

# Comparison of Agglomerative Hierarchical Clustering Methods for Grouping Indonesian Provinces Based on Community Literacy Development Index

Olga Afrilly Putri\*, Bunga Nafandra, Zamahsary Martha

Departemen Statistika, Universitas Negeri Padang, Indonesia

\*Corresponding author: [olgafrillyputri@gmail.com](mailto:olgafrillyputri@gmail.com)

Submitted : 10 Februari 2026

Revised : 16 Maret 2026

Accepted : 17 Maret 2026

## ABSTRACT

*Community literacy development is a key indicator in enhancing the quality of human resources in Indonesia. This research aims to group provinces in Indonesia based on the Community Literacy Development Index by considering the equity of library services, the adequacy of library collections, and the level of community visits per day. The method used is agglomerative hierarchical cluster analysis. Prior to clustering, the data are standardized to address variations in measurement units and variable scales. The most appropriate clustering method is identified using the cophenetic correlation coefficient, while the optimal number of clusters is determined based on the silhouette method. The findings indicate that the Average Linkage method yields the most reliable hierarchical clustering solution, resulting in four distinct clusters. Each cluster demonstrates unique characteristics, representing variations in literacy levels, service performance, adequacy of collections, and library visitation rates across provinces. The results reveal disparities in literacy development among regions in Indonesia. Consequently, this study provides valuable insights that may support the formulation of more targeted and effective literacy and library development policies.*

**Keywords:** *Agglomerative Hierarchical Method, Clustering Analysis, Community Literacy Development Index*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Di dalam era digital sekarang ini, literasi masyarakat sangat penting dalam pembangunan sumber daya manusia dan berpengaruh terhadap kualitas sosial serta ekonomi. Literasi masyarakat dalam penelitian ini merujuk pada kemampuan masyarakat dalam mengakses, memahami, dan memanfaatkan berbagai sumber informasi serta pengetahuan secara efektif untuk meningkatkan kualitas hidup. Inisiasi literasi masyarakat merupakan sebuah usaha untuk meningkatkan kesadaran dan keterampilan literasi dalam masyarakat (Syarifudin et al., 2023).

Untuk meningkatkan pemahaman masyarakat dan pemerataan kualitas hidup dalam hal kemudahan mengakses informasi serta ilmu pengetahuan, diperlukan suatu pengukuran nilai Indeks Pembangunan Literasi Masyarakat (IPLM). Berdasarkan Badan Perpustakaan Nasional Republik Indonesia (2024), nilai Indeks Pembangunan Literasi Masyarakat (IPLM) antarprovinsi di Indonesia menunjukkan ketimpangan atau keberagaman yang cukup tinggi. Untuk itu diperlukan analisis yang mampu menggambarkan pola kemiripan dan perbedaan antarwilayah agar dapat mengidentifikasi struktur pembangunan literasi masyarakat secara lebih komprehensif serta mengungkap ketimpangan Indeks Pembangunan Literasi Masyarakat antarprovinsi.

Pendekatan analisis yang relevan untuk mengklasifikasikan provinsi-provinsi di Indonesia berdasarkan karakteristik Indeks Pembangunan Literasi Masyarakat (IPLM) ialah pendekatan klasterisasi, khususnya klasterisasi hierarki. Karena klasterisasi hierarki mampu menggambarkan hubungan antarprovinsi secara visual dan interpretatif. Untuk data yang sangat banyak, klasterisasi dapat dilakukan dengan menerapkan algoritma khusus, mengingat pengelompokan data tidaklah praktis jika dilakukan secara manual dengan memeriksa setiap objek data (Hartono et al., 2025). Sebuah algoritma yang sesuai untuk mencapai tujuan tersebut ialah algoritma aglomeratif. *Hierarchical Agglomerative Clustering* (HAC) beroperasi dengan menyatukan objek-objek yang paling mirip secara bertahap. Pada tahap awal, setiap objek diperlakukan sebagai klaster tersendiri, kemudian secara berkelanjutan digabungkan hingga seluruh objek tergabung dalam satu klaster besar (Rahmawati & Fallo, 2025).

Penelitian sebelumnya yang membahas analisis klater dengan pendekatan hierarki diantaranya dilakukan oleh Novaldi & Wijayanto (2023) yang mengkaji pengelompokan provinsi di Indonesia berdasarkan kualitas pemuda tahun 2022 di Indonesia menggunakan metode *hierarchical agglomerative clustering*. Hasil penelitian tersebut menunjukkan bahwa pendekatan ini efektif untuk mengidentifikasi pola kesamaan dan perbedaan antarwilayah secara objektif. Namun demikian, penelitian tersebut masih berfokus pada indikator kualitas pemuda dan belum mengeksplorasi aspek pembangunan literasi masyarakat secara komprehensif. Oleh sebab itu, penelitian ini dilakukan untuk memperluas kajian dengan membandingkan beberapa metode klasterisasi hierarki aglomeratif dalam pengelompokan provinsi di Indonesia berdasarkan Indeks Pembangunan Literasi Masyarakat (IPLM), serta menganalisis struktur pengelompokan yang lebih komprehensif dan informatif.

Berdasarkan isu yang dibahas, penelitian ini menggunakan metode klasterisasi hierarki aglomeratif untuk mengelompokkan provinsi-provinsi di Indonesia menurut karakteristik Indeks Pembangunan Literasi Masyarakat (IPLM). Dalam penelitian ini dilakukan perbandingan beberapa algoritma pada metode klasterisasi hierarki, yaitu *single linkage*, *complete linkage*, dan *average linkage*. Perbandingan tersebut dilakukan untuk mengetahui metode yang menghasilkan struktur pengelompokan yang paling baik dalam menggambarkan kemiripan karakteristik IPLM antar provinsi di Indonesia.

Melalui analisis ini diharapkan dapat diperoleh pengelompokan provinsi yang memiliki karakteristik pembangunan literasi masyarakat yang serupa, sehingga dapat memberikan gambaran mengenai struktur pembangunan literasi masyarakat secara nasional serta menjadi dasar dalam perumusan kebijakan yang lebih tepat sasaran.

## II. METODE PENELITIAN

### A. Data Penelitian

Data dalam penelitian ini bersumber dari data sekunder yang diperoleh melalui laman resmi Badan Pusat Statistik (BPS) Indonesia. Data yang dianalisis berupa Indeks Pembangunan Literasi Masyarakat (IPLM) menurut provinsi di Indonesia tahun 2024. Data tersebut digunakan sebagai input dalam analisis klaster untuk mengelompokkan provinsi berdasarkan tingkat pembangunan literasi masyarakat. Rincian variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala
X1	Indeks Pembangunan Literasi Masyarakat	Numerik
X2	Pemerataan Layanan Perpustakaan	Numerik
X3	Ketercukupan Koleksi Perpustakaan	Numerik
X4	Tingkat Kunjungan Masyarakat per Hari	Numerik

Variabel yang digunakan dalam penelitian ini terdiri dari Indeks Pembangunan Literasi Masyarakat, pemerataan layanan perpustakaan, ketercukupan koleksi perpustakaan, serta kunjungan masyarakat per hari. Pemilihan variabel tersebut didasarkan pada indikator yang digunakan dalam pengukuran Indeks Pembangunan Literasi Masyarakat (IPLM) yang dikembangkan oleh Perpustakaan Nasional Republik Indonesia. Indikator-indikator tersebut merepresentasikan aspek ketersediaan layanan literasi, akses terhadap sumber bacaan, serta pemanfaatan fasilitas literasi oleh masyarakat.

### B. Teknik Analisis Data

Metode yang diterapkan dalam penelitian ini ialah metode analisis perbandingan klasterisasi hierarki aglomeratif yang meliputi metode *single linkage*, *complete linkage*, dan *average linkage*. Dari tiga metode itu dibandingkan untuk memilih metode yang paling efektif dalam mengelompokkan provinsi-provinsi di Indonesia didasarkan Indeks Pembangunan Literasi Masyarakat. Langkah-langkah analisis dalam penelitian ini dilakukan melalui tahapan sebagai berikut:

#### 1) Standarisasi Data

Standarisasi data diperlukan ketika terdapat perbedaan satuan pengukuran yang cukup mencolok antarvariabel penelitian, dengan tujuan meminimalkan dominasi satu variabel terhadap variabel lainnya serta mencegah terjadinya bias dalam hasil analisis (Irwan et al., 2024). Proses standarisasi dilakukan dengan mentransformasikan data ke dalam nilai *z-score* yang memiliki nilai rata-rata sebesar 0 ( $\mu = 0$ ) dan simpangan baku sebesar 1 ( $\sigma = 1$ ). Adapun formula yang digunakan dalam standarisasi data mengacu pada Whendasmoro & Joseph (2022):

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Keterangan:

$Z_i$  = nilai standarisasi data ke- $i$

$x_i$  = nilai data ke- $i$

$\mu$  = rata-rata

$\sigma$  = standar deviasi

## 2) Membentuk Matriks Jarak *Euclidean*

Ukuran jarak digunakan untuk merepresentasikan tingkat kemiripan antarobjek dalam suatu analisis. Nilai jarak yang semakin besar menunjukkan tingkat ketidakmiripan yang semakin tinggi antarobjek, sedangkan nilai jarak yang semakin kecil mencerminkan tingkat kesamaan yang lebih besar (Irwan et al., 2024). Salah satu metode pengukuran jarak yang paling umum diterapkan adalah jarak *Euclidean*, diperoleh melalui perhitungan selisih antar dua objek yang kemudian dikuadratkan dan dijumlahkan sesuai dengan dimensinya (Pangestu & Fitriani, 2022).

$$d(i, j) = \sum_{i=1}^p (x_{ik} - x_{jk})^2 \quad (2)$$

Keterangan:

$d(i, j)$  : jarak *euclidean* antara objek ke- $x$  dengan objek ke- $y$

$p$  : banyak variabel kluster

$x_{ik}$  : data pada objek ke- $i$  variabel ke- $k$

$x_{jk}$  : data pada objek ke- $j$  variabel ke- $k$

## 3) Analisis Kluster Hierarki Metode Aglomeratif

Analisis kluster merupakan suatu metode pengelompokan sejumlah objek ke dalam kelompok-kelompok tertentu, di mana objek-objek yang berada dalam satu kluster memiliki tingkat kemiripan yang tinggi, sedangkan objek-objek antar kluster memiliki tingkat perbedaan yang besar (Jollyta et al., 2021). Analisis kluster hierarki dilakukan dengan cara menggabungkan dua atau lebih objek yang memiliki tingkat kesamaan paling tinggi, kemudian proses pengelompokan dilanjutkan secara bertahap dengan objek lain yang memiliki jarak terdekat hingga seluruh objek tergabung ke dalam kluster dan membentuk struktur bertingkat (hierarki) yang dapat direpresentasikan dalam bentuk dendrogram (Mubarok & Rusyiana, 2020). Analisis kluster hierarki metode aglomeratif merupakan teknik pengelompokan data yang bersifat *bottom-up*, di mana setiap objek pengamatan pada awalnya dianggap sebagai satu kluster tunggal, kemudian secara bertahap digabungkan dengan objek atau kluster lain yang memiliki tingkat kemiripan paling tinggi hingga membentuk struktur kluster bertingkat (hierarki) (Jollyta et al., 2021). Metode klusterisasi hierarki dinilai efektif untuk pengelompokan data dengan jumlah objek kurang dari 100 (Ghaisani et al., 2019).

Dalam klusterisasi hierarki aglomeratif terdapat beberapa metode yang dapat digunakan, yaitu *single linkage*, *average linkage*, dan *complete linkage*.

### a. Metode *Single Linkage*

Metode ini merupakan teknik klusterisasi yang menetapkan penggabungan kluster berdasarkan jarak paling dekat antar objek. Dua objek atau kluster akan digabungkan apabila memiliki jarak minimum yang paling kecil dibandingkan pasangan lainnya, sehingga proses pengelompokan dilakukan secara bertahap dengan menggabungkan objek-objek yang saling berdekatan (Prabowo et al., 2020). Secara matematis, jarak antar kluster pada metode *single linkage* untuk menentukan jarak kluster ( $i, j$ ) dengan kluster  $k$  dapat dirumuskan sebagai berikut (Iis et al., 2022):

$$d_{(ij)k} = \min (d_{ik}, d_{jk}) \quad (3)$$

Keterangan:

$d_{ik}$  = jarak antara kelompok  $i$  dan  $k$

$d_{jk}$  = jarak antara kelompok  $j$  dan  $k$

$d_{(ij)k}$  = jarak antara kelompok  $ij$  dan kelompok ke  $k$

b. Metode *Complete Linkage*

Metode ini merupakan teknik klusterisasi yang menentukan penggabungan kluster berdasarkan jarak maksimum antar objek. Dua objek atau kluster akan digabungkan apabila jarak terjauh antar anggota kluster tersebut paling kecil dibandingkan pasangan lainnya, sehingga proses pengelompokan cenderung menghasilkan kluster yang lebih kompak dan homogen (Prabowo et al., 2020). Secara matematis, jarak antar kluster pada metode *Complete Linkage* untuk menentukan jarak antara kluster  $(i,j)$  dengan kluster  $k$  dirumuskan sebagai berikut (Iis et al., 2022).

$$d_{(ij)k} = \max (d_{ik}, d_{jk}) \tag{4}$$

Keterangan:

$d_{ik}$  = jarak antara kelompok  $i$  dan  $k$

$d_{jk}$  = jarak antara kelompok  $j$  dan  $k$

$d_{(ij)k}$  = jarak antara kelompok  $ij$  dan kelompok ke  $k$

c. Metode *Average Linkage*

Metode ini merupakan metode klusterisasi yang mendasarkan proses penggabungan kluster pada nilai jarak rata-rata antar objek (Prabowo et al., 2020). Jarak antara dua kluster dihitung sebagai rata-rata jarak seluruh pasangan objek yang berasal dari kluster-kluster yang akan digabungkan, dengan tujuan meminimalkan nilai rata-rata jarak tersebut sehingga diperoleh struktur kluster yang relatif seimbang (Asiska et al., 2019). Rumusan matematis metode *Average Linkage* untuk menentukan jarak antara kluster  $(i,j)$  dengan kluster  $k$  disajikan sebagai berikut (Iis et al., 2022).

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{n_{(ij)}n_k} \tag{5}$$

Keterangan:

$d_{ab}$  = jarak antara objek  $a$  kluster  $(ij)$  dan objek  $b$  kluster  $k$

$n_{(ij)}$  = jumlah objek pada kluster  $(ij)$

$n_k$  = jumlah objek pada kluster  $k$

**4) Menentukan Metode Klustering**

Koefisien korelasi *cophenetic* merupakan salah satu indikator yang digunakan untuk menguji validitas hasil klusterisasi (Pratiwi et al., 2019). Koefisien korelasi *cophenetic* digunakan untuk menilai tingkat kesesuaian antara nilai-nilai pada matriks ketidakmiripan awal, seperti matriks jarak *Euclidean*, dengan nilai-nilai yang dihasilkan pada dendrogram atau matriks *cophenetic* setelah proses klusterisasi dilakukan (Iis et al., 2022). Nilai koefisien korelasi *cophenetic* yang semakin mendekati angka satu menunjukkan bahwa struktur dendrogram yang dihasilkan mampu merepresentasikan data dengan baik, sehingga solusi klustering yang diperoleh dapat dikatakan memiliki kualitas yang memadai (Iis et al., 2022).

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d}) (d_{c_{ik}} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2] [\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2]}} \tag{6}$$

Keterangan:

$r_{coph}$  = koefisien korelasi *cophenetic*

$d_{ik}$  = jarak asli (*euclidean*) antara objek ke- $i$  dengan objek ke- $k$

$\bar{d}$  = rata-rata  $d_{ik}$

$d_{c_{ik}}$  = jarak *cophenetic* objek ke- $i$  dengan objek ke- $k$

$\bar{d}_c$  = rata-rata  $d_{c_{ik}}$

**5) Pemilihan Jumlah Kluster Optimal**

a. Metode *Elbow* dengan *Sum of Square Error* (SSE)

Metode *Elbow* ini adalah salah satu teknik yang digunakan untuk menentukan jumlah kluster yang paling optimal dengan mengevaluasi perbandingan variasi antar kluster pada setiap alternatif jumlah kluster yang diuji. Metode ini memanfaatkan visualisasi grafik untuk menunjukkan perubahan nilai variasi tersebut, di mana titik yang membentuk pola menyerupai siku (*elbow*) mengindikasikan jumlah kluster yang paling sesuai. Semakin jelas perubahan kemiringan grafik pada titik tersebut, maka jumlah kluster yang dihasilkan dapat dianggap sebagai jumlah kluster yang optimal (Jollyta et al., 2021).

b. Metode *Silhouette Index*

Metode ini ialah salah satu pendekatan validasi kluster yang berfungsi untuk mengevaluasi keakuratan hasil pengelompokan data, baik pada tingkat individu, kluster tunggal, maupun keseluruhan kluster yang terbentuk. Indeks ini menilai seberapa dekat setiap objek dengan kluster yang menaunginya dibandingkan dengan kluster terdekat lainnya, sehingga dapat menggambarkan tingkat kohesi dalam kluster serta pemisahan antar kluster. Selain itu, *Silhouette Index* dapat dimanfaatkan sebagai alat visual untuk menilai kesesuaian parameter pengelompokan, seperti penentuan jumlah kluster yang optimal (Jollyta et al., 2021).

### III. HASIL DAN PEMBAHASAN

#### A. Statistika Deskriptif

Tabel 2 memuat rangkuman statistik deskriptif dari seluruh variabel penelitian.

**Tabel 2.** Statistika Deskriptif Variabel Penelitian

Variabel	Mean	Standar Deviasi	Min	Max
X1	68,18	10,8137	35,25	88,24
X2	0,427	0,1804	0,026	0,763
X3	0,387	0,2051	0,033	0,848
X4	0,323	0,2384	0,038	1,000

Berdasarkan Tabel 2, variabel Indeks Pembangunan Literasi Masyarakat (X1) memiliki rata-rata sebesar 68,18 dengan standar deviasi 10,8137, yang menunjukkan adanya variasi tingkat literasi masyarakat antar wilayah. Variabel Pemerataan Layanan Perpustakaan (X2) memiliki nilai rata-rata 0,427, mengindikasikan bahwa distribusi layanan perpustakaan belum merata secara optimal. Selanjutnya, Ketercukupan Koleksi Perpustakaan (X3) memiliki rata-rata 0,387, yang mencerminkan perbedaan kecukupan koleksi antar wilayah. Sementara itu, Tingkat Kunjungan Masyarakat per Hari (X4) memiliki nilai rata-rata 0,323, yang menunjukkan bahwa intensitas kunjungan masyarakat ke perpustakaan masih relatif rendah dan bervariasi. Secara umum, seluruh variabel menunjukkan adanya variasi yang cukup besar sehingga mendukung perlunya analisis lanjutan.

#### B. Standarisasi Data

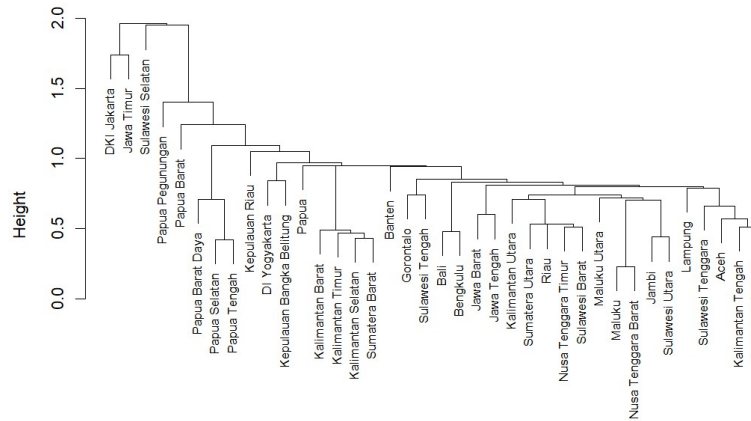
Data yang digunakan dalam penelitian ini memiliki perbedaan satuan dan skala pengukuran antarvariabel, sehingga diperlukan proses standarisasi sebelum dilakukan analisis kluster. Hasil data yang telah distandarisasi selanjutnya disajikan pada Tabel 3.

**Tabel 3.** Data Setelah Standarisasi

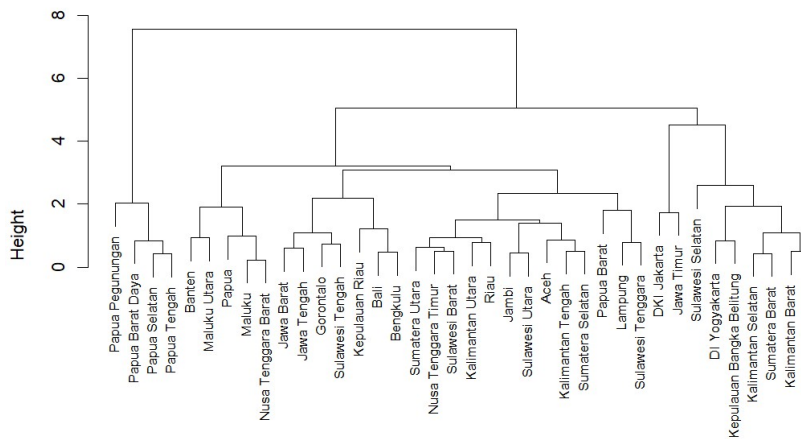
Provinsi	X1	X2	X3	X4
Aceh	0,3920	-0,5841	-0,0287	0,1824
Bali	-0,1971	1,4119	-0,4358	-0,5852
Banten	-0,5827	-0,5924	-1,2451	1,1791
...	...	...	...	...
Sumatera Barat	1,3214	1,0001	1,7376	0,4152
Sumatera Selatan	0,3752	0,1470	0,1521	-0,2467
Sumatera Utara	-0,5356	-0,3657	-0,1657	-0,4279

#### C. Analisis Klustering Hierarki

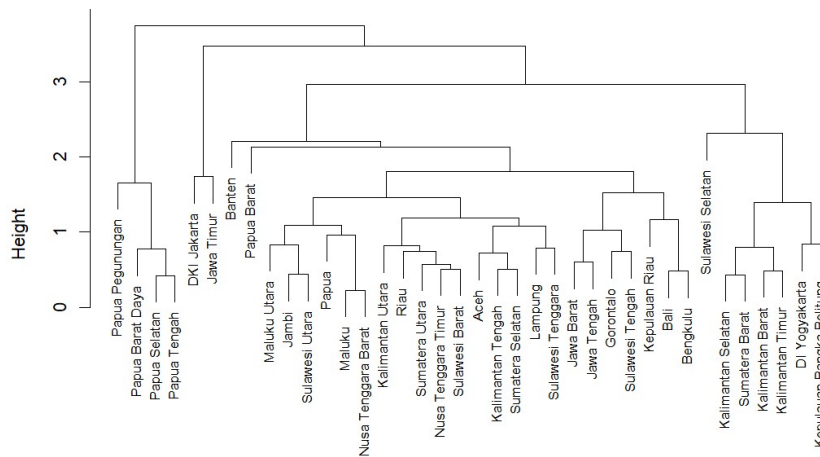
Pembentukan kluster dilakukan dengan 3 metode kluster hierarki aglomeratif yang berbeda, yaitu *single linkage*, *complete linkage*, dan *average linkage*. Adapun hasil dendrogram untuk ketiga metode tersebut disajikan pada Gambar 1, Gambar 2, dan Gambar 3.



Gambar 1. Dendrogram Metode *Single Linkage*



Gambar 2. Dendrogram Metode *Complete Linkage*



Gambar 3. Dendrogram Metode *Average Linkage*

Berdasarkan Gambar 1, Gambar 2, dan Gambar 3 yang menampilkan dendrogram hasil analisis *hierarchical clustering* menggunakan metode *single linkage*, *complete linkage*, dan *average linkage*, terlihat bahwa ketiga metode tersebut menghasilkan struktur pengelompokan wilayah yang berbeda. Perbedaan ini disebabkan oleh perbedaan

kriteria penggabungan jarak kluster pada masing-masing metode, dimana *single linkage* menggunakan jarak minimum antar anggota kluster, *complete linkage* menggunakan jarak maksimum, sedangkan *average linkage* menggunakan rata-rata jarak antar anggota kluster.

Selain itu, dendrogram yang dihasilkan belum menunjukkan secara langsung jumlah kluster yang paling optimal. Dendrogram hanya menggambarkan proses penggabungan objek secara bertahap berdasarkan jarak kedekatan antar objek. Oleh karena itu, diperlukan tahap evaluasi lebih lanjut untuk menentukan metode klustering yang menghasilkan pengelompokan terbaik sekaligus menentukan jumlah kluster yang optimal berdasarkan ukuran validitas kluster.

#### D. Menentukan Metode Klustering Terbaik

Penentuan metode klustering hierarki yang paling sesuai dilakukan dengan menggunakan nilai koefisien korelasi *cophenetic* sebagai kriteria evaluasi. Metode yang menghasilkan nilai koefisien korelasi *cophenetic* tertinggi dianggap sebagai metode yang paling optimal dan selanjutnya dianalisis lebih lanjut dalam penelitian ini. Nilai koefisien korelasi *cophenetic* untuk masing-masing metode klustering hierarki aglomeratif dapat dilihat pada Tabel 4.

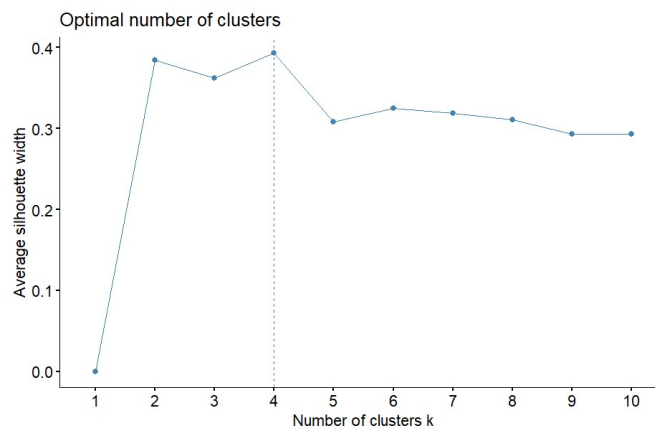
**Tabel 4.** Nilai Koefisien Korelasi *Cophenetic*

Metode	Koefisien Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0,6097
<i>Complete Linkage</i>	0,7045
<i>Average Linkage</i>	0,7425

Berdasarkan hasil yang ditunjukkan pada Tabel 4, metode *Average Linkage* menghasilkan nilai koefisien korelasi *cophenetic* paling tinggi, yakni sebesar 0,7425. Temuan ini mengindikasikan bahwa metode tersebut memberikan struktur kluster yang lebih representatif dibandingkan metode *single linkage* dan *complete linkage*. Oleh karena itu, metode *Average Linkage* dipilih dan digunakan dalam penelitian ini untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan Indeks Pembangunan Literasi Masyarakat.

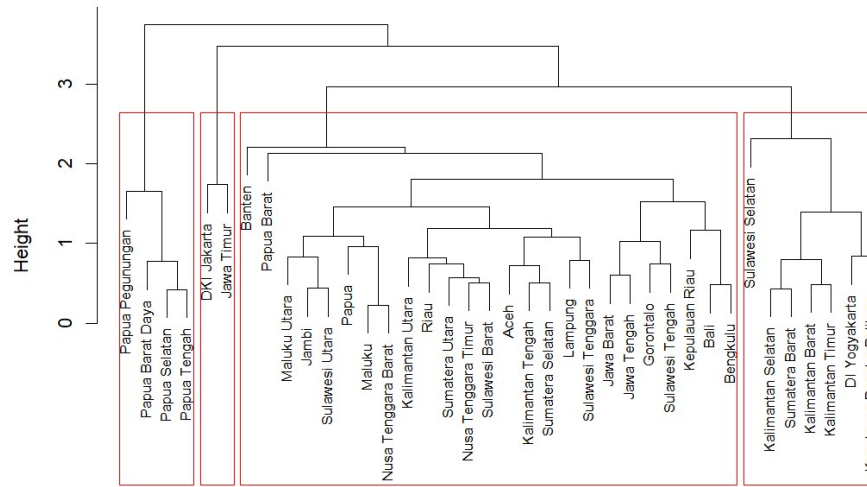
#### E. Menentukan Jumlah Kluster Optimum

Penentuan jumlah kluster optimum dilakukan setelah proses pembentukan kluster hierarki. Gambar 4 menampilkan grafik hasil penerapan metode *silhouette* yang digunakan untuk mengidentifikasi jumlah kluster yang paling tepat. Metode ini mengevaluasi kualitas pengelompokan berdasarkan tingkat kedekatan objek dalam satu kluster serta tingkat pemisahannya dengan kluster lain.



**Gambar 4.** Grafik Metode *Silhouette*

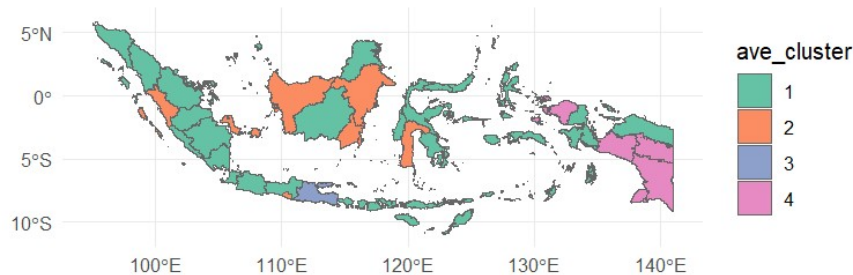
Berdasarkan grafik pada Gambar 4, nilai *silhouette* rata-rata tertinggi diperoleh pada pembentukan empat kluster, sehingga jumlah tersebut dianggap sebagai jumlah kluster yang paling optimal untuk digunakan pada tahap analisis selanjutnya. Adapun dendrogram metode *average linkage* dengan 4 kluster disajikan pada Gambar 5.



Gambar 5. Dendrogram Metode *Average Linkage* dengan 4 Kluster

F. Interpretasi Hasil

Gambar 6 merepresentasikan peta sebaran kluster provinsi-provinsi di Indonesia didasarkan Indeks Pembangunan Literasi Masyarakat.



Gambar 6. Peta Sebaran Wilayah Indonesia Berdasarkan Indeks Pembangunan Literasi Masyarakat

Tabel 5 merangkum provinsi-provinsi di Indonesia berdasarkan hasil analisis kluster *Average Linkage*, beserta karakteristik antar klasternya.

Tabel 5. Hasil Klasterisasi Provinsi-provinsi di Indonesia Berdasarkan Indeks Pembangunan Literasi Masyarakat

Kluster	X1	X2	X3	X4
Indonesia	68,2	0,427	0,387	0,323
1 Aceh, Bali, Banten, Bengkulu, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Selatan, Sumatera Utara	67,0	0,417	0,349	0,269
2 DI Yogyakarta, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Kepulauan Bangka Belitung, Sulawesi Selatan, Sumatera Barat	82,3	0,629	0,718	0,459
3 DKI Jakarta, Jawa Timur	75,8	0,493	0,282	1,000
4 Papua Barat Daya, Papua Pegunungan, Papua Selatan, Papua Tengah	47,0	0,097	0,105	0,086

Berdasarkan Tabel 5, hasil klasterisasi provinsi-provinsi di Indonesia berdasarkan Indeks Pembangunan Literasi Masyarakat menghasilkan empat kluster dengan karakteristik yang berbeda. Secara umum, rata-rata nasional menunjukkan nilai Indeks Pembangunan Literasi Masyarakat (X1) sebesar 68,2, pemerataan layanan perpustakaan (X2) sebesar 0,427, ketercukupan koleksi perpustakaan (X3) sebesar 0,387, serta tingkat kunjungan masyarakat per hari (X4) sebesar 0,323.

Klaster 1 yang terdiri dari Aceh, Bali, Banten, Bengkulu, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Selatan, dan Sumatera Utara memiliki nilai rata-rata  $X_1 = 67,0$ ,  $X_2 = 0,417$ ,  $X_3 = 0,349$ , dan  $X_4 = 0,269$ . Klaster ini mencerminkan provinsi dengan tingkat literasi masyarakat dan pemanfaatan perpustakaan yang relatif mendekati rata-rata nasional, namun masih menunjukkan keterbatasan dalam kecukupan koleksi dan tingkat kunjungan.

Klaster 2 yang meliputi DI Yogyakarta, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Kepulauan Bangka Belitung, Sulawesi Selatan, dan Sumatera Barat menunjukkan nilai rata-rata tertinggi pada hampir seluruh variabel, yaitu  $X_1 = 82,3$ ,  $X_2 = 0,629$ ,  $X_3 = 0,718$ , dan  $X_4 = 0,459$ . Hal ini mengindikasikan bahwa provinsi dalam klaster ini memiliki tingkat literasi masyarakat yang tinggi, layanan perpustakaan yang lebih merata, koleksi yang relatif memadai, serta intensitas kunjungan masyarakat yang lebih tinggi dibandingkan klaster lainnya.

Klaster 3, yang terdiri dari DKI Jakarta dan Jawa Timur, memiliki nilai rata-rata  $X_1 = 75,8$ ,  $X_2 = 0,493$ ,  $X_3 = 0,282$ , dan  $X_4 = 1,000$ . Klaster ini ditandai oleh tingkat kunjungan masyarakat ke perpustakaan yang sangat tinggi, meskipun kecukupan koleksi perpustakaan relatif lebih rendah dibandingkan klaster dengan tingkat literasi tinggi.

Sementara itu, klaster 4 mencakup Papua Barat Daya, Papua Pegunungan, Papua Selatan, dan Papua Tengah memiliki nilai rata-rata terendah pada seluruh variabel, yaitu  $X_1 = 41,0$ ,  $X_2 = 0,097$ ,  $X_3 = 0,089$ , dan  $X_4 = 0,086$ . Kondisi ini menunjukkan masih rendahnya tingkat literasi masyarakat, keterbatasan pemerataan layanan perpustakaan, kecukupan koleksi, serta rendahnya tingkat kunjungan masyarakat di wilayah-wilayah tersebut.

#### IV. KESIMPULAN

Berdasarkan hasil analisis klaster hierarki terhadap provinsi di Indonesia menggunakan Indeks Pembangunan Literasi Masyarakat beserta indikator-indikator pendukungnya, diperoleh bahwa metode *Average Linkage* merupakan metode paling optimal berdasarkan nilai koefisien korelasi *cophenetic* tertinggi. Berdasarkan metode *silhouette*, jumlah klaster yang paling tepat untuk dibentuk adalah empat klaster. Hasil klasterisasi mengungkapkan adanya perbedaan karakteristik yang cukup signifikan antar klaster, baik dari sisi tingkat literasi masyarakat, pemerataan layanan perpustakaan, kecukupan koleksi, maupun tingkat kunjungan masyarakat. Temuan ini mengindikasikan masih adanya ketimpangan pembangunan literasi antar wilayah di Indonesia, sehingga hasil pengelompokan yang diperoleh dapat menjadi dasar dalam perumusan kebijakan pengembangan literasi dan perpustakaan yang lebih terarah dan sesuai dengan karakteristik masing-masing klaster.

Berdasarkan hasil penelitian ini, disarankan kepada pemerintah dan pemangku kepentingan terkait untuk merumuskan kebijakan pengembangan literasi masyarakat yang lebih spesifik sesuai dengan kondisi masing-masing klaster, khususnya bagi wilayah yang memiliki tingkat pembangunan literasi relatif rendah. Selain itu, penelitian selanjutnya disarankan untuk mempertimbangkan penggunaan variabel tambahan yang dapat merepresentasikan literasi masyarakat secara lebih komprehensif, seperti indikator literasi digital atau akses terhadap sumber informasi lainnya, serta mengkaji penggunaan metode klasterisasi lainnya guna memperoleh hasil pengelompokan yang lebih optimal.

#### DAFTAR PUSTAKA

- Asiska, N., Satyahadewi, N., & Perdana, H. (2019). Pencarian Cluster Optimum Pada Single Linkage, Complete Linkage dan Average Linkage. *Buletin Ilmiah Math, Stat, Dan Terapannya (Bimaster)*, 08, 393–398. <https://doi.org/https://doi.org/10.26418/bimst.v8i3.33173>
- Badan Perpustakaan Nasional Republik Indonesia. (2024). *Laporan Akhir Kajian Indeks Pembangunan Literasi Masyarakat (IPLM) Tahun 2024*. Perpustakaan Nasional Republik Indonesia.
- Ghaisani, S. Y., Hikmah, N., Prasetyo, A. H., & Widodo, E. (2019). Analisis Cluster Hirarki untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016. *Konferensi Nasional Penelitian Matematika Dan Pembelajarannya*, 1–11.
- Hartono, B., Lusiana, V., & Al Amin, H. (2025). Perbandingan Proses Klasterisasi Data Menggunakan K-Means Clustering dan Agglomerative Hierarchical Clustering. *Jurnal Riset Komputer*, 12(4), 2407–389. <https://doi.org/10.30865/jurikom.v12i4.8766>
- Iis, Yahya, I., Wibawa, G. N. A., Baharuddin, Ruslan, & Laome, L. (2022). Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020. *Prosiding Seminar Nasional Sains Dan Terapan*, 1–16.

- Irwan, Sanusi, W., & Hasanah, A. (2024). Perbandingan Analisis Cluster Metode Complete Linkage dan Metode Ward dalam Pengelompokan Indeks Pembangunan Manusia di Sulawesi Selatan. *Journal of Mathematics, Computations, and Statistics*, 7(1), 75–86. <http://www.ojs.unm.ac.id/jmathcos>
- Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). Teknik Evaluasi Cluster. In *Teknik Evaluasi Cluster Solusi Menggunakan Phyton dan Rapidminer* (pp. 1–9). Deepublish.
- Mubarok, M. I., & Rusyiana, A. (2020). Pemetaan Zonal Risiko Penularan Covid-19 di Sulawesi Selatan Menggunakan Plot Dendogram Hierarchical Clustering. *Seminar Nasional Official Statistics*, (1), 55–63.
- Novaldi, J., & Wijayanto, A. W. (2023). Analisis Cluster Kualitas Pemuda di Indonesia pada Tahun 2022 dengan Agglomerative Hierarchical dan K-Means. *Komputika: Jurnal Sistem Komputer*, 12(2), 91–99. <https://doi.org/10.34010/komputika.v12i2.10348>
- Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, 19(2), 141–155. <https://doi.org/10.30595/sainteks.v19i2.14495>
- Prabowo, R. A., Nisa, K., Faisal, A., & Setiawan, E. (2020). Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage Dan Ward Pada Data Yang Mengandung Masalah Multikolinearitas. *Jurnal Siger Matematika*, 01(02).
- Pratiwi, S. I., Widiharih, T., & Hakim, A. R. (2019). Analisis Klaster Metode Ward dan Average Linkage dengan Validasi Dunn Index dan Koefisien Korelasi Cophenetic. *Jurnal Gaussian*, 8(4), 486–495.
- Rahmawati, F., & Fallo, S. I. (2025). Hierarchical Agglomerative Clustering Dengan Metode Ward untuk Pemetaan Pasar Tenaga Kerja Pascapandemi di Jawa Tengah: Pendekatan Machine Learning Berbasis Klasterisasi. *Leibniz: Jurnal Matematika*, 5, 65–77. <https://doi.org/https://doi.org/10.59632/leibniz.v5i01.573>
- Syaifudin, A., Auliyah Sofyanti, D., Irnadianis Ivada, F., Krisna Bagus Sajiwo, K., Fakhri Zamzami, M., Samiyah Iroyna, N., Hasanah M. Zach, N., & Aripriatiwi, R. A. (2023). Gerakan Literasi Masyarakat : Penguatan Literasi Untuk Membangun Masyarakat Literat. *Jurnal Pengabdian Masyarakat Indonesia*, 3(6), 773–778. <https://doi.org/10.52436/1.jpmi.1751>
- Whendasmoro, R. G., & Joseph. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872–876. <https://doi.org/10.30865/jurikom.v9i4.4526>