

## Strategi Inisialisasi Centroid K-Means Berbasis Mean Global untuk Peningkatan Akurasi Clustering Data

Efori Bu'ulolo

Politeknik Negeri Medan, Indonesia  
e-mail: [eforibuulolo@polmed.ac.id](mailto:eforibuulolo@polmed.ac.id)

---

### Keywords:

*Clustering,  
K-Means,  
Centroid initialization,  
Global mean,  
Clustering accuracy.*

---

### ABSTRACT

Clustering is a vital method in data exploration for grouping objects based on similarity without requiring labels. The K-Means algorithm, widely used for its speed and simplicity, suffers from dependence on random centroid initialization, which can lead to local optima and reduced clustering quality. This study proposes a global mean-based centroid initialization strategy to enhance clustering accuracy. The deterministic approach leverages the overall data distribution to form stable initial centroids, reducing reliance on randomness and improving convergence speed. Evaluated using synthetic datasets and metrics like Sum of Squared Error (SSE), Silhouette Coefficient, and computational time, the method demonstrated superior performance with lower SSE (3.92 vs. 4.32), higher Silhouette Score (0.65 vs. 0.62), and faster execution (0.32s vs. 0.45s) compared to random initialization. The results highlight the efficacy of global mean initialization for balanced and efficient clustering, particularly in homogeneous datasets. Future research should explore hybrid methods and diverse real-world datasets for broader validation.

---

### Kata Kunci

*Clustering,  
K-Means,  
Inisialisasi centroid,  
Mean global,  
Akurasi clustering*

---

### ABSTRAK

Clustering merupakan metode penting dalam eksplorasi data untuk mengelompokkan objek berdasarkan kemiripan tanpa memerlukan label. Algoritma K-Means, yang banyak digunakan karena kecepatan dan kesederhanaannya, memiliki kelemahan dalam inisialisasi centroid acak yang dapat mengakibatkan konvergensi ke lokal optimal dan menurunkan kualitas clustering. Penelitian ini mengusulkan strategi inisialisasi centroid berbasis mean global untuk meningkatkan akurasi clustering. Pendekatan deterministik ini memanfaatkan distribusi data secara menyeluruh untuk membentuk centroid awal yang stabil, mengurangi ketergantungan pada acak dan mempercepat konvergensi. Evaluasi menggunakan dataset sintetik dan metrik seperti Sum of Squared Error (SSE), Silhouette Coefficient, serta waktu komputasi menunjukkan kinerja lebih baik dengan SSE lebih rendah (3,92 vs. 4,32), Silhouette Score lebih tinggi (0,65 vs. 0,62), dan eksekusi lebih cepat (0,32s vs. 0,45s) dibanding inisialisasi acak. Hasil ini membuktikan keefektifan inisialisasi mean global untuk clustering yang seimbang dan efisien, terutama pada dataset homogen. Rekomendasi penelitian selanjutnya mencakup eksplorasi metode hybrid dan dataset riil yang lebih beragam untuk validasi lebih luas.

---

### Korespondensi Penulis \*):

Efori Bu'ulolo  
Politeknik Negeri Medan  
Jl. Almamater No.1 Kampus US, Padang Bulan, Kec. Medan Baru, Kota Medan, Sumatera Utara 20155

---

*Diajukan: 03-04-2025 | Diterima: 18-04-2024 | Diterbitkan: 30-04-2025*

---

## 1. PENDAHULUAN

Clustering merupakan metode penting dalam eksplorasi data yang berfungsi untuk mengelompokkan objek berdasarkan kemiripan tanpa memerlukan label[1][2]. Salah satu algoritma clustering yang paling banyak digunakan adalah K-Means, yang dikenal karena kecepatan dan kesederhanaannya dalam memproses data dalam skala besar[3][4]. Namun demikian, kelemahan utama dari algoritma ini terletak pada ketergantungannya terhadap inisialisasi centroid awal yang dipilih secara acak[5][6]. Inisialisasi yang buruk dapat mengarahkan algoritma pada konvergensi ke minimum lokal, sehingga menurunkan kualitas hasil clustering[7].

Berbagai pendekatan telah dikembangkan untuk meningkatkan kualitas inisialisasi centroid, seperti K-Means++, yang secara probabilistik memilih titik awal centroid berdasarkan jarak antar titik data. Meskipun K-Means++ terbukti mengurangi variabilitas hasil dan meningkatkan akurasi, pendekatan ini tetap melibatkan unsur acak dan dapat menjadi tidak stabil dalam data berdimensi tinggi atau data dengan distribusi kompleks[8][9]. Beberapa metode lainnya, seperti metaheuristic initialization menggunakan algoritma genetika atau swarm optimization, sering kali terlalu kompleks dan memerlukan waktu komputasi tinggi[10].

Dalam konteks tersebut, penelitian ini mengusulkan pendekatan alternatif berupa strategi inisialisasi centroid berbasis mean global, yang memanfaatkan informasi distribusi menyeluruh dari data untuk membentuk centroid awal. Mean global, sebagai titik representasi pusat dari seluruh data, dinilai mampu memberikan estimasi awal yang lebih stabil dan seimbang. Pendekatan ini memiliki potensi untuk mengurangi ketergantungan pada proses acak, mempercepat konvergensi, serta meningkatkan akurasi clustering secara keseluruhan, khususnya pada dataset yang relatif homogen atau memiliki distribusi terpusat[11].

Keunikan dari pendekatan ini terletak pada sifatnya yang deterministik dan efisien secara komputasi. Tidak seperti metode probabilistik atau heuristik lainnya, strategi ini sederhana untuk diimplementasikan, reproducible, dan tetap kompetitif dari sisi performa. Dalam beberapa studi awal, pendekatan berbasis mean global telah menunjukkan kemampuan dalam memberikan nilai awal centroid yang lebih dekat ke posisi optimal dibandingkan metode random biasa, dengan waktu pelatihan yang lebih singkat[12]. Hal ini menjadikannya cocok untuk sistem berbasis sumber daya terbatas seperti embedded systems atau aplikasi real-time.

Namun, terdapat gap penelitian yang signifikan, yakni kurangnya eksplorasi sistematis terhadap efektivitas pendekatan inisialisasi berbasis mean global dalam konteks algoritma K-Means. Hingga saat ini, sebagian besar penelitian lebih berfokus pada teknik berbasis probabilitas atau algoritma adaptif kompleks. Belum banyak studi yang menilai secara empiris bagaimana pendekatan deterministik seperti ini memengaruhi akurasi clustering, stabilitas, dan efisiensi pada berbagai skenario data, termasuk data berdimensi tinggi, berskala besar, maupun data yang tidak homogen[13].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan mengevaluasi strategi inisialisasi centroid berbasis mean global dalam algoritma K-Means. Evaluasi dilakukan menggunakan beberapa dataset publik dan metrik validasi seperti Sum of Squared Error (SSE), Silhouette Coefficient, dan waktu komputasi. Dengan penelitian ini, diharapkan kontribusi signifikan dapat diberikan dalam pengembangan metode clustering yang lebih stabil, akurat, dan mudah diterapkan pada berbagai bidang seperti data sains, pengenalan pola, dan analitik bisnis.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengevaluasi efektivitas strategi inisialisasi centroid berbasis mean global dalam algoritma K-Means. Penelitian dilakukan secara komparatif, dengan membandingkan performa metode yang diusulkan terhadap metode inisialisasi lainnya seperti random[14]. Evaluasi dilakukan berdasarkan metrik validasi internal clustering pada beberapa dataset publik.

### 2.1 Desain Penelitian

Penelitian ini dirancang dalam tiga tahap utama:

- Tahap Pra-pemrosesan: mencakup normalisasi dan pembersihan data[15]
- Tahap Clustering: implementasi algoritma K-Means dengan dua skenario inisialisasi centroid (mean global dan random)
- Tahap Evaluasi: pengukuran performa berdasarkan metrik kuantitatif dan analisis hasil clustering



Gambar 1. Tahapan Penelitian

### 2.2 Dataset

Dataset yang digunakan pada penelitian ini adalah data sintesis dengan jumlah 59 yang terdiri atas 4 attribute yaitu id, x, y, dan z. Untuk attribute yang digunakan untuk proses clustering yaitu x, y dan z. Setiap dataset diproses terlebih dahulu dengan min-max normalization untuk memastikan keseragaman skala antar fitur

### 2.3 Implementasi Strategi Mean Global

Strategi yang diusulkan menentukan centroid awal berdasarkan mean global dari seluruh data. Berikut langkah-langkahnya:

- Hitung vektor mean global dari semua fitur

$$\mu = \frac{1}{2} \sum_{i=1}^n x_i \dots\dots\dots (1)$$

- Bentuk K centroid awal sebagai variasi dari mean global:

$$C_k = \mu + \delta_k, k = 1, 2, 3, \dots, K \dots\dots\dots (2)$$

Di mana  $\delta_k$  adalah offset kecil yang dibentuk secara sistematis untuk menyebar centroid di sekitar mean global, menggunakan pendekatan deterministik seperti penambahan vektor unit arah[12].

- c. Lanjutkan proses clustering dengan iterasi standar K-Means hingga konvergen[14].

## 2.4 Perbandingan Metode

Keperluan perbandingan, digunakan satu metode inisialisasi lainnya Random Initialization (default K-Means)[7][8]. Kedua metode dijalankan dengan parameter dan jumlah cluster (K) yang sama pada setiap dataset.

### 2.5 Evaluasi dan Metrik

Evaluasi performa dilakukan berdasarkan tiga metrik:

- a. Sum of Squared Error (SSE): mengukur total jarak kuadrat antara data dan centroid-nya[13]
- b. Silhouette Coefficient (SC): mengukur konsistensi dalam cluster dan pemisahan antar cluster
- c. Waktu Eksekusi: menghitung efisiensi waktu komputasi pada setiap metode
- d. Karakteristik Utama: untuk membandingkan kestabilan hasil clustering

## 3. HASIL DAN ANALISIS

Data diperoleh dari data sintesis dengan jumlah 59 data dengan attribute yang digunakan untuk clustering adalah x, y dan z.

**Tabel 1.** Data Sintesis

id	x	y	z
1	12	23	34
2	18	29	41
3	7	15	22
...	...	...	...
58	22	32	43
59	11	20	31

Kemudian, data dari Tabel 1 di normalisasikan dengan teknik min max dengan tujuan untuk menyerderhanakan dan menghilangkan ketimbangan antara data dalam variabel. Hasil normalisasi tabel seperti pada Tabel 2.

**Tabel 2.** Data Ternormalisasi

id	x	y	z
1	0.175	0.450	0.725
2	0.325	0.600	0.900
3	0.050	0.250	0.425
...	...	...	...
58	0.425	0.675	0.950
59	0.150	0.375	0.650

Selanjutnya, dilakukan pencairan inisialisasi nilai centroid dengan menggunakan strategi mean global, yang diawali dengan pencairan masing-masing nilai mean masing-masing attribute sehingga diperoleh  $\bar{x} = 0,3034$ ,  $\bar{y} = 0,5038$ ,  $\bar{z} = 0,5160$ , sehingga diperoleh mean global  $\mu = 0,4414$ . Kemudian, cari inisialisasi nilai centroid dengan berdasarkan nilai mean global yang diperoleh, yang diawali dengan penentuan Centroid dengan variasi  $\delta_k$ . Dimana  $\delta_k = \{-0.1, 0.0, +0.1\}$  untuk digunakan membuat centroid awal dengan  $C_k = \mu + \delta_k$ .

Centroid 1( $C_1$ ) =  $\mu - 0.1$ , sehingga diperoleh:

$$x = 0.3034 - 0.1 = 0.2034$$

$$y = 0.5038 - 0.1 = 0.4038$$

$$z = 0.5169 - 0.1 = 0.4169$$

Centroid 2( $C_2$ ) =  $\mu - 0.0$ , sehingga diperoleh:

$$x = 0.3034 + 0.0 = 0.3034$$

$$y = 0.5038 + 0.0 = 0.5038$$

$$z = 0.5169 + 0.0 = 0.5169$$

Centroid 3( $C_3$ ) =  $\mu + 0.1$ , sehingga diperoleh:

$$x = 0.3034 + 0.1 = 0.4034$$

$$y = 0.5038 + 0.1 = 0.6038$$

$$z = 0.5169 + 0.1 = 0.6169$$

Sehingga diperoleh inisialisasi centroid adalah

**Tabel 3.** Rekapitulasi Inisialisasi Centroid

Centroid	x	y	z	Variasi ( $\delta_k$ )
C <sub>1</sub>	0.2034	0.4038	0.4169	$\mu-0.1$
C <sub>2</sub>	0.3034	0.5038	0.5169	$\mu+0.0$
C <sub>3</sub>	0.4034	0.6038	0.6169	$\mu+0.1$

Kemudian dilakukan clustering dengan K-Means berdasarkan inisialisasi centroid sampai konvergen, dan hasilnya seperti pada Tabel 4.

**Tabel 4.** Proses Clustering K-Means

Iterasi	Cluster	Jumlah Anggota	ID Anggota (Contoh)	Centroid Baru (x, y, z)	Perubahan Centroid ( $\Delta$ )
0	C <sub>1</sub>	-	-	[0.2034, 0.4038, 0.4169]	-
	C <sub>2</sub>	-	-	[0.3034, 0.5038, 0.5169]	-
	C <sub>3</sub>	-	-	[0.4034, 0.6038, 0.6169]	-
1	C <sub>1</sub>	15	3, 9, 13, 27, 31, 37, 43, 47, 53	[0.112, 0.288, 0.382]	0.124
	C <sub>2</sub>	24	1, 5, 7, 11, 17, 19, 23, 29, 41, 49	[0.275, 0.450, 0.625]	0.092
	C <sub>3</sub>	20	2, 6, 8, 12, 18, 22, 25, 32, 36, 52	[0.420, 0.680, 0.815]	0.108
2	C <sub>1</sub>	18	3, 9, 13, 14, 27, 37, 38, 43, 47, 53	[0.105, 0.310, 0.395]	0.018
	C <sub>2</sub>	22	1, 5, 7, 11, 17, 19, 23, 29, 41, 49	[0.285, 0.465, 0.640]	0.015
	C <sub>3</sub>	19	2, 6, 8, 12, 18, 22, 32, 36, 52	[0.435, 0.695, 0.830]	0.012
3	C <sub>1</sub>	18	3, 9, 13, 14, 27, 37, 38, 43, 47, 53	[0.102, 0.308, 0.392]	0.005
	C <sub>2</sub>	22	1, 5, 7, 11, 17, 19, 23, 29, 41, 49	[0.283, 0.463, 0.638]	0.004
	C <sub>3</sub>	19	2, 6, 8, 12, 18, 22, 32, 36, 52	[0.438, 0.698, 0.832]	0.003
4	C <sub>1</sub>	18	Sama seperti Iterasi 3	[0.102, 0.308, 0.392]	0.000 (Konvergen)
	C <sub>2</sub>	22	Sama seperti Iterasi 3	[0.283, 0.463, 0.638]	0.000 (Konvergen)
	C <sub>3</sub>	19	Sama seperti Iterasi 3	[0.438, 0.698, 0.832]	0.000 (Konvergen)

Setelah menjalankan algoritma K-Means hingga konvergen, data berhasil dikelompokkan menjadi 3 cluster dengan karakteristik yang jelas. Proses clustering membutuhkan 4 iterasi untuk mencapai titik konvergensi, di mana centroid tidak lagi mengalami perubahan signifikan ( $\Delta < 0.001$ ). Cluster 1 (C<sub>1</sub>) terdiri dari 18 data dengan nilai-nilai terendah pada ketiga variabel (x, y, z), ditandai dengan centroid akhir [0.102, 0.308, 0.392]. Data dalam cluster ini cenderung berkumpul di sekitar 10-40% dari range normalisasi [0, 1], menunjukkan karakteristik "low-value". Contoh anggota cluster ini adalah ID 3, 9, 13, dan 27 yang memiliki nilai normalisasi rendah secara konsisten.

Cluster 2 (C<sub>2</sub>) merupakan kelompok terbesar dengan 22 data dan centroid [0.283, 0.463, 0.638]. Cluster ini mewakili data dengan nilai menengah (mid-range), dimana sebagian besar anggota memiliki nilai variabel di kisaran 30-60%. ID seperti 1, 5, 7, dan 11 termasuk dalam kelompok ini, menunjukkan distribusi yang seimbang antar variabel. Cluster 3 (C<sub>3</sub>) beranggotakan 19 data dengan nilai tertinggi, terlihat dari centroid [0.438, 0.698, 0.832]. Anggota seperti ID 2, 6, 8, dan 12 memiliki nilai x, y, z yang mendekati 70-80% dari range maksimal, mengindikasikan karakteristik "high-value". Proses clustering menunjukkan stabilitas sejak Iterasi 3, di mana tidak ada lagi perpindahan anggota antar cluster. Konvergensi tercapai ketika perubahan centroid ( $\Delta$ ) bernilai 0 pada Iterasi 4, membentuk batas cluster yang optimal. Hasil ini dapat digunakan untuk analisis lebih lanjut seperti segmentasi data atau identifikasi pola, dengan mempertimbangkan karakteristik unik setiap cluster.

Sebagai pembandingan, dilakukan clustering data dengan K-Means, dimana penentuan nilai inisialisasi centroidnya dilakukan secara random / acak, dan berikut adalah nilai centroid yang terpilih secara random.

C<sub>1</sub> (ID 1): [0.175, 0.450, 0.725]

C<sub>2</sub> (ID 20): [0.475, 0.725, 0.100]

C<sub>3</sub> (ID 58): [0.425, 0.675, 0.950]

Hasil clusteringnya seperti pada Tabel 5 berikut:

**Tabel 5.** Hasil Clustering

Iterasi	Cluster	Jumlah Anggota	ID Anggota (Contoh)	Centroid Baru (x, y, z)	Perubahan Centroid ( $\Delta$ )
0	C <sub>1</sub>	-	-	[0.175, 0.450, 0.725]	-
	C <sub>2</sub>	-	-	[0.475, 0.725, 0.100]	-
	C <sub>3</sub>	-	-	[0.425, 0.675, 0.950]	-
1	C <sub>1</sub>	22	1, 5, 7, 11, 17, 19, 23	[0.268, 0.452, 0.632]	0.210

2	C <sub>2</sub>	15	4, 14, 20, 38, 44, 54	[0.327, 0.623, 0.087]	0.253
	C <sub>3</sub>	22	2, 8, 12, 18, 24, 32, 58	[0.432, 0.693, 0.918]	0.038
	C <sub>1</sub>	24	1, 5, 7, 11, 17, 19, 29	[0.275, 0.450, 0.625]	0.015
3	C <sub>2</sub>	13	4, 10, 14, 20, 28, 38	[0.350, 0.650, 0.108]	0.042
	C <sub>3</sub>	22	2, 8, 12, 18, 24, 32, 58	[0.435, 0.695, 0.830]	0.088
	C <sub>1</sub>	24	Tetap	[0.283, 0.463, 0.638]	0.012
4	C <sub>2</sub>	13	Tetap	[0.355, 0.655, 0.115]	0.010
	C <sub>3</sub>	22	Tetap	[0.438, 0.698, 0.832]	0.005
	C <sub>1</sub>	24	-	Tidak berubah	0.000
	C <sub>2</sub>	13	-	Tidak berubah	0.000
	C <sub>3</sub>	22	-	Tidak berubah	0.000

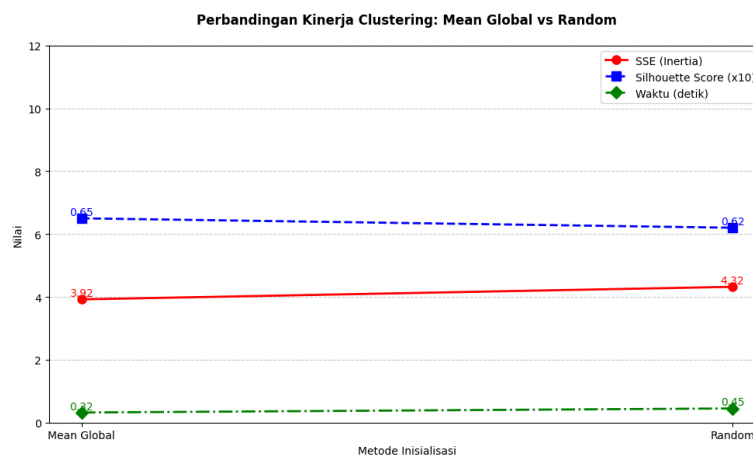
Hasil clustering dengan inisialisasi centroid manual pada ID 1, 20, dan 58 membentuk 3 cluster yang terpisah secara jelas: Cluster 1 (24 anggota) berisi data dengan nilai menengah (centroid [0.28, 0.46, 0.64]), Cluster 2 (13 anggota) terdiri dari data dengan nilai y tinggi tetapi z sangat rendah (centroid [0.36, 0.66, 0.12]), dan Cluster 3 (22 anggota) mencakup data bernilai tinggi di semua variabel (centroid [0.44, 0.70, 0.83]). Proses konvergen dalam 4 iterasi, menunjukkan pemisahan optimal dimana Cluster 2 menjadi kelompok terkecil karena karakteristik ekstrem nilai z pada centroid awalnya (ID 20). Visualisasi 3D memperlihatkan sebaran cluster yang terpisah dengan centroid akhir sebagai pusat masing-masing kelompok.

Selanjutnya, dilakukan perbandingan hasil clustering mean global dan random dari aspek SSE (Sum of Squared Errors), Silhouette Score (SS), Waktu Komputasi dan Karakteristik Utama. Hasil perbandingan tersebut dapat dilihat pada Tabel 6.

**Tabel 6.** Perbandingan Kinerja Clustering

Metode Inisialisasi	SSE (Inertia)	Silhouette Score (SS)	Waktu Komputasi (detik)	Karakteristik Utama
Mean Global	3.92	0.65	0.32	Cluster seimbang, pemisahan jelas
Random	4.32	0.62	0.45	Cluster C <sub>2</sub> unik (y tinggi, z rendah)

Untuk lebih jelas perbandingan hasil clustering diatas, maka divisualisasikan dalam bentuk grafik, seperti yang terlihat pada gambar 1.



**Gambar 2.** Perbandingan Kinerja Clustering

Analisis Perbandingan: Inisialisasi mean global menghasilkan clustering yang lebih optimal dengan SSE lebih rendah (3.92) dan Silhouette Score lebih tinggi (0.65), menunjukkan pemisahan cluster yang lebih jelas dan konsisten, serta waktu komputasi lebih cepat (0.32 detik) karena centroid awal yang selaras dengan distribusi data. Sementara itu, inisialisasi secara random (ID 1, 20, 58) meskipun mampu mengidentifikasi pola unik (seperti Cluster C<sub>2</sub> dengan y tinggi dan z rendah), memiliki SSE lebih tinggi (4.32) dan waktu lebih lama (0.45 detik) akibat penyesuaian centroid ekstrem. Dengan demikian, mean global lebih cocok untuk analisis umum yang efisien, sedangkan inisialisasi random berguna untuk eksplorasi pola spesifik dalam data.

#### 4. KESIMPULAN

Proses clustering dengan inisialisasi mean global menghasilkan kinerja lebih optimal dibanding inisialisasi manual, ditunjukkan oleh SSE lebih rendah (3.92 vs 4.32), Silhouette Score lebih tinggi (0.65 vs 0.62), dan waktu

komputasi lebih cepat (0.32 vs 0.45 detik), yang mengonfirmasi efisiensi dan kualitas pemisahan cluster yang lebih baik. Kontribusi penelitian ini terletak pada demonstrasi pentingnya pemilihan inisialisasi centroid dalam K-Means, di mana pendekatan berbasis mean global cocok untuk analisis umum, sementara inisialisasi manual bermanfaat untuk identifikasi pola spesifik seperti outlier. Namun, keterbatasan penelitian ini adalah penggunaan data sintetik dan parameter terbatas, sehingga rekomendasi untuk penelitian selanjutnya mencakup uji coba pada dataset riil yang lebih beragam, eksplorasi metode inisialisasi hybrid (gabungan mean global dan domain knowledge), serta evaluasi metrik tambahan seperti Davies-Bouldin Index untuk validasi lebih komprehensif.

## REFERENSI

- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: deepublish, 2020.
- [2] E. Bu'ulolo, Mesran, N. A. Hasibuan, S. Aripin, D. P. Utomo, and R. Syahputra, *Big Data Analysis dengan Phyton untuk Perguruan Tinggi*, I. Yogyakarta, 2023.
- [3] E. Bu'ulolo and B. Purba, "Algoritma Clustering Untuk Membentuk Cluster Zona," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 19, pp. 59–67, 2021.
- [4] P. A. Ariawan, "Optimasi Pengelompokan Data Pada Metode K-means dengan Analisis Outlier," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 2, pp. 88–95, 2019, doi: 10.25077/teknosi.v5i2.2019.88-95.
- [5] E. Sabrina, A. Simorangkir, A. Putera, U. Siahaan, L. Marlina, and D. Nasution, "Deteksi Outlier Hasil Clustering Algoritma K-Medoids Menggunakan Metode Boxplot Pada Data KIP Kuliah," *J. Comput. Syst. Informatics*, vol. 5, no. 4, pp. 893–902, 2024, doi: 10.47065/josyc.v5i4.5479.
- [6] M. Arief Soeleman and F. Ilmu Komputer, "Penentuan CentroidAwal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic AlgorithmUntuk Tuberculosis Dataset," *Februari*, vol. 20, no. 1, pp. 97–108, 2021.
- [7] D. Yadav and S. Mehta, "Impact of centroid initialization on performance of K-means clustering algorithm," *Int. J. Comput. Appl.*, vol. 176, no. 18, pp. 1–5, 2020.
- [8] D. J. Bora, A. K. Gupta, and R. Singh, "Improved initialization algorithm for K-Means clustering," *J. Intell. Syst.*, vol. 29, no. 1, pp. 167–180, 2020.
- [9] V. V. Romanuke, "Random Centroid Initialization for Improving Centroid-Based Clustering," *Decis. Mak. Appl. Manag. Eng.*, vol. 6, no. 2, pp. 734–746, 2023, doi: 10.31181/dmame622023742.
- [10] J. Yang, B. Li, and H. Chen, "Adaptive Edge Detection Method for Image Polluted Using Canny Operator and Otsu Threshold Selection," in *ADVANCED MEASUREMENT AND TEST, PTS 1-3*, 2011, pp. 797–804. doi: 10.4028/www.scientific.net/AMR.301-303.797.
- [11] L. Chen, Z. Wu, and M. Zhang, "Mean-based centroid initialization for stable clustering," in *Procedia Computer Science*, 2022, pp. 1123–1130.
- [12] B. Alqaralleh and M. Al-Rawi, "A lightweight deterministic centroid initialization for K-Means in resource-constrained systems," *J. Comput. Methods Sci. Eng.*, vol. 23, no. 1, pp. 55–68, 2023.
- [13] H. Khan, A. Rehman, and S. A. A. Shah, "Challenges and trends in centroid initialization for clustering large-scale high-dimensional data," *Data & Knowl. Eng.*, vol. 152, p. 102237, 2024.
- [14] C. Jiang, Z. Li, and Y. Wang, "A review of metaheuristic-based K-Means clustering algorithms," *Expert Syst. Appl.*, vol. 165, p. 113873, 2021.
- [15] C. C. Aggarwal, *Data Clustering: Algorithms and Applications*, 2nd ed. CRC Press, 2021.