

# Analisis Klaster Menggunakan Metode Average Linkage dengan Validasi Multiscale Bootstrap (Studi Kasus: Indikator Pendidikan di Indonesia Tahun 2021)

## *Cluster Analysis Using the Average Linkage Method with Multiscale Bootstrap Validation (Case Study: Education Indicators in Indonesia in 2021)*

Fikri Rasidia<sup>1a)</sup>, Rito Goejantoro<sup>2</sup>, M. Fathurahman<sup>3</sup>

<sup>1, 2, 3</sup>Laboratorium Statistika Komputasi, FMIPA Universitas Mulawarman, Indonesia

<sup>1, 2, 3</sup>Program Studi S1 Statistika, FMIPA Universitas Mulawarman, Indonesia

<sup>a)</sup>Corresponding author: [Fikrirasidia86@gmail.com](mailto:Fikrirasidia86@gmail.com)

### ABSTRACT

The average linkage method is one of the hierarchical cluster analyses, where the clustering process starts by finding two objects that have the closest distance to the average rule of the two groups. The multiscale bootstrap method is a method used to see the validity of the cluster analysis results. This study aims to determine the clusters formed using the average linkage method, as well as to determine the validity of the clusters formed based on education indicators in each province in Indonesia. The result of the study is one cluster with AU (Approximately Unbiased)  $\geq 0.95$  so that the cluster is considered to be able to represent the actual population.

Keywords: average linkage, cluster analysis, education, multiscale bootstrap

## 1. Pendahuluan

Klasterisasi adalah teknik statistika yang bertujuan untuk mengelompokkan objek berdasarkan kesamaan di antara mereka. Objek yang memiliki kemiripan akan dikelompokkan ke dalam satu atau beberapa klaster, sehingga objek dalam satu klaster memiliki kemiripan antara satu sama lain (Kuncoro, 2003). Terdapat dua pendekatan umum dalam klasterisasi, yaitu hierarki dan non-hierarki. Dalam klasterisasi hierarki terdapat berbagai pendekatan seperti pautan tunggal, pautan lengkap, pautan rata-rata, dan *Ward's method*. Sementara itu, dalam klasterisasi non-hierarki, terdapat pendekatan seperti K-Means (Supranto, 2004).

Dalam teknik pautan tunggal, jarak antar objek diukur dengan menemukan dua objek yang memiliki jarak terkecil. Setelah itu, kedua objek tersebut digabungkan menjadi satu grup. Sebaliknya, teknik pautan lengkap menggunakan jarak maksimum antar grup, dan prosesnya mirip dengan pautan tunggal. Dalam teknik pautan rata-rata, proses dimulai dengan menemukan dua objek dengan jarak terdekat berdasarkan rata-rata jarak antar grup. *Ward's method* mengukur jarak antara dua grup yang terbentuk dengan jumlah kuadrat perbedaan antara grup-grup tersebut (Supranto, 2004).

Salah satu kekurangan klasterisasi adalah ketidakmampuannya untuk menilai sejauh mana hasil klasterisasi tersebut dapat dipercaya, sehingga diperlukan teknik tambahan untuk mengevaluasi hasil tersebut. Salah satu teknik yang dapat digunakan adalah *multiscale bootstrap* yang berfungsi untuk menilai sejauh mana hasil klasterisasi sebelumnya dapat diandalkan dengan menghitung nilai *Approximately Unbiased* (AU). Teknik ini menerapkan pendekatan *bootstrap resampling* untuk setiap grup yang terbentuk (Efron dan Tibshirani, 1993).

Beberapa penelitian sebelumnya telah menerapkan *Hierarchical Clustering Multiscale Bootstrap*. Sebagai contoh, Anuraga (2015) dalam penelitiannya tentang klasterisasi data kemiskinan di Jawa Timur menggunakan teknik ini dan menemukan kestabilan grup pada jumlah  $B = 500$ . Safarina (2019) menggunakan teknik serupa untuk data Indeks Pembangunan Manusia (IPM) di Kalimantan Timur tahun 2017, dan menemukan tiga grup yang dapat diandalkan dengan nilai  $AU \geq 0,95$ .

Penelitian ini bertujuan untuk mengidentifikasi kelompok yang terbentuk dengan teknik *average linkage* berdasarkan indikator pendidikan di setiap provinsi di Indonesia dan untuk menilai sejauh mana kelompok yang terbentuk dapat diandalkan dengan menggunakan *multiscale bootstrap* berdasarkan indikator pendidikan di setiap provinsi di Indonesia.

## 2. Tinjauan Pustaka

### 2.1. Analisis Multivariat

Berdasarkan banyaknya variabel yang dianalisis, maka analisis statistika dibagi menjadi analisis univariat, bivariat, dan multivariat. Analisis multivariat merupakan pengembangan lanjutan dari analisis univariat maupun bivariat yang dapat diartikan sebagai teknik statistika tertentu yang menganalisis banyak

variabel secara simultan (Kuncoro, 2003).

Analisis multivariat terbagi menjadi dua yakni metode dependensi dan metode interdependensi. Ciri dari metode dependensi yaitu terdapat satu atau beberapa variabel yang memiliki fungsi sebagai variabel terikat dan variabel bebas. Metode-metode yang tergolong kelompok metode dependensi yaitu analisis regresi berganda, analisis diskriminan berganda, dan analisis korelasi kanonikal. Sedangkan ciri dari metode interdependensi yaitu seluruh variabelnya merupakan variabel bebas. Metode-metode yang termasuk dalam metode interdependensi adalah analisis klaster, analisis faktor, dan multidimensional scaling (Laeli, 2014).

### 2.1.1. Metode Dependensi

Pendekatan dependensi bertujuan untuk menguraikan atau memperkirakan nilai variabel terikat dengan menggunakan lebih dari satu variabel bebas yang mempengaruhinya. Dalam pendekatan ini, terdapat berbagai jenis teknik yang dikategorikan berdasarkan jumlah variabel terikat serta jenis pengukuran data dari variabel terikat dan variabel bebas (Supranto, 2004).

Beberapa jenis metode multivariat yang termasuk dalam kelompok metode dependensi dapat dilihat pada Tabel 1.

**Tabel 1.** Jenis metode multivariat dependensi

| Jenis Analisis Multivariat | Variabel Dependensi |                | Jenis Variabel Independensi |
|----------------------------|---------------------|----------------|-----------------------------|
|                            | Jumlah Variabel     | Jenis Variabel |                             |
| 1. Regresi                 | 1                   | Metrik         | Metrik/Nonmetrik            |
| 2. Regresi Logistik        | 1                   | Nonmetrik      | Metrik/Nonmetrik            |
| 3. Analisis Diskriminan    | 1                   | Nonmetrik      | Metrik/Nonmetrik            |
| 4. Analisis Konjoin        | 1                   | Nonmetrik      | Nonmetrik                   |
| 5. Analisis Kanonikal      | >1                  | Metrik         | Metrik/Nonmetrik            |
| 6. MANOVA                  | >1                  | Metrik         | Nonmetrik                   |

### 2.1.2. Metode Interdependensi

Analisis interdependensi bertujuan untuk mengetahui mengapa dan bagaimana variabel yang ada dapat saling berhubungan. Pembagian metode interdependensi ini didasarkan pada jenis variabel yang ada, apakah bersifat metrik atau nonmetrik (Widarjono, 2002).

Beberapa jenis metode interdependensi dapat dilihat pada Tabel 2.

**Tabel 2.** Jenis metode multivariat interpendensi

| Jenis Analisis                              | Jenis Variabel |
|---|----------------|
| 1. Analisis Faktor                          | Metrik         |
| 2. Analisis Klaster                         | Metrik         |
| 3. Analisis Korespondensi                   | Nonmetrik      |
| 4. Analisis <i>Multidimensional Scaling</i> | Metrik         |

## 2.2. Statistika Deskriptif

Statistika deskriptif bertujuan untuk mengumpulkan, menyusun, dan memproses data agar dapat dipresentasikan mengenai kondisi atau kejadian tertentu. Ukuran pemusatan adalah angka yang menggambarkan kecenderungan nilai-nilai dalam suatu distribusi data untuk berkumpul di sekitar titik tertentu. Ukuran pemusatan mencakup rata-rata, nilai maksimum, nilai minimum, dan Ukuran penyebaran mencakup *range*, varian, deviasi kuartil dan deviasi standar (Ramadhani, Purnamasari, Amijaya, 2018).

### 2.3. Standardisasi Data

Standardisasi data dapat diterapkan baik pada data numerik maupun kategorik apabila terdapat perbedaan ukuran satuan yang signifikan antar nilai yang diamati. Perbedaan ukuran yang besar dapat membuat perhitungan menjadi tidak akurat. Standardisasi data dalam bentuk *z-score* dapat dijelaskan sebagai berikut (Yulianto & Hidayatullah, 2014).

$$z_{jk} = \frac{x_{jk} - \bar{x}_k}{s_k} \tag{1}$$

di mana

- $z_{jk}$  : Standardisasi data untuk data ke- $j$  variabel ke- $k$
- $x_{jk}$  : data ke- $j$  variabel ke- $k$
- $\bar{x}_k$  : rata-rata variabel ke- $k$
- $s_k$  : deviasi standar variabel ke- $k$ .

**2.4. Pendeteksian Multikolinieritas**

Untuk mengetahui apakah di dalam suatu model regresi terdapat korelasi antar variabel atau tidak, maka perlu dilakukan uji multikolinieritas. Suatu model regresi dapat dikatakan terdapat multikolinieritas jika memiliki nilai *Variance Inflation Factors* (VIF) > 10 dengan persamaan berikut (Nachrowi & Usman, 2006).

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{2}$$

dengan,

$$R_j^2 = 1 - \frac{\sum_{i=1}^n (x_{ji} - \hat{x}_{ji})^2}{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2} \tag{3}$$

di mana  $R_j^2$  merupakan nilai koefisien determinasi hasil dari regresi antara satu variabel bebas  $x_j$  yang berperan sebagai variabel terikat dengan variabel bebas lainnya (Nachrowi & Usman, 2006).

**2.5. Analisis Klaster**

Klasterisasi adalah teknik multivariat yang digunakan untuk mengelompokkan beberapa variabel berdasarkan kesamaan yang dimilikinya. Dalam pendekatan pengelompokan hierarki, terdapat dua metode utama yakni metode *agglomerative* (penyatuan) dan metode *divisive* (pembagian). Sebuah klaster dianggap baik jika memiliki tingkat kesamaan yang tinggi di antara anggotanya dan perbedaan yang jelas antara klaster-klaster yang berbeda (Prasetyo, 2012).

Jarak Euclid untuk mengukur jumlah kuadrat perbedaan nilai masing-masing variabel dinotasikan sebagai berikut (Supranto, 2004).

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{4}$$

dengan bentuk umum sebagai berikut

$$\text{Jarak Kemiripan} = \begin{pmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \dots & d_{nn} \end{pmatrix} \tag{5}$$

Jika jarak Euclid diakarkan, maka pada jarak Euclid kuadrat diformulasikan sebagai berikut

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \tag{6}$$

*Cityblock* atau biasa disebut dengan jarak Manhattan merupakan jumlah selisih mutlak untuk setiap variabel yang dinyatakan sebagai berikut

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|. \tag{7}$$

**2.6. Metode Klaster Hierarki**

Dalam metode klasterisasi hierarki, proses dimulai dengan mengidentifikasi dua atau lebih objek yang memiliki kesamaan paling dekat, kemudian perhitungan dilanjutkan dengan objek-objek lain yang paling mendekati, dan seterusnya. Ada empat teknik dalam pengelompokan aglomeratif, yaitu (Ranher, 2002).

a. *Single Linkage*

Pautan tunggal dilakukan dengan mengelompokkan jarak terkecil antara klaster pertama dengan klaster kedua, dinotasikan sebagai berikut

$$D(A, B) = \min\{d(y_i, y_j) \text{ untuk } y_i \text{ di } A \text{ dan } y_j \text{ di } B\} \tag{8}$$

b. *Complete Linkage*

Pengelompokan jarak antar klaster satu dengan klaster lain berdasarkan objek yang memiliki jarak terbesar, dinyatakan sebagai berikut

$$D(A, B) = \max\{d(y_i, y_j) \text{ untuk } y_i \text{ di } A \text{ dan } y_j \text{ di } B\} \tag{9}$$

c. *Average Linkage*

Pengelompokan jarak antar kluster satu dengan kluster lain berdasarkan jarak rata-rata antar objek, dengan persamaan sebagai berikut

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j) \tag{10}$$

d. *Ward's Method*

Pengelompokan objek berdasarkan kenaikan nilai *Sum of Squared Error* (SSE) terkecil, dinotasikan sebagai berikut

$$SSE = \sum_{i=1}^n (y_i - \bar{y})^T (y_i - \bar{y}) \tag{11}$$

**2.7. Metode Bootstrap**

*Bootstrap* adalah teknik yang dirancang untuk menyederhanakan proses statistik dengan memanfaatkan komputer modern. Umumnya, menggunakan  $B = 25$  *bootstrap* sudah cukup untuk mendapatkan informasi yang bermanfaat, sedangkan dalam klusterisasi, jumlah pengulangan yang diperlukan berkisar antara  $B = 100$  hingga  $B = 500$ . Penggunaan  $B \geq 200$  jarang diperlukan untuk estimasi deviasi standar, namun jumlah yang lebih besar sering dibutuhkan untuk membangun interval kepercayaan *bootstrap* (Ramadhani, Purnamasari, Amijaya, 2018).

**2.8. Multiscale Bootstrap**

Cara kerja *multiscale bootstrap* pada analisis kluster adalah dengan memilih salah satu data pengamatan dari hasil *resampling bootstrap* yang dianggap telah mewakili distribusi data asli dengan menghitung nilai *Approximately Unbiased* (AU) dan *Bootstrap Probability* (BP). Suatu kluster dikatakan valid jika nilai  $AU \geq 0,95$  (Shimodaira, 2004).

Berikut merupakan contoh penerapan *multiscale bootstrap* pada analisis kluster:

1. Sampel data diambil secara acak sehingga diperoleh data sampel baru yang didefinisikan sebagai  $\mathbf{d}'$  dan dilakukan secara berulang sebanyak  $B$  kali sehingga didapatkan himpunan data bootstrap ( $\mathbf{d}'_1, \mathbf{d}'_2, \dots, \mathbf{d}'_B$ ) dimana setiap sampel merupakan sampel acak yang saling independensi.
2. Menentukan  $N'$  dengan cara sebagai berikut:

- a.  $N'$  yang pertama =  $\frac{N}{2}$  (12)

- b.  $N'$  yang kedua =  $N'$  pertama +  $\left(\frac{N}{p}\right)$  (13)

- c.  $N'$  yang ketiga =  $N'$  kedua +  $\left(\frac{N}{p}\right)$  (14)

⋮

- d.  $N'$  yang ketiga puluh empat =  $N'$  ketiga puluh tiga +  $\left(\frac{N}{p}\right)$  (15)

3. Nilai  $v$  dan  $c$  yang diperoleh dari hasil perhitungan  $z(N'_i)$ , dinyatakan sebagai berikut

$$z(N'_i) = v\sqrt{N'_i/N} + c\sqrt{N_i/N'} \tag{16}$$

di mana

- $z(N'_i)$  : Nilai  $z$
- $N$  : Jumlah data pengamatan
- $N'$  : Jumlah data pengamatan *bootstrap*
- $v$  dan  $c$  : Parameter untuk AU

4. Nilai AU dinotasikan sebagai berikut

$$AU = \Phi(-v + c) \tag{17}$$

Nilai BP dinyatakan sebagai berikut

$$BP = \Phi|v + c| \tag{18}$$

di mana

- $\Phi$  : fungsi normal standar.

**2.9. Pendidikan**

Berdasarkan Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, pendidikan merupakan upaya yang dilakukan dengan sengaja dan terstruktur untuk menciptakan lingkungan belajar serta proses pembelajaran. Tujuannya adalah agar peserta didik dapat secara aktif mengembangkan potensi diri, sehingga mereka memiliki pengetahuan spiritual keagamaan, kemampuan mengendalikan diri, kecerdasan,

akhlak yang baik, serta keterampilan yang diperlukan untuk dirinya sendiri, masyarakat, bangsa, dan negara (Nuzleha, 2021). Pendidikan adalah bentuk investasi yang bertujuan untuk meningkatkan kualitas sumber daya manusia dan mempersiapkan tenaga kerja agar semakin berkualitas. Selain itu, pendidikan dapat berkontribusi pada pertumbuhan pendapatan nasional serta kesejahteraan masyarakat. Untuk menilai kualitas pendidikan yang ada, dibutuhkan indikator pendidikan (Mulyadi, 2017). Indikator pendidikan berfungsi sebagai alat ukur untuk mengevaluasi sejauh mana mutu pendidikan yang tersedia. Pentingnya indikator ini muncul karena adanya berbagai masalah dalam pendidikan, seperti tingginya angka putus sekolah, ketidakesetaraan dalam akses pendidikan, dan kualitas pendidikan yang perlu memenuhi tuntutan lapangan kerja yang semakin kompetitif (Mulyadi, 2017).

**3. Metodologi**

Variabel penelitian yang digunakan adalah jumlah perpustakaan sekolah, jumlah laboratorium, angka partisipasi sekolah, angka partisipasi murni, angka partisipasi kasar, angka mengulang, angka putus sekolah, rasio siswa per guru, dan jumlah ruang kelas dalam kondisi baik. Teknik penentuan sampel yang digunakan dalam penelitian ini adalah *purposive sampling*. Metode pengumpulan data dalam penelitian ini dengan cara mengambil data sekunder melalui laman resmi Badan Pusat Statistik Provinsi dan laman resmi Statistik Pendidikan yaitu <https://statistik.data.kemendikbud.go.id/>. Penelitian ini dilaksanakan pada bulan Maret 2024 sampai dengan Juni 2024. Berikut ilustrasi analisis data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis statistika deskriptif terhadap variabel penelitian
2. Melakukan standarisasi data dengan menggunakan persamaan (1)
3. Melakukan pendeteksian multikolinieritas dengan menggunakan persamaan (2)
4. Membuat kluster dengan menggunakan metode *hierarchical clustering average linkage* untuk mengelompokkan provinsi berdasarkan indikator pendidikan dengan langkah-langkah sebagai berikut:
  - a. Membuat matriks jarak berdasarkan persamaan (4)
  - b. Menghitung jarak kluster gabungan berdasarkan jarak terdekat berdasarkan persamaan (10)
  - c. Membuat matriks jarak terbaru berdasarkan perhitungan sebelumnya
  - d. Mengulangi langkah (b) dan (c) hingga terbentuk kluster hasil pengelompokan
  - e. Menginterpretasi hasil kluster
5. Melakukan teknik *multiscale bootstrap* dengan tahap sebagai berikut:
  - a. Melakukan *resampling bootstrap* sebanyak  $B = 500$
  - b. Menghitung nilai  $N'$  berdasarkan persamaan (12) hingga (15)
  - c. Menghitung nilai AU berdasarkan persamaan (17) dan menghitung nilai BP berdasarkan persamaan (18)
6. Melakukan interpretasi dengan melihat hasil analisis yang dilakukan, kemudian menarik kesimpulan dari hasil analisis yang telah dilakukan.

**4. Hasil dan Pembahasan**

**4.1. Deskripsi Data**

Data yang digunakan merupakan data 9 indikator pendidikan di 34 provinsi di Indonesia pada tahun 2021 yang dinotasikan sebagai  $x_k$ . Untuk melihat hasil analisis statistika deskriptif dari data yang diperoleh dapat dilihat pada Tabel 3.

**Tabel 3.** Statistika deskriptif data variabel pengamatan

| Variabel Penelitian | Rata-Rata | Minimum | Maksimum | Simpangan Baku |
|---------------------|-----------|---------|----------|----------------|
| $X_1$               | 432,32    | 63      | 1.682    | 379,277        |
| $X_2$               | 126,29    | 6       | 441      | 109,124        |
| $X_3$               | 74,97     | 64,83   | 88,95    | 5,954          |
| $X_4$               | 62,32     | 44,73   | 73,45    | 6,162          |
| $X_5$               | 87,08     | 73,35   | 98,31    | 6,103          |
| $X_6$               | 310,38    | 3       | 1162     | 285,889        |
| $X_7$               | 414,76    | 8       | 3416     | 705,060        |
| $X_8$               | 14,04     | 8,891   | 18,265   | 2,030          |
| $X_9$               | 3.270     | 350     | 15.495   | 3312,229       |

Berdasarkan Tabel 3. terlihat bahwa jumlah perpustakaan sekolah tertinggi terdapat pada Provinsi Jawa Barat sebesar 1.682 perpustakaan sekolah, dan jumlah terendah terdapat pada Provinsi Kalimantan Utara sebesar 63 perpustakaan sekolah. Jumlah laboratorium sekolah tertinggi terdapat pada Provinsi Sumatera sebesar 441 laboratorium, dan jumlah terendah terdapat pada Provinsi Gorontalo sebesar 6 laboratorium sekolah. Angka

partisipasi sekolah tertinggi terdapat pada Provinsi D.I. Yogyakarta sebesar 88,95% dan angka partisipasi sekolah terendah terdapat pada Provinsi Papua sebesar 64,83%. Angka partisipasi murni tertinggi terdapat pada Provinsi Kepulauan Riau sebesar 73,45% dan angka partisipasi murni terendah terdapat pada Provinsi Papua sebesar 44,73%. Angka partisipasi kasar tertinggi terdapat pada Provinsi Kalimantan Utara sebesar 98,31% dan angka partisipasi kasar terendah terdapat pada Provinsi Banten sebesar 73,35%. Jumlah angka mengulang tertinggi terdapat pada Provinsi Papua sebesar 1.162 kasus mengulang dan angka mengulang terendah terdapat pada Provinsi Bali sebesar 3 kasus mengulang. Jumlah angka putus sekolah tertinggi terdapat pada Provinsi Sumatera Utara sebesar 3.416 kasus putus sekolah dan angka putus sekolah terendah terdapat pada Provinsi D.I. Yogyakarta sebesar 8 kasus putus sekolah. Rasio siswa per guru tertinggi terdapat pada Provinsi Jawa Barat sebesar 18,265 dan rasio siswa per guru terendah terdapat pada Provinsi Aceh sebesar 8,891 siswa per guru. Jumlah ruang kelas dalam kondisi baik tertinggi terdapat pada Provinsi Jawa Barat sebesar 15.495 dan jumlah ruang kelas dalam kondisi baik terendah terdapat pada Provinsi Kalimantan Utara sebesar 350 ruangan dalam kondisi baik.

**4.2. Metode Average Linkage**

Jarak pada metode *average linkage* dinyatakan sebagai jarak rata-rata dari semua anggota data pengamatan dalam satu kluster dengan semua anggota data pengamatan dalam kluster yang lain. Tahap pertama yang dilakukan pada metode ini yaitu mencari jarak minimum antar data pengamatan sebagai berikut:

1. Sebelum melakukan pengelompokan menggunakan *average linkage*, dilakukan perhitungan matriks jarak antar data dengan menggunakan jarak Euclid, dan karena telah dilakukan proses standardisasi data, simbol variabel dinotasikan sebagai  $Z_k$ . Contoh perhitungan jarak antara data pengamatan 1 dan 2 adalah sebagai berikut:

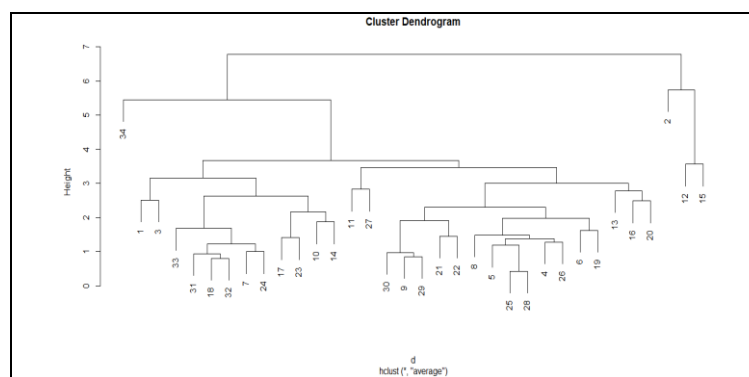
$$d_{12} = \sqrt{\sum_{k=2}^9 (Z_{1k} - Z_{2k})^2} = 6,654$$

Melakukan perhitungan matriks jarak dengan cara yang sama terhadap data pengamatan 1 dan 3, data pengamatan 1 dan 4, dan seterusnya

2. Menentukan jarak terkecil dari matriks jarak
3. Menghitung jarak gabungan menggunakan metode *average linkage* sebagai berikut:

$$D(1(25,28)) = \frac{D_{(25,1)} + D_{(28,1)}}{2 \times 1} = \frac{3,439 + 3,328}{2} = 3,383$$

4. Membuat matriks jarak terbaru
5. Mengulangi langkah (2) hingga langkah (4) hingga terbentuk kluster hasil pengelompokan. Untuk memperjelas proses penggabungan data pengamatan satu per satu dapat digambarkan pada dendrogram berikut:



**Gambar 1.** Dendrogram metode average linkage.

Tahap selanjutnya setelah kluster terbentuk adalah memberi ciri spesifik kluster untuk menggambarkan anggota dari masing-masing kluster yang terbentuk dengan melihat nilai rata-rata variabel dari masing-masing kluster pada Tabel 4.

**Tabel 4.** Anggota Kluster yang terbentuk

| Notasi | Variabel            | Kluster |    |     |    |
|--------|---------------------|---------|----|-----|----|
|        |                     | 1       | 2  | 3   | 4  |
| $Z_2$  | Jumlah Laboratorium | 416     | 98 | 441 | 80 |

| Notasi         | Variabel                              | Klaster |        |        |        |
|----------------|---------------------------------------|---------|--------|--------|--------|
|                |                                       | 1       | 2      | 3      | 4      |
| Z <sub>3</sub> | Angka Partisipasi Sekolah             | 70,395  | 75,509 | 78,21  | 64,830 |
| Z <sub>4</sub> | Angka Partisipasi Murni               | 60      | 62,868 | 68     | 44,730 |
| Z <sub>5</sub> | Angka Partisipasi Kasar               | 81,750  | 87,537 | 94,68  | 76,550 |
| Z <sub>6</sub> | Angka Mengulang                       | 744     | 254    | 273    | 1162   |
| Z <sub>7</sub> | Angka Putus Sekolah                   | 820     | 288    | 3416   | 407    |
| Z <sub>8</sub> | Rasio Siswa per guru                  | 17      | 13,749 | 15,879 | 14,320 |
| Z <sub>9</sub> | Jumlah ruang kelas dalam kondisi baik | 13633   | 2502   | 7712   | 1164   |

Berdasarkan Tabel 4 dapat diperoleh karakteristik dari empat klaster yang terbentuk sebagai berikut:

1. Provinsi pada klaster 3 memiliki nilai jumlah laboratorium (Z<sub>2</sub>), angka partisipasi sekolah (Z<sub>3</sub>), angka partisipasi murni (Z<sub>4</sub>), angka partisipasi kasar (Z<sub>5</sub>), angka putus sekolah (Z<sub>7</sub>) yang lebih tinggi dibandingkan klaster 1, 2, dan 4.
2. Provinsi pada klaster 4 memiliki nilai angka mengulang (Z<sub>6</sub>) yang lebih tinggi dibandingkan klaster 1, 2, dan 3.
3. Provinsi pada klaster 1 memiliki nilai rasio siswa per guru (Z<sub>8</sub>), dan jumlah ruang kelas dalam kondisi baik (Z<sub>9</sub>) yang lebih tinggi dibandingkan klaster 2, 3, dan 4.

**4.3. Metode Multiscale Bootstrap**

Metode *multiscale bootstrap* ini digunakan untuk mengetahui keakuratan hasil dari analisis klaster sebelumnya. Suatu klaster dapat dianggap akurat jika nilai  $AU \geq 95$ . Berikut merupakan tahapan perhitungan *multiscale bootstrap*

1. Melakukan *resampling bootstrap* dengan mencari nilai  $N'$  menggunakan persamaan (12) hingga (15)
2. Menghitung nilai  $z(N')$ , nilai  $v$  dan nilai  $c$  dengan persamaan (16). Hasil perhitungan nilai  $v$  dan  $c$  dapat dilihat pada Tabel 5.
3. Menghitung nilai  $AU$  dan  $BP$  menggunakan persamaan (17) dan (18), berikut merupakan contoh menghitung nilai  $AU$  dan  $BP$

$$\begin{aligned}
 AU &= \phi(-(-2,695) + 0,689) \\
 &= \phi(3,384) \\
 &= 0,996 \text{ (dilihat pada tabel Z)}.
 \end{aligned}$$

Nilai  $AU$  sebesar  $0,996 \geq 0,95$  yang menyatakan bahwa klaster pada iterasi pertamam valid. Selanjutnya adalah menghitung nilai  $BP$  dengan cara berikut

$$\begin{aligned}
 BP &= \phi|(-2,695 + 0,689)| \\
 &= \phi|-2,006| \\
 &= 0,977 \text{ (diperoleh dari tabel Z)}.
 \end{aligned}$$

Nilai  $BP$  sebesar 0,977 menyatakan bahwa peluang klaster pada iterasi pertama untuk menjadi klaster yang valid adalah sebesar 0,977 atau 97%. Berikut merupakan hasil dari perhitungan secara lengkap pada Tabel 5.

**Tabel 5.** Hasil metode *multiscale bootstrap*

| Iterasi | AU    | BP    | SE.AU | SE.BP | Nilai v | Nilai c |
|---------|-------|-------|-------|-------|---------|---------|
| 1       | 1,000 | 0,977 | 0,000 | 0,003 | -2,695  | 0,689   |
| 2       | 0,838 | 0,371 | 0,028 | 0,008 | -0,329  | 0,658   |
| 3       | 0,843 | 0,445 | 0,026 | 0,008 | -0,447  | 0,560   |
| 4       | 0,937 | 0,299 | 0,016 | 0,008 | -0,500  | 1,028   |
| 5       | 0,985 | 0,681 | 0,004 | 0,008 | -1,317  | 0,848   |
| 6       | 0,759 | 0,274 | 0,037 | 0,007 | -0,051  | 0,651   |
| 7       | 0,709 | 0,239 | 0,041 | 0,007 | 0,080   | 0,629   |
| 8       | 0,993 | 0,331 | 0,003 | 0,008 | -1,001  | 1,439   |
| 9       | 0,787 | 0,205 | 0,038 | 0,007 | 0,015   | 0,810   |
| 10      | 0,887 | 0,122 | 0,033 | 0,005 | 0,003   | 1,161   |
| 11      | 0,955 | 0,563 | 0,010 | 0,008 | -0,927  | 0,786   |
| 12      | 0,859 | 0,239 | 0,028 | 0,007 | -0,183  | 0,891   |
| 13      | 0,971 | 0,149 | 0,101 | 0,006 | -0,431  | 1,472   |
| 14      | 0,665 | 0,195 | 0,045 | 0,007 | 0,217   | 0,644   |
| 15      | 0,972 | 0,499 | 0,007 | 0,008 | -0,891  | 1,019   |
| 16      | 0,823 | 0,154 | 0,036 | 0,006 | 0,046   | 0,973   |

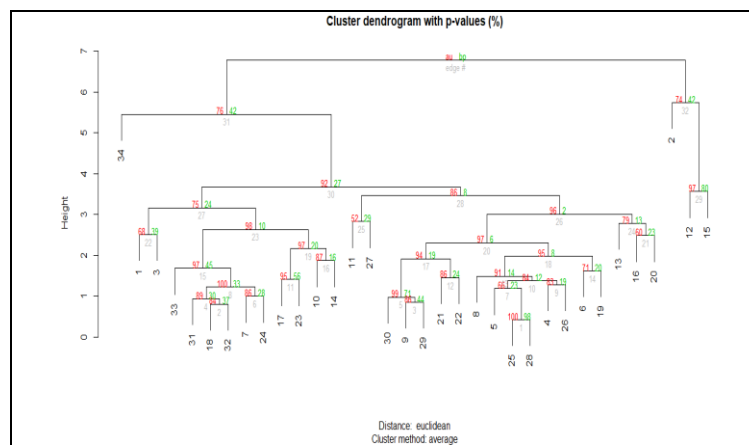
| Iterasi | AU    | BP    | SE.AU | SE.BP | Nilai v | Nilai c |
|---------|-------|-------|-------|-------|---------|---------|
| 17      | 0,934 | 0,183 | 0,018 | 0,006 | -0,299  | 1,204   |
| 18      | 0,985 | 0,079 | 0,008 | 0,004 | -0,383  | 1,796   |
| 19      | 0,885 | 0,182 | 0,028 | 0,006 | -0,146  | 1,054   |
| 20      | 0,969 | 0,064 | 0,016 | 0,004 | -0,168  | 1,694   |
| 21      | 0,701 | 0,238 | 0,042 | 0,007 | 0,092   | 0,621   |
| 22      | 0,591 | 0,366 | 0,042 | 0,008 | 0,057   | 0,286   |
| 23      | 0,936 | 0,096 | 0,025 | 0,005 | -0,111  | 1,413   |
| 26      | 0,980 | 0,019 | 0,022 | 0,002 | 0,020   | 2,065   |
| 24      | 0,734 | 0,131 | 0,047 | 0,006 | 0,284   | 0,872   |
| 25      | 0,515 | 0,286 | 0,045 | 0,008 | 0,264   | 0,302   |
| 26      | 0,980 | 0,019 | 0,022 | 0,002 | 0,020   | 2,065   |
| 27      | 0,756 | 0,240 | 0,038 | 0,007 | 0,005   | 0,700   |
| 28      | 0,886 | 0,077 | 0,038 | 0,004 | 0,111   | 1,315   |
| 29      | 0,968 | 0,803 | 0,008 | 0,007 | -1,353  | 0,500   |
| 30      | 0,921 | 0,274 | 0,018 | 0,007 | -0,404  | 1,006   |
| 31      | 0,765 | 0,417 | 0,033 | 0,008 | -0,255  | 0,466   |
| 32      | 0,738 | 0,419 | 0,035 | 0,008 | -0,217  | 0,422   |
| 33      | 1,000 | 1,000 | 0,000 | 0,000 | 0,000   | 0,000   |

Keterangan:

SE.AU : *Standard Error Approximately Bootstrap*

SE.BP : *Standard Error Bootstrap Probability.*

Hasil dari metode *multiscale bootstrap* menggunakan *software R* dapat dilihat pada Gambar 2.



Gambar 2. Dendrogram *Multiscale bootstrap*

Berdasarkan hasil uji validitas dengan metode *multiscale bootstrap* pada Gambar 2. dapat diketahui bahwa nilai AU pada klaster pertama diperoleh pada iterasi ke-29 yaitu sebesar 0,97, sehingga menandakan bahwa klaster tersebut telah akurat sehingga dapat menggambarkan populasi yang sebenarnya. Pada klaster kedua, nilai AU diperoleh pada iterasi ke-30 dengan nilai 0,92 yang menandakan bahwa klaster kedua tidak valid sehingga tidak menggambarkan populasi yang sebenarnya. Pada klaster ketiga, nilai AU yang diperoleh pada iterasi ke-32 sebesar 0,73 yang menandakan bahwa klaster ketiga tidak valid sehingga tidak menggambarkan populasi yang sebenarnya. Pada klaster keempat, nilai AU yang diperoleh pada iterasi ke-31 sebesar 0,76 yang menandakan bahwa klaster tersebut tidak valid sehingga tidak menggambarkan populasi yang sebenarnya.

### 5. Kesimpulan

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa klaster yang terbentuk dengan menggunakan metode *average linkage* berdasarkan indikator pendidikan di Indonesia adalah 4 klaster, di mana pada klaster pertama beranggotakan Provinsi Jawa Barat dan Jawa Timur yang memiliki rasio jumlah siswa per guru dan jumlah ruang kelas dalam kondisi baik yang lebih tinggi dibandingkan klaster lainnya. Klaster kedua beranggotakan Provinsi Aceh, 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 Selatan, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo,



Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat yang memiliki jumlah laboratorium, APS, APM, APK, dan jumlah ruang kelas dalam kondisi baik yang lebih tinggi dibandingkan dengan klaster 4, namun lebih rendah dibandingkan dengan klaster 3. Klaster 3 beranggotakan Provinsi Sumatera Utara yang memiliki jumlah laboratorium, angka partisipasi sekolah, angka partisipasi kasar, angka partisipasi murni, dan angka putus sekolah yang lebih tinggi dibandingkan klaster lainnya. Klaster keempat memiliki anggota Provinsi Papua yang memiliki angka mengulang yang lebih tinggi dibandingkan klaster lainnya.

Validitas dari klaster yang terbentuk dengan menggunakan metode *multiscale bootstrap* adalah satu klaster valid yang menandakan bahwa klaster tersebut dapat menggambarkan populasi yang sebenarnya, dan karena hanya satu klaster yang valid, dapat disimpulkan bahwa metode yang digunakan belum optimal untuk mengelompokkan indikator pendidikan di Indonesia.

## 6. Daftar Pustaka

- Anuraga, G. 2015. Hierarchical Clustering Multiscale Bootstrap Untuk Pengelompokan Kemiskinan di Jawa Timur. *Jurnal Statistika Universitas Muhammadiyah Semarang*. 3(1), 27-33.
- Efron, B. dan Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. New York: Chapman Hall
- Kuncoro, M. 2003. *Metode Riset Untuk Bisnis Dan Ekonomi*. Jakarta: Erlangga.
- Laeli, S. 2014. Analisis Klaster dengan Average Linkage dan Wards Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link. Yogyakarta: Universitas Negeri Yogyakarta.
- Mulyadi, S. 2017. *Ekonomi Sumber Daya Manusia dan Perspektif Pembangunan*. Jakarta: PT Raja Grafindo Persada.
- Nachrowi, N. D., & Usman, H. 2006. *Pendekatan Populer dan Praktis Ekonometrika Untuk Analisis Ekonomi dan Keuangan*. Jakarta: Fakultas Ekonomi Universitas Indonesia.
- Nuzleha. 2021. Analisis Tingkat Pendidikan Terhadap Kinerja Pegawai Dinas Kependudukan dan Pencatatan Sipil Provinsi Lampung. *MOTIVASI: Jurnal Manajemen dan Bisnis*. 6(2), 117-125.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Penerbit ANDI.
- 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). *Jurnal EKSPONENSIAL*. 9(1), 1-10.
- Rencher, A. 2002. *Method of Multivariate Analysis Second Edition*. Canada: John Wiley & Sons.
- Safarina, Y., Hayati, M. N., Nasution, Y. N. 2019. Penerapan Metode Hierarchical Klastering Multiscale Bootstrap (Studi Kasus: Indeks Pembangunan Manusia (IPM) di Kalimantan Timur Tahun 2017). *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*. 29-35.
- Shimodaira, H. 2004. Approximately Unbiased Test of Regions Using Multistep-Multiscale Bootstrap Resampling. *The Annals of Statistics*. 32(6), 2616-2641.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- Widarjono, A. 2002. *Analisis Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.
- Yulianto, S & Hidayatullah, K. H. 2014. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jateng Berdasarkan Indikator Kesejahteraan Rakyat Semarang. *Jurnal Akademik Statistika Muhammadiyah Semarang*. 2(1), 58-59.