RESEARCH ARTICLE • OPEN ACCESS

Pengelompokan Desa di Jawa Barat Berdasarkan Indeks Desa Membangun (IDM) Menggunakan ALgoritma Clustering Large Application (CLARA)

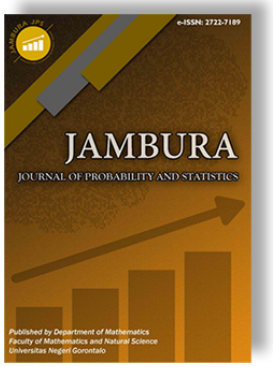
Muhammad Rifqi Hendriawan, and Reny Rian Marliana

Volume 6, Issue 1, Pages 35–41, May 2025

Received 09 September 2024, Revised 01 May 2025, Accepted 31 May 2025, Published Online 31 May 2025

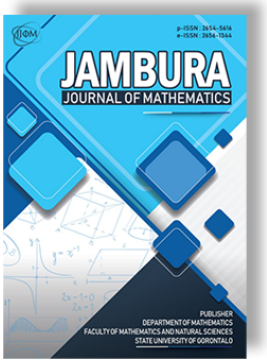
To Cite this Article : M. R. Hendriawan, and R. R. Marliana,“ Pengelompokan Desa di Jawa Barat Berdasarkan Indeks Desa Membangun (IDM) Menggunakan ALgoritma Clustering Large Application (CLARA) ”, *Jambura J. Probab. Stat.*, vol. 6, no. 1, pp. 35–41, 2025, *https://doi.org/10.34312/jjps.v4i1.27450*

JOURNAL INFO • JAMBURA JOURNAL OF PROBABILITY AND STATISTICS



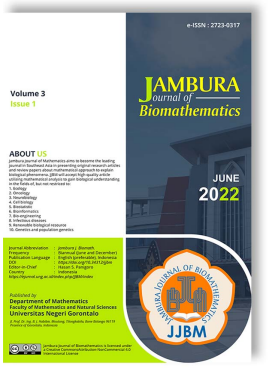
Homepage	:	https://ejurnal.ung.ac.id/index.php/jps/index
Journal Abbreviation	:	Jambura J. Probab. Stat.
Frequency	:	Biannual (May and November)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.34312/jjps
Online ISSN	:	2722-7189
Editor-in-Chief	:	Ismail Djakaria
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jps/oai
Google Scholar ID	:	kWdujzMAAAAJ
Email	:	redaksi.jjps@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

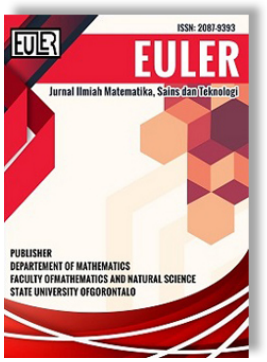


Jambura Journal of Mathematics

Jambura Journal of Mathematics Education



Jambura Journal of Biomathematics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Jambura Journal of Probability and Statistics, Volume 6, Issue 1, Pages 35–41, May 2025 https://doi.org/10.34312/jjps.v4i1.27450

Research Article

Check for updates

Pengelompokan Desa di Jawa Barat Berdasarkan Indeks Desa Membangun (IDM) Menggunakan ALgoritma Clustering Large Application (CLARA)

Muhammad Rifqi Hendriawan [1], **Reny Rian Marliana** [2*]

[1,2] *Program Studi Statistika, Fakultas MIPA, Universitas Islam Bandung*

ARTICLE HISTORY
Received 09 September 2024
Revised 01 May 2025
Accepted 31 May 2025
Published 31 May 2025

KATA KUNCI
CLARA
Clustering
Indeks Desa Membangun.

KEYWORDS
CLARA
Clustering
Development Village Index.

ABSTRAK. *Indeks Desa Membangun (IDM) adalah alat penting untuk mengukur perencanaan dan pembangunan desa, mendukung pemerintah dalam menangani desa tertinggal dan meningkatkan desa mandiri. IDM mengukur pembangunan melalui tiga kategori, yaitu Indeks Ketahanan Ekonomi (IKE), Indeks Ketahanan Sosial (IKS), dan Indeks Ketahanan Lingkungan (IKL). Tujuan penelitian ini adalah mengelompokkan desa-desa di Provinsi Jawa Barat menggunakan algoritma Clustering Large Application (CLARA). Hasil pengelompokkan digunakan untuk mempermudah identifikasi desa tertinggal. Algoritma CLARA merupakan pengembangan dari K-Medoids, digunakan untuk analisis clustering non-hierarki dengan medoid sebagai pusat cluster. Data yang digunakan adalah IKE, IKS dan IKL dari desa-desa di Provinsi Jawa Barat tahun 2023. Hasil analisis menghasilkan banyaknya cluster yang ideal untuk Desa di Provinsi Jawa Barat adalah tiga cluster yang dikategorikan sebagai desa mandiri (cluster 1), desa maju (cluster 2), dan desa berkembang (cluster 3). Berdasarkan Kemendes terdapat 1856 desa pada cluster 1, 2494 desa pada cluster 2, dan 961 desa pada cluster 3. Sedangkan pengelompokan berdasarkan algoritma CLARA terdapat 1644 desa pada cluster 1, 1573 desa pada cluster 2, dan 2094 desa pada cluster 3. Nilai silhouette coefficient global (SC) pada algoritma CLARA menunjukan nilai SC sebesar 0,3614. Ini menandakan bahwa kriteria pengelompokan termasuk kedalam kategori yang lemah, karena berada diantara nilai 0,26-0,50.*

ABSTRACT. *The Village Development Index (IDM) is an important instrument for measuring village planning and development, supporting the government in dealing with underdeveloped villages and advanced villages. IDM assesses development through three indexes, namely the Economic Resilience Index (IKE), Social Resilience Index (IKS), and Environmental Resilience Index (IKL). The aim of the study is to establish clusters of villages in West Java Province using the Clustering Large Application (CLARA) algorithm. This clustering was used to facilitate the identification of underdeveloped villages. The CLARA algorithm, which is a development of K-Medoids, is used for non-hierarchical clustering analysis with medoids as the cluster center. This study used data of IKE, IKS, and IKL from villages in West Java in 2023. The results produce three clusters, namely independent villages (cluster 1), developed villages (cluster 2), and developing villages (cluster 3). Based on the Ministry of Village, there are 1856 villages in cluster 1, 2494 villages in cluster 2, and 961 villages in cluster 3. Meanwhile, based on the CLARA algorithm, there are 1644 villages in cluster 1, 1573 villages in cluster 2, and 2094 villages in cluster 3. The global silhouette coefficient (SC) value in the CLARA algorithm shows an SC value of 0,3614. This indicates that the grouping criteria are included in the weak category, because they are between the values of 0,26-0,50.*

1. Pendahuluan

Sejak awal kemerdekaan Indonesia, pembangunan desa telah menjadi fokus utama pemerintah. Tujuan pembangunan desa telah dinyatakan secara normatif dalam Undang-Undang Nomor 6 Tahun 2014 tentang desa (selanjutnya disebut Undang-Undang Desa). Berdasarkan undang-undang tersebut, tujuan Pembangunan desa adalah untuk meningkatkan kesejahteraan masyarakat desa, kualitas hidup manusia dan menanggulangi kemiskinan. Pembangunan desa merupakan salah satu fondasi utama dalam membangun negara yang kuat dan stabil, langkah-langkah strategis seperti pembangunan infrastruktur, menjaga kualitas lingkungan, akses pendidikan dan kesehatan, serta memperkuat ekonomi merupakan bentuk komitmen yang telah dilakukan oleh pemerintah. Namun pembangunan desa bukanlah hal yang harus dilakukan oleh pemerintah saja, tetapi menjadi perhatian dan tanggung jawab dari seluruh elemen masyarakat. Masyarakat desa sendiri memiliki peranan yang penting dalam mengidentifikasi, merencanakan, dan melaksanakan berbagai program pembangunan sesuai dengan kondisi dan potensi desa mereka sendiri. Di sisi lain, partisipasi aktif dari berbagai pihak seperti organisasi non-pemerintahan, swasta, dan relawan sangat

* Corresponding Author.

Email : *renyrianmarliana@unisba.ac.id* (
) Homepage : https://ejurnal.ung.ac.id/index.php/jps/index / E-ISSN : 2722-7189

dibutuhkan untuk mendukung proses pembangunan suatu desa. Dengan demikian adanya kerjasama dan kolaborasi dari pemerintah dan berbagai *stakeholders* ini, upaya pembangunan desa dapat menjadi lebih efektif dan berkelanjutan, serta mampu memenuhi kebutuhan dan aspirasi masyarakat.

Pembangunan pedesaan merupakan sebuah kunci keberhasilan pembangunan desa yang akan berdampak langsung terhadap keberhasilan pembangunan perekonomian negara. Sesuai dengan Rencana Kerja Pemerintah (RKP) Tahun 2024, salah satu dari tujuh Prioritas Nasional (PN) guna mengukur status kemandirian suatu desa, yaitu memperkuat ketahanan ekonomi untuk pertumbuhan berkualitas dan berkeadilan. Keberhasilan pembangunan suatu desa dapat dilihat melalui perkembangan status kemandirian yang diukur melalui Indeks Desa Membangun (IDM). IDM mampu melihat posisi dan status desa serta arah tingkat kemajuan dan kemandirian desa. IDM tidak hanya mengukur kemajuan infrastruktur saja, namun melibatkan aspek sosial, ekonomi, dan lingkungan [1]. IDM diukur melalui rata-rata dari Indeks Ketahanan Ekonomi (IKE), Indeks Ketahanan Sosisal (IKS), dan Indeks Ketahanan Lingkungan (IKL). IDM mengelompokkan desa-desa di Indonesia berdasarkan, status kemajuan dan kemandirian desa yang diklasifikasikan menjadi Desa Mandiri, Desa Maju, Desa Berkembang, Desa Tertinggal, dan Desa Sangat Tertinggal.

Klasifikasi membantu pemerintah dalam menetapkan status kemajuan dan kemandirian desa berdasarkan kriteria tertentu. Namun pemerintah dan berbagai *stakeholders* perlu mengidentifikasi pola kebutuhan pembangunan desa yang serupa, sehingga rencana dan kebijakan pembangunan dapat berjalan lebih efektif sesuai dengan karakteristik setiap desa. Klasifikasi adalah proses di mana sebuah model atau fungsi ditemukan untuk menjelaskan dan menggambarkan konsep atau kelas data tertentu, dengan tujuan tertentu [2]. Klasifikasi membantu dalam mengorganisir data, namun diperlukan teknik analisis *cluster* untuk membagi populasi menjadi unit-unit penarikan sampel (*cluster*). Teknik *cluster* membantu pemerintah dan berbagai *stakeholders* dalam melakukan analisis dan pemahaman data dengan cara yang lebih terstruktur. *Clustering* merupakan sebuah proses pengelompokan objek-objek ke dalam kelompok-kelompok, di mana objek-objek yang berada dalam satu kelompok memiliki kesamaan di antara mereka sendiri dan berbeda dari objek-objek yang termasuk dalam kelompok lain [3].

Pengelompokan desa di Provinsi Jawa Barat perlu dilakukan sebagai bahan dalam proses perencanaan, pelaksanaan, serta pemantauan dan evaluasi pembangunan desa. Hal ini pun dilakukan agar penetapan status perkembangan desa lebih terarah sesuai dengan IKE, IKS, dan IKL. Oleh karena itu, dibutuhkan sebuah algoritma untuk mengelompokan desa berdasarkan kemiripan atau kesamaan karakteristik pembangunan menggunakan teknik *cluster*. Terdapat dua metode dalam analisis *cluster*, yaitu analisis berhierarki dan non-hierarki. Metode hierarki digunakan ketika jumlah *cluster* yang akan dibentuk tidak diketahui sebelumnya dan kurang efisien untuk data berukuran besar. Sedangkan metode non-hierarki digunakan ketika jumlah *cluster* yang akan dibentuk telah ditentukan sebelumnya.

Dalam pengelompokan data, metode non-hierarki yang sering digunakan adalah *K-Means Clustering*, *K-Medoid*, *Clustering Large Applications* (CLARA), dan *Clustering Large Applications based on Range Search* (CLARANS). *K-Means* sangat populer karena kesederhanaanya, algoritma ini dimulai dengan menentukan titik pusat untuk setiap *cluster* secara acak. Kemudian, setiap objek dikelompokan ke dalam *cluster* terdekat berdasarkan jarak ke pusat *cluster* tersebut [4]. Sedangkan *K-Medoid* atau yang lebih dikenal sebagai *Partitioning Around Medoids* (PAM) dikembangkan untuk meningkatkan efektivitas *K-Means*. Algoritma ini memiliki struktur mikro yang serupa dengan *K-Means* namun menggunakan *medoid* sebagai pusat dari *cluster* [5]. Penggunaan *medoid* pun dinilai lebih efektif karena dapat mengurangi kesalahan yang dihitung oleh rumus tertentu dan tahan terhadap *outlier*. Algoritma *K-Medoid* cenderung kurang efisien dalam menangani data yang besar. Performa *K-Means* maupun PAM tidak terlalu baik, bahkan kurang praktis saat digunakan pada dataset berukuran besar karena keduanya mengharuskan penentuan jumlah *cluster* secara tetap sejak awal. Seiring bertambahnya jumlah titik data, jumlah *cluster* yang mungkin juga meningkat secara eksponensial. Masalah ini dapat diatasi dengan menggunakan metode CLARA [6]. Algoritma CLARA dan CLARANS dapat menghasilkan solusi lebih efisien dalam menangani kumpulan data besar [5], [7], [8], [9]. Algoritma ini merupakan pengembangan dari *K-Medoid* dengan menggunakan teknik sampling dalam proses pengelompokan data untuk mengatasi keterbatasan algoritma *K-Medoid* dalam memproses data berukuran besar [4], [5], [10]. Pengambilan sampel dalam algoritma CLARA dinilai lebih efisien dan mudah dipahami secara komputasi dibandingkan dengan pendekatan pencarian secara acak dinamis pada CLARANS. Selain itu, algoritma CLARA lebih mampu mendeteksi outlier dan noise, serta lebih robust dibandingkan dengan *K-Means* [7], [11].

Selama ini, proses klasifikasi status IDM dilakukan setiap tahun dengan metode konvensional, yang memerlukan waktu cukup lama, yaitu sekitar 8 hingga 9 bulan. Banyaknya jumlah desa menjadi kendala tersendiri karena tidak memungkinkan untuk dilakukan klasifikasi secara bersamaan, sehingga proses menjadi lebih lambat dan data yang dihasilkan kurang mutakhir. Dalam skala besar seperti Provinsi Jawa Barat yang memiliki 5.311 desa, penggunaan algoritma CLARA menjadi solusi yang tepat untuk mengelompokkan desa dalam rangka klasifikasi status IDM secara lebih efisien.

Oleh karena itu, pada penelitian kali ini akan dilakukan pengelompokan desa di Provinsi Jawa Barat menggunakan algoritma CLARA. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik terkait pengelompokan status desa, serta memberi rekomendasi kebijakan untuk meningkatkan efektivitas kerjasama dan evaluasi program pemerintah di masa yang akan datang. Hasil penelitian ini diharapkan dapat memberikan kontribusi sebagai bahan dalam proses perencanaan, pelaksanaan, serta pemantauan dan evaluasi pembangunan desa yang berkelanjutan di provinsi Jawa Barat.

2. Metode Penelitian

Penelitian ini menggunakan data sekunder mengenai IKE, IKS dan IKL desa-desa di Provinsi Jawa Barat tahun 2023. Dataset ini diperoleh dari Dinas Pemberdayaan Masyarakat dan Desa (DPMD) yang diperoleh dari website Open Data Jawa Barat, yakni https://opendata.jabarprov.go.id/id/dataset. Analisis data dilakukan menggunakan Microsoft Excel serta *software RStudio* dengan *package cluster*, *factorextra*, *tydiverse*, dan *openxlxs*.

2.1. Manhattan Distance

Manhattan Distance adalah sebuah matrik jarak yang menghitung total perbedaan absolut (mutlak) antara koordinat sepasang objek. *Manhattan distance* lebih tahan terhadap pencilan (*outlier*) dibandingkan dengan *euclidean distance*. Dalam *euclidean distance*, jarak dihitung dengan mengkuadratkan perbedaan pada setiap dimensi, yang membuat nilai jarak dari *outlier* menjadi jauh lebih besar, sedangkan *manhattan distance* hanya mengambil nilai absolut tanpa kuadrat. *Manhattan distance* lebih cocok digunakan untuk data yang memiliki pencilan atau kasus dengan distribusi data yang tidak seragam, di mana perubahan signifikan pada satu dimensi tidak seharusnya berdampak besar pada keseluruhan jarak, sehingga efek pencilan lebih kecil. Oleh karena data pada penelitian ini memiliki outlier (Gambar 1) maka *Manhattan Distance* digunakan untuk menghitung matriks jarak. *Manhattan Distance* dinyatakan dalam persamaan berikut [12].

$$d(x, y) = |x_i - y_i| \tag{1}$$

Dimana x_i merupakan objek atau suatu data pengamatan ke-i dan y_i medoid data pengamatan ke-i

2.2. Clustering Large Application (CLARA)

Clustering Large Applications (CLARA) merupakan teknik analisis *clustering* non-hierarki atau biasa dikenal dengan partitional *clustering* dengan *medoid* sebagai pusat *cluster*. *Medoid* merupakan objek atau suatu titik data yang terletak di pusat *cluster*, atau dengan kata lain *medoid* adalah objek yang mewakili anggotanya dan memiliki perbedaan rata-rata (*dissimilarity*) yang paling kecil dibandingkan anggota lainnya. Nilai *dissimilarity* ini mengacu pada ukuran perbedaan antara dua titik data, ada beberapa cara yang digunakan seperti menggunakan jarak *Manhattan*, *Euclidean*, dan *Minkowski*. Apabila *medoid* dirumuskan dalam formula matematika didefinisikan sebagai berikut [13].

$$medoids_G = y_{im}; dengan \quad m = \arg\min \textstyle\sum_i^N d(x_i, y_{im}) \tag{2}$$

CLARA merupakan pengembangan dari metode *K-Medoids* yang menggunakan sampling untuk mengelola kumpulan data berukuran besar yang dikembangkan oleh Kaufman dan Rousseeuw [14]. Teknik sampling digunakan untuk memilih sampel dari data yang berukuran besar guna memperoleh *medoid* awal yang akan digunakan dalam proses *clustering*. Teknik sampling yang digunakan dalam CLARA adalah *simple random sampling* [15]. Ukuran sampel yang diambil pada algoritma CLARA yaitu $40+2k$, dimana angka 40 adalah ukuran minimum dari sampel dasar yang dipilih, sementara $2k$ adalah tambahan sampel yang diambil berdasarkan jumlah *cluster* yang diinginkan, di mana k adalah jumlah *cluster*. Kemudian menjalankan metode *K-Medoid* pada sampel tersebut.

Perbedaaan algoritma CLARA dengan *K-Medoid* adalah terletak pada proses pencarian *medoid*. CLARA memlih *medoid* terbaik berdasarkan sampel yang dipilih dari suatu dataset, sedangkan *K-Medoid* memilih *medoid* diantara kumpulan data yang ada. Selain itu CLARA tahan terhadap *outlier* dan dapat digunakan pada data dalam jumlah besar. CLARA lebih efisien dalam hal waktu komputasi dan penyimpanan saat menangani kumpulan data besar. Kualitas dari *medoid* diukur dengan menghitung rata-rata perbedaan jarak antar setiap objek dalam dataset. Adanya pengambilan sampel secara acak, diharapkan *medoid* dari sampel akan mendekati nilai *medoid* dari keseluruhan dataset. Adapun tahapan yang digunakan dalam algoritma CLARA adalah sebagai berikut [16]:

1. Menentukan jumlah *cluster (K)* yang akan dibentuk.
   Pada penelitian ini, penentuan nilai K tersebut mengikuti status IDM di Provinsi Jawa Barat tahun 2023, dimana di provinsi Jawa Barat terdapat tiga kategori dalam status IDM yakni Desa Mandiri, Desa Maju, dan Desa Berkembang.
2. Mengambil ukuran sampel dari data asli menggunakan pendekatan sampling sebanyak $40+2K$. Hal ini biasa disebut dengan m sampel.
3. Memilih pusat *cluster* awal (*medoid*) dari m sampel, dengan *medoid* memiliki perbedaan rata-rata (*dissimilarity*) yang paling kecil dibandingkan anggota lainnya.
4. Menghitung jarak menggunakan *manhattan distance* dari setiap objek *non-medoid* ke *medoid* pada tiap *cluster*. Kemudian menempatkan tiap objek *non-medoid* tersebut ke *medoid* terdekat, lalu hitung total jaraknya.
5. Memilih secara acak objek *non-medoid* pada populasi sebagai kandidat *medoid* baru.
6. Menghitung jarak setiap objek *non-medoid* ke *medoid* baru dan menempatkan tiap objek *non-medoid* tersebut ke kandidat *medoid* terdekat, lalu hitung total jaraknya.
7. Menghitung simpangan total (S) atau selisih total jarak, dengan total jarak pada kandidat *medoid* baru dikurangi dengan total jarak pada *medoid* lama.

   $$S = \sum d(x, y_m) - d(x, y_{ma}) \tag{3}$$

8. Jika nilai $S < 0$ maka ulangi tahapan ke-6 sampai ke-7, namun jika nilai $S > 0$ maka iterasi dihentikan.
9. Distribusikan setiap data ke dalam *cluster-cluster* yang jaraknya paling kecil.
10. Validasi hasil *cluster* dengan menghitung nilai *silhouette coefficient* pada masing-masing *cluster* yang terbentuk.

2.3. Silhouette coefficient

Koefisien silhouette merupakan satu ukuran yang dapat digunakan untuk menentukan jumlah kelompok (k) optimal. Adapun langkah untuk menghitung *silhouette score* adalah sebagai berikut [17].

1. Menghitung rata-rata jarak data ke-i terhadap semua data lain dalam satu *cluster*

   $$a(i) = \frac{1}{m_j - 1} \textstyle\sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^j) \tag{4}$$

   Dengan i = index data ($i = 1, 2, \ldots m_j$), $a(i)$ = rata-rata jarak data ke-i terhadap semua data dalam satu *cluster*, j = *cluster*, m_j = jumlah data dalam *cluster* ke-j, dan $d(x_i^j, x_r^j)$ = jarak data ke-i dengan data ke-r dalam satu *cluster* j.
2. Rata-rata jarak data ke-i terhadap *cluster* yang berbeda

   $$d_i(j, n) = \frac{1}{m_n} \textstyle\sum_{r=1}^{m_n} d(x_i^j, x_r^n) \tag{5}$$

   Kemudian mengambil nilai minimum $d_i\,(j, n)$ menggunakan persamaan (6)

   $$b(i) = \min_{\substack{n=1,\ldots,K \\ n \neq j}} \{d_i(j, n)\} \tag{6}$$

Dengan i = index data, j = *cluster*, $d_i(j,n)$ = rata-rata jarak data ke-i terhadap semua data yang tidak dalam satu *cluster* dengan data ke-i, m_n = jumlah data dalam *cluster* ke-n, dan $d(x_i^j, x_r^n)$ = jarak data ke-i dengan data ke-r dalam satu *cluster* n.

3. Menghitung nilai dari *silhouette index* untuk setiap data ke-i Dalam menghitung nilai indeks *silhouette* dari sebuah data ke-i, terdapat dua komponen yang terlibat yaitu $a(i)$ dan $b(i)$. Koefisien ini membandingkan rata-rata kemiripan setiap objek terhadap objek lainnya dalam satu kelompok ($a(i)$) dengan minimum rata-rata kemiripan setiap objek terhadap objek lainnya di luar kelompok ($b(i)$). Nilai $s(i)$ diperoleh dengan menggabungkan $a(i)$ dan $b(i)$ sebagai berikut:

$$\begin{aligned} s_i &= 1 - \frac{a(i)}{b(i)}, && \text{jika } a(i) < b(i) \\ s_i &= 0, && \text{jika } a(i) = b(i) \\ s_i &= \frac{b(i)}{a(i)} - 1, && \text{jika } a(i) > b(i) \end{aligned} \tag{7}$$

Kemudian menghitung koefisien *silhouette* untuk objek ke-i yang berada dalam satu *cluster* menggunakan persamaan 8.

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{7}$$

4. Menghitung nilai *index silhouette* pada seluruh data dalam satu *cluster*.

$$SI = \frac{1}{m_j} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} si \tag{8}$$

Dengan SI merupakan rata-rata *index silhouette* pada *cluster* k.

5. Menghitung nilai *silhouette coefficient global*

$$SC = \frac{1}{K} \sum_{j=1}^{K} SI \tag{9}$$

Dengan K ialah jumlah *cluster* dan SC ialah rata-rata *silhouette coefficient* dan dari seluruh dataset.

3. Hasil dan Pembahasan

3.1. Statistik Deskriptif Data

Table 1. Statistik Deskriptif

Variabel	Minimum	Maksimum	Rata-rata	Simpangan Baku
Nilai IKE	0,3500	1	0,7245	0,1148
Nilai IKS	0,5800	1	0,8463	0,0673
Nilai IKL	0,2700	1	0,7693	0,1342

Tabel 1 menyajikan informasi mengenai nilai minimun, maksimum, rata-rata, dan simpangan baku dari 5311 desa yang terdapat di Provinsi Jawa Barat. Desa Dewisari menjadi desa dengan Nilai IKE paling kecil yakni 0,35. Desa Cintanagara menjadi desa dengan Nilai IKS paling kecil yakni 0,58. Desa Cinangka, Simpang, dan Tamanmekar menjadi desa dengan Nilai IKL paling kecil yakni 0,27. Namun, masing-masing variabel memiliki nilai maksimum yang serupa yakni sebesar 1.

3.2. Pengujian Pencilan

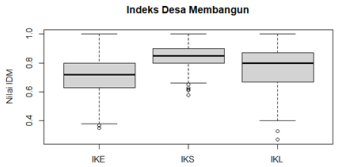


Figure 1. Hasil Deteksi Outlier

Pada Gambar 1 terlihat masing-masing variabel memiliki *outlier*. Dimana *outlier* pada Nilai IKE memiliki *outlier* sebanyak 3 buah, 10 *outlier* pada Nilai IKS, dan 9 *outlier* pada Nilai IKL. Oleh karena itu, matriks jarak pada penelitian ini dihitung menggunakan *Manhattan distance*.

3.3. Clustering menggunakan CLARA

Pada penelitian ini ditentukan jumlah *cluster* untuk algoritma CLARA sebanyak tiga buah *cluster* (K=3). Untuk memilih sampel, digunakan rumus $40+2K$ [16] ukuran sampel yang disebut dengan m sampel. Sehingga diperoleh 46 ukuran sampel dari 5311 data penelitian. Sampel tersebut dipilih secara acak menggunakan *software* RStudio. Kemudian menentukan *medoid* awal *cluster* dengan bantuan *software* Microsoft Excel. Hasilnya dapat dilihat pada tabel 2

Table 2. Medoid Awal Cluster

No	Nama Desa	Nilai IKE	Nilai IKS	Nilai IKL	Jumlah	Rata-Rata
1	WARNAJATI	0,72	0,87	0,80	10,49	0,2280
2	SUKADAMI	0,67	0,87	0,80	10,73	0,2333
3	BAGENDIT	0,72	0,87	0,87	11,05	0,2402

Langkah selanjutnya adalah menghitung jarak kedekatan menggunakan *manhattan distance* antara *medoids* awal dengan semua data penelitian. Adapun *medoid* awal yang terpilih ialah Desa Warnajati, Desa Sukadami, dan Desa Bagendit karena memiliki perbedaan rata-rata (*dissimilarity*) yang paling kecil dibandingkan anggota lainnya.

Table 3. Jumlah dan Total Kedekatan Pada Masing-masing Cluster

Cluster	Jumlah Desa	Total Kedekatan
1	1776	451,22
2	1462	358,38
3	2073	385,75
Total	**5311**	**1195,35**

Berdasarkan tabel 3, *Cluster* 3 memiliki jumlah terbanyak yakni 2073 desa, sedangkan *cluster* 2 memiliki jumlah terkecil yakni 1462 desa. Sedangkan untuk total kedekatan *cluster* didapatkan 1195,35. Selanjutnya mencari nilai *medoid* baru pada iterasi kedua. Pada tahap ini *medoid* dipilih secara acak, dengan syarat objek yang telah terpilih sebelumnya menjadi *medoid* tidak boleh terpilih lagi. Hasilnya dapat ditunjukkan pada tabel 4

Langkah selanjutnya adalah menghitung jarak kedekatan menggunakan *manhattan distance* antara *medoids* baru dengan semua data penelitian. Adapun *medoid* baru yang terpilih ialah Desa Sukasari, Desa Jambugeulis, dan Desa Balongsari karena memi-

Table 4. Medoid Baru Iterasi Kedua

No	Nama Desa	Nilai IKE	Nilai IKS	Nilai IKL
1	SUKASARI	0,83	0,91	0,87
2	JAMBUGEULIS	0,65	0,81	0,87
3	BALONGSARI	0,68	0,83	0,67

liki perbedaan rata-rata (*dissimilarity*) yang paling kecil dibandingkan anggota lainnya. *Cluster* 3 memiliki jumlah terbanyak yakni 2094 desa, sedangkan *cluster* 2 memiliki jumlah terkecil yakni 1573 desa. Sedangkan untuk total kedekatan *cluster* didapatkan 879,34. Selanjutnya menghitung simpangan total (S) yang didapatkan dari selisih total jarak, dimana total jarak pada kandidat *medoid* baru (iterasi kedua) dikurangi dengan total jarak pada *medoid* awal (iterasi pertama).

$$S = \sum d(x, y_m) - d(x, y_{ma})$$
$$S = -316{,}01$$

Didapatkan nilai simpangan total dengan ukuran sampel 46 antara medoid baru dengan *medoid* lama, yakni -316,01. Karena nilai $S < 0$ maka iterasi dilanjutkan dengan melakukan iterasi ketiga.

Langkah selanjutnya adalah menghitung jarak kedekatan menggunakan *manhattan distance* antara *medoids* baru dengan semua data penelitian. Adapun medoid baru yang terpilih ialah Desa Simpang, Desa Hegarmanah, dan Desa Cimanuk karena memiliki perbedaan rata-rata (*dissimilarity*) yang paling kecil dibandingkan anggota lainnya.

Table 5. Jumlah & Total Kedekatan Pada Masing-masing Cluster

Cluster	Jumlah Desa	Total Kedekatan
1	1806	479,83
2	2412	496,04
3	1093	191,71
Total	**5311**	**1167,58**

Berdasarkan tabel 5, *Cluster* 2 memiliki jumlah terbanyak yakni 2412 desa, sedangkan *cluster* 3 memiliki jumlah terkecil yakni 1093 desa. Sedangkan untuk total kedekatan *cluster* didapatkan 1167,58. Selanjutnya menghitung simpangan total (S).

$$S = \sum d(x, y_m) - d(x, y_{ma})$$
$$S = 288{,}24$$

Didapatkan nilai simpangan total dengan ukuran sampel 46 antara *medoid* baru dengan *medoid* lama, yakni 288,24. Karena nilai $S > 0$ maka iterasi dihentikan. Dengan demikian iterasi kedua dianggap sebagai *medoid* yang optimal, maka hasil akhir *cluster* didasarkan pada medoid dari iterasi kedua.

3.4. *Perbandingan Karakteristik Pengelompokan*

Dalam konteks ini, hasil *cluster* yang telah dilakukan oleh KEMENDES akan dibandingkan dengan hasil *cluster* menggunakan algoritma CLARA. Berdasarkan hasil pemetaan pada gambar 2, diperoleh jumlah *cluster* yang terbentuk adalah tiga buah *cluster*. Berdasarkan hasil *clustering* menggunakan algoritma CLARA diperoleh 1644 desa yang termasuk dalam *cluster* 1 (Desa Mandiri), 1573 desa yang termasuk dalam *cluster* 2 (Desa Maju), dan 2094 desa yang termasuk dalam *cluster* 3 (Desa Berkembang). Sedangkan hasil *clustering* berdasarkan KEMENDES diperoleh 1856 desa yang termasuk dalam *cluster* 1 (Desa Mandiri), 2494 desa yang termasuk dalam cluster 2 (Desa Maju), dan 961 desa yang termasuk dalam *cluster* 3 (Desa Berkembang).

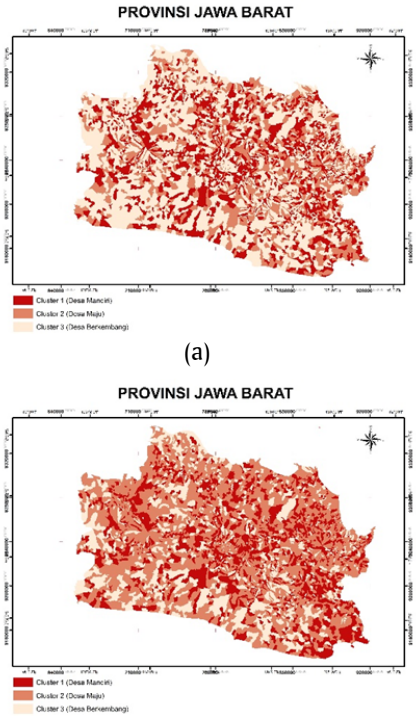


Figure 2. Peta hasil pengelompokan (a). Algoritma CLARA dan (b). KEMENDES

3.5. *Silhouette coefficient*

Diperoleh rata-rata index *silhouette* (SI) pada masing-masing *cluster* sebagai berikut.

Table 6. Index Silhouette

Cluster	Status Desa	Jumlah Cluster	SI
1	Mandiri	1644	0,2730
2	Maju	1573	0,4629
3	Berkembang	2094	0,3482
Total		**5311**	**1,0842**

Nilai SI merupakan rata-rata *index silhouette* pada *cluster* K, dimana terdapat 3 *cluster*. Pada Tabel 8 didapatkan nilai *index silhouette* sebesar 1,0842, dimana nilai ini didapatkan dari hasil penjumlahan masing-masing *cluster*. Selanjutnya menghitung nilai *silhouette coefficient global* (SC).

$$SC = \sum_{j=1}^{K} SI$$

$$SC = \frac{1{,}0842}{3} = 0,3614$$

Dari hasil perhitungan diatas, didapatkan nilai *silhouette coefficient global* (SC) sebesar 0,3614. Ini menandakan bahwa kriteria pengelompokan termasuk kedalam kategori yang lemah, karena berada diantara nilai 0,26-0,50. Berikut ini merupakan perbandingan karakteristik pengelompokan pada algoritma CLARA dan pengelompokan berdasarkan KEMENDES.

Table 7. Karakteristik Pengelompokan pada Algoritma CLARA dan KEMENDES

Karakteristik	Cluster	Status Desa	Rata-rata Nilai IKE	Rata-rata Nilai IKS	Rata-rata Nilai IKL	Jumlah Anggota Cluster
Algoritma CLARA	1	Mandiri	0,8424	0,9040	0,8374	1644
	2	Maju	0,6497	0,8177	0,8734	1573
	3	Berkembang	0,6881	0,8225	0,6376	2094
KEMENDES	1	Mandiri	0,8161	0,8963	0,8691	1856
	2	Maju	0,6984	0,8343	0,7465	2494
	3	Berkembang	0,6151	0,7810	0,6357	961

Berdasarkan Tabel 7, terdapat perbedaan jumlah desa dalam cluster yang terbentuk berdasarkan Algoritma CLARA dan hasil klasifikasi KEMENDES. Perbedaan yang cukup tinggi terlihat pada status Desa Berkembang. Algoritma CLARA menghasilkan 2094 desa di Provinsi Jawa Barat yang diklasifikasikan sebagai Desa Berkembang, sementara hasil klasifikasi KEMENDES menunjukkan bahwa hanya 961 desa di Provinsi Jawa Barat yang diklasifikasikan sebagai Desa Berkembang. Selain itu, Algoritma CLARA mengklasifikan sebuah desa sebagai Desa Berkembang dengan rata-rata nilai IKE 0,6881, rata-rata nilai IKS 0,8225 dan rata-rata nilai IKL 0,6376. Sedangkan KEMENDES mengklasifikasikan sebuah desa sebagai Desa Berkembang dengan rata-rata nilai IKE, IKS, dan IKL yang lebih rendah dibandingkan algoritma CLARA yaitu, 0,6151, 0,7810, dan 0,6357. Berdasarkan hal tersebut, dapat diasumsikan bahwa Algoritma CLARA lebih sensitif terhadap perbedaan karakteristik nilai IKE, IKS dan IKL dengan standar nilai yang lebih tinggi dibandingkan dengan KEMENDES dalam mengklasifikasikan sebuah desa di Provinsi Jawa Barat sebagai desa Mandiri, Maju dan/atau berkembang.

4. Kesimpulan

Algoritma CLARA merupakan algoritma pengelompokkan yang dirancang untuk menangani dataset besar dengan mengadaptasi metode PAM. CLARA bekerja dengan cara mengambil beberapa sampel acak dari dataset dengan ketentuan $40+2K$ untuk menghindari beban komputasi tinggi. Pada setiap sampel, algoritma PAM digunakan untuk mencari medoid terbaik dan membentuk klaster. Hasil dari beberapa sampel dibandingkan berdasarkan nilai perbedaan rata-rata (*dissimilarity*) dan solusi terbaik dari seluruh iterasi dipilih sebagai hasil akhir. Dengan pendekatan ini, CLARA mampu memberikan hasil pengelompokkan yang mendekati optimal untuk dataset besar tanpa harus memproses seluruh data secara langsung.

Pada penelitian ini, algoritma CLARA memberikan hasil pengelompokkan desa di Provinsi Jawa Barat yang berbeda dengan hasil pengelompokkan oleh KEMENDES. Pada hasil pengelompokan berdasarkan KEMENDES terdapat 1856 desa yang diklasifikasikan sebagai desa Mandiri, 2494 desaMaju, dan 961 desa Berkembang, sedangkan pengelompokan berdasarkan algoritma CLARA terdapat 1644 desa Mandiri, 1573 desa Maju, dan 2094 desa Berkembang. Nilai *silhouette coefficient global* (SC) pada algoritma CLARA menunjukan nilai SC sebesar 0,3614. Ini menandakan bahwa kriteria pengelompokan hasil algoritma CLARA termasuk kedalam kategori yang lemah, karena berada diantara nilai 0,26-0,50.

Oleh karena itu, dengan hasil pengelompokkan yang lemah, maka algoritma CLARA perlu dikembangkan kembali terutama dalam hal penentuan ukuran sampel dan teknik sampling. Pemilihan pusat klaster (Medoid) sangat bergantung pada banyaknya ukuran sampel dan teknik sampling yang digunakan. Sehingga agar lebih refresentaif penentuan ukuran sampel tidak hanya berdasarkan $40+2K$ tetapi perlu disesuaikan dengan teknik sampling yang digunakan yaitu *simple random sampling*.

References

[1] Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia, "Peraturan menteri desa, pembangunan daerah tertinggal, dan transmigrasi republik indonesia nomor 2 tahun 2016 tentang indeks desa membangun," http://jdih.kemendesa.go.id/katalog/peraturan_menteri_desa_pembangunan_daerah_tertinggal_dan_transmigrasi_nomor_2_tahun_2016, 2016, diundangkan pada tanggal 24 Februari 2016 dalam Berita Negara Republik Indonesia Tahun 2016 Nomor 300. [Online]. Available: http://jdih.kemendesa.go.id/katalog/peraturan_menteri_desa_pembangunan_daerah_tertinggal_dan_transmigrasi_nomor_2_tahun_2016

[2] S. Defiyanti, "Integrasi metode clustering dan klasifikasi untuk data numerik," *Citee, no. July*, pp. 256–261, 2017.

[3] R. D. Kusumah, B. Warsito, and M. A. Mukid, "Perbandingan metode k–means dan self organizing map (studi kasus: pengelompokan kabupaten/kota di jawa tengah berdasarkan indikator indeks pembangunan manusia 2015)," *Jurnal Gaussian*, vol. 6, no. 3, pp. 429–437, 2017.

[4] Y. Aboubi, H. Drias, and N. Kamel, "Bat-clara: Bat-inspired algorithm for clustering large applications," *IFAC-PapersOnLine*, vol. 49, no. 12, pp. 243–248, 2016.

[5] R. Ardhani, S. Marshelle, A. Fitrianto, E. Erfiani, and L. R. D. Jumansyah, "Perbandingan k-medoids dan clara (clustering large application) pada data populasi ternak di indonesia," *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 3, pp. 1744–1763, 2024.

[6] A. Alam, M. Muqeem, and S. Ahmad, "Comprehensive review on clustering techniques and its application on high dimensional data," *International Journal of Computer Science & Network Security*, vol. 21, no. 6, pp. 237–244, 2021.

[7] T. Gupta and S. P. Panda, "A comparison of k-means clustering algorithm and clara clustering algorithm on iris dataset," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 4766–4768, 2018.

[8] J. Swarndeep Saket and S. Pandya, "An overview of partitioning algorithms in clustering techniques," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 6, pp. 1943–1946, 2016.

[9] T. Mohammadi, F. D'Ascenzo, M. Pepe, S. B. Zanghì, M. Bernardi, L. Spadafora, G. Frati, M. Peruzzi, G. M. De Ferrari, and G. Biondi-Zoccai, "Unsupervised machine learning with cluster analysis in patients discharged after an acute coronary syndrome: insights from a 23,270-patient study," *The American Journal of Cardiology*, vol. 193, pp. 44–51, 2023.

[10] L. Kuang and L. Zhang, "A scheduling algorithm based on clara clustering," in *AIP Conference Proceedings*, vol. 1864, no. 1. AIP Publishing, 2017.

[11] P. R. Fitrayana and D. R. S. Saputro, "Algoritme clustering large application (clara) untuk menanganidata outlier," in *Prisma, Prosiding Seminar Nasional Matematika*, vol. 5, 2022, pp. 721–725.

[12] M. Nishom, "Perbandingan akurasi euclidean distance, minkowski distance, dan manhattan distance pada algoritma k-means clustering berbasis chi-square," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 1, pp. 20–24, 2019.

[13] E. N. Fitriyani and A. I. Achmad, "Penerapan analisis k-medoids cluster untuk mengelompokkan wilayah di provinsi jawa barat berdasarkan fasilitas kesehatan tahun 2021," in *Bandung Conference Series: Statistics*, vol. 3, no. 2, 2023, pp. 283–293.

[14] E. Schubert and P. J. Rousseeuw, "Fast and eager k-medoids clustering: O (k) runtime improvement of the pam, clara, and clarans algorithms," *Information Systems*, vol. 101, p. 101804, 2021.

[15] W. Cao, S. Wang, and M. Xu, "Optimal scheduling of virtual power plant based on latin hypercube sampling and improved clara clustering algorithm," *Processes*, vol. 10, no. 11, p. 2414, 2022.

[16] K. Setiawan, Kastum, and Y. P. Pratama, "K-means clustering analysis of poverty data in cilacap district," *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 5, no. 1, pp. 53–62, 2025, published online April 2025. [Online]. Available: https://doi.org/10.35870/ijsecs.v5i1.3759

[17] S. Petrovic, "A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters," in *Proceedings of the 11th Nordic workshop of secure IT systems*, vol. 2006. Citeseer, 2006, pp. 53–64.