



Pengelompokan Provinsi berdasarkan Aspek Pembangunan Pendidikan di Indonesia Tahun 2023 menggunakan Analisis *Cluster*

Nurfajriyani^{1*}, Dentina Dewi Amaliana², Sri Pingit Wulandari³

^{1,2,3}Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember (ITS), Indonesia

Korespondensi penulis: nurfajriyani.qrn@gmail.com*

Abstract : *Improving the quality of Human Resources (HR) is a major challenge in facing global competition. Education as the main means of improving the quality of HR in Indonesia is still faced with the problem of inequality of access and quality between regions. This inequality causes disparities in educational development between urban and remote areas. This study focuses on grouping provinces in Indonesia based on aspects of educational development in 2023, using cluster analysis. Secondary data from the Central Statistics Agency (BPS) is used as the basis for analysis, including variables of average length of schooling, Gross Participation Rate (APK), Pure Participation Rate (APM), number of senior high schools, and community literacy development index. This study uses hierarchical and non-hierarchical cluster analysis methods to group provinces in Indonesia. The results of the hierarchical cluster analysis using the average linkage method show the most optimal cluster with the formation of three clusters. The first cluster consists of 31 provinces, the second cluster consists of 2 provinces, and the third cluster consists of 1 province. Data characteristics show large variations in the number of senior high schools and relative homogeneity in the average length of schooling between provinces.*

Keywords : *Cluster Analysis, Average Linkage, Educational Development*

Abstrak : Peningkatan kualitas Sumber Daya Manusia (SDM) merupakan tantangan utama dalam menghadapi persaingan global. Pendidikan sebagai sarana utama dalam peningkatan kualitas SDM di Indonesia masih dihadapkan pada permasalahan ketimpangan akses dan mutu antar wilayah. Ketimpangan ini menyebabkan disparitas pembangunan pendidikan antara daerah perkotaan dan daerah terpencil. Penelitian ini berfokus pada pengelompokan provinsi di Indonesia berdasarkan aspek pembangunan pendidikan pada tahun 2023, dengan menggunakan analisis *cluster*. Data sekunder dari Badan Pusat Statistik (BPS) digunakan sebagai basis analisis, mencakup variabel rata-rata lama sekolah, Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), jumlah sekolah menengah atas, dan indeks pengembangan literasi masyarakat. Penelitian ini menggunakan metode analisis *cluster* hierarki dan non-hierarki untuk mengelompokkan provinsi di Indonesia. Hasil analisis *cluster* hierarki dengan metode *average linkage* menunjukkan *cluster* paling optimum dengan pembentukan tiga *cluster*. *Cluster* pertama terdiri dari 31 provinsi, *cluster* kedua terdiri dari 2 provinsi, dan *cluster* ketiga terdiri dari 1 provinsi. Karakteristik data menunjukkan adanya variasi besar dalam jumlah sekolah SMA dan homogenitas relatif pada rata-rata lama sekolah antar provinsi.

Kata Kunci : Analisis Cluster, Keterkaitan rata-rata, Pengembangan Pendidikan

1. PENDAHULUAN

Pendidikan merupakan salah satu sarana untuk meningkatkan kualitas Sumber Daya Manusia (SDM). Untuk meningkatkan efektivitas dan efisiensi penyelenggaraan pendidikan, perlu ditingkatkan kualitas manajemen pendidikan. Peningkatan kualitas pendidikan ini menjadi landasan utama dalam mempersiapkan Indonesia menghadapi masa depan yang kompetitif dan dinamis [1]. Di era globalisasi dan perkembangan informasi teknologi yang pesat, pendidikan harus mampu memberikan keterampilan dan pengetahuan yang relevan serta adaptif terhadap perubahan. Perlunya manajemen pendidikan yang baik tidak hanya mencakup peningkatan sarana-prasarana yang baik,

tetapi juga optimalisasi sistem pembelajaran, pengembangan tenaga pendidik, serta peningkatan partisipasi masyarakat dalam mendukung proses pendidikan. Manajemen pendidikan yang efektif akan menciptakan lingkungan belajar yang kondusif, di mana setiap siswa akan mendapatkan akses setara terhadap pendidikan berkualitas. Sayangnya, distribusi sumber daya pendidikan yang belum merata mengakibatkan terjadinya disparitas pendidikan antar wilayah di Indonesia. Sesuai dengan undang-undang dasar negara Republik Indonesia tahun 1945, pasal 31 menyatakan bahwa setiap orang berhak mendapatkan pendidikan yang layak dan setara. Namun, pada kenyataannya di Indonesia yang wilayahnya sangat luas di merupakan negara kepulauan dan angka kelahirannya tinggi tidak sesuai dengan adanya pasal tersebut. Di Indonesia, masalah pemerataan pendidikan muncul ketika masih banyaknya warga negara khususnya usia anak sekolah yang tidak masuk ke dalam lembaga pendidikan karena faktor ekonomi, kekurangan guru, dan anak-anak putus sekolah atau bahkan tidak pernah merasakan bangku sekolah. Disparitas ini mengakibatkan siswa di daerah terpencil mengalami kesulitan dalam memahami materi pembelajaran, yang berbeda jauh dari siswa di daerah perkotaan yang lebih maju [2]. Mempertimbangkan faktor perbedaan kualitas pendidikan di Indonesia, maka dalam penelitian ini dilakukan analisis *cluster* sebagai jawaban permasalahan tersebut.

Analisis *cluster* adalah pengelompokan individu yang memiliki karakteristik serupa di dalam suatu kelompok, dan berbeda karakteristik dengan kelompok lain. Dengan terbentuknya kelompok-kelompok demikian, identifikasi karakteristik kelompok mudah dilakukan sehingga diperoleh gambaran karakteristik setiap kelompok [3]. Tujuan utama analisis *cluster* adalah mengklasifikasi objek (kasus/elemen) seperti manusia, produk (barang), toko, perusahaan, wilayah ke dalam kelompok-kelompok yang relatif homogen berdasarkan pada suatu set variabel yang dipertimbangkan untuk diteliti. Objek di dalam kelompok harus relatif mirip/sama. Metode pembentukan *cluster* terdiri dari metode hierarki dan non hierarki. Metode hierarki mengelompokkan berdasarkan kemiripan sifatnya yang belum diketahui jumlah *cluster* yang terbentuk dan digunakan pada data yang memiliki struktur hierarki atau terdiri dari beberapa sub-kelompok yang saling berkaitan dan umumnya jumlah sampel yang digunakan relatif kecil [4]. Sedangkan, metode non hierarki dimulai dengan menentukan jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, kemudian proses *cluster* dilakukan tanpa mengikuti proses hierarki [5]. Penelitian ini menggunakan analisis *cluster* untuk mengelompokkan provinsi di Indonesia untuk memberi fokus kepada kelompok provinsi yang harus diprioritaskan

upaya pemerataan faktor pendidikannya. Dengan metode tersebut, diharapkan dapat memudahkan pemerintah untuk merumuskan kebijakan terbaru yang telah disesuaikan.

Penelitian terdahulu yang menggunakan metode serupa oleh Hikmah, dkk., (2022) diperoleh hasil 3 *cluster* kecamatan berdasarkan indikator pendidikan di Sulawesi Barat. Dari 3 *cluster* tersebut tersusun atas *cluster* 1 yang terdiri dari 3 kecamatan, *cluster* 2 yang terdiri dari 19 kecamatan, dan *cluster* 3 yang terdiri dari 30 kecamatan. Selain itu, dalam penelitian ini juga menggunakan penelitian yang digunakan oleh Kurniawan, dkk., (2021) sebagai referensi penelitian terdahulu. Berdasarkan penelitian tersebut diperoleh hasil 3 *cluster* kecamatan yang bergabung berdasarkan tingkat pendidikan. Dari 3 *cluster* tersebut tersusun atas *cluster* 0 berjumlah 7 kecamatan, *cluster* 1 berjumlah 17 kecamatan, dan *cluster* 2 berjumlah 20 kecamatan. Berdasarkan hasil tersebut, Kurniawan, dkk., menyarankan pemerintah untuk memberikan perhatian khusus berupa bantuan agar tidak tertinggal oleh wilayah lainnya. Berdasarkan dua penelitian tersebut, dapat dilihat bahwa penggunaan analisis *cluster* dalam mengelompokkan daerah berdasarkan indikator pendidikan dan tingkat pendidikan. Penelitian tersebut dapat digunakan sebagai acuan untuk menjadi dasar merumuskan kebijakan yang lebih terarah dalam mengatasi disparitas pendidikan.

Tujuan penelitian ini adalah untuk mengetahui hasil pengelompokkan pembangunan pendidikan di Indonesia tahun 2023 menggunakan analisis *cluster* hierarki dan non hierarki. Manfaat bagi penulis adalah dapat menerapkan dan memahami konsep analisis *cluster*. Manfaat bagi pembaca adalah dapat mengetahui penerapan analisis *cluster* untuk pemerataan pembangunan pendidikan di Indonesia tahun 2023 berdasarkan indikator kualitas pendidikan serta dapat menerapkan saran dari hasil analisis di kehidupan sehari-hari. Manfaat bagi pemerintah atau institusi terkait adalah agar dapat mengeluarkan kebijakan terkait upaya mengoptimalkan pemerataan pembangunan pendidikan di Indonesia berdasarkan karakteristik tiap *cluster* yang berbeda.

2. TINJAUAN PUSTAKA

Uji Asumsi Normal Multivariat

Uji normalitas dimaksudkan untuk mendeteksi data memenuhi asumsi normalitas atau tidak. Uji normalitas pada dasarnya melakukan perbandingan antara data yang dimiliki dengan data yang berdistribusi normal memiliki nilai rata-rata dan standar deviasi yang sama dengan data yang diujikan [8]. Untuk memeriksa data berdistribusi normal multivariat, dapat dilihat dari *Q-Q plot* antara *square distance* (d_j^2) dengan nilai *quartile*

dari distribusi $\chi^2 \left(\frac{r-0,5}{n} \right)$. Jika hasil plot menggambarkan garis lurus maka data tersebut dapat dinyatakan sebagai normal multivariat [9]. Adapun hipotesis yang digunakan untuk uji asumsi distribusi normal multivariat adalah sebagai berikut [10].

Hipotesis :

H_0 : Data berdistribusi normal multivariat

H_1 : Data tidak berdistribusi normal multivariat

Taraf Signifikan : α

Daerah kritis : Tolak H_0 jika $r_q < r_{(n;\alpha)}$ atau $P\text{-value} > \alpha$

Statistik uji :

$$r_q = \frac{\sum_{j=1}^n (x_j - \bar{x})(q_j - \bar{q})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_j - \bar{q})^2}} \quad (1)$$

Keterangan :

r_q = koefisien korelasi pearson antara mahalanobis dan
quartile

x_j = nilai mahalanobis data ke-j

$\bar{x}$ = rata-rata nilai mahalanobis

q_j = nilai *quartile* dari data ke-j

$\bar{q}$ = rata-rata nilai *quartile*

Uji Independensi

Uji independensi digunakan untuk mengetahui ada tidaknya hubungan antar dua variabel yang telah ditetapkan [11]. Uji *Bartlett* adalah prosedur inferensial yang digunakan untuk menilai kesetaraan varians dalam populasi yang berbeda. Dalam beberapa kasus, pelanggaran asumsi ini mungkin tidak memengaruhi kesimpulan peneliti secara substantif, namun dalam kasus lain, pelanggaran asumsi ini dapat berdampak signifikan terhadap hasil penelitian [12]. Pengujian asumsi independensi menggunakan uji *Bartlett* dijelaskan sebagai berikut.

Hipotesis :

H_0 : $\rho = I$ (Matriks korelasi identik dengan matriks identitas atau antar variabel independen)

H_1 : $\rho \neq I$ (Matriks korelasi tidak identik dengan matriks identitas atau antar variabel dependen)

Taraf Signifikan : α

Daerah kritis : Tolak H_0 jika $\chi^2 > \chi^2_{(\alpha; df)}$ atau $P\text{-value} < \alpha$

Statistik uji :

$$\chi^2 = (\ln n) \left[B - \sum_{k=1}^p d_k \ln s_k^2 \right] \quad (2)$$

$$B = \sum_{k=1}^p d_k \ln s_k^2 \quad (3)$$

$$s^2 = \frac{\sum_{k=1}^p d_k s_k^2}{\sum_{k=1}^p d_k} \quad (4)$$

Keterangan :

n = jumlah data

s_k^2 = varians data variabel ke- k

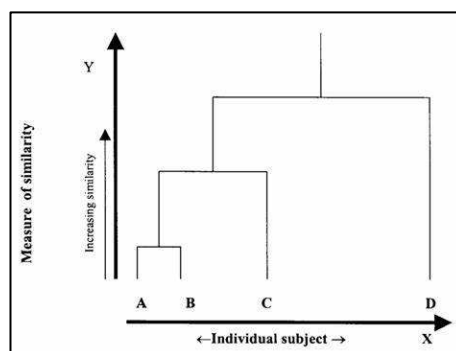
d_k = derajat kebebasan untuk variabel ke- k

Analisis Cluster

Clustering adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau *cluster*. *Clustering* membagi data ke dalam grup-grup yang mempunyai objek yang karakteristiknya sama [13].

Metode Hierarki

Analisis *cluster* dengan metode hierarki digunakan sebagai pengelompokan objek secara hierarki berdasarkan kemiripan sifatnya yang belum diketahui jumlah *cluster* yang terbentuk. Metode hierarki digunakan pada data yang memiliki struktur hierarkis atau terdiri dari beberapa sub-kelompok yang saling berkaitan dan umumnya jumlah sampel yang digunakan relatif kecil. Hasil pengelompokan dengan metode hierarki ditampilkan dalam sebuah diagram pohon yang disebut dendrogram [4].



Gambar 1. Contoh Dendrogram

Terdapat dua pendekatan dasar dalam metode hierarki yaitu *agglomerative* (pemusatan) dan *disive* (penyebaran). Pendekatan *agglomerative* melibatkan pembentukan *cluster* baru dengan menggabungkan objek individu berdasarkan kedekatan satu sama lain. Dengan menentukan jarak antar *cluster* maka kedekatan dapat ditentukan [13]. Perhitungan jarak *encludiean* dalam analisis *cluster* metode hierarki adalah sebagai berikut.

$$d_{ij} = \sum_{k=1}^p (X_{ik} - X_{jk})^2 \quad (5)$$

Keterangan :

p = banyaknya variabel

d_{ij} = jarak antara objek ke- i dan objek ke- j

X_{ik} = data dari objek ke- i pada variabel ke- k

X_{jk} = data dari objek ke- j pada variabel ke- k

Terdapat beberapa metode *cluster* hierarki dengan pendekatan *agglomeratif* yaitu *single linkage*, *complete linkage*, dan *average linkage* yang dijelaskan sebagai berikut.

Single Linkage

Single linkage adalah teknik pengelompokan yang didasarkan pada jarak terdekat antara dua objek. Pada awalnya, pengelompokan *single linkage* memilih jarak terkecil dan menggabungkan objek yang sesuai [14]. Berikut rumus perhitungan jarak yang digunakan.

$$d_{(uv)w} = \min(d_{uw}, d_{vw}) \quad (6)$$

Keterangan :

$d_{(uv)w}$ = jarak minimum antara kelompok (UV) dan (W)

d_{uw} = jarak antara tetangga terdekat dari *cluster* U dan W

d_{vw} = jarak antara tetangga terdekat dari *cluster* V dan W

Complete Linkage

Jarak maksimum. Jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan objek yang mempunyai jarak terjauh [14]. Rumus perhitungan jarak dengan *complete linkage* adalah sebagai berikut.

$$d_{(uv)w} = \max(d_{uw}, d_{vw}) \quad (7)$$

Keterangan :

$d_{(uv)w}$ = jarak maksimum antara kelompok (UV) dan (W)

d_{uw} = jarak terjauh dari *cluster* U dan W

d_{vw} = jarak terjauh dari *cluster* V dan W

Average Linkage

Average linkage merupakan variasi dari algoritma *single linkage* dan *complete linkage* yaitu pengelompokan yang dibentuk berdasarkan koefisien rata-rata pada jarak seluruh individu dalam satu kelompok dengan rata-rata jarak seluruh individu pada kelompok lainnya [14]. Berikut rumus untuk perhitungan jarak *average linkage*.

$$d_{(UV),W} = \frac{\sum_{i=1}^n \sum_{k=1}^n d_{ik}}{N_{(UV)}N_W} \quad (8)$$

Keterangan :

$d_{(UV),W}$ = jarak antara *cluster* gabungan (UV) dan *cluster* W

d_{ik} = jarak titik data i di *cluster* (UV) dan titik data k di *cluster* W

$N_{(UV)}$ = jumlah titik data dalam *cluster* gabungan (UV)

N_W = jumlah titik data dalam *cluster* W

Metode Non-Hierarki

Analisis *cluster* non-hierarki dimulai dengan terlebih dahulu menentukan jumlah *cluster* yang diinginkan peneliti (dua, tiga, atau lebih). Setelah menentukan *cluster*, menempatkan objek-objek ke dalam *cluster* dan tidak mengikuti proses hierarki. Dalam analisis *cluster* non-hierarki terdapat metode yang sering digunakan, yaitu metode *K-Means* [15].

a. K-Means Clustering

K-Means clustering adalah metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya [13]. Langkah dalam menggunakan metode *k-means* adalah sebagai berikut.

- 1) Menentukan banyak *k-cluster* yang ingin dibentuk
- 2) Membangkitkan nilai random untuk pusat *cluster* awal (*centroid*) sebanyak *k-cluster*
- 3) Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidian* hingga ditemukan jarak yang paling

dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidian Distance*:

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2} \quad (9)$$

Keterangan :

$d(x_i, \mu_i)$ = jarak antara *cluster* x dengan pusat *cluster* μ pada data ke-i

x_i = bobot data ke-i pada *cluster* yang ingin dicari jaraknya

μ_i = bobot data ke-i pada pusat *cluster*

- 4) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil)
- 5) Mengupdate nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} d_i \quad (10)$$

Keterangan :

C_k = centroid baru untuk *cluster* ke-k

n_k = jumlah data dalam *cluster* ke-k

d_i = jumlah dari nilai jarak yang masuk dalam masing- masing *cluster*

- 6) Melakukan perulangan dari langkah 2 hingga 5 sampai anggota tiap *cluster* tidak ada yang berubah
- 7) Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

Cluster Optimum

Pendekatan yang biasa digunakan untuk menentukan jumlah *cluster* optimal adalah dengan menggunakan statistik *Pseudo-F*. Statistik *Pseudo-F* sering dianggap sebagai metode yang andal dan menunjukkan kinerja yang unggul dibandingkan pendekatan lainnya [16]. Rumus untuk menghitung *Pseudo-F* dapat dilakukan dengan Persamaan berikut.

$$Pseudo - F = \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \quad (11)$$

dengan,

$$R^2 = \frac{SST - SSW}{SST} \quad (12)$$

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \quad (13)$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_k)^2 \quad (14)$$

Keterangan :

SST = jumlah total kuadrat jarak sampel ke rata-rata keseluruhan

SSW = jumlah total dari kuadrat jarak sampel ke rata-rata *cluster*

n = jumlah sampel

c = jumlah variabel

n_c = banyaknya sampel ke- i pada kelompok ke- j

p = banyaknya variabel

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

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

$\bar{x}_k$ = rata-rata sampel pada variabel ke- k

Setelah mendapatkan jumlah kelompok yang optimum dalam setiap metode. Selanjutnya, proses analisis dilanjutkan dengan menentukan metode pengelompokan yang paling optimum berdasarkan nilai *internal cluster dispersion rate (icdrate)* minimum.

$$icdrate = 1 - R^2 \quad (15)$$

Keterangan :

R^2 = proporsi jumlah kuadrat jarak antar pusat kelompok dengan jumlah kuadrat sampel terhadap rata-rata keseluruhan

Indikator Pendidikan

Pendidikan merupakan aspek mendasar dalam membangun suatu negara. Melalui pendidikan masyarakat dapat terangkat harkat, martabat, serta kesejahteraannya [17]. Berapa indikator pendidikan di Indonesia yaitu Rata-Rata Lama Sekolah (RLS), Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM) dan Indeks Pembangunan Literasi Masyarakat (IPLM).

1) Rata-rata Lama Sekolah

Rata-rata lama sekolah (RLS) adalah angka yang menggambarkan lamanya (tahun) masa sekolah yang dialami penduduk usia 25 tahun ke atas. Badan Pusat Statistik (2020) mengemukakan bahwa RLS didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. RLS dapat digunakan

untuk mengetahui tingkat dan kualitas pendidikan masyarakat dalam suatu wilayah [18]. Rata-rata lama sekolah dapat dihitung dengan Persamaan berikut.

$$RLS = \frac{1}{n} \times \sum_{i=1}^n x_i \quad (16)$$

Keterangan :

x_i = lama sekolah penduduk ke- i yang berusia 25 tahun

n = jumlah penduduk usia 25 tahun ke atas

2) Angka Partisipasi Kasar

Angka partisipasi kasar (APK) adalah persentase perbandingan antara jumlah penduduk yang masih bersekolah di jenjang pendidikan tertentu (tanpa memandang usia penduduk tersebut) dengan jumlah penduduk yang memenuhi syarat resmi penduduk usia sekolah di jenjang pendidikan yang sama. APK digunakan sebagai tolak ukur kesuksesan program pendidikan yang dicanangkan pemerintah dalam upaya memberikan kesempatan yang lebih luas bagi penduduk untuk mengenyam pendidikan [19]. APK dapat dihitung menggunakan Persamaan berikut.

$$APK = \frac{\text{Jumlah Siswa Terdaftar}}{\text{Populasi Usia Sekolah}} \times 100\% \quad (17)$$

3) Angka Partisipasi Murni

Angka Partisipasi Murni (APM) adalah perbandingan antara siswa usia sekolah tertentu pada jenjang pendidikan dengan penduduk usia yang sesuai dan dinyatakan dalam persentase. APM digunakan untuk mengetahui banyaknya anak usia sekolah yang bersekolah pada jenjang yang sesuai. Semakin tinggi APM berarti semakin banyak anak usia sekolah yang bersekolah sesuai usia resmi di jenjang pendidikan tertentu [20]. APM dapat dihitung menggunakan Persamaan berikut.

$$APK (SMA) = \frac{\text{Jumlah Siswa SMA Terdaftar Usia 16-18 Tahun}}{\text{Jumlah Penduduk Usia 16-18 Tahun}} \times 100\% \quad (18)$$

4) Indeks Pembangunan Literasi Masyarakat

Indeks Pembangunan Literasi Masyarakat (IPLM) adalah data tingkat pembangunan literasi masyarakat yang diperoleh dari komponen Unsur Pembangun Literasi Masyarakat (UPLM) dan Aspek Masyarakat (AM) dalam upaya pembinaan dan pengembangan perpustakaan sebagai wahana belajar sepanjang hayat guna meningkatkan literasi [21]. IPLM dapat dihitung menggunakan Persamaan berikut.

$$IPLM = \sum_{i=1}^n \frac{UPLM_i}{AM} \times 100 \quad (19)$$

Keterangan :

$UPLM_i$ = variabel komponen pembentuk indeks dari Unsur

Pembangunan Literasi Masyarakat

AM= jumlah populasi sesuai segmentasi berdasarkan jenis perpustakaan

3. METODOLOGI PENELITIAN

Sumber Data

Data yang digunakan merupakan data sekunder yang diunduh melalui *website* Badan Pusat Statistik (BPS) dengan sumber data yaitu Statistik Indonesia 2023. Data yang digunakan menurut provinsi di Indonesia yang terdiri sebanyak 34 data provinsi untuk masing-masing variabel.

Identifikasi Variabel dan Skala Pengukuran

Tabel 1. Variabel Penelitian

Variabel yang digunakan pada penelitian ini disajikan pada Tabel 1.

Variabel	Keterangan	Skala Pengukuran	Satuan
X_1	Rata-rata Lama Sekolah	Interval	Tahun
X_2	Angka Partisipasi Kasar	Rasio	Persen
X_3	Angka Partisipasi Murni	Rasio	Persen
X_4	Jumlah Sekolah SMA	Rasio	Sekolah
X_5	Indeks Pembangunan Literasi Masyarakat	Rasio	-

Struktur Data

Tabel 2. Struktur Data

Struktur data yang digunakan pada penelitian ini disajikan pada Tabel 2.

Provinsi ke-	X_1	X_2	X_3	X_4	X_5
1	X_{11}	X_{21}	X_{31}	X_{41}	X_{51}
2	X_{12}	X_{22}	X_{32}	X_{42}	X_{52}
3	X_{13}	X_{23}	X_{33}	X_{43}	X_{53}
:	:	:	:	:	:
34	$X_{1\ 34}$	$X_{2\ 34}$	$X_{3\ 34}$	$X_{4\ 34}$	$X_{5\ 34}$

Langkah Analisis

Langkah analisis dalam penelitian ini adalah sebagai berikut.

- a. Mendeskripsikan karakteristik data indikator kualitas pendidikan di Indonesia tahun 2023.
- b. Melakukan pengujian asumsi distribusi normal multivariat dan uji dependensi pada indikator kualitas pendidikan di Indonesia tahun 2023.
- c. Melakukan analisis *cluster* dengan metode hierarki dan non hierarki pada indikator kualitas pendidikan di Indonesia tahun 2023, dengan langkah sebagai berikut.
 - 1) Menentukan jumlah *cluster* optimum
 - 2) Menghitung nilai *Pseudo-F* dan *ICD rate* pada setiap *cluster*
 - 3) Menentukan metode dan jumlah *cluster* optimal pada *cluster* hierarki dan non hierarki berdasarkan nilai *Pseudo-F* dan *ICD rate*
- d. Melakukan pemetaan *cluster* hierarki dan non hierarki optimal.
- e. Menginterpretasikan hasil analisis.
- f. Menarik kesimpulan dan saran.

4. ANALISIS DAN PEMBAHASAN

Bab ini membahas mengenai analisis karakteristik data, pengujian asumsi, dan analisis pengelompokan provinsi berdasarkan indikator kualitas pendidikan di Indonesia menggunakan *clustering* hierarki dan non-hierarki.

Karakteristik Indikator Kualitas Pendidikan Indonesia Tahun 2023

Karakteristik indikator kualitas pendidikan di Indonesia tahun 2023 menggunakan

Tabel 3. Karakteristik Data Indikator Kualitas Pendidikan di Indonesia

statistika deskriptif dijelaskan pada Tabel 3.

Variabel	Rata-Rata	Standar Deviasi	Minimum	Maksimum
X ₁	9,30	0,82	7,34	11,42
X ₂	88,72	6,17	73,90	98,02
X ₃	63,89	5,92	48,32	76,37
X ₄	424	386	68	1768
X ₅	65,80	8,14	47,57	86,74

Tabel 3 menunjukkan data kualitas pendidikan di Indonesia tahun 2023. Dapat diketahui bahwa rata-rata lama sekolah (X₁) mencapai 9,3 tahun, setara dengan jenjang

SMP. Angka partisipasi kasar (X_2) berada pada 88,7%, mencerminkan mayoritas penduduk mengikuti pendidikan, namun angka partisipasi murni (X_3) yang lebih rendah, yaitu 63,9%, menunjukkan ketidaksesuaian usia ideal pada jenjang pendidikan tertentu. Jumlah sekolah SMA per provinsi (X_4) sangat bervariasi, dengan rata-rata 424 sekolah, mencerminkan kesenjangan signifikan dalam akses pendidikan antarwilayah. Indeks literasi masyarakat (X_5) rata-rata sebesar 65,8, yang mengindikasikan kemampuan literasi cukup baik secara nasional, meskipun beberapa wilayah masih membutuhkan perhatian lebih untuk peningkatan literasi. Variabel dengan variasi yang besar dan cenderung tidak homogen adalah jumlah sekolah SMA (X_4) dari setiap provinsi di Indonesia dan variabel dengan variasi yang relatif kecil dan cenderung homogen adalah lama periode penduduk bersekolah (X_1) dari setiap provinsi.

Pengujian Asumsi *Cluster*

Sebelum melakukan analisis *cluster*, diperlukan pemenuhan asumsi distribusi normal multivariat dan dependensi. Hasil pengujian asumsi-asumsi tersebut adalah sebagai berikut.

a) Uji Distribusi Normal Multivariat

Pengujian distribusi normal multivariat dilakukan untuk memastikan bahwa data pada variabel indikator kualitas pendidikan di Indonesia tahun 2023 memenuhi asumsi normalitas multivariat. Pengujian dilakukan dengan menggunakan korelasi *pearson* antara jarak mahalanobis dengan *quartile*. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : Data indikator kualitas pendidikan di Indonesia tahun 2023 berdistribusi normal multivariat

H_1 : Data indikator kualitas pendidikan di Indonesia tahun 2023 tidak berdistribusi normal multivariat

Taraf signifikan yang digunakan adalah 0,05, dengan daerah penolakan H_0 jika $r_Q < r_{(\alpha;n)}$ atau $P\text{-value} > 0,05$. Hasil pengujian diperoleh nilai r_Q sebesar 0,979 yang lebih besar dari nilai $r_{(0,05;34)}$ sebesar 0,967 dan diperkuat dengan $P\text{-value}$ sebesar 0,000 yang lebih kecil dari taraf signifikan α sebesar 0,05. Sehingga diputuskan gagal tolak H_0 , artinya data kualitas pendidikan di Indonesia tahun 2023 memenuhi asumsi distribusi normal multivariat.

b) Uji Independensi

Pengujian independensi bertujuan untuk mengetahui hubungan atau ketergantungan antar variabel. Uji independensi dilakukan melalui uji *Bartlett* pada data indikator kualitas pendidikan Indonesia tahun 2023, dengan hipotesis sebagai berikut.

H_0 : Data indikator kualitas pendidikan di Indonesia tahun 2023 bersifat independen

H_1 : Data indikator kualitas pendidikan di Indonesia tahun 2023 bersifat dependen

Taraf signifikan yang digunakan adalah 0,05, dengan daerah penolakan H_0 jika $\chi^2_{hitung} > \chi^2_{(\alpha; df)}$ atau *P-value* < 0,05. Hasil pengujian didapatkan nilai χ^2_{hitung} sebesar 38,667 yang lebih besar dari nilai $\chi^2_{(0,05;10)}$ sebesar 18,307 dan diperkuat dengan *P-value* sebesar 0,000 yang lebih kecil dari taraf signifikan α sebesar 0,05. Sehingga, diputuskan tolak H_0 , yang artinya data indikator kualitas pendidikan di Indonesia tahun 2023 memenuhi asumsi dependen.

Analisis Cluster

Analisis *cluster* dilakukan dengan membandingkan antara dua metode yaitu analisis *cluster* hierarki dengan menggunakan *single average*, *average linkage*, dan *complete linkage* serta analisis *cluster* non hierarki dengan menggunakan *K-Means* akan dijelaskan sebagai berikut.

a. Penentuan Jumlah Cluster Optimum

Pembentukan lima *cluster* menggunakan analisis hierarki dan non hierarki pada pengelompokan kualitas pendidikan di Indonesia tahun 2023 menghasilkan nilai

Tabel 4. Penentuan Jumlah Cluster Hierarki dan Non Hierarki

Pseudo-F dan *ICD Rate* masing-masing *cluster* disajikan pada Tabel 6.

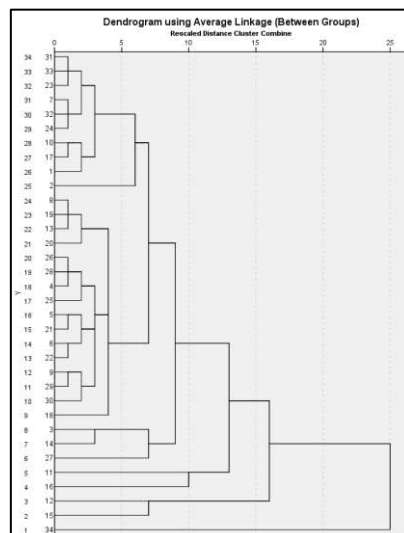
Jenis Metode	Jumlah Cluster	<i>Pseudo-F</i>	ICD Rate
<i>Single Linkage</i>	2	0,170	0,995
	3	9,456	0,621
	4	6,167	0,619
	5	4,672	0,608
<i>Average Linkage</i>	2	0,170	0,995
	3	27,412	0,361
	4	19,087	0,344

	5	23,961	0,232
Complete Linkage	2	0,170	0,995
	3	0,120	0,992
	4	17,724	0,361
	5	13,901	0,343
	2	3,348	0,905
K-Means	3	2,072	0,882
	4	18,710	0,348
	5	13,318	0,352

Berdasarkan Tabel 6 menunjukkan bahwa hasil analisis *cluster* hierarki dan non hierarki pada pengelompokan Provinsi di Indonesia berdasarkan kualitas pendidikan tahun 2023 dengan metode *single linkage*, *average linkage*, *complete linkage*, dan *K-Means*. Menentukan jumlah *cluster* yang optimal pada masing-masing metode, dipertimbangkan nilai *Pseudo-F* terbesar dan *ICD Rate* terendah. Berdasarkan empat metode tersebut, metode *average linkage* dipilih sebagai metode yang optimal karena memiliki nilai *Pseudo-F* terbesar dan *ICD Rate* terendah, sehingga dapat disimpulkan jumlah *cluster* optimum yang terbentuk adalah 3 *cluster*.

b. Dendrogram

Dendrogram merupakan diagram yang digunakan untuk menggambarkan hasil *clustering* dalam bentuk pohon dengan cabang yang lebih panjang menunjukkan jarak yang lebih jauh antara kelompok. Sedangkan, cabang yang lebih pendek menunjukkan jarak yang lebih dekat antara kelompok. Dendrogram hasil analisis *cluster* hierarki pengelompokan Provinsi di Indonesia berdasarkan kualitas pendidikan tahun 2023 disajikan pada Gambar 2.



Gambar 2. Dendrogram Analisis Cluster Hierarki

Gambar 2 menunjukkan dendrogram dengan pembagian provinsi di Indonesia menjadi tiga *cluster* berdasarkan kemiripan karakteristik. Pembentukan tiga *cluster* atau kelompok akan menghasilkan 31 provinsi pada *cluster* pertama, 2 provinsi pada *cluster* kedua, dan 1 provinsi pada *cluster* ketiga.

Hasil *cluster* optimum pada pengelompokan Provinsi di Indonesia berdasarkan kualitas pendidikan tahun 2023 menggunakan analisis *cluster* hierarki metode *average linkage* menjadi lima *cluster* disajikan pada Tabel 7.

Tabel 5. Hasil *Cluster* Hierarki Optimum

<i>Cluster</i>	Jumlah Provinsi	Provinsi
C ₁	31	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
C ₂	2	Jawa Barat, Jawa Timur
C ₃	1	Papua

Berdasarkan Tabel 7 menunjukkan bahwa hasil *cluster* optimum menggunakan analisis *cluster* hierarki metode *average linkage* sebanyak 3 *cluster*, dengan *cluster* pertama terdapat 31 Provinsi terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, *cluster* kedua terdapat 2 Provinsi terdiri dari Jawa Barat dan Jawa Timur, *cluster* ketiga terdapat 1 Provinsi terdiri dari Papua.

c. Pemetaan Hasil Cluster Optimum

Pemetaan hasil *cluster* optimum pada kualitas pendidikan di Indonesia tahun 2023 ditunjukkan pada Gambar 3.



Gambar 3. Pemetaan Hasil *Cluster* Menggunakan *Average Linkage* Kualitas Pendidikan

Gambar 3 menunjukkan pemetaan *cluster* kualitas pendidikan di Indonesia tahun 2023 berdasarkan hasil analisis *cluster average linkage*, di mana setiap provinsi dikelompokkan ke dalam tiga *cluster* yang berbeda. Setiap warna pada peta merepresentasikan *cluster* dengan karakteristik kualitas pendidikan yang serupa. *Cluster 1*, ditandai dengan warna biru tua, mencakup 31 provinsi dengan kualitas pendidikan yang relatif baik, tersebar di wilayah Sumatera, Jawa, Bali, Kalimantan, Sulawesi, Maluku, hingga Papua Barat. *Cluster 2*, yang ditandai dengan warna jingga, mencakup Jawa Barat dan Jawa Timur. Kualitas pendidikan di *cluster 2* berada pada tingkat menengah, kemungkinan dipengaruhi oleh tingginya jumlah penduduk di kedua provinsi tersebut. *Cluster 3*, ditandai dengan warna hijau, hanya mencakup wilayah Papua, dan merepresentasikan wilayah dengan kualitas pendidikan rendah atau tertinggal. Pemetaan ini menggambarkan adanya kesetaraan di sebagian besar wilayah, namun juga menunjukkan ketimpangan yang signifikan pada daerah terpencil seperti Papua.

d. Karakteristik *Cluster* Optimum

Karakteristik masing-masing *cluster* berdasarkan indikator kualitas pendidikan di Indonesia tahun 2023 ditunjukkan pada Tabel 8.

Tabel 6. Karakteristik Cluster Hierarki

Variabel	Cluster	Mean	Minimum	Maksimum
Rata-rata Lama Sekolah	1	9,40	8,17	11,42
	2	8,85	8,53	9,16
	3	7,34	7,34	7,34
Angka Partisipasi Kasar	1	89,50	75,02	98,02
	2	84,17	79,07	89,26
	3	73,90	73,90	73,90
Angka Partisipasi Murni	1	64,59	54,76	76,37
	2	60,80	59,01	62,59
	3	48,32	48,32	48,32
Jumlah Sekolah SMA	1	351	68	1067
	2	1642	1517	1768
	3	268	268	268
Indeks Pembangunan Literasi Masyarakat	1	66,28	52,50	86,74
	2	67,60	60,02	75,18
	3	47,57	47,57	47,57

Berdasarkan Tabel 8, Karakteristik *cluster* hierarki kualitas pendidikan di Indonesia tahun 2023 menunjukkan perbedaan mencolok antar kelompok. *Cluster* 1, dengan rata-rata lama sekolah tertinggi (9,4 tahun), angka partisipasi kasar (89,5%), dan angka partisipasi murni (64,59%), menggambarkan wilayah dengan tingkat pendidikan relatif baik, meski distribusi jumlah sekolah SMA masih belum merata. *Cluster* 2 memiliki rata-rata lama sekolah (8,85 tahun) dan partisipasi pendidikan lebih rendah dibanding *cluster* 1, namun jumlah sekolah SMA terbanyak (rata-rata 1642), yang mencerminkan dominasi wilayah dengan penduduk padat. *Cluster* 3, yang terdiri dari Provinsi Papua, menampilkan tantangan signifikan, dengan rata-rata lama sekolah terendah (7,34 tahun), partisipasi pendidikan terendah, dan indeks literasi hanya 47,57, mengindikasikan perlunya peningkatan akses dan kualitas pendidikan. Secara keseluruhan, perbedaan ini menegaskan adanya ketimpangan yang membutuhkan perhatian khusus untuk pemerataan pembangunan pendidikan.

5. KESIMPULAN

Kesimpulan yang didapatkan dari hasil analisis dan pembahasan yaitu karakteristik jumlah sekolah SMA memiliki variasi cukup besar dan cenderung tidak homogen dan rata-

rata lama sekolah memiliki variasi cenderung kecil dan cenderung homogen. Analisis *cluster* hierarki dan non hierarki menunjukkan bahwa, secara optimum terbentuk 3 *cluster* dengan *cluster* pertama (C_1) terdiri dari 31 provinsi, *cluster* kedua (C_2) terdiri dari 2 provinsi, dan *cluster* ketiga (C_3) terdiri dari 1 provinsi. Berdasarkan nilai *Pseudo-F* tertinggi dan *ICD Rate* terendah, didapatkan hasil pemilihan metode terbaik, yakni metode hierarki dengan *average linkage* sebanyak 3 *cluster*.

DAFTAR PUSTAKA

- Abels, P., Nguyen, H. D. M., Kawamura, H., & Chikada, M. (2021). Global quality assurance system of higher education in the United States, Vietnam, and Japan: Possible? *European Journal of Educational Sciences*, 8(3), 63–80. <https://doi.org/10.19044/ejes.v8no3a63>
- Amah, N., Wahyuningsih, S., Deny, F., & Amijaya, T. (2017). Analisis cluster non-hirarki dengan menggunakan metode K-Modes pada mahasiswa program studi statistika angkatan 2015 FMIPA Universitas Mulawarman. *Jurnal Eksponensial*, 8(1), 9–16.
- Apriliana, T., & Widodo, E. (2023). Analisis cluster hierarki untuk pengelompokan provinsi di Indonesia berdasarkan jumlah base transceiver station dan kekuatan sinyal. *KONSTELASI Konvergensi Teknologi dan Sistem Informasi*, 3(2), 286–296.
- Arsham, M., & Lovric, M. (2011). Bartlett's test. *Statistical Management Science*, 1–2.
- Faradiba, & Nomleni, M. V. (2024). Analisis indikator pendidikan di Indonesia periode 1994–2022. *Jurnal Penelitian Inovasi*, 4(1), 121–128.
- Ginting, E. V., Ginting, R. R., Hasibuan, R. J., & Perangin-angin, L. M. (2022). Analisis faktor tidak meratanya pendidikan di SDN0704 Sungai Korang. *Jurnal Pendidikan Indonesia*, 3(4), 407–416.
- Hermanto, B. I. P., Gio, D., Nazriani, P. U., Bahri, S., Sirait, R., Syahputra, K. J., Shara, Y., & Lubis, I. T. (2024). *Aplikasi metode statistika korelasi Pearson & Spearman di artikel jurnal nasional bidang psikologi dengan SPSS* (1st ed.). Uwais Inspirasi Indonesia.
- Hikmah, F., Fardinah, L., Qadrini, L., & Tande, E. (2022). Analisis klaster pengelompokan kecamatan di Sulawesi Barat berdasarkan indikator pendidikan. *SAINTIFIK Jurnal Matematika Sains, dan Pembelajarannya*, 8(2), 188–196.
- Imro'ah, N., Ayuningtias, I., & Debataraja, N. N. (2019). Analisis cluster non-hirarki dengan metode K-Modes. *Bimaster Bul. Ilmiah Matematika Statistika dan Terapan*, 8(4), 909–916.
- Kurniawan, R., Mukarrobin, M., & Mahradianur. (2021). Klastering tingkat pendidikan di DKI Jakarta pada tingkat kecamatan menggunakan algoritma K-Means. *Technologia*, 12(4), 234–240.

- Mukhaiyar, U., Rontos, F., Handoko, K., & Kardiyanti, S. (2022). Analisis faktor-faktor yang memengaruhi angka partisipasi kasar SMA/ sederajat di Indonesia menggunakan regresi ridge. *Euler Jurnal Ilmiah Matematika Sains dan Teknologi*, 10(2), 222–234.
- Rahmana, S. P., Firda, A., & Rahayu, A. (2013). Analisis pola hubungan kerugian negara akibat korupsi dengan demografi koruptor di Jawa Timur. *Jurnal Sains dan Seni POMITS*, 2(1), 305–306.
- Rahmawati, L., Sihwi, S. W., & Suryani, E. (2016). Analisa clustering menggunakan metode K-Means dan hierarchical clustering (Studi kasus: Dokumen skripsi jurusan kimia, FMIPA, Universitas Sebelas Maret). *Jurnal Teknologi Informasi ITSmart*, 3(2), 66.
- Ramadhani, L., Purnamasari, I., & Amijaya, F. D. T. (2018). Penerapan metode complete linkage dan metode hierarchical clustering multiscale bootstrap (Studi kasus: Kemiskinan di Kalimantan Timur tahun 2016). *Eksponensial*, 9(1), 1–10.
- Romadon, R. R., Darmawan, H., & Oktaviani, I. N. (2021). *Pedoman pengukuran indeks pembangunan literasi masyarakat*. Perpustakaan PRESS.
- Rusli, J., Gaffar, J., Bakhri, S., & Ahmad, M. A. (2014). Pengujian normal multivariat dan homoskedastisitas matriks varians-kovarians pada prestasi belajar dan kredit lulus mahasiswa UNM angkatan 2014. *Scribd*.
- Sabrina, R., Manurung, A. I., & Sirait, B. A. (2022). Peningkatan rata-rata lama sekolah (RLS) dari harapan lama sekolah (HLS) di Sumatera Utara. *Jurnal Pendidikan Tambusai*, 6(1), 4784–4792.
- Safira, N., & Wibowo, Y. H. (2021). Angka partisipasi kasar dan angka partisipasi murni sebagai indikator keberhasilan pendidikan Indonesia. *Jurnal PAKAR Pendidik*, 19(2), 101–115.
- Setyawan, D. A., Fahradsina, N., Pd, M., Faelasofi, R., Widayarsi, T., & Mawardati, R. (2022). *Statistika Terapan*. Tahta Media.
- Sinaga, S. J., Satyahadewi, N., & Perdana, H. (2023). Determining the optimum number of clusters in hierarchical clustering using Pseudo-F. *Euler Jurnal Ilmiah Matematika Sains dan Teknologi*, 11(2), 372–382.
- Suhaeni, C., Kurnia, A., & Ristiyanti, R. (2018). Perbandingan hasil pengelompokan menggunakan analisis cluster berhirarki, K-Means cluster, dan cluster ensemble (Studi kasus data indikator pelayanan kesehatan ibu hamil). *Jurnal Media Infotama*, 14(1).