

PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS PEMBANGUNAN LITERASI MASYARAKAT 2023 MENGGUNAKAN METODE KLASTER HIERARKI DAN NON HIERARKI

Akhlaqul Karimah*, Ika Nur Laily Fitriana

Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka

Penulis korespondensi: akhlaqul.karimah18@gmail.com

ABSTRAK

Pembangunan literasi merupakan salah satu indikator penting dalam mengukur kemajuan suatu bangsa di Indonesia. Tujuan penelitian ini adalah untuk mengelompokkan provinsi di Indonesia berdasarkan indikator Indeks Pembangunan Literasi Masyarakat yang bersumber dari Satu Data Perpustakaan tahun 2023 dan Statistik Indonesia (Badan Pusat Statistik) tahun 2023. Pengelompokan dilakukan untuk mengetahui provinsi-provinsi mana yang memiliki kualitas literasi rendah sehingga diperlukan perbaikan akses dan fasilitas literasi yang lebih memadai. Pada penelitian ini dilakukan perbandingan metode kluster hierarki dan non hierarki. Metode kluster hierarki yang digunakan yaitu *single linkage* dan *average linkage*, sedangkan metode kluster non hierarki yang digunakan yaitu *k-means*. Metode yang terbaik antara kluster hierarki dan non hierarki digunakan dalam pengelompokan provinsi berdasarkan indikator Pembangunan Literasi Masyarakat. Pemilihan metode terbaik berdasarkan nilai *icdrate* dan *pseudo-f statistics*. Metode terbaik antara kluster hierarki dan non hierarki yaitu metode k-means dengan 5 kluster dihasilkan nilai *Pseudo-F statistics* sebesar 90,05 dan *icdrate* sebesar 0,075. Kebaikan hasil kluster (R^2) sebesar 92,5%. Setiap kluster memiliki karakteristik masing – masing. Kluster 1 yang beranggotakan Provinsi DI Yogyakarta, dan Sulawesi Selatan merupakan kluster yang memiliki Indeks Pembangunan Literasi Masyarakat sangat tinggi. Kluster 2 termasuk dalam kategori yang memiliki Indeks Pembangunan Literasi Masyarakat tinggi. Kluster 3 termasuk dalam kategori yang memiliki Indeks Pembangunan Literasi Masyarakat sedang/cukup. Kluster 4 termasuk dalam kategori rendah. Sedangkan kluster 5 merupakan kluster yang memiliki Indeks Pembangunan Literasi Masyarakat sangat rendah. Berdasarkan penelitian hingga 5 kluster, hal ini menunjukkan terjadinya kesenjangan atau perbedaan pembangunan literasi di wilayah Indonesia.

Kata kunci: Analisis Kluster; *Icdrate*; Indeks Pembangunan Literasi Masyarakat; *Pseudo-F Statistics*, R^2

1 PENDAHULUAN

Pengembangan literasi adalah elemen penting untuk mengukur kemajuan bangsa. Indonesia, dengan populasi besar dan keanekaragaman sosial budaya di 34 provinsi, menghadapi tantangan besar dalam meningkatkan budaya literasi. Data UNESCO tahun 2022 menunjukkan minat baca masyarakat hanya 0,001%, dengan peringkat literasi siswa Indonesia di posisi ke-60 dari 70 negara. Indeks Pembangunan Literasi Masyarakat (IPLM) digunakan sebagai alat ukur mencakup literasi digital, finansial, dan informasi, yang relevan di era *Society 5.0*. Namun, perbedaan atau ketidaksetaraan signifikan terlihat antara provinsi di Pulau Jawa serta kota besar dengan wilayah timur Indonesia, menunjukkan ketimpangan akses fasilitas pendidikan dan infrastruktur literasi.

Penelitian sebelumnya yang dilakukan oleh Sari et al., (2023) menunjukkan bahwa metode *K-Means* merupakan metode terbaik dalam mengelompokkan kota di Indonesia berdasarkan indikator inflasi tahun 2021 dengan nilai simpangan baku terkecil dan menghasilkan 3 kluster. Penelitian Az-Zahra & Wijayanto (2024) juga menunjukkan metode *K-Means* merupakan metode terbaik dalam mengelompokkan kesejahteraan di daerah perbatasan Indonesia dengan ukuran 2 kluster. Penelitian oleh Auliya (2021) menunjukkan bahwa metode *single linkage* merupakan metode terbaik dalam mengklasterkan curah hujan yang terjadi di Jawa Timur dari 2016 sampai 2020 dengan jumlah kluster adalah 3. Penelitian sebelumnya yang telah dilakukan oleh Widyadhana et al. (2021) menunjukkan bahwa metode *Average Linkage* adalah metode terbaik dalam menklasterkan data kemiskinan di Jawa tengah dengan nilai *Silhouette Coefficient* sebesar 0.35 dan jumlah kluster adalah 2. Penelitian oleh Butar (2023) menunjukkan bahwa metode *average linkage* merupakan metode terbaik yang mampu mengelompokkan provinsi berdasarkan potensi ekonomi kelautan dengan nilai *Internal Cluster Dispersion Rate (Icdrate)* paling kecil dibandingkan model kluster lainnya dan menghasilkan 2 kluster.

Penelitian ini bertujuan untuk menganalisis Indeks Pembangunan Literasi Masyarakat (IPLM) di provinsi Indonesia pada tahun 2023 menggunakan metode *kluster*, sehingga dapat diperoleh rumus rekomendasi kebijakan untuk pengembangan literasi yang lebih merata berdasarkan hasil kluster. Pada penelitian ini dilakukan perbandingan dengan metode *Single Linkage*, *Average Linkage*, dan *K-Means*. Penelitian ini diharapkan dapat menjadi acuan dalam pengambilan kebijakan dan perencanaan program pengembangan literasi yang lebih efektif dan tepat sasaran, serta memberikan kontribusi dalam upaya mengurangi kesenjangan literasi antar wilayah di Indonesia.

2 METODE PENELITIAN

2.1 Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2023, yaitu data Indeks Pembangunan Literasi Masyarakat (IPLM) di 34 provinsi Indonesia pada tahun 2023. Terdapat 5 variabel dengan tipe data numerik dan skala rasio yang dianalisis, yang termasuk dalam IPLM. Meskipun BPS tahun 2023 mencakup 8 indikator utama literasi, penelitian ini hanya menganalisis 5 variabel karena keterbatasan data sekunder. Kelima variabel ini dianggap paling representatif dalam menggambarkan IPLM dan mencakup indikator utama yang relevan. Variabel - variabel yang dianalisis dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
X1	Indeks Pembangunan Literasi Masyarakat	Skala 0 – 100
X2	Pemerataan Pelayanan Perpustakaan	Presentase
X3	Ketercukupan Koleksi Perpustakaan	Presentase
X4	Rasio Ketercukupan Tenaga Perpustakaan	Presentase
X5	Tingkat Kunjungan Masyarakat Per Hari	Presentase

2.2 Analisis Kluster Hierarki

Analisis kluster merupakan metode dalam analisis multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimilikinya (Putri, 2024). Metode ini mengelompokkan individu atau objek penelitian sehingga objek yang paling mirip satu sama lain

akan berada dalam kluster yang sama (Fitriana, 2021; Johnson & Wichern, 2014). Analisis kluster terbagi menjadi dua jenis, yaitu metode hierarki dan non-hierarki (Gorunescu, 2011; Hair et al., 2014; Han & Kamber, 2006). *Hierarchical clustering* adalah metode pengelompokan yang hasilnya disajikan secara berjenjang, mulai dari n objek hingga menjadi satu kluster. Metode ini terbagi menjadi dua jenis, yaitu *agglomerative* dan *divisive*. *Agglomeration Schedule* adalah daftar yang memberikan informasi tentang objek yang akan dikelompokkan pada setiap tahap proses analisis kluster menggunakan metode hierarki (Laraswati, 2014). Beberapa teknik dalam metode *agglomerative* termasuk *single linkage*, *complete linkage*, *average linkage*, dan *ward's* (Johnson & Wichern, 2014). Hasil klusterisasi hierarki biasanya direpresentasikan dalam bentuk pohon (dendrogram). Berbagai ukuran pendekatan untuk mengukur kesamaan jarak, seperti jarak Euclidean dan Manhattan, dapat digunakan (Ningsih & Wijayanto, 2023). Pada metode kluster hierarki, jumlah kluster tidak ditentukan di awal, melainkan berdasarkan dendrogram yang dihasilkan, yang kemudian digunakan untuk menentukan jumlah kluster yang akan dibuat (Butar, 2023).

2.2.1. Single Linkage

Single Linkage adalah pengelompokan yang dilakukan dengan menggabungkan pengamatan objek yang memiliki kesamaan terdekat. Berdasarkan prosedur pembentukan kluster di atas, untuk metode single linkage (*Nearest Neighbor*) menggunakan kriteria berikut (Johnson & Wichern, 2014).

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1)$$

Keterangan:

d_{UW} : jarak antara U dan W

d_{VW} : jarak antara V dan W

$d_{(UV)W}$: jarak minimum antara objek UV dan W.

2.2.2. Average Linkage

Metode *average linkage* merupakan metode berdasarkan pada jarak rata-rata antar seluruh objek dalam sebuah kluster dengan seluruh objek pada kluster lain (Widyadhana et al., 2021). Berdasarkan prosedur pembentukan kluster tersebut, untuk metode *average linkage (between method)* menggunakan kriteria berikut (Wahyuni & Wulandari, 2022).

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ij}}{n_{(UV)}n_W} \quad (2)$$

Keterangan:

d_{ij} : jarak antara objek i di dalam *klaster UV* dan objek k di dalam *klaster W*

$n_{(UV)}$ dan n_W : nomor item dari *klaster UV* dan *W*, berturut-turut

Semakin besar nilai ukuran ketakmiripan antara dua objek, maka semakin besar pula perbedaan antara kedua objek tersebut sehingga semakin cenderung untuk tidak menggabungkannya dalam satu kelompok yang sama (Rumiyati & Otok, 2023).

2.3 Metode Analisis Kluster Non Hierarki *K-Means*

Non-hierarchical clustering adalah metode pengelompokan di mana objek dikelompokkan ke dalam K kelompok yang telah ditentukan sebelumnya. *K-Means* adalah salah satu metode pengelompokan data non-hierarki yang membagi data ke dalam kelompok berdasarkan kesamaan

karakteristik, dengan data yang serupa dikelompokkan bersama dan yang berbeda ditempatkan dalam kelompok terpisah (Nafis et al., 2019). Dalam metode *K-Means*, pengelompokan bergantung pada nilai awal titik pusat (*centroid*), yang memengaruhi *centroid* berikutnya dan penentuan kluster selanjutnya. Proses dihentikan ketika pola kluster tidak berubah (Faujia et al., 2022).

2.4 Perbedaan Algoritma K-Means Klustering dan Hierarchical Klustering

Dalam perhitungan K-Means clustering, data perlu ditransformasikan terlebih dahulu agar dapat diproses dengan mudah. Proses ini melibatkan perhitungan jarak antara data dan *centroid*, kemudian memilih jarak terdekat antara data dan *centroid* tersebut. Setelah jarak dihitung, *centroid* baru ditentukan dan proses ini diulang hingga ditemukan kluster yang konsisten dengan *centroid* yang sama antara data terakhir dan sebelumnya. Sementara itu, pada hierarchical clustering, meskipun metode perhitungan jaraknya mirip dengan *K-Means*, jarak dihitung antara data satu dengan data lainnya, bukan dengan *centroid*. Data yang memiliki jarak terdekat digabung menjadi satu kluster, dan proses ini dilanjutkan hingga jumlah kluster yang diinginkan tercapai (Setiawan et al., 2022). Tabel kelebihan dan kekurangan metode *single linkage*, *average linkage*, dan *k-means* disajikan dalam Tabel 2 (Widyadhana et al., 2021).

Tabel 2. Kelebihan dan Kekurangan Setiap Metode

Metode Kluster	Kelebihan	Kekurangan
<i>K-Means</i>	<ul style="list-style-type: none"> • Kompleksitas rendah • Perhitungannya cepat • Dapat menangani data yang besar • Anggota dalam kluster dapat disesuaikan 	<ul style="list-style-type: none"> • Perlu menentukan jumlah kluster terlebih dahulu • Sensitif terhadap pencilan • Tidak mampu digunakan untuk kluster yang bervariasi • Sensitif terhadap skala data Sentroid awal yang berbeda menghasilkan hasil yang berbeda
<i>Average Linkage</i>	<ul style="list-style-type: none"> • Tidak mengharuskan penentuan berapa banyak kluster • Dendrogram memberikan gambaran grafis • Dapat mendeteksi macam-macam bentuk dan ukuran kluster 	<ul style="list-style-type: none"> • Kompleksitas tinggi • Proses perhitungannya lambat • Setelah kluster terbentuk tidak dapat dilakukan penyesuaian • Sulit dalam menentukan kluster yang dianggap tidak bermakna • Kluster bergantung pada jarak metrik yang digunakan
<i>Single Linkage</i>	<ul style="list-style-type: none"> • Sederhana dan mudah di implementasikan • Efisien untuk dataset kecil • Tidak memerlukan jumlah kluster awal • Dapat menangkap struktur data memanjang 	<ul style="list-style-type: none"> • <i>Chaining effect</i>: Menghasilkan kluster yang panjang dan tidak kompak • Sangat sensitif terhadap pencilan • Kurang efisien untuk dataset besar • <i>Kluster</i> hasil tidak selalu representatif terhadap keseluruhan struktur data

Selanjutnya, untuk jumlah kelompok optimum dapat diketahui dengan kriteria nilai *Pseudo F-statistics*. Berikut persamaan yang digunakan untuk menghitung nilai *Pseudo F-statistics* (Dewi, 2015).

$$Pseudo F - statistic = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (3)$$

dengan,

$$R^2 = \frac{(SST-SSW)}{SST} \quad (4)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2 \quad (5)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \quad (6)$$

Keterangan:

SST (*Sum Square Total*): Total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan

SSW (*Sum Square Within*): Total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya

n: banyaknya sampel

c: banyaknya variabel

nc: Banyaknya data pada kelompok nc

p: banyaknya kelompok

x_{ijk} : sampel ke-i pada variabel ke-j kelompok ke-k

\bar{x}_j : rata-rata seluruh sampel pada variabel ke-j

\bar{x}_{jk} : rata-rata sampel pada variabel ke-j dan kelompok ke-k

Nilai *Pseudo F-statistics* tertinggi menunjukkan jumlah kelompok yang optimal untuk mempartisi data, di mana keragaman dalam kelompok sangat homogen dan antar kelompok sangat heterogen. Salah satu metode untuk memilih klaster terbaik adalah dengan menghitung rata-rata persebaran *Internal Cluster Dispersion Rate (icdrate)* terhadap seluruh partisi. *Icdrate* menggambarkan tingkat perbedaan dalam setiap klaster. Semakin kecil nilai *icdrate*, semakin baik hasil pengelompokannya (Fitriana, 2021). Formula untuk menghitung *icdrate* terdapat di persamaan (7).

$$icdrate = 1 - \frac{SSB}{SST} = 1 - \frac{SST-SSW}{SST} = 1 - R^2 \quad (7)$$

Keterangan:

SSB: *Sum Square Between* (SST-SSW)

R^2 : *Recovery Rate* (SSB/SST)

2.5 Indeks Pembangunan Literasi Masyarakat

Indeks Pembangunan Literasi Masyarakat (IPLM) adalah ukuran yang menilai upaya pemerintah daerah (provinsi dan kabupaten/kota) dalam mengelola dan mengembangkan perpustakaan sebagai sarana untuk mendukung pembelajaran sepanjang hayat, guna mencapai budaya literasi di masyarakat (Perpustakaan, 2023).

2.6 Langkah Analisis

Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Mengumpulkan data sekunder yang diambil dari website bps.go.id.
2. Menganalisis karakteristik data dengan menggunakan statistika deskriptif

3. Melakukan pengelompokan provinsi di Indonesia dengan metode *Single Linkage*, *Average Linkage*, dan *K-Means*.
4. Menghitung dan membandingkan nilai *Pseudo F-statistics*, *icdrate*, dan R^2 pada masing – masing klaster *single linkage*, *average linkage*, dan *K-Means*.
5. Melakukan profiling terhadap hasil klaster terbaik.

3 HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Data Indeks Pembangunan Literasi Masyarakat

Karakteristik dari data Indeks Pembangunan Literasi Masyarakat perlu diketahui sebelum melakukan analisis, untuk mengetahui gambaran umum tentang data Indeks Pembangunan Literasi Masyarakat di Indonesia berdasarkan Provinsi, dilakukan analisis statistika deskriptif.

Tabel 3. Statistika Deskriptif

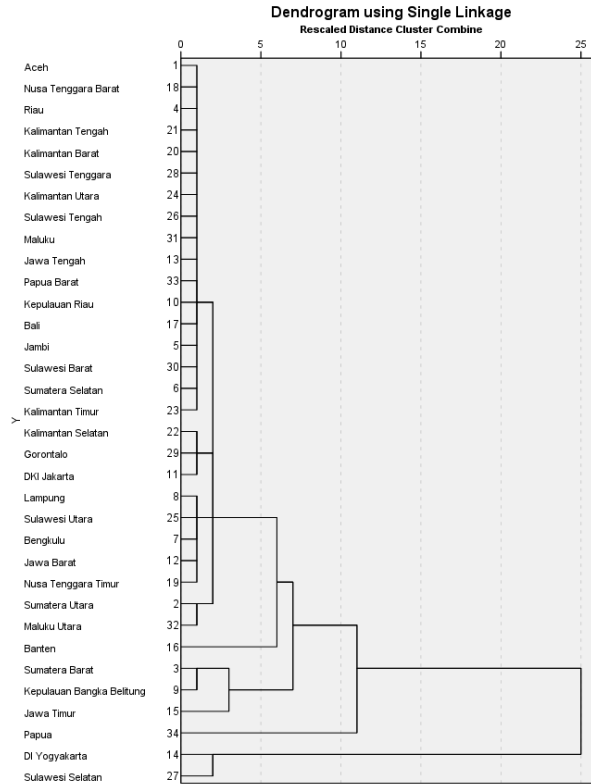
Variabel	N	Min	Maks	Mean	Varians
Indeks Pembangunan Literasi Masyarakat	34	47,57	86,74	65,80	66,20
Pemerataan Layanan Perpustakaan	34	0,19	0,78	0,45	0,02
Ketercukupan Koleksi Perpustakaan	34	0,09	0,84	0,35	0,03
Rasio Ketercukupan Tenaga Perpustakaan	34	0,20	1,00	0,64	0,09
Tingkat Kunjungan Masyarakat per hari	34	0,03	1,00	0,20	0,07

Nilai minimum dan maksimum menunjukkan nilai terkecil dan terbesar dari keseluruhan data di setiap provinsi untuk masing-masing variabel. Berdasarkan Tabel 3 dapat diketahui bahwa masing-masing Indeks Indeks Pembangunan Literasi Masyarakat memiliki rata-rata yang berbeda-beda. Rata-rata yang terendah yaitu Tingkat Kunjungan Masyarakat per hari (X5). Hal tersebut menunjukkan bahwa masih sedikit minat masyarakat untuk berkunjung ke Perpustakaan dalam setiap harinya. Sedangkan, rata-rata yang tertinggi yaitu Indeks Pembangunan Literasi Masyarakat (X1). Hal ini menunjukkan bahwa Pembangunan Literasi Masyarakat sudah baik.

Varians merupakan ukuran variabilitas data, yang berarti semakin besar nilai varians berarti semakin tinggi fluktuasi data antara satu data dengan data yang lain. Berdasarkan Tabel 3 menunjukkan bahwa Pemerataan Layanan Perpustakaan (X2) memiliki varians atau ragam yang paling kecil, sehingga dapat dikatakan bahwa masing-masing provinsi di Indonesia memiliki layanan Perpustakaan yang cukup atau tersebar secara merata. Namun untuk variabel Indeks Pembangunan Literasi Masyarakat (X1) memiliki varians yang tinggi sehingga persebarannya tidak merata.

3.2 Hasil Analisis Kluster Hierarki dan Non-Hierarki Indeks Pembangunan Literasi Masyarakat

Pada penelitian ini akan dilakukan analisis kluster menggunakan metode hierarki *single linkage* pada data Indikator Indeks Pembangunan Literasi Masyarakat di Indonesia tahun 2023 menjadi tiga, empat, dan lima kluster. Hasil pengelompokan kluster dapat dilihat pada Gambar 1 dendogram berikut ini.



Gambar 1. Dendogram *Single Linkage*

Dari Gambar 1. jika diambil batas klaster sebanyak 3, 4 dan 5 maka pengelompokan provinsi pada data indikator Indeks Pembangunan Literasi Masyarakat di Indonesia tahun 2023 adalah sebagai berikut pada Tabel 4.

Tabel 4. Pengelompokan Metode *Single Linkage*

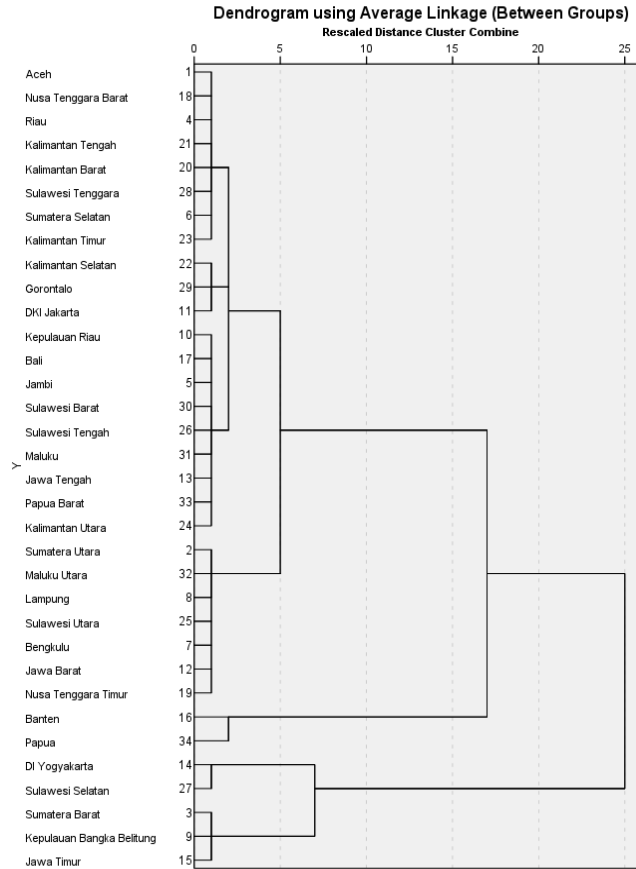
Jumlah Klaster	Pengelompokan Klaster	Jumlah Anggota	Anggota
3	Klaster 1	31	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.

Tabel 4. Pengelompokan Metode *Single Linkage* (Lanjutan)

Jumlah Klaster	Pengelompokan Klaster	Jumlah Anggota	Anggota
4	Klaster 2	1	Papua
	Klaster 3	2	DI Yogyakarta, Sulawesi Selatan
	Klaster 1	28	Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.
	Klaster 2	3	Jawa Timur, Sumatera Barat, Kepulauan Bangka Belitung
5	Klaster 3	1	Papua
	klaster 4	2	DI Yogyakarta, Sulawesi Selatan
	Klaster 1	27	Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.
	Klaster 2	2	Banten, Sumatera Barat
	Klaster 3	2	Jawa Timur, Kepulauan Bangka Belitung
	Klaster 4	1	Papua
	Klaster 5	2	DI Yogyakarta, Sulawesi Selatan

Berdasarkan Tabel 4 menyajikan pembagian anggota untuk 3 klaster adalah kelompok 1 terdapat 31 provinsi, kelompok 2 terdapat 1 provinsi, kelompok 3 terdapat 2 provinsi. Kemudian untuk 4 klaster adalah kelompok 1 terdapat 28 provinsi, kelompok 2 terdapat 3 provinsi, kelompok 3 terdapat 1 provinsi, kelompok 4 terdapat 2 provinsi. Sedangkan untuk 5 klaster adalah kelompok 1 terdapat 27 provinsi, kelompok 2 terdapat 2 provinsi, kelompok 3 terdapat 2 provinsi, kelompok 4 terdapat 1 provinsi, kelompok 5 terdapat 2 provinsi.

Selanjutnya dilakukan analisis klaster menggunakan metode hierarki *average linkage* pada data Indikator Indeks Pembangunan Literasi Masyarakat di Indonesia tahun 2023 sehingga diperoleh pengelompokan sebanyak 3 pengelompokan yaitu pengelompokan menjadi tiga, empat, dan lima klaster. Hasil pengelompokan klaster dapat dilihat pada Gambar 2 dendogram berikut ini.



Gambar 2. Dendrogram *Average Linkage*

Dari Gambar 2. jika diambil batas klaster sebanyak 3, 4 dan 5 maka pengelompokkan provinsi pada data indikator Indeks Pembangunan Literasi Masyarakat di Indonesia tahun 2023 adalah sebagai berikut pada Tabel 5.

Tabel 5. Pengelompokan Metode *Average Linkage*

Jumlah Klaster	Pengelompokan Klaster	Jumlah Anggota	Anggota
3	Klaster 1	27	Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.
	Klaster 2	2	Banten, Papua
	Klaster 3	5	DI Yogyakarta, Sulawesi Selatan, Sumatera Barat, Kepulauan Bangka Belitung, Jawa Timur

Tabel 5. Pengelompokan Metode *Average Linkage* (Lanjutan)

Jumlah Klaster	Pengelompokan Klaster	Jumlah Anggota	Anggota
4	Klaster 1	27	Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.
	Klaster 2	2	Banten, Papua
	Klaster 3	2	DI Yogyakarta, Sulawesi Selatan
	klaster 4	3	Sumatera Barat, Kepulauan Bangka Belitung, Jawa Timur
5	Klaster 1	20	Aceh, Riau, Jambi, Sumatera Selatan, Kepulauan Riau, DKI Jakarta, Jawa Tengah, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat.
	Klaster 2	7	Sumatera Utara, Maluku Utara, Lampung, Sulawesi Utara, Bengkulu, Jawa Barat, Nusa Tenggara Timur
	Klaster 3	2	Banten, Papua
	Klaster 4	2	DI Yogyakarta, Sulawesi Selatan
	Klaster 5	3	Sumatera Barat, Kepulauan Bangka Belitung, Jawa Timur

Berdasarkan Tabel 5 menyajikan pembagian anggota untuk 3 klaster adalah kelompok 1 terdapat 27 provinsi, kelompok 2 terdapat 2 provinsi, kelompok 3 terdapat 5 provinsi. Kemudian untuk 4 klaster adalah kelompok 1 terdapat 27 provinsi, kelompok 2 terdapat 2 provinsi, kelompok 3 terdapat 2 provinsi, kelompok 4 terdapat 4 provinsi. Sedangkan untuk 5 klaster adalah kelompok 1 terdapat 20 provinsi, kelompok 2 terdapat 7 provinsi, kelompok 3 terdapat 2 provinsi, kelompok 4 terdapat 2 provinsi, kelompok 5 terdapat 3 provinsi.

Pada penelitian pengelompokan menggunakan metode klaster hierarki dan non-hierarki masing-masing akan dibentuk menjadi 3, 4 dan 5 klaster. Pembentukan menjadi 3, 4 dan 5 klaster ini berdasarkan kelompok indikator yang akan dibentuk. Jika klaster yang optimum adalah 3, maka pengelompokan Indeks Pembangunan Literasi Masyarakat menjadi kategori rendah, sedang, dan tinggi. Jika klaster optimum yang terbentuk adalah 4, maka kategori Indeks Pembangunan Literasi Masyarakat menjadi sangat rendah, rendah, sedang, dan tinggi. Jika klaster optimum yang terbentuk sebanyak 5 klaster, maka kategori Indeks Pembangunan Literasi Masyarakat yang terbentuk yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi.

Pemilihan banyak kluster yang optimum menggunakan *Pseudo-F Statistics*. Perbandingan nilai *Pseudo-F Statistics* metode kluster hierarki dan non-hierarki dengan banyak kelompok 3,4 dan 5 ditampilkan pada Tabel 6.

Tabel 6. Perbandingan Nilai *Pseudo-F Statistics*

Jumlah Kluster	<i>Single Linkage</i>	<i>Average Linkage</i>	<i>K-Means</i>
3	17,37	36,70	67,30
4	27,25	36,96	63,36
5	13,60	77,86	90,05

Berdasarkan Tabel 6, Nilai *Pseudo-F Statistics* terbesar dari semua kombinasi metode dan jumlah kluster adalah sebesar 90,05 yang terdapat pada kelompok dengan 5 kluster dengan menggunakan metode *K-means*. Nilai *Pseudo F-statistics* yang tinggi menunjukkan bahwa dengan 5 kluster dapat menghasilkan keragaman dalam kelompok sangat homogen sedangkan antar kelompok sangat heterogen. Hal ini menunjukkan bahwa kluster optimum yang terbentuk adalah 5 kluster. Kategori 5 kluster yang terbentuk yaitu Indeks Pembangunan Literasi Masyarakat sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Selanjutnya, dengan jumlah 5 kluster dilakukan pengelompokan menggunakan metode *Single Linkage*, *Average Linkage*, dan *K-Means*. Lalu dihitung nilai *icdrate* untuk menentukan metode yang terbaik. Metode yang terbaik memiliki nilai *icdrate* terkecil dan R^2 terbesar. Hasil perbandingan *icdrate* dan R^2 ditampilkan pada Tabel 7.

Tabel 7. *Icdrate* dan R^2

	<i>Single Linkage</i>	<i>Average Linkage</i>	<i>K-Means</i>
<i>Icdrate</i>	0,268	0,085	0,075
R^2	0,731	0,914	0,925

Berdasarkan Tabel 7, metode *k-means* memiliki *icdrate* terkecil yaitu sebesar 0,075 dan R^2 terbesar yaitu sebesar 0,925. Dispersi atau perbedaan dalam kluster sebesar 7,5% dan kebaikan hasil kluster sebesar 92,5%. Hasil kebaikan kluster dikatakan sangat baik karena mendekati 100%, sehingga reliabel untuk diinterpretasikan lebih lanjut. Semakin kecil nilai *icdrate* maka semakin baik hasil pengelompokannya. Metode terbaik dalam mengelompokkan provinsi di Indonesia berdasarkan Indeks Pembangunan Literasi Masyarakat dengan jumlah pengelompokan optimal adalah 5 kluster dengan metode non-hierarki (*k-means*). Penggunaan metode *k-means* didapatkan anggota masing-masing kluster berdasarkan kedekatan (*similarity*) disajikan pada Tabel 8.

Tabel 8. Rincian Anggota Klaster

Nomor Klaster	Klaster	Anggota Banyak	Anggota dalam Klaster
1	Sangat Tinggi	2	DI Yogyakarta, Sulawesi Selatan
2	Tinggi	2	Banten, Papua
3	Sedang/Cukup	3	Sumatera Barat, Kepulauan Bangka Belitung, Jawa Timur
4	Rendah	12	Aceh, Riau, Sumatera Selatan, DKI Jakarta, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tenggara, Gorontalo
5	Sangat Rendah	15	Sumatera Utara, Jambi, Bengkulu, Lampung, Kepulauan Riau, Jawa Barat, Jawa Tengah, Bali, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat

Berdasarkan Tabel 8 menunjukkan bahwa klaster dengan kategori sangat tinggi terdapat 2 anggota. Klaster dengan kategori tinggi terdapat 2 anggota. Klaster dengan kategori sedang/cukup terdapat 3 anggota. Sedangkan klaster dengan kategori rendah dan sangat rendah masing - masing terdapat 12 dan 15 anggota. Berikut ini terdapat rata-rata Indeks Pembangunan Literasi Masyarakat setiap klaster pada Tabel 9.

Tabel 9. Rata – Rata Indeks Pembangunan Literasi Masyarakat Setiap Klaster

Variabel	Rata – Rata				
	Sangat Tinggi (Klaster 1)	Tinggi (Klaster 2)	Sedang/Cukup (Klaster 3)	Rendah (Klaster 4)	Sangat Rendah (Klaster 5)
X1	85,92	76,66	68,02	61,28	50,04
X2	0,66	0,57	0,43	0,43	0,25
X3	0,70	0,49	0,31	0,33	0,13
X4	1,00	0,87	0,84	0,40	0,47
X5	0,66	0,44	0,18	0,14	0,04
Rata - Rata Total	17,79	15,81	13,96	12,52	10,19

Karakteristik setiap klaster yang terbentuk dengan metode non-Hierarki (*k-means*) terdapat pada Tabel 9. Klaster 1 memiliki nilai paling tinggi di antara klaster yang lain. Hasil rata-rata total menunjukkan bahwa klaster 1 mempunyai rata-rata Indeks Pembangunan Literasi Masyarakat (IPLM) paling tinggi.

Klaster 2 memiliki karakteristik variabel yang nilainya bervariasi, dari yang rendah sampai tertinggi. Klaster 2 memiliki 1 indikator IPLM yang paling tinggi dari pada lainnya yaitu Indeks Pembangunan Literasi Masyarakat (X1). Klaster ini juga memiliki 1 Indikator IPLM yang paling

rendah yaitu Tingkat Kunjungan Masyarakat per Hari (X5). Secara keseluruhan klaster 2 merupakan klaster dengan Indikator IPLM tinggi.

Secara keseluruhan klaster 3 menjadi klaster dengan IPLM yang sedang/cukup dengan rata-rata total sebesar 13,96. Pada klaster 4 dengan Indikator IPLM yang rendah dengan rata-rata total sebesar 12,52. Sedangkan klaster 5 memiliki 1 Indikator IPLM yang paling rendah yaitu Tingkat Kunjungan Masyarakat per Hari (X5) diantara klaster lain. Hasil rata-rata total menunjukkan bahwa klaster 5 mempunyai rata-rata IPLM sangat rendah.

4 KESIMPULAN

Pengelompokan provinsi di Indonesia berdasarkan Indikator Pembangunan Literasi Masyarakat (IPLM) dengan analisis klaster hierarki dan non-hierarki menghasilkan 5 klaster. Metode terbaik antara klaster hierarki dan non hierarki yaitu metode *k-means* dengan 5 klaster dihasilkan nilai *Pseudo-F statistics* sebesar 90,05 dan *icdrate* sebesar 0,075. Kebaikan hasil klaster (R^2) sebesar 0,925 atau 92,5%. Setiap klaster memiliki karakteristik masing – masing. Klaster 1 yang beranggotakan Provinsi DI Yogyakarta, dan Sulawesi Selatan merupakan klaster yang memiliki IPLM paling tinggi. Klaster 2 termasuk dalam kategori yang memiliki IPLM tinggi. Klaster 3 termasuk dalam kategori yang memiliki IPLM sedang atau cukup. Klaster 4 termasuk dalam kategori IPLM rendah. Sedangkan klaster 5 merupakan klaster yang memiliki IPLM sangat rendah dengan anggota klasternya yaitu provinsi Sumatera Utara, Jambi, Bengkulu, Lampung, Kepulauan Riau, Jawa Barat, Jawa Tengah, Bali, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat. Berdasarkan penelitian hingga 5 klaster, hal ini menunjukkan bahwa ada terjadinya kesenjangan atau perbedaan pembangunan literasi di wilayah Indonesia.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Ibu Ika Nur Laily Fitriana, M.Stat., dosen pembimbing mata kuliah Karya Ilmiah di Universitas Terbuka, atas bimbingan dan dukungan selama penyusunan karya ilmiah ini. Terima kasih juga kepada keluarga dan teman-teman yang memberikan motivasi. Karya ilmiah ini, yang berjudul "Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Literasi Masyarakat 2023 Menggunakan Metode Klaster Hierarki dan Non Hierarki," tidak akan terwujud tanpa bantuan banyak pihak. Kekurangan dalam karya ilmiah ini sepenuhnya menjadi tanggung jawab penulis. Penulis berharap karya ini dapat memberikan kontribusi positif bagi ilmu pengetahuan dan kebijakan literasi di Indonesia.

DAFTAR PUSTAKA

- Auliya, E. R. (2021). PENGKLASTERAN DAERAH DI JAWA TIMUR BERDASARKAN CURAH HUJAN. *MATH UNESA (Jurnal Ilmiah Matematika)*, 09(02), 431 - 436.
- Az-Zahra, A., & Wijayanto, A. W. (2024). Tinjauan Kesejahteraan di Daerah Perbatasan Republik Indonesia Tahun 2021: Penerapan Analisis Klaster K-Means dan Hierarki. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 12(1), 55-64. <https://doi.org/10.26418/justin.v12i1.69040>
- Butar, R. P. B. (2023). Analisis Hierarchical Dan Non Hierarchical Clustering Untuk Pengelompokan Potensi Ekonomi Kelautan Indonesia 2021. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 11(3), 543-553. <https://doi.org/10.26418/justin.v11i3.67283>

- Dewi, A. R. (2015). *Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan Menggunakan Metode C-Means Dan Fuzzy C-Means Clustering*. Institut Teknologi Sepuluh Nopember]. Surabaya.
- Faujia, R. A., Setianingsih, E. S., & Pratiwi, H. (2022). *Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia* Seminar Nasional Official Statistics Jakarta.
- Fitriana, I. N. L. (2021). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Keluarga Sehat Menggunakan Metode Klaster Hirarki dan Non Hirarki. *Jurnal Paradigma: Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, 2(1), 27-36.
- Gorunescu, F. (2011). *Data Mining Concepts and Techniques, Second Edition*. (Vol. 12) [Intelligent Systems Reference Library]. Springer.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate Data Analysis. Ed ke-12*. Pearson.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques, Second Edition*. Morgan Kauffman Publishers.
- Johnson, R., & Wichern, D. (2014). *Applied Multivariate Statistical Analysis (Sixth Edition)*. Pearson.
- Laraswati, T. F. (2014). *Perbandingan Kinerja Metode Complete Linkage, Metode Average Linkage, Dan Metode K-Means Dalam Menentukan Hasil Analisis Cluster*. Universitas Negeri Yogyakarta]. Yogyakarta.
- Nafis, C., Prima, R., & Ridho, A. (2019). *Analisis Hierarchical Dan Non-Hierarchical Clustering Pada Data Indikator Kesejahteraan Rakyat Di Indonesia Tahun 2017*. [Laporan Praktikum].
- Ningsih, I. K. M. M., & Wijayanto, A. W. (2023). Komparasi Metode Clustering Pada Provinsi di Indonesia Berdasarkan Pertumbuhan Ekonomi Tahun 2022. *Komputika: Jurnal Sistem Komputer*, 12(2), 103 - 112. <https://doi.org/10.34010/komputika.v13i1.10520>
- Perpustakaan, K. P. (2023). *Indeks Pembangunan Literasi Masyarakat (IPLM) 2023*. Satu Data Perpusnas. Retrieved October 22, 2024, from <https://satudata.perpusnas.go.id/dataset/indeks-pembangunan-literasi-masyarakat-iplm-2023>.
- Putri, D. A. V. (2024). *Pengelompokan Provinsi Di Indonesia Berdasarkan Angka Partisipasi Kasar Perguruan Tinggi Tahun 2022*. [Karya Ilmiah]. Universitas Terbuka.
- Rumiyati, A. T., & Otok, B. W. (2023). *Pengantar Sosiometri Edisi 2*. Universitas Terbuka.
- Sari, L. P., Fanani, A., & Asyhar, A. H. (2023). Analisis Perbandingan Pengelompokan Kota di Indonesia Berdasarkan Indikator Inflasi Tahun 2021 dengan Metode Ward dan K-Means. *Jurnal Sains Matematika dan Statistika*, 9(2), 108-118. <https://doi.org/https://dx.doi.org/10.24014/jsms.v9i2.21100>
- Setiawan, H., Karamitta, K., Hosea, K., & Hakim, L. (2022). *Analisa Perbandingan Metode Hierarchical dan K-means dalam Clustering Data Terhadap Penjualan Jajansamavivi*. Seminar Nasional Multimedia & Artificial Intelligence, Jakarta.
- Wahyuni, I., & Wulandari, S. P. (2022). Pemetaan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Analisis Cluster Hierarki. *JURNAL SAINS DAN SENI ITS*, 11(1), 70-75.
- Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). *Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah*. PRISMA, Prosiding Seminar Nasional Matematika, Semarang.