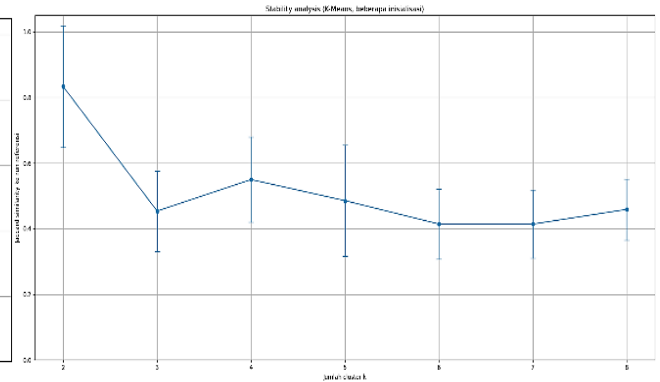
(a) Jaccard Stability pada datauji20



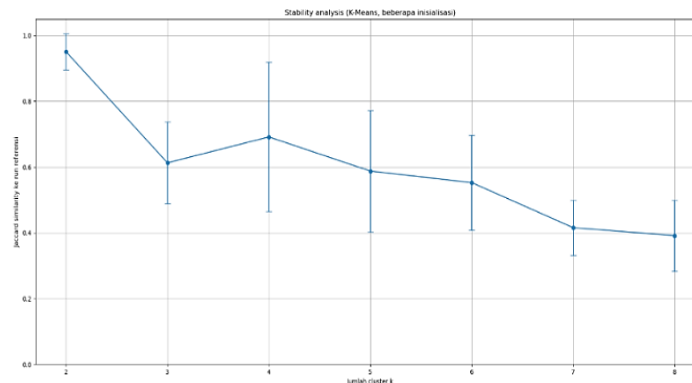
(b) Jaccard Stability pada datauji30



(c) Jaccard Stability pada datauji40



(d) Jaccard Stability pada datauji50



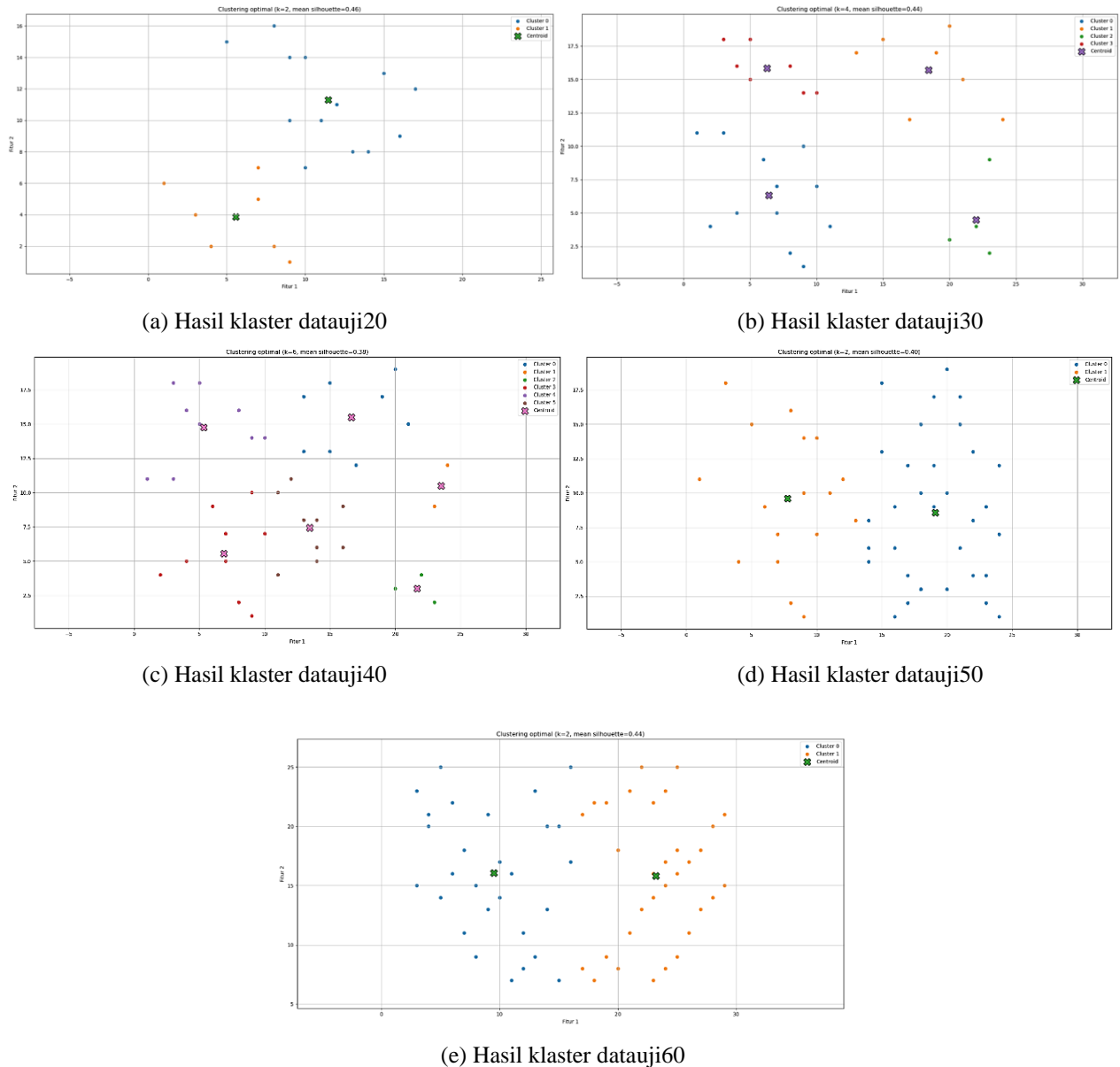
(e) Jaccard Stability cluster pada datauji60

**Gambar 5.** Hasil Metode Jaccard Stability

Pada Gambar 5(c), untuk datauji40 maka nilai Jaccard stability tertinggi yaitu 0.5072 dengan K=5, sedikit lebih tinggi daripada K=6 yaitu 0.4881, sehingga K=5 menjadi kandidat terbaik. Dari sisi stabilitas maka K=5 sedikit lebih unggul, namun dari sisi kualitas maka K=6 adalah lebih baik apabila mempertimbangkan nilai *mean* Silhouette yang tertinggi.

Pada Gambar 5(a), untuk datauji20 maka nilai Jaccard stability tertinggi yaitu 0.819 untuk K=2, sehingga ini menjadi kandidat K terbaik. Pada Gambar 5(d), untuk datauji50 maka nilai Jaccard stability tertinggi adalah 0.8332 untuk K=2. Pada Gambar 5(e), untuk datauji60 maka nilai Jaccard stability tertinggi yaitu 0.9507 untuk K=2, sehingga ini menjadi kandidat K terbaik. Ketiga data uji ini memiliki nilai Jaccard stability yang cukup tinggi yaitu mendekati 1 sehingga hasil kluster relatif stabil.

### 3.3 Hasil Kluster Data



**Gambar 6.** Hasil Kluster Data

Hasil kluster data (*plotting data*) untuk seluruh data uji dapat dilihat pada Gambar 6. Pada Gambar 6(b) dapat dilihat hasil kluster datauji30, kluster yang optimal adalah 4 dengan setiap anggota dicatat pada Tabel 5. Kluster 0, 1, 2, dan 3, masing-masing memiliki anggota 12, 7, 4, dan 7 data. Pada Gambar 6(d) dapat dilihat hasil kluster datauji50, kluster yang optimal adalah 2 dengan setiap anggota dicatat pada Tabel 6. Kluster 0 memiliki anggota 33 dan kluster 1 memiliki anggota 17 data.

Seluruh hasil kluster yang optimal untuk setiap data uji dirangkum pada Tabel 7. Pada datauji20, datauji30, dan datauji40 kluster optimal adalah 2, 4, dan 6. Melalui hasil percobaan diperoleh jumlah kluster terbanyak adalah untuk datauji40 yaitu 6 kluster. Pada datauji 50 dan datauji60 kluster optimal adalah sama yaitu 2 kluster.

**Tabel 5.** Anggota Kluster Datauji30

Kluster 0		Kluster 1		Kluster 2		Kluster 3	
Centroid: (6.42, 6.33)		Centroid (18.43, 15.71)		Centroid (22.0, 4.5)		Centroid (6.29, 15.86)	
No	Fitur (x, y)	No	Fitur (x, y)	No	Fitur (x, y)	No	Fitur (x, y)
1	data# 1: (8, 2)	1	data# 4: (17, 12)	1	data#16: (22, 4)	1	data# 2: (3, 18)
2	data# 3: (1, 11)	2	data#17: (20, 19)	2	data#18: (23, 9)	2	data# 7: (9, 14)
3	data# 5: (10, 7)	3	data#19: (24, 12)	3	data#20: (23, 2)	3	data#11: (5, 15)
4	data# 6: (6, 9)	4	data#22: (21, 15)	4	data#21: (20, 3)	4	data#12: (8, 16)
5	data# 8: (7, 5)	5	data#23: (15, 18)			5	data#15: (10, 14)
6	data# 9: (9, 1)	6	data#24: (19, 17)			6	data#28: (4, 16)



Klaster 0 Centroid: (6.42, 6.33)		Klaster 1 Centroid (18.43, 15.71)		Klaster 2 Centroid (22.0, 4.5)		Klaster 3 Centroid (6.29, 15.86)	
No	Fitur (x, y)	No	Fitur (x, y)	No	Fitur (x, y)	No	Fitur (x, y)
7	data#10: (4, 5)	7	data#27: (13, 17)			7	data#29: (5, 18)
8	data#13: (9, 10)						
9	data#14: (7, 7)						
10	data#25: (2, 4)						
11	data#26: (3, 11)						
12	data#30: (11, 4)						

**Tabel 6.** Anggota Klaster Datauji50

Klaster 0 Centroid (19.09, 8.58)		Klaster 1 Centroid (7.76, 9.59)	
No	Fitur (x, y)	No	Fitur (x, y)
1	data# 4: (14, 8)	1	data# 1: (8, 2)
2	data# 6: (15, 13)	2	data# 2: (3, 18)
3	data# 8: (17, 12)	3	data# 3: (1, 11)
4	data#21: (22, 4)	4	data# 5: (11, 10)
5	data#22: (20, 19)	5	data# 7: (12, 11)
6	data#23: (23, 9)	6	data# 9: (10, 7)
7	data#24: (24, 12)	7	data#10: (6, 9)
8	data#25: (23, 2)	8	data#11: (9, 14)
9	data#26: (20, 3)	9	data#12: (7, 5)
10	data#27: (21, 15)	10	data#13: (9, 1)
11	data#28: (15, 18)	11	data#14: (4, 5)
12	data#29: (16, 9)	12	data#15: (5, 15)
13	data#30: (19, 17)	13	data#16: (13, 8)
14	data#31: (17, 2)	14	data#17: (8, 16)
15	data#32: (18, 3)	15	data#18: (9, 10)
16	data#33: (18, 15)	16	data#19: (7, 7)
17	data#34: (14, 6)	17	data#20: (10, 14)
18	data#35: (16, 6)		
19	data#36: (17, 4)		
20	data#37: (23, 4)		
21	data#38: (14, 5)		
22	data#39: (16, 1)		
23	data#40: (18, 3)		
24	data#41: (21, 17)		
25	data#42: (24, 7)		
26	data#43: (24, 1)		
27	data#44: (18, 10)		
28	data#45: (22, 8)		
29	data#46: (20, 10)		
30	data#47: (22, 13)		
31	data#48: (19, 9)		
32	data#49: (19, 12)		
33	data#50: (21, 6)		

Hasil percobaan menunjukkan bahwa metode Elbow dan Silhouette dapat digunakan secara komplementer dalam menentukan K yang optimal. Kurva SSE memberikan indikasi kandidat K melalui titik pelandaian (*elbow*) yang mencerminkan K telah optimal. Nilai Silhouette menegaskan kualitas struktur klaster berdasarkan kekompakan intra-klaster dan keterpisahan antar-klaster. Pada beberapa data uji, kedua metrik menghasilkan rekomendasi K yang konsisten, sedangkan pada data uji tertentu Elbow dapat memberikan beberapa kandidat sehingga diperlukan validasi menggunakan nilai Silhouette dan Jaccard stability untuk memilih K yang optimal.

Pada Tabel 7 mencatat nilai tertinggi untuk rata-rata Silhouette score setiap datauji. Nilai metrik Elbow dengan nilai SSE yang mulai melandai dicatat pada kolom 5. Pada kolom terakhir mencatat jumlah data anggota setiap klaster yang optimal untuk setiap data uji.

**Tabel 7.** Hasil Klaster

No	Data Uji	n	Jumlah Klaster (K)	Elbow SSE (K)	Mean Silhouette (K)	Jumlah Data Setiap Klaster
1	datauji20	20	2	328.571	0.462	C0=13; C1=7



No	Data Uji	n	Jumlah Kluster (K)	Elbow SSE (K)	Mean Silhouette (K)	Jumlah Data Setiap Kluster
2	datauji30	30	4	470.012	0.443	C0=12; C1=7; C2=4; C3=7
3	datauji40	40	6	452.653	0.380	C0=8; C1=2; C2=3; C3=9; C4=9; C5=9
4	datauji50	50	2	1747.964	0.401	C0=33; C1=17
5	datauji60	60	2	2606.607	0.437	C0=28; C1=32

#### 4.KESIMPULAN

Pemilihan jumlah kluster K yang optimal pada K-Means menggunakan metode Elbow (SSE) dan nilai Silhouette secara komplementer. Penelitian ini menambahkan uji kestabilan kluster menggunakan nilai Jaccard Stability. Kurva Elbow memberikan indikasi kandidat K melalui pola penurunan SSE yang mulai melandai. Pada beberapa data uji, kedua metrik mengarah pada kandidat nilai K yang sama. Pada datauji30 diperoleh K optimal adalah 4, yaitu baik dari SSE maupun nilai Silhouette. Hasil perbandingan antar ukuran data memperlihatkan bahwa semakin beragam sebaran data, kurva SSE dapat menampilkan lebih dari satu bagian yang tampak melandai, sehingga penentuan elbow secara visual berpotensi menjadi kurang relevan. Nilai Silhouette berperan sebagai validator membantu memilih K yang optimal. Pada datauji40, Elbow berada di kisaran K=5-6 dan nilai Silhouette mendukung K=6. Pendekatan yang hanya mengandalkan selisih SSE terbesar tidak sesuai dengan konsep Elbow. Keputusan yang lebih tepat adalah memilih titik perubahan laju penurunan SSE (elbow) kemudian dikonfirmasi dengan nilai Silhouette yang paling tinggi.

#### REFERENCES

- Ashari, I. F., Nugroho, E. D., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*, 7(1), 95–103. <https://doi.org/10.30871/jaic.v7i1.4947>
- Chrisamudra, R. W., & Abdurrahman, S. (2025). Pemanfaatan Algoritma K-Means untuk Klastering Spasial Beban Kasus Pneumonia pada Kelompok Balita di Wilayah dengan Kepadatan Populasi Tinggi. *TIN: Terapan Informatika Nusantara*, 6(6), 761–768. <https://doi.org/10.47065/tin.v6i6.8598>
- Diantika, M. W., Rusgiyono, A., & Saputra, B. A. (2025). Perbandingan Metode Optimasi Silhouette, Elbow, Dan Gap Statistics Dalam Menentukan Nilai K Terbaik Pada Analisis K-Means Clustering. *Jurnal Gaussian*, 14(2), 335–344. <https://doi.org/10.14710/j.gauss.14.2.335-344>
- Guntara, M., & Lutfi, N. (2023). Optimasi Cacah Kluster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coeficient dan Elbow Method. *Jurnal Teknologi Informasi*, 2(1), 43–52. <https://doi.org/10.26798/juti.v2i1.944>
- Han, J., Pei, J., & Tong, H. (2023). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Hartono, B., Eniyati, S., & Hadiono, K. (2023). Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(3), 503–509. <https://doi.org/10.30865/json.v4i3.6021>
- Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN. *KAKIFIKOM*, 6(1), 60–74. <https://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3938>
- Juanita, S., & Cahyono, R. D. (2024). K-Means Clustering With Comparison Of Elbow And Silhouette Methods For Medicines Clustering Based On User Reviews. *Jurnal Teknik Informatika (Jutif)*, 5(1), 283–289. <https://doi.org/10.52436/1.jutif.2024.5.1.1349>
- Kurniawan, Y. I., Anugrah, P. R., Sugihono, R. M., Abimanyu, F. A., & Afuan, L. (2023). Pengelompokan Prioritas Negara Yang Membutuhkan Bantuan Menggunakan Clustering K-Means dengan Elbow Dan Silhouette. *Jurnal Pendidikan dan Teknologi Indonesia*, 3(10), 455–463. <https://doi.org/10.52436/1.jpti.343>
- Kuswardana, D. A., Prasetya, D. A., Trimono, T., & Diyasa, I. G. S. M. (2025). Comparison of Elbow and Silhouette Methods in Optimizing K-Prototype Clustering for Customer Transactions. *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, 12(1), 43–48. <https://doi.org/10.21107/edutic.v12i1.29744>
- Lusiana, V., Amin, I. H. A., & Hartono, B. (2024). Pengaruh Ekstraksi Fitur Tekstur Pada Hasil Klastering Data Citra Buah Menggunakan Metode K-Means Cluster. *Journal of Computer System and Informatics (JoSYC)*, 5(4), 933–941. <https://doi.org/10.47065/josyc.v5i4.5770>
- Maori, N. A., & Evanita, E. (2023). Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2), 277–288. <https://doi.org/10.24176/simet.v14i2.9630>
- Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids. *Jurnal Gaussian*, 13(1), 13–24. <https://doi.org/10.14710/j.gauss.13.1.13-24>
- Ritonga, P. K., & Hasibuan, M. S. (2025). Analisis Perbandingan Silhouette dengan Elbow pada Algoritma K-Means dan DBSCAN. *METIK JURNAL*, 9(1), 64–71. <https://doi.org/10.47002/metik.v9i1.1027>



- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2021(1), 31. <https://doi.org/10.1186/s13638-021-01910-w>
- Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 8(1), 56–65. <https://doi.org/10.30998/string.v8i1.16388>
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. *Entropy*, 23(6), 759. <https://doi.org/10.3390/e23060759>
- Syahkur, M. R., Hartama, D., & Solikhun, S. (2024). Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita. *JST (Jurnal Sains Dan Teknologi)*, 13(3), 487–496. <https://doi.org/10.23887/jstundiksha.v13i3.86419>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining*. Pearson.
- Vania, P., & Sari, B. N. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. *Jurnal Ilmiah Wahana Pendidikan*, 9(21), 547–558. <https://doi.org/10.5281/zenodo.10081332>
- Wicaksono, A. P., Widjaja, S., Nugroho, M. F., & Putri, C. P. (2024). Elbow and Silhouette Methods for K Value Analysis of Ticket Sales Grouping on K-Means. *Sistemasi: Jurnal Sistem Informasi*, 13(1), 28–38. <https://doi.org/10.32520/stmsi.v13i1.2889>