

Pengelompokan Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma *Centroid Linkage*

Yossy Candra, Rito Goejantoro, dan Andrea Tri Rian Dani

Volume 12, Issue 1, Pages 9–15, June 2024

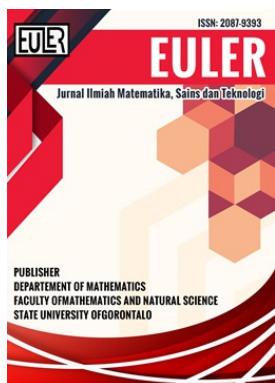
Diterima 4 April 2024, Direvisi 6 Mei 2024, Disetujui 11 Mei 2024, Diterbitkan 16 Mei 2024



To Cite this Article : Y. Candra, dkk., "Pengelompokan Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma *Centroid Linkage*", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 12, no. 1, pp. 9–15, 2024, <https://doi.org/10.37905/euler.v12i1.24887>

© 2024 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI

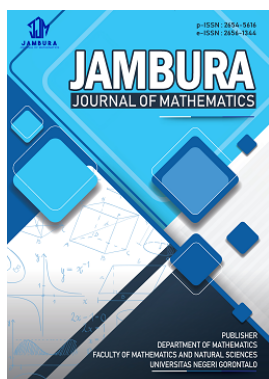


	Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	:	Biannual (June and December)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/euler
	Online ISSN	:	2776-3706
	Editor-in-Chief	:	Resmawan
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	:	QF_r_gAAAAJ
	Email	:	euler@ung.ac.id

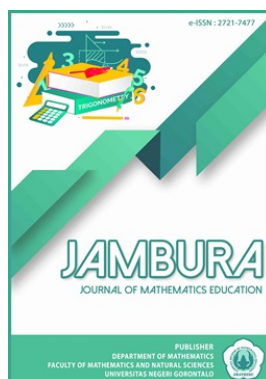
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



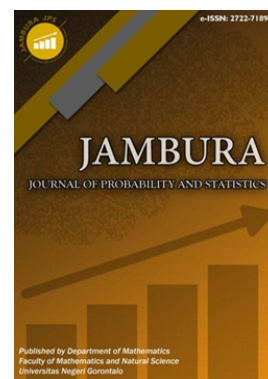
Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Pengelompokan Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma *Centroid Linkage*

Yossy Candra¹, Rito Goejantoro^{1,*}, dan Andrea Tri Rian Dani¹

¹Program Studi Statistika, Universitas Mulawarman, Indonesia

ARTICLE HISTORY

Diterima 4 April 2024
Direvisi 6 Mei 2024
Disetujui 11 Mei 2024
Diterbitkan 16 Mei 2024

KATA KUNCI

Analisis Kluster
Centroid Linkage
Silhouette Coefficient
Social Science

KEYWORDS

Centroid Linkage
Cluster Analysis
Silhouette Coefficient
Social Science

ABSTRAK. Dengan keanekaragaman budaya yang kaya dan potensi sumber daya alam yang melimpah, Indonesia masih dihadapkan pada berbagai permasalahan *social science* (ilmu sosial). Ketimpangan ekonomi, rendahnya kualitas pendidikan, akses kesehatan yang terbatas, dan tingginya angka kriminalitas adalah masalah sosial yang melanda berbagai Provinsi di Indonesia. Penelitian ini dilakukan untuk mengelompokkan Provinsi-Provinsi di Indonesia berdasarkan indikator sosial yang meliputi ekonomi, pendidikan, kesehatan, dan kriminalitas. Penelitian ini menggunakan analisis kluster dengan algoritma *Centroid Linkage* untuk mengelompokkan Provinsi-Provinsi di Indonesia. Algoritma *Centroid Linkage* dipilih karena kelebihannya dalam menghasilkan pengelompokan yang optimal. Uji validitas kluster menggunakan *Silhouette Coefficient* (SC). Studi kasus yang digunakan adalah variabel-variabel yang diduga terkait dengan masalah ekonomi, kesehatan, pendidikan, dan kriminalitas pada 34 Provinsi di Indonesia Tahun 2021. Berdasarkan analisis, hasil pengelompokan dengan algoritma *Centroid Linkage* menunjukkan jumlah kluster yang optimal adalah 2 kluster, dengan nilai SC sebesar 0,538. Kluster 1 terdiri dari 33 Provinsi dan kluster 2 terdiri dari hanya satu Provinsi yaitu DKI Jakarta.

ABSTRACT. With its rich cultural diversity and abundant natural resource potential, Indonesia still faces various *social science* problems. Economic inequality, low quality of education, limited access to health, and high crime rates are social problems that hit various provinces in Indonesia. This research was conducted to group provinces in Indonesia based on social indicators, which include economy, education, health, and crime. This research uses cluster analysis with the *Centroid Linkage* algorithm to group provinces in Indonesia. The *Centroid Linkage* algorithm was chosen because of its advantages in producing optimal grouping. Test cluster validity using the *Silhouette Coefficient* (SC). The case studies used are variables that are thought to be related to economic, health, education, and crime problems in 34 provinces in Indonesia in 2021. Based on the analysis, the grouping results using the *Centroid Linkage* algorithm show that the optimal number of clusters is 2, with an SC value of 0.538. Cluster 1 consists of 33 provinces, and Cluster 2 consists of only one province, DKI Jakarta.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Editorial of EULER: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Ekonomi, pendidikan, kesehatan, dan kriminalitas merupakan empat faktor dalam *social science* yang mempengaruhi tingkat kesejahteraan dan keamanan masyarakat. *Social science* (Ilmu sosial) mempelajari bagaimana interaksi dan pengaruh yang terjadi antara faktor-faktor ini, serta bagaimana mereka mempengaruhi kualitas hidup dan stabilitas sosial [1].

Ekonomi memiliki peran penting dalam menentukan kesejahteraan masyarakat. Akses terhadap sumber daya dan kemampuan keluarga untuk memenuhi kebutuhan dasar seperti makanan, pakaian, dan tempat tinggal merupakan faktor kunci dalam mewujudkan kesejahteraan [2]. Pendidikan merupakan faktor penentu lainnya dalam meningkatkan kesejahteraan. Akses terhadap pendidikan berkualitas tinggi memberikan individu

kesempatan untuk mengembangkan keterampilan dan pengetahuan yang dibutuhkan untuk mendapatkan pekerjaan yang lebih baik dan meningkatkan taraf hidup mereka [3]. Kesehatan memainkan peran penting dalam meningkatkan kesejahteraan masyarakat. Memfasilitasi pembangunan dan pengembangan infrastruktur kesehatan yang ada, meningkatkan ketersediaan sumber daya manusia dan teknologi medis, serta meningkatkan kualitas dan aksesibilitas layanan kesehatan bagi seluruh lapisan masyarakat [4]. Kriminalitas merupakan faktor yang dapat menurunkan kesejahteraan dan keamanan masyarakat. Tingkat kejahatan yang tinggi dapat menciptakan rasa takut dan ketidakamanan, yang dapat berdampak negatif pada kualitas hidup dan menghambat pembangunan ekonomi [5].

Melalui analisis menyeluruh terhadap keempat faktor ini, dapat diperoleh pemahaman yang lebih mendalam tentang dinamika sosial suatu wilayah. Hal ini memungkinkan untuk mengi-

*Penulis Korespondensi.

identifikasi pola-pola yang mungkin ada dan merumuskan strategi yang lebih efektif dalam meningkatkan kesejahteraan dan keamanan masyarakat. Penelitian ini dilakukan untuk mengelompokkan Provinsi di Indonesia berdasarkan keempat faktor tersebut, menggunakan pendekatan *data mining*.

Data mining, yang dikenal dengan istilah *Knowledge Discovery in Database (KDD)*, yaitu proses menyaring dan menganalisis sejumlah data untuk mengekstrak dan menemukan pengetahuan yang relevan dan berguna [6]. Hal ini bertujuan membantu dalam membuat keputusan yang benar dan tepat berdasarkan informasi yang diperoleh melalui proses tersebut. *Data mining* berdasarkan jenis tugas diklasifikasikan menjadi model prediksi, analisis klasifikasi, analisis asosiasi, deteksi anomali, dan analisis kluster [7]. Berdasarkan tujuan penelitian, metode yang sesuai untuk digunakan adalah analisis kluster.

Analisis kluster adalah metode pengelompokan objek-objek berdasarkan tingkat kemiripan. Objek dalam satu kluster memiliki tingkat kemiripan yang tinggi (homogenitas), sedangkan antar kluster memiliki tingkat ketidakmiripan yang signifikan (heterogenitas) [8]. Terdapat dua metode utama dalam analisis kluster, yaitu metode non-hierarki dan metode hierarki [9]. Metode non-hierarki adalah metode analisis kluster yang dimulai dengan menentukan jumlah kluster iterasi awal [10]. Sementara itu, metode hierarki adalah pendekatan dalam analisis kluster yang mengatur objek-objek ke dalam tingkatan-tingkatan berdasarkan tingkat kesamaan sifat di antara objek tersebut [11]. Pada penelitian ini menggunakan metode hierarki. Kelebihan metode hierarki adalah representasi visual yang lebih kaya melalui dendrogram, yang menggambarkan bagaimana setiap objek bergabung dalam satu kluster. Dendrogram dapat membantu dalam pemahaman yang lebih baik tentang struktur kluster atau struktur data. Pada metode hierarki, terdapat berbagai algoritma yang dapat digunakan. Salah satu algoritma dalam pengelompokan metode hierarki adalah *Centroid Linkage*.

Centroid Linkage adalah bagian dari proses kluster yang menggunakan jarak antara *centroid* sebagai dasar. Pendekatan ini efektif dalam mengurangi variasi di dalam kluster karena memasukkan *centroid* saat menggabungkan kluster [12]. Salah satu keuntungan dari metode ini adalah *robust* terhadap pencilan (*outlier*) [13]. *Outlier* adalah data yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data yang mengganggu proses analisis data [14]. *Centroid Linkage* memiliki keuntungan *robust* terhadap *outlier* karena dalam proses pengelompokan, algoritma ini menggunakan rata-rata *centroid* dari dua kluster yang digabungkan, sehingga pengaruh *outlier* terhadap hasil pengelompokan menjadi lebih kecil.

Penelitian yang menggunakan algoritma *Centroid Linkage* yang telah dilakukan dalam banyak kasus. Silvi [15] melakukan analisis *Cluster* dengan data *outlier* menggunakan *Centroid Linkage* dan *K-Means Clustering* untuk mengelompokkan indikator HIV/AIDS di Indonesia. Penelitiannya menghasilkan 7 kluster dengan dimana algoritma *Centroid Linkage* menghasilkan pengelompokan yang lebih homogen dibandingkan *K-Means*. Selanjutnya, Dzikrullah [16] melakukan pengelompokan Provinsi berdasarkan kualitas jaringan internet dengan metode *Centroid Linkage*. Penelitiannya menghasilkan 4 kluster, dengan Provinsi Papua serta Papua Barat menempati kluster yang cenderung menunjukkan kualitas jaringan internet yang rendah. Dalam kasus lain, Pus-

pita [17] membandingkan metode *Centroid* dan *Ward* dalam pengelompokan tingkat penyelesaian pendidikan di Indonesia. Hasil penelitiannya menunjukkan *Centroid Linkage* sebagai metode yang lebih baik dengan profilisasi kluster yang optimal sebanyak 3 kluster. Adapun penelitian yang dilakukan oleh Radja, dkk [18] mengimplementasikan algoritma *Centroid Linkage* dan *K-Medoids* dalam mengelompokkan Kabupaten/Kota di Sulawesi Selatan berdasarkan indikator pendidikan. Hasil penelitiannya menunjukkan bahwa algoritma *Centroid Linkage* sebagai metode terbaik yang menghasilkan 2 kluster dengan rasio deviasi standar dalam kluster dan deviasi standar antar kluster yang minimum. Berdasarkan hasil kajian dari beberapa penelitian sebelumnya, ditunjukkan bahwa algoritma *Centroid Linkage* efektif dalam mengelompokkan data dengan menghasilkan pengelompokan yang optimal, dan dapat diterapkan dalam berbagai bidang.

Pada penelitian ini, digunakan algoritma *Centroid Linkage* dengan jarak pengukuran kemiripan menggunakan jarak kuadrat Euclid. Pengukuran jumlah kluster optimal menggunakan *Silhouette Coefficient*. Penelitian ini untuk mengelompokkan wilayah-wilayah yang mengalami masalah sosial dengan indikator-indikator yang meliputi masalah ekonomi, pendidikan, kesehatan, dan kriminalitas. Hasil penelitian ini diharapkan dapat memberikan pandangan mendalam tentang perbedaan dan kesamaan karakteristik antarprovinsi, serta mendukung pengembangan kebijakan yang lebih efektif di Indonesia.

2. Metode

Penelitian ini menggunakan data indikator ekonomi, pendidikan, kesehatan, dan kriminalitas yang diperoleh dari Badan Pusat Statistika (BPS). Variabel penelitian yang digunakan sebanyak 20 variabel yang disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
x_1	Persentase Penduduk Miskin
x_2	Gini Ratio
x_3	Laju Pertumbuhan Ekonomi
x_4	Rata-Rata Lama Sekolah
x_5	Tingkat Pengangguran Terbuka
x_6	Tingkat Partisipasi Angkatan Kerja
x_7	Laju Pertumbuhan Penduduk
x_8	Kepadatan Penduduk
x_9	Angka Harapan Hidup
x_{10}	PDRB Per Kapita Harga Konstan
x_{11}	Indeks Kebahagiaan
x_{12}	Persentase Penduduk Korban Kejahatan
x_{13}	Jumlah Kejahatan yang Dilaporkan
x_{14}	Risiko Penduduk Terkena Kejahatan Per 100.000 Penduduk
x_{15}	Persentase Penyelesaian Kejahatan
x_{16}	Proporsi Penduduk yang Menjadi Korban Kejahatan
x_{17}	Jumlah Desa yang Pernah Ada Kejadian Kejahatan Pembunuhan
x_{18}	Persentase Desa yang Ada Kejadian Konflik Massal
x_{19}	Jumlah Desa yang Membangun Poskamling
x_{20}	Persentase Desa Menurut Keberadaan Pos Polisi

Tahapan analisis data diuraikan dengan prosedur berikut:

1. Melakukan eksplorasi data dengan statistika deskriptif.
2. Melakukan standardisasi Z_{score} menggunakan persamaan

(1):

$$Z_{i,p} = \frac{x_{i,p} - \bar{x}_p}{S_p} \quad (1)$$

dengan rata-rata untuk variabel ke- p (\bar{x}_p):

$$\bar{x}_p = \frac{1}{n} \sum_{i=1}^n x_{i,p} \quad (2)$$

dan standar deviasi untuk variabel ke- p (S_p):

$$S_p = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{i,p} - \bar{x}_p)^2} \quad (3)$$

dengan $x_{i,p}$ adalah data ke- i pada variabel ke- p .

3. Melakukan pengelompokan data dengan menerapkan algoritma *Centroid Linkage*:

(1) Menghitung jarak kemiripan dengan perhitungan jarak kuadrat Euclid, menggunakan persamaan (4):

$$d^2(x_p, y_p) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_j - y_j)^2 \quad (4)$$

(2) Menggabungkan kluster U dan V menjadi kluster baru (UV) memiliki ukuran jarak terkecil dengan cara menghapus baris dan kolom kluster U dan V , dan menambahkan satu baris dan kolom untuk kluster UV

(3) Menghitung jarak rata-rata berdasarkan algoritma *Centroid Linkage*, menggunakan persamaan (5):

$$d_{(UV)} = \frac{n_U \bar{x}_U + n_V \bar{x}_V}{n_U + n_V} \quad (5)$$

(4) Menghitung kembali ukuran jarak kemiripan menggunakan perhitungan jarak kuadrat Euclid dari *centroid* yang bergabung.

(5) Menggabungkan kluster dengan ukuran jarak terkecil.

(6) Mengulangi langkah (1) hingga (5) untuk mendapatkan pasangan kluster terdekat berikutnya, hingga kluster yang diinginkan terpenuhi.

(7) Menginterpretasikan hasil kluster

4. Menghitung nilai *Silhouette Coefficient* (SC) untuk melihat jumlah kluster yang optimal:

(1) Menghitung rata-rata jarak dari suatu objek ke- i dengan semua objek yang berada pada kluster yang sama, menggunakan persamaan (6):

$$a_i = \frac{1}{n_k - 1} \sum_{j=1}^{n_k-1} d_{i,j}, j \neq i. \quad (6)$$

dengan n_k adalah banyak data pada kluster k .

(2) Menghitung rata-rata jarak dari suatu objek ke- i dengan semua objek yang berada pada kluster yang berbeda, menggunakan persamaan (7):

$$b_i = \frac{1}{n_k} \sum_{c=1}^{n_k} d_{i,c}. \quad (7)$$

(3) Menghitung nilai *Silhouette Coefficient*, menggunakan persamaan (8):

$$SC(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (8)$$

dengan

a_i : Rata-rata jarak objek ke- i dengan semua objek yang berada pada kluster yang sama

b_i : Rata-rata jarak objek ke- i dengan semua objek yang berada pada kluster yang berbeda.

5. Profilisasi dan interpretasi hasil kluster.

3. Hasil dan Pembahasan

3.1. Deskripsi Data Penelitian

Deskripsi data penelitian digunakan untuk mengetahui gambaran awal data penelitian yang digunakan. Diberikan contoh perhitungan rata-rata dan deviasi standar pada variabel pertama yaitu persentase penduduk miskin (x_1) berdasarkan Persamaan (2) dan (3), sehingga diperoleh nilai rata-rata 10,427% dan deviasi standar sebesar 5,412%. Selanjutnya perhitungan dilakukan untuk keseluruhan variabel dan diperoleh deskripsi data penelitian yang disajikan pada Tabel 2.

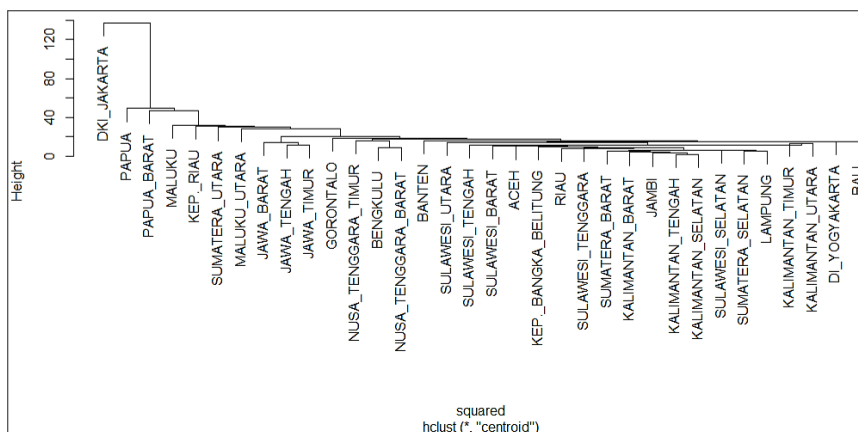
Tabel 2. Deskripsi Data Penelitian

Variabel	Minimum	Maksimum	Rata-Rata	Deviasi Standar
x_1	4,560	27,380	10,427	5,412
x_2	0,247	0,436	0,346	0,042
x_3	-2,460	16,790	4,204	3,626
x_4	6,760	11,170	8,716	0,928
x_5	3,010	9,910	5,492	1,819
x_6	62,150	73,780	67,818	3,028
x_7	0,600	3,480	1,504	0,553
x_8	9	15,978	745	2.721,058
x_9	65,290	75,080	70,196	2,513
x_{10}	13.092,810	174.941,720	43.557,160	32.795,195
x_{11}	68,080	76,340	72,611	2,098
x_{12}	0,200	0,980	0,520	0,166
x_{13}	971	36.534	7.044	7.702,057
x_{14}	15	289	131	67,991
x_{15}	32,830	100	67,617	17,416
x_{16}	0,010	0,120	0,034	0,023
x_{17}	3	90	27	20,534
x_{18}	0	17,600	2,436	3,605
x_{19}	132	5.171	1.224	1.371,925
x_{20}	4,380	83,150	15,555	13,353

Berdasarkan Tabel 2, terlihat jika setiap variabel penelitian memiliki karakteristik yang berbeda-beda. Sebagai contoh variabel pertama yaitu persentase penduduk miskin (x_1) terendah berasal dari Provinsi Kalimantan Selatan, yaitu sebesar 4,560%, sedangkan persentase penduduk miskin tertinggi berasal dari Provinsi Papua, yaitu sebesar 27,380%. Nilai rata-rata persentase penduduk miskin pada 34 Provinsi di Indonesia tahun 2021 adalah sebesar 10,427%.

3.2. Standardisasi Data

Standardisasi data bertujuan untuk membuat rentang nilai yang sama pada semua variabel yang digunakan [19]. Pada penelitian ini standardisasi yang digunakan adalah standardisasi Z_{score} berdasarkan pada Persamaan (1). Diberikan contoh perhitungan standardisasi Z_{score} pada variabel pertama persentase penduduk miskin (x_1), diperoleh untuk data Provinsi Aceh sebesar 0,943, Provinsi Sumatera Utara sebesar -0,358, Provinsi Sumatera Barat sebesar -0,811, hingga Provinsi Papua sebesar 3,133. Selanjutnya



Gambar 1. Dendrogram hasil pengelompokan menggunakan algoritma Centroid Linkage

perhitungannya dilakukan untuk keseluruhan variabel dan diperoleh data hasil standardisasi yang disajikan pada Tabel 3.

Tabel 3. Hasil Standardisasi Data

Provinsi	Z ₁	Z ₂	Z ₃	...	Z ₂₀
Aceh	0,943	-0,538	-0,390	...	-0,759
Sumatera Utara	-0,358	-0,775	-0,440	...	-0,578
Sumatera Barat	-0,811	-1,083	-0,252	...	0,163
⋮	⋮	⋮	⋮	⋮	⋮
Papua	3,133	1,188	3,021	...	-0,837

Data hasil standardisasi Z_{score} pada Tabel 3, yang selanjutnya digunakan pada proses pengelompokan dengan menggunakan Algoritma Centroid Linkage.

3.3. Pengelompokan Menggunakan Algoritma Centroid Linkage

Centroid Linkage adalah proses perhitungan rata-rata dari seluruh objek dalam suatu kluster. Metode ini fokus pada nilai rata-rata dari seluruh objek dalam kluster. Pada proses identifikasi kluster, jarak diukur sebagai kedekatan dengan centroid yang terbentuk. Centroid yang terbentuk adalah nilai tengah dari observasi pada variabel yang membentuk kluster tersebut. Pada pendekatan ini, setiap kali terbentuk kluster baru, pusat kluster dihitung kembali hingga tercapai suatu kluster yang diinginkan [20]. Setelah melalui beberapa tahapan sesuai algoritma Centroid Linkage, diperoleh dendrogram hasil pengelompokan yang disajikan pada Gambar 1.

Berdasarkan Gambar 1, dapat diketahui bahwa Provinsi yang pertama kali bergabung adalah Kalimantan Tengah dan Kalimantan Selatan. Provinsi selanjutnya adalah Jambi bergabung dengan Kalimantan Tengah dan Kalimantan Selatan. Proses pengelompokan terus berlanjut sampai dengan semua provinsi bergabung dalam satu kluster.

3.4. Validasi Hasil Pengelompokan Menggunakan Silhouette Coefficient

Pada penelitian ini dilakukan perhitungan nilai Silhouette Coefficient (SC) untuk mengetahui kualitas dari setiap hasil pengelompokan. Nilai SC yang paling maksimum menunjukkan pengelompokan yang paling optimal. Diberikan contoh perhitungan untuk jumlah dua kluster, dengan tahapan perhitungan SC dilakukan sebagai berikut:

1. Menghitung rata-rata jarak dari suatu objek ke- i dengan semua objek yang berada pada kluster yang sama. Contoh perhitungan rata-rata jarak dari data pertama yaitu Provinsi Aceh berdasarkan Persamaan (6) adalah sebagai berikut:

$$a_{Aceh} = \frac{1}{32} [6,020 + 4,009 + 4,145 + \dots + 8,056] = 4,888.$$

2. Menghitung jarak rata-rata dari Provinsi Aceh dengan semua objek pada kluster yang berbeda berdasarkan Persamaan (7), sehingga perhitungannya adalah sebagai berikut:

$$b_{Aceh} = \frac{1}{1} (12,404) = 12,404.$$

3. Nilai a_{Aceh} dan b_{Aceh} yang telah diperoleh, selanjutnya digunakan untuk menghitung SC berdasarkan Persamaan (8):

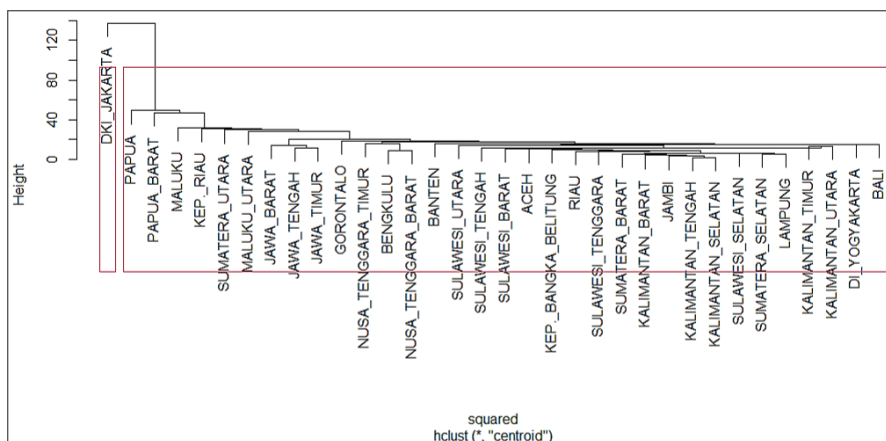
$$SC_{Aceh} = \frac{b_{Aceh} - a_{Aceh}}{\max\{a_{Aceh}, b_{Aceh}\}} = \frac{12,404 - 4,888}{\max\{4,888, 12,404\}} = 0,606.$$

Nilai SC untuk Provinsi lainnya dapat dihitung dengan formula yang sama dengan Provinsi Aceh. Nilai SC untuk tiap Provinsi disajikan pada Tabel 4.

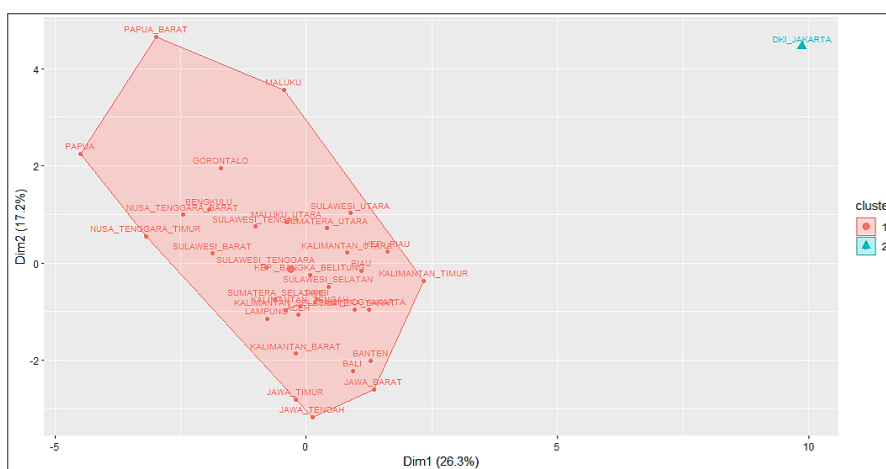
Tabel 4. Perhitungan SC Lokal Setiap Provinsi

Provinsi	Nilai SC
Aceh	0,606
Sumatera Utara	0,434
Sumatera Barat	0,585
⋮	⋮
Papua	0,467
Rata-Rata	0,538

Berdasarkan Tabel 4, diperoleh nilai SC untuk jumlah kluster sebanyak dua kluster, sebesar 0,538. Perhitungan yang sama dilakukan untuk jumlah tiga kluster, empat kluster, dan lima kluster. Perhitungan SC untuk jumlah kluster sebanyak dua, tiga, empat, dan lima kluster disajikan pada Tabel 5.



Gambar 2. Classification Tree dengan Dendrogram hasil pengelompokan optimal menggunakan algoritma Centroid Linkage



Gambar 3. Classification Tree dengan Cluster plot hasil pengelompokan optimal menggunakan algoritma Centroid Linkage

Tabel 5. Perbandingan Hasil Validasi Kluster Berdasarkan Nilai SC

Banyak Kluster	Nilai SC
2	0,538*
3	0,294
4	0,265
5	0,190

Catatan: (*) Nilai SC tertinggi

Berdasarkan hasil perhitungan pada Tabel 5, diketahui nilai SC yang paling optimal adalah sebanyak 2 kluster sebesar 0,538, sehingga bisa disimpulkan pengelompokan Provinsi di Indonesia berdasarkan indikator ekonomi, kesehatan, pendidikan, dan kriminalitas adalah sebanyak 2 kluster.

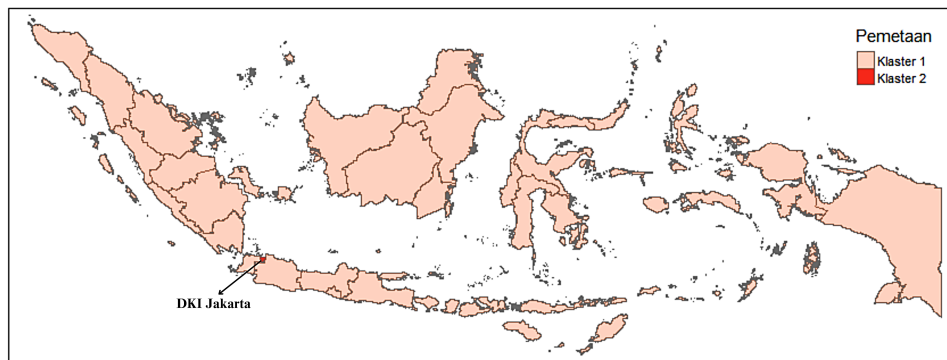
Dendrogram hasil pengelompokan optimal yaitu sebanyak 2 kluster ditampilkan pada Gambar 2 dan cluster plot ditampilkan pada Gambar 3. Berdasarkan hasil pengelompokan dua kluster, dilakukan perhitungan rata-rata variabel untuk masing-masing kluster. Hasil perhitungan secara lengkap disajikan pada Tabel 6.

Berdasarkan Tabel 6, terlihat bahwa kluster 1 dan kluster 2 memiliki pembeda ciri yang berbeda berdasarkan masing-masing variabel. Kluster 1 merupakan kluster dengan variabel persentase penduduk miskin, laju pertumbuhan ekonomi, tingkat partisipa-

Tabel 6. Nilai Rata-Rata Variabel Untuk Masing-Masing Kluster

Variabel	Kluster	
	1	2
x_1	10,602*	4,670
x_2	0,344	0,411*
x_3	4,223*	3,560
x_4	8,642	11,170*
x_5	5,401	8,500*
x_6	68,292*	62,630
x_7	1,532*	0,600
x_8	283	15,978*
x_9	70,109	73,060*
x_{10}	39.575,810	174.941,720*
x_{11}	72,669*	70,680
x_{12}	0,523*	0,400
x_{13}	6.376	29.103*
x_{14}	126	277*
x_{15}	66,697	97,990*
x_{16}	0,033	0,050*
x_{17}	28*	11
x_{18}	1,976	17,600*
x_{19}	1,255*	227
x_{20}	13,507	83,150*

Catatan: (*) Rata-rata kluster yang tertinggi



Gambar 4. Peta profilisasi hasil pengelompokan optimal

si angkatan kerja, laju pertumbuhan penduduk, indeks kebahagiaan, persentase penduduk korban kejahatan, jumlah desa yang pernah ada kejadian kejahatan pembunuhan, dan jumlah desa yang membangun poskamling yang lebih tinggi jika dibandingkan klaster 2. Klaster 2 merupakan klaster dengan variabel gini ratio, rata-rata lama sekolah, tingkat pengangguran terbuka, kepadatan penduduk, angka harapan hidup, PDRB per kapita harga konstan, jumlah kejahatan yang dilaporkan, risiko penduduk terkena kejahatan per 100.000 penduduk, persentase penyelesaian kejahatan, proporsi penduduk yang menjadi korban kejahatan, persentase desa yang ada kejadian konflik massal, dan persentase desa menurut keberadaan pos polisi yang lebih tinggi jika dibandingkan dengan klaster 1.

3.5. Profilisasi dan Interpretasi Hasil Klaster

Pada penelitian ini dilakukan profilisasi hasil klaster yang paling optimal, dimana pada penelitian ini dilakukan uji validitas klaster menggunakan *Silhouette Coefficient* untuk melihat banyak klaster yang menghasilkan pengelompokan optimal. Jumlah klaster yang paling optimal pada penelitian ini adalah sebanyak 2 klaster. Peta hasil pengelompokan ditampilkan pada Gambar 4.

Berdasarkan Gambar 4, dapat dilihat pada klaster 1 yang berwarna merah muda ditempati oleh sebagian besar Provinsi yakni sebanyak 33 Provinsi, sedangkan pada klaster 2 yang berwarna merah hanya diisi 1 Provinsi yaitu DKI Jakarta. DKI Jakarta memiliki profil yang berbeda dengan provinsi-provinsi lainnya, meskipun unggul dalam pendidikan dan kesehatan, DKI Jakarta juga menghadapi tantangan ekonomi dan kriminalitas yang mungkin lebih kompleks. Hal ini menunjukkan bahwa dinamika dan karakteristik provinsi Ibu Kota Negara, yaitu DKI Jakarta berdasarkan indikator ekonomi, pendidikan, kesehatan, dan kriminalitas membedakannya secara signifikan dari wilayah-wilayah lain di Indonesia.

4. Kesimpulan

Berdasarkan hasil analisis dapat disimpulkan bahwa pengelompokan dengan algoritma *Centroid Linkage* menunjukkan jumlah klaster yang optimal adalah 2 klaster, dengan nilai *Silhouette Coefficient* (SC) sebesar 0,538. Klaster 1 terdiri dari 33 Provinsi yang cenderung memiliki perekonomian yang lebih baik dan tingkat kriminalitas yang rendah, sedangkan klaster 2 terdiri dari hanya satu Provinsi yaitu DKI Jakarta yang walaupun menonjol dalam aspek pendidikan dan kesehatan, namun dihadapkan pada masalah kriminalitas. Selanjutnya hasil ini dapat dijadikan refe-

rensi bagi pemerintah sebagai perumus kebijakan dan pemangku kepentingan untuk menyelesaikan masalah ekonomi, pendidikan, kesehatan, dan kriminalitas sesuai dengan prioritasnya dengan tujuan meningkatkan kesejahteraan dan keamanan dalam masyarakat.

Kontribusi Penulis. Yossy Candra: Identifikasi masalah, pemilihan studi kasus, metodologi, analisis data, dan penulisan naskah. Rito Goejantoro: Konseptualisasi, tinjauan penulisan, dan supervisi. Andrea Tri Rian Dani: Konseptualisasi, tinjauan penulisan, dan supervisi. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis menyampaikan terima kasih kepada editor dan reviewer atas pembacaan yang cermat, kritik yang mendalam, dan rekomendasi yang praktis untuk meningkatkan kualitas tulisan ini. Ucapan terima kasih juga disampaikan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman yang telah memfasilitasi terlaksananya penelitian ini.

Pembiayaan. Penelitian ini tidak menerima pembiayaan eksternal

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Referensi

- [1] N. M. A. Aditia, and N. P. M. Dewi, "Pengaruh Pengeluaran Pemerintah di Bidang Pendidikan, Kesehatan, dan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat di Provinsi Bali," *Jurnal Ekonomi Pembangunan*, vol. 7, no. 2, pp. 212-242, 2018. DOI: <https://erepo.unud.ac.id/eprint/19459>.
- [2] D. P. Wardana, "Pengaruh Pembangunan Ekonomi Terhadap Pembangunan Manusia di Kalimantan Timur," *Jurnal Ekonomi Keuangan dan Manajemen*, vol. 12, no. 2, pp. 179-191, 2016. DOI: <https://doi.org/10.30872/jinv.v12i2.804>.
- [3] E. A. Aini, I. Isnaini, S. Sukanti, and L. N. Amalia, "Pengaruh Tingkat Pendidikan Terhadap Tingkat Kesejahteraan Masyarakat di Kelurahan Kesatrian Kota Malang," *Technomedia Journal*, vol. 3, no. 1, pp. 58-72, 2018. DOI: <https://doi.org/10.33050/tmj.v3i1.333>.
- [4] B. Tambaip, A. P. Tjilen, and Y. Ohoiwutun, "Peran Fasilitas Kesehatan untuk Kesejahteraan Masyarakat," *Jurnal Kebijakan Publik*, vol. 14, no. 2, pp. 189-196, 2023. DOI: <http://dx.doi.org/10.31258/jkp.v14i2.8245>.
- [5] Sugiharti, Lilik, M. A. Esquivas, M. S. Shaari, L. Agustini, and H. Rohmawati, "Criminality and Income Inequality in Indonesia," *Social Science*, vol. 11, no. 3, pp. 142, 2022. DOI: <https://doi.org/10.3390/socsci11030142>.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques Third Edition*. Morgan Kaufman: Massachusetts (US), 2012.
- [7] B. Santoso, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta, 2007.
- [8] A. T. R. Dani, F. B. Putra, M. Fauziyah, Sifriyani, Suyitno, and M. Fathurahman, "K-Means Algorithm For Grouping Provinces In Indonesia Based On Macroeconomic And Criminality Indicators," *Jurnal Statistika*, vol. 11, no. 2, pp. 12-21, 2023. DOI: <https://doi.org/10.26714/jsunimus.11.2.2023.12-21>.

- [9] M. Yafi, R. Goejantoro, and A. T. R. Dani, "Pengelompokan Algoritma K-Medoids dengan Principal Component Analysis (PCA)," *Prosiding Seminar Nasional Matematika, Statistika, Aplikasinya*, vol. 3, no. 1, pp. 183-195, 2023. ISSN 2657-232X.
- [10] Widyawati, W. L. Saptomo, and Y. R. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *Jurnal Ilmiah Sinus (JIS)*, vol. 18, no. 1, pp. 75-87, Jan. 2020. Doi:<http://dx.doi.org/10.30646/sinus.v18i1.448>.
- [11] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2017.
- [12] F. Liantoni, and H. Nugroho, "Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *Jurnal Simantec*, vol. 5, no. 1, pp. 9-16, Dec. 2015. DOI: <https://doi.org/10.21107/simantec.v5i1.1009>.
- [13] F. Nielsen, *Introduction to HPC with MPI for Data Science*. Springer, 2016.
- [14] Suyanti, and Y. L. Sukestiyarno, "Deteksi Outlier Menggunakan Diagnosa Regresi Berbasis Estimator Parameter Robust," *Jurnal Matematika*, vol. 3, no. 2, pp. 118-125, 2014. DOI: <https://doi.org/10.15294/UJM.V3I2.4341>.
- [15] R. Silvi, "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia," *Jurnal Matematika MANTIK*, vol. 4, no. 1, pp. 22-31, 2018. DOI: <https://doi.org/10.15642/mantik.2018.4.1.22>.
- [16] D. Dzirkullah, "Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet dengan Metode Centroid Linkage," *Jurnal of Mathematics, Computations, and Statistics*, vol. 5, no. 1, pp. 48-57, 2022. DOI: <https://doi.org/10.35580/jmathcos.v5i1.32363>.
- [17] P. Puspita, "Perbandingan Metode Centroid dan Ward dalam Pengelompokan Tingkat Penyelesaian Pendidikan di Indonesia," *Jurnal Ilmiah Pendidikan Matematika dan Statistika*, vol. 3, no. 3, pp. 501-510, 2022. DOI: <https://doi.org/10.46306/lb.v3i3.159>.
- [18] N. A. Raja, G. M. Tinungki, and N. Sirajang, "Implementasi Algoritma Centroid Linkage dan K-Medoids dalam Mengelompokkan Kabupaten/Kota di Sulawesi Selatan Berdasarkan Indikator Pendidikan," *ESTIMASI: Journal of Statistics and Its Application*, vol. 5, no. 1, pp. 61-74, 2024. DOI:<https://doi.org/10.20956/ejsa.v5i1.13605>.
- [19] I. Permana, and F. N. Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation," *IJRSE*, vol. 2, no. 1, pp. 67-72, 2022. DOI: <https://doi.org/10.57152/ijrse.v2i1.311>.
- [20] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, no. 2, pp. 64-78, 2019. DOI: <https://doi.org/10.34312/jjom.v1i2.2354>.