

# **Perbandingan Analisis *Agglomerative Hierarchical Clustering* Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat**

## **(Studi Komparatif Penentuan Kelompok Kabupaten/Kota Menggunakan Metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward*)**

Asma' Khoirunnisa<sup>1</sup>, Firdaus Amruzain Satiranandi Wibowo<sup>2</sup>, Kismiantini<sup>3</sup>  
Universitas Negeri Yogyakarta<sup>1,2,3</sup>  
[asmakhoirunnisa.2021@student.uny.ac.id](mailto:asmakhoirunnisa.2021@student.uny.ac.id)

**Abstrak**—Bidang pendidikan pada saat ini merupakan salah satu prioritas yang menjadi urgensi pemerintah. Berdasarkan Badan Pusat Statistik 2021, tercatat bahwa angka putus sekolah di Provinsi Jawa Barat menduduki peringkat pertama dari 34 provinsi di Indonesia. Fenomena ini tentunya menurunkan indeks pendidikan serta bertolak belakang dengan intensi penting yang menjadi fokus pembangunan pemerintah yakni pemerataan pendidikan. Pada penelitian ini akan dilakukan pemetaan wilayah di Provinsi Jawa Barat berdasarkan indeks pendidikan sehingga didapatkan perencanaan dan kebijakan pemerintah yang lebih akurat. Salah satu algoritma yang dapat digunakan dalam pengelompokan kabupaten berdasarkan indikator pendidikan adalah *Agglomerative Hierarchical Clustering* (AHC). Beberapa metode pengelompokan yang digunakan pada penelitian ini adalah AHC dengan metode *single linkage*, metode *complete linkage*, metode *average linkage*, dan metode *ward*. Hasil dari penelitian ini adalah perbandingan metode yang dilakukan berdasarkan nilai koefisien korelasi *cophenetic*. Berdasarkan analisis pada penelitian ini, diketahui gambaran pada sektor pendidikan di Provinsi Jawa Barat tahun 2021, yaitu masih terdapat indikasi belum tercapainya pemerataan pendidikan khususnya pada jenjang sekolah menengah atas hingga perguruan tinggi serta terpusatnya jumlah sekolah dan guru di klaster tertentu. Tinjauan indeks pendidikan ini dibentuk menjadi 4 klaster pada setiap masing-masing metode. Berdasarkan hasil perbandingan nilai koefisien korelasi *cophenetic* dari keempat metode pada analisis AHC, metode *ward* adalah metode terbaik dengan koefisien korelasi *cophenetic* terbesar (0,952).

**Kata kunci:** *Indeks Pendidikan, Analisis Cluster, Agglomerative Hierarchy Clustering*

### **I. PENDAHULUAN**

Pendidikan di Indonesia telah diatur pada Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Pendidikan adalah upaya sadar dan terencana dalam mewujudkan suasana belajar dan proses pembelajaran agar peserta didik dapat secara aktif mengembangkan potensi dalam dirinya untuk menyandang kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dalam dirinya dan masyarakat (Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional). Melalui pendidikan, kualitas sumber daya manusia yang akan menjadi sebuah kunci pertumbuhan negara dapat dengan mudah dikendalikan. Akan tetapi, masa pandemi menjadi tantangan besar yang telah mengubah situasi ideal. Terjadinya pandemi COVID-19 memaksa instansi pendidikan melakukan perubahan metode pembelajaran [1]. Pandemi COVID-19 telah menciptakan tantangan berikutnya dalam dunia pendidikan Indonesia karena pemerintah mengeluarkan kebijakan untuk menghilangkan pembelajaran tatap muka sementara. [2]

Mengingat pentingnya peran pendidikan terhadap kemajuan bangsa, maka dirasa perlu untuk melakukan pengukuran dan kalkulasi terhadap beberapa hal yang menjadi indikator pendidikan pada tahun 2021 untuk melihat sejauh mana kemajuan pemerataan pendidikan yang ada, utamanya di Provinsi Jawa Barat. Apabila ditinjau berdasarkan indeks pendidikan, Provinsi Jawa Barat memerlukan perhatian lebih mengingat beberapa waktu ini terjadi peningkatan angka pada jumlah siswa yang tidak melanjutkan pendidikan sekolahnya dikarenakan akses yang sulit untuk dijangkau [3]. Peristiwa ini tentu berimbas dengan menurunnya tingkat sumber daya manusia (SDM) yang ada di Provinsi Jawa Barat dan beragam

masalah turunan lainnya. Untuk mengentas permasalahan SDM di Indonesia ini, transformasi pendidikan merupakan suatu keniscayaan karena dengan sasaran ini, pendidikan di Indonesia seutuhnya dapat terlaksana. Penurunan jumlah sekolah di semua jenjang pada tahun ajaran 2020/2021 secara umum dapat terjadi di sekolah swasta terutama pada jenjang pendidikan menengah [4]. Beberapa wilayah atau kawasan bahkan masih terdapat pengetahuan yang minim dikarenakan kurangnya hal dalam kesadaran akan pentingnya pendidikan. Oleh sebab itu, penulis melihat adanya urgensi dalam memodelkan pengelompokan wilayah berdasarkan indeks pendidikan yang ada.

Pengelompokan wilayah berdasarkan indeks pendidikan di Provinsi Jawa Barat pernah dilakukan oleh Alfianti pada tahun 2021 [3] tetapi menggunakan algoritma *K-means* dengan variabel jumlah lembaga sekolah dasar dan menengah di Provinsi Jawa Barat. Dalam penelitian ini, penulis akan melakukan studi komparasi pengelompokan wilayah berdasarkan indeks pendidikan di Provinsi Jawa Barat dengan menggunakan *Agglomerative Hierarchy Clustering (Single Linkage, Complete Linkage, Average Linkage, dan Ward)* untuk mengetahui wilayah dengan tingkat pendidikan yang rendah tentunya membutuhkan atensi khusus oleh pemerintah khususnya instansi berupa Dinas Pendidikan Provinsi Jawa Barat. Peneliti menggunakan AHC daripada *non-hierarchical clustering* seperti K-means karena metode pengelompokan *hierarchical* tidak memerlukan pengetahuan apriori tentang jumlah cluster atau partisi awal sehingga ini merupakan keuntungan yang pasti dibandingkan metode *non-hierarchical* [5]. Studi komparatif permodelan *Agglomerative Hierarchy Clustering (AHC)* ini dilakukan untuk mengetahui sebaran tingkat pendidikan masyarakat dalam rangka membantu pembuatan kebijakan.

## II. METODE PENELITIAN

Analisis *cluster* merupakan teknik multivariat dengan tujuan mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang dimilikinya. Algoritma-algoritma *clustering* sering dimanfaatkan untuk kompresi data dan konstruksi model, disamping mengklasifikasikan data. Metode *agglomerative hierarchical clustering (AHC)* adalah metode analisis *cluster* dengan tujuan mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Hal yang menjadi pembeda dalam algoritma ini adalah tidak ditentukannya jumlah *cluster* di awal [6]. Metode AHC terbagi menjadi beberapa algoritma, di antaranya metode *single linkage, complete linkage, average linkage, dan ward* [7].

### A. Data Penelitian

Data penelitian adalah data sekunder yang diperoleh dari publikasi Provinsi Jawa Barat dalam studi kasus berdasarkan data dari BPS Provinsi Jawa Barat. Sejumlah 27 kabupaten/kota di Provinsi Jawa Barat digunakan dalam penelitian ini.

### B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel yang berfungsi pada pengelompokan kabupaten/kota berdasarkan indikator pendidikan di Provinsi Jawa Barat. Adapun indikator-indikatornya dapat dilihat pada Tabel 1 dengan total variabel penelitian sejumlah 21.

TABEL 1. VARIABEL PENELITIAN

Variabel	Keterangan	Deskripsi
$X_1$	APK SD/MI/Paket A	Angka Partisipasi Kasar
$X_2$	APK SMP/MTS/Paket B	
$X_3$	APK SMA/MA/Paket C	
$X_4$	APK Perguruan Tinggi	
$X_5$	APM SD/MI/Paket A	Angka Partisipasi Murni
$X_6$	APM SMP/MTS/Paket B	
$X_7$	APM SMA/MA/Paket C	
$X_8$	APM Perguruan Tinggi	
$X_9$	APS 7-12	Angka Partisipasi Sekolah
$X_{10}$	APS 13-15	
$X_{11}$	APS 16-18	
$X_{12}$	APS 19-24	

Variabel	Keterangan	Deskripsi
$X_{13}$	Jumlah SD	Negeri dan Swasta
$X_{14}$	Jumlah SMP	
$X_{15}$	Jumlah SMA	
$X_{16}$	Jumlah Murid SD	Negeri dan Swasta
$X_{17}$	Jumlah Murid SMP	
$X_{18}$	Jumlah Murid SMA	
$X_{19}$	Jumlah Guru SD	Negeri dan Swasta
$X_{20}$	Jumlah Guru SMP	
$X_{21}$	Jumlah Guru SMA	

### C. Teknik Analisis Data

#### 1. Analisis Multivariat

Analisis multivariat adalah analisis statistik yang digunakan untuk menganalisis data yang terdiri dari beberapa variabel dan variabel-variabel tersebut saling berkorelasi satu sama lain. Data dalam analisis multivariat dapat dinyatakan dalam bentuk matriks dimana terdapat  $n$  objek dan  $p$  variabel [5]. Misalkan pada suatu pengamatan terdapat  $p$  variabel yaitu  $X = (X_1, X_2, \dots, X_p)$  dari banyaknya  $n$  objek, dalam hal ini data dapat direpresentasikan dalam bentuk matriks seperti berikut:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nj} & \dots & X_{np} \end{bmatrix} \quad (1)$$

#### 2. Clustering

Analisis *cluster* adalah teknik analisis multivariat dengan tujuan mengelompokkan data observasi atau variabel pengamatan ke dalam cluster sedemikian sehingga masing-masing bersifat homogen. Proses pembentukan *cluster* dapat dilakukan dengan dua cara, yaitu dengan metode *hierarchical* dan *non-hierarchical*. Pada metode hierarki terdiri dari metode *agglomerative* dan metode defisit. Metode *agglomerative* sendiri terdiri dari 3 metode, yaitu metode *linkage*, metode *variance*, dan metode *centroid*, dimana *linkage* terdiri dari metode *single linkage*, *complete linkage*, dan *average linkage*. Sedangkan pada metode *variance* terdiri dari metode *Ward*. Metode *non-hierarchical* terdiri dari 3 metode yaitu, metode *sequential threshold*, metode *parallel*, dan metode *optimizing partitioning* [8].

#### 3. Metode Hierarki (*Hierarchical Clustering*)

Metode *hierarchical* adalah metode analisis *cluster* yang membentuk level tertentu seperti pada struktur pohon dikarenakan prosesnya yang berlangsung secara bertahap. Metode pengelompokan *hierarchical* tidak memerlukan pengetahuan apriori tentang jumlah *cluster* atau partisi awal sehingga ini merupakan keuntungan yang pasti dibandingkan metode *non-hierarchical* [5]. Hasil pengklasteran dengan metode *hierarchical* dapat disajikan dalam bentuk dendrogram. Dendrogram adalah representasi visual dari setiap langkah dari analisis *cluster* dengan nilai koefisien jarak pada setiap langkah. Angka di sebelah kanan adalah objek penelitian, di mana objek-objek tersebut dihubungkan oleh garis dengan objek yang lain sehingga pada akhirnya akan membentuk satu [8]. Tahap-tahap pengklusteran data dengan menggunakan metode *hierarchical* adalah [9]:

- a. Mulai dengan  $N$  klaster, masing-masing berisi satu entitas dan matriks simetris  $N \times N$  jarak (atau kesamaan)  $D = \{d_{ik}\}$ .
- b. Mencari matriks jarak untuk pasangan cluster terdekat (paling serupa). Misalkan jarak antara klaster "paling serupa"  $U$  dan  $V$  menjadi  $d_{UV}$ .
- c. Gabungkan klaster  $U$  dan  $V$ . Beri label  $(UV)$  pada klaster yang baru terbentuk. Perbarui entri pada matriks jarak dengan :
  - (a) menghapus baris dan kolom yang sesuai dengan klaster  $U$  dan  $V$ ; dan
  - (b) menambahkan baris dan juga kolom yang memberikan jarak antara klaster  $(UV)$  dengan cluster yang tersisa.
- d. Mengulangi langkah-langkah b dan c sebanyak  $N - 1$  kali. Berdasarkan bahwa semua objek akan berada dalam satu klaster setelah algoritma berakhir. Mencatat identitas tiap klaster yang digabungkan dan level (jarak atau similiaritas) tempat dimana fusi berlangsung.

Metode-metode yang bisa digunakan dalam metode hierarki adalah metode *agglomerative (agglomerative method)* dan metode defisif (*divisive method*). Beberapa *Hierarchical Clustering* yaitu:

a. Metode *Single Linkage*

Disebutkan dalam Buku *Applied Multivariate Techniques*, langkah awal yang harus dilakukan pada metode ini adalah menemukan jarak terpendek dalam  $D = \{d_{ik}\}$  kemudian menggabungkan objek-objek yang bersesuaian misalnya,  $U$  dan  $V$ , untuk mendapatkan *cluster (UV)* [5]. Selanjutnya, jarak antara  $(UV)$  dan *cluster W* yang lain dihitung berdasarkan formulasi jarak *euclidean* seperti berikut:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \tag{2}$$

dengan :

$d_{UW}$  adalah jarak terpendek antara *cluster-cluster U* dan  $W$

$d_{VW}$  adalah jarak terpendek antara *cluster-cluster V* dan  $W$

b. Metode *Complete Linkage*

Metode *complete linkage* didasarkan pada jarak *euclidean* maksimum. Jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan obyek yang mempunyai jarak terjauh. Pada awal perhitungan, terlebih dahulu mencari nilai minimum dalam  $D = \{d_{ij}\}$  dan menggabungkan obyek-obyek yang bersesuaian, misalnya  $U$  dan  $V$ , untuk mendapatkan *cluster (UV)*. Pada langkah (c) dari algoritma yang dijelaskan sebelumnya, jarak antara  $(UV)$  dan *cluster* lain  $W$ , dihitung dengan cara [6]:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \tag{3}$$

dengan:

$d_{UW}$  adalah jarak terpendek antara *cluster-cluster U* dan  $W$

$d_{VW}$  adalah jarak terpendek antara *cluster-cluster V* dan  $W$

c. Metode *Average Linkage*

*Average linkage* memperlakukan jarak *euclidean* antara dua *cluster* sebagai jarak rata-rata antara semua pasangan objek ketika salah satu anggota dari pasangan tersebut kepunyaan tiap *cluster*. Langkah ini dimulai dengan mencari matriks jarak  $D = \{d_{ik}\}$  untuk memperoleh objek-objek paling dekat (paling mirip) misalnya  $U$  dan  $V$ . Objek objek ini digabungkan untuk membentuk *cluster (UV)*. Untuk langkah ketiga dari algoritma di atas jarak-jarak antara  $(UV)$  dan *cluster W* yang lain ditentukan oleh [5]:

$$d_{(UV)W} = \frac{\sum_a \sum_b d_{ab}}{N_{UV}N_W} \tag{4}$$

dengan:

$d_{(UV)W}$  adalah jarak *euclidean* antara *cluster (UV)* dan *cluster W*

$d_{ab}$  adalah jarak antara objek  $a$  pada *cluster* ( $UV$ ) dan objek  $b$  pada *cluster*  $W$   
 $N_{UV}$  adalah jumlah item pada *cluster* ( $UV$ )  
 $N_W$  adalah jumlah item pada *cluster* ( $UV$ ) dan  $W$

d. Metode *Ward*

Metode *Ward* merupakan suatu metode pembentukan *cluster* yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada mean *cluster* untuk setiap pengamatan. *Error sum of squares* (SSE) digunakan sebagai fungsi objektif. Dua objek akan digabungkan jika mempunyai fungsi objektif terkecil di antara kemungkinan yang ada [5].

$$SSE = \sum_{j=1}^p \left( \sum_{i=1}^n X_{ij}^2 - \frac{1}{n} \left( \sum_{i=1}^n X_{ij} \right)^2 \right) \quad (5)$$

dengan:

$X_{ij}$  adalah nilai untuk objek ke- $i$  pada *cluster* ke- $j$

$p$  adalah banyaknya variabel yang diukur

$n$  adalah banyaknya objek dalam *cluster* yang terbentuk

### III. HASIL DAN PEMBAHASAN

Proses pengelompokkan indeks pendidikan daerah berdasarkan pada variabel penelitian yang telah ditentukan di Provinsi Jawa Barat dengan menggunakan algoritma *agglomerative hierarchical clustering* (AHC). Pada penelitian ini terdapat 21 variabel penelitian dengan 27 Kabupaten/Kota di Provinsi Jawa Barat.

#### A. Statistika Deskriptif

Penelitian ini menggunakan basis data sekunder yang diperoleh dari publikasi BPS Provinsi Jawa Barat [4]. Tabel 2 menunjukkan *summary statistics* dari variabel-variabel yang digunakan dalam studi ini.

TABEL 2. STATISTK DESKRIPTIF

Variabel	Minimum	Mean	Maksimum	Simpangan Baku
$X_1$	99,28	104,09	110,25	2,4584
$X_2$	83,51	93,72	111,75	6,303031
$X_3$	52,97	80,63	114,29	14,50465
$X_4$	6,86	19,85	41,09	9,545617
$X_5$	94,79	97,73	99,43	1,549419
$X_6$	69,96	82,01	89,61	4,714607
$X_7$	44,53	60,59	77,18	9,051963
$X_8$	5,27	15,79	33,78	7,794899
$X_9$	98,78	99,43	99,85	0,256625
$X_{10}$	89,15	95,45	98,88	2,517466
$X_{11}$	55,52	70,02	83,76	7,399365
$X_{12}$	8,47	22,02	47,63	10,21692
$X_{13}$	86,00	727,0	977,0	467,2072
$X_{14}$	26,00	210,50	724,00	154,9011
$X_{15}$	5,00	61,74	192,00	46,41087
$X_{16}$	14792	164925	530093	117474,8
$X_{17}$	4114	27619	134053	26401,77
$X_{18}$	3243	28283	82495	19225,01
$X_{19}$	1014	7766	19848	4530,19

$X_{20}$	567	3535	9589	2093,413
$X_{21}$	193	1408	3671	912,9827

Dari Tabel 2 diperoleh jumlah minimum dan maksimum APK SD/MI/Paket A berturut-turut 99,28 dan 110,25 dengan rerata 104,09. Diketahui hal ini berarti mengindikasikan bahwa tingkat partisipasi penduduk (APK) pada jenjang SD/MI/Paket A relatif baik karena nilai rata-rata lebih dari 100% yaitu sebesar 104,9% yang berarti proporsi penduduk yang. Sedang sekolah pada jenjang derajat SD/MI/Paket A lebih banyak dibandingkan dengan jumlah penduduk yang berusia di luar batas 7 s/d 12 tahun atau dapat dikatakan bahwa tingkat partisipasi penduduk secara umum pada jenjang SD/MI/Paket A telah cukup baik. Tingkat partisipasi penduduk pada jenjang pendidikan SMP di Provinsi Jawa Barat hampir sama dengan jenjang SD/SMP/Paket A memiliki nilai standar deviasi yang cukup tinggi sebesar 14,50465, hal ini mengindikasikan bahwa pemerataan pertumbuhan pendidikan formal jenjang SMA sederajat belum merata di setiap Kabupaten/Kota pada Provinsi Jawa Barat. Nilai APK tingkat SMA tertinggi berada pada Kota Bekasi dan terendah pada Kota Tasikmalaya.

Nilai APS terendah di Kabupaten/Kota pada Provinsi Jawa Barat pada usia 0-6 tahun terdapat pada Kabupaten Cianjur sebesar 8,47%, hal ini mengindikasikan bahwa pada masyarakat Kabupaten Cianjur sebagian kecil masyarakat yang usia 0-6 tahun atau usia sekolah PAUD yang telah memanfaatkan fasilitas pendidikan. Pada APS usia 7-12 tahun memiliki nilai standar deviasi rendah yaitu sebesar 0,256% maka dapat dikatakan bahwa penduduk usia 7-12 tahun pada Provinsi Jawa Barat secara merata di Kabupaten/Kota telah memanfaatkan fasilitas pendidikan yang tersedia sehingga nilai kesenjangannya kecil. Begitu juga nilai APS usia 13-15 yang memiliki nilai standar deviasi yang kecil yaitu 2,517%. Namun, hal yang berbeda tama pada nilai APS usia 16-18 tahun dengan nilai standar deviasi sebesar 7,399365 atau kesenjangan di kabupaten/Kota Provinsi Jawa Barat pada usia tersebut cukup tinggi. Nilai APS usia 16-18 tahun tertinggi didapatkan oleh Kabupaten Garut sebesar 99,85 dan terendah 8,47 pada Kabupaten Cianjur. Dari data tersebut dapat dijelaskan bawa sebanyak 99,85% masyarakat Kabupaten Cianjur yang berusia 16-18 tahun telah mendapatkan hak atas pendidikan atau bersekolah sedangkan pada Kabupaten Cianjur masih sebesar 8,47% masyarakat yang berusia 16-18 tahun yang telah memiliki peluang untuk bersekolah.

Berdasarkan Tabel 2, nilai rata-rata APM berusia 7-12 tahun diketahui sebesar 97,73 yang berarti bahwa 98% penduduk di Kabupaten/Kota pada Provinsi Jawa Barat yang berusia 7-12 tahun telah bersekolah tepat waktu di jenjang SD/MI/Paket A. Sekitar 82% penduduk di kabupaten/Kota Provinsi Jawa Barat yang berusia 13-15 tahun telah bersekolah tepat waktu di jenjang SMP/MTS/Paket B serta sekitar 61% penduduk di Kabupaten/Kota Provinsi Jawa Bara yang berusia 16-18 tahun telah bersekolah tepat waktu di jenjang SMA/MA/Paket C. Dan untuk perguruan tinggi sekitar 14,79 yang menamakan jenjangnya. Nilai APM tertinggi pada usia 13-15 tahun yaitu sebesar 89,61% pada Kabupaten Ciamis yang berarti bahwa sekitar 90% penduduk Kabupaten Ciamis pada usia 13-15 tahun telah memanfaatkan fasilitas pendidikan pada jenjang SMP/Sederajat dan bersekolah tepat waktu sedangkan nilai terendah didapatkan pada Kabupaten Cirebon yaitu sebesar 69,96% yang berarti sekitar 70% penduduk Kabupaten Cirebon yang berusia 16-18 tahun bisa merasakan fasilitas pendidikan tepat waktu.

Pada variabel rata-rata jumlah siswa diperoleh nilai maksimum yaitu sebesar 1847,0 berada pada kabupaten Bogor yang artinya, penduduk Bogor yang berusia 15 tahun ke atas secara rata-rata jumlah siswa sebesar 1847,0 pada kelas XI atau SMA/ sederajat. Hal berbeda terjadi pada Kab Banjar dengan jumlah siswa sebesar 86,00 yang mengindikasikan bahwa jumlah siswa masyarakat di Kab Banjar masih rendah karena rata-rata jumlah penduduk yang mengindikasikan hanya tamat SD/ sederajat. Untuk variabel Jumlah Murid SD rata-rata penduduk Provinsi 164925 yang artinya murid SD rata-rata penduduk Provinsi Jawa Barat yang usia 10 tahun keatas sejumlah 164925. Hal yang menarik adalah terlihat bahwa jumlah siswa SD begitu banyak dengan nilai maksimum dan minimum secara berturut-turut adalah 530093 dan 14792. Akan tetapi jumlah sekolah dan guru yang ada masih cenderung sedikit yakni dengan nilai maksimum masing-masing senilai 977 dan 19878.

Dalam melakukan pengelompokan kabupaten/kota berdasarkan indikator Pendidikan yang terdapat di Provinsi Jawa Barat tahun 2021 digunakan analisis *Cluster*. Pada pembahasan laporan ini akan digunakan analisis AHC dengan pengujian-pengujian asumsi yang menyertainya. Berdasarkan statistik deskriptif di atas didapatkan bahwa terdapat variabel yang memiliki selisih nilai maksimum dan minimum yang terlalu tinggi, sehingga dapat menyebabkan ketimpangan pendidikan. Oleh karena itu, perlu dilakukan

pengelompokan menggunakan analisis AHC. Selain itu, variabel-variabel yang digunakan masih memiliki satuan yang berbeda maka harus dilakukan standarisasi. Langkah ini ditulis pada bagian metode dan penelitian (*Analisis Agglomerative Hierarchical Cluster (AHC)*).

B. *Analisis Agglomerative Hierarchical Cluster*

1. *Melakukan Proses Standarisasi Data*

Proses standarisasi dapat dilakukan jika pada setiap variabel penelitian terdapat perbedaan satuan pengukuran yang signifikan. Hal ini penting, karena tanpa standarisasi, titik data yang diukur menggunakan skala yang berbeda kemungkinan besar tidak akan memberikan kontribusi yang sama untuk analisis [10]. Perbedaan tersebut dapat mengakibatkan hasil pada analisis cluster tidak valid sehingga perlu melakukan proses standarisasi dengan cara mentransformasi data asli agar menjadi seragam satuannya. Mengingat data yang terkumpul mempunyai variabilitas satuan, maka perlu dilakukan langkah standarisasi atau transformasi terhadap variabel yang relevan ke bentuk z-score, sebagai berikut [10]:

$$X_{stand} = \frac{X - rata - rata(X)}{simpangan baku (X)} \quad (6)$$

dengan:

$X_{stand}$  adalah nilai variabel yang sudah di standarisasi

$X$  adalah variabel sebelum standarisasi

$rata - rata(X)$  adalah nilai rata-rata variabel  $X$

$simpangan baku (X)$  adalah nilai standar deviasi variabel  $X$

Berdasarkan pada data awal, terlihat bahwa setiap variabel dapat memiliki satuan yang berbeda dan memiliki selisih yang sangat tinggi jika dibandingkan pada variabel satu dengan variabel yang lain. Sebagai contoh komparasi pada variabel  $x_8$  yaitu APM perguruan tinggi dengan variabel  $x_{16}$  yaitu jumlah siswa sekolah dasar. Jika tidak dilakukan standarisasi, maka kemungkinan besar variabel  $x_{16}$  cenderung akan mendominasi dalam proses pengclusteran nantinya. Proses standarisasi dilakukan menggunakan pemanfaatan perangkat lunak R Studio Versi 4.2.1 dengan mengetikkan `install.packages("mvnrmtest")`.

2. *Memeriksa Korelasi Antar Variabel Menggunakan Korelasi Pearson*

Dalam analisis kluster Asumsi multikolinearitas bertujuan untuk mendeteksi apakah terdapat korelasi antar kelima variabel. Korelasi yang kuat dibuktikan dengan nilai koefisien korelasi ( $r$ ) yang mendekati nilai 1 sehingga mengindikasikan adanya korelasi positif yang kuat. Sementara itu, apabila koefisien korelasi mendekati -1 mengindikasikan bahwa terdapat korelasi negatif yang kuat. Uji asumsi ini dapat ditinjau dengan adanya nilai korelasi antar variabel yang cenderung rendah.

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (7)$$

dengan:

$r_{xy}$  adalah koefisien korelasi variabel  $X$  dan  $Y$

$x$  adalah deviasi dari rata-rata untuk nilai variabel  $X$

$y$  adalah deviasi dari rata-rata untuk nilai variabel  $Y$

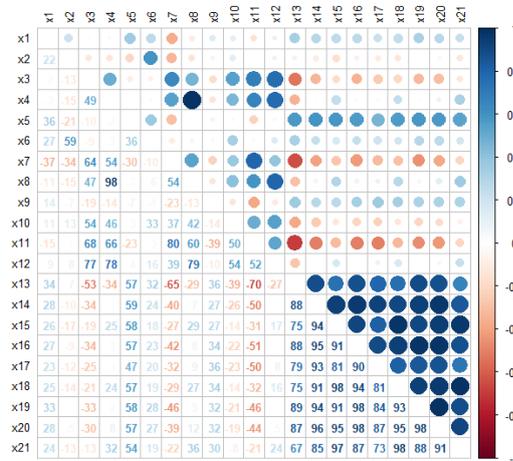
$\sum xy$  adalah jumlah perkalian antara nilai  $X$  dan  $Y$

$x^2$  adalah kuadrat dari nilai  $x$

$y^2$  adalah kuadrat nilai  $y$

Berdasarkan Gambar 1, kelima variabel menunjukkan kecenderungan memiliki koefisien korelasi yang terindikasi relatif rendah. Salah satu korelasi paling tinggi ditunjukkan di antara

variabel  $x_{16}$  (jumlah murid sekolah dasar) dengan variabel  $x_{19}$  (jumlah guru sekolah dasar) dengan nilai koefisien sebesar -0,98. Jika diamati dari diagram pencar, data tiap variabel cenderung menyebar dan tidak membentuk pola linear sedikitpun. Hal ini dapat dikatakan cukup memberikan indikasi bahwa variabel tersebut independent dan tidak saling berpengaruh satu sama lain.



GAMBAR 1. MATRIKS KORELASI ANTAR VARIABEL INDEKS PENDIDIKAN PROVINSI JAWA BARAT TAHUN 2021

3. *Jarak Menentukan Ukuran Kemiripan atau Ketakmiripan Antara Dua Objek*

Analisis *cluster* bertujuan untuk memilah dari sekumpulan objek menjadi beberapa kelompok yang berdasarkan bentuk ukuran kemiripan dilihat dari sebuah karakteristik pada variabel penyusunnya. Semakin dekat dari jarak suatu objek terhadap objek lain, maka semakin tinggi pula kemiripan objek tersebut sehingga dapat dimasukkan ke dalam kelompok yang sama. Ukuran yang digunakan pada penelitian ini dapat menentukan besaran jarak itu sendiri yaitu jarak *Euclidean*. Jarak *Euclidean* adalah jumlah akar dari total kuadrat selisih nilai antar dua objek dari setiap variabel faktor [11].

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{8}$$

dengan:

$d_{euc}(x, y)$  adalah jarak euclid antar dua objek

$x_i$  adalah variabel kabupaten  $x$  dengan variabel ke- $i$

$y_i$  adalah variabel kabupaten  $y$  dengan variabel ke- $i$

$n$  adalah banyaknya variabel

$i$  adalah urutan variabel faktor dari 1 hingga  $n$

Hasil dari perhitungan jarak antar objek berjumlah 27 kabupaten membentuk matriks  $27 \times 27$ .

4. *Proses Klasterisasi*

Peneliti menggunakan koefisien korelasi *cophenetic* sebagai acuan dalam mengetahui tingkat ketelitian pada metode cluster yang dipakai. Nilai koefisien korelasi *cophenetic* yang bernilai tinggi mendekati 1 mengindikasikan bahwa solusi yang dihasilkan dari proses clustering sudah baik. Koefisien korelasi *cophenetic* merupakan koefisien korelasi di antara elemen-elemen asli matriks jarak *Euclidian* dan setiap elemen-elemen yang dihasilkan oleh matriks *cophenetic* [12].

Dari Tabel 3 dapat dilihat nilai korelasi *cophenetic* yang tertinggi adalah pada metode *ward* sehingga dapat dikatakan metode *ward* merupakan metode cluster yang terbaik. Proses identifikasi koefisien korelasi *cophenetic* ini dilakukan menggunakan pemanfaatan perangkat lunak R Studio Versi 4.2.1 dengan mengetikkan `install.packages("cluster")`, `install.packages("purrr")`, dan `install.packages("factoextra")` seperti pada Gambar 2.

```
library(cluster)

## Warning: package 'cluster' was built under R version 4.2.2

library(purrr)
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.2.2

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##   %+%, alpha

## Welcome! Want to learn more? See two factoextra-related books at https://go
o.gl/ve3WBa

# vector of methods to compare
m <- c( "average", "single", "complete", "ward")
names(m) <- c( "average", "single", "complete", "ward")

# function to compute coefficient
ac <- function(x) {
  agnes(cluster_jabar, method = x)$ac
}
map_dbl(m, ac)

##   average   single  complete    ward
## 0.9185234 0.8860535 0.9413720 0.9525536
```

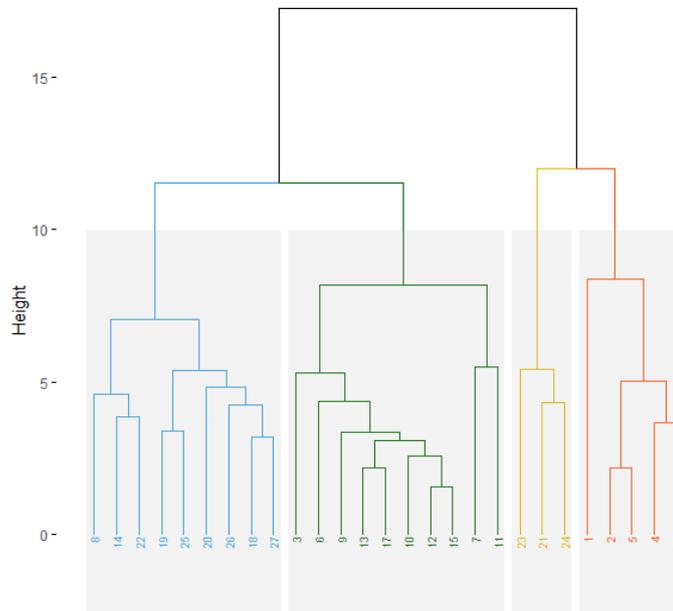
GAMBAR 2. TEKS KOEFISIEN KORELASI *COPHENETIC*

TABEL 3. HASIL KORELASI *COPHENETIC*

Metode <i>Single Linkage</i>	Metode <i>Complete Linkage</i>	Metode <i>Average Linkage</i>	Metode <i>Ward</i>
0,918	0,8860	0,941	0,952

Selanjutnya adalah melakukan analisis *cluster* dengan menggunakan metode *ward*. Hasil pengclusteran dengan program R dapat dilihat pada Gambar 4 dendrogram dengan metode *ward*. Dari gambar tersebut terlihat bahwa terdapat empat kluster yang terbentuk dengan menggunakan metode *ward*. Selanjutnya, dengan menggunakan program R, diperoleh hasil pada Gambar 3 dan Gambar 4. Berdasarkan Tabel 4 terdapat empat kluster beserta anggotanya, kluster 1 terdiri atas 9 anggota yaitu Kuningan, Purwakarta, Kota Cirebon, Kota Tasikmalaya, Pangandaran, Kota Banjar, Kota Sukabumi, Kota Bogor, dan Kota Cimahi. Selanjutnya kluster 2 terdiri atas 10 anggota yaitu Cirebon, Ciamis, Subang, Bandung Barat, Majalengka, Indramayu, Karawang,

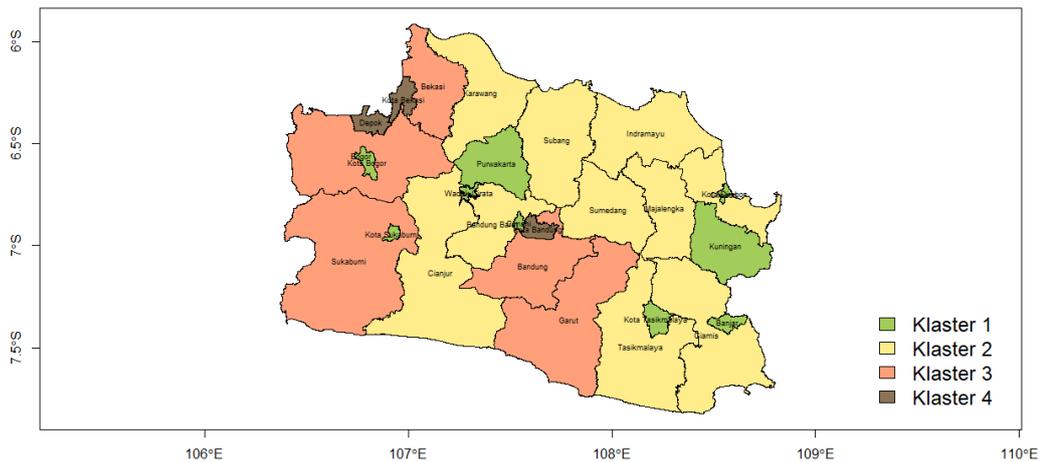
Cianjur, Sumedang, dan Tasikmalaya. Kluster 3 berisikan 5 anggota yaitu Bogor, Sukabumi, Garut, Bandung, dan Bekasi. Terakhir yaitu kluster 4 berisi 3 anggota yaitu Kota Bekasi, Kota Bandung, dan Kota Depok.



GAMBAR 3. DENDROGRAM METODE WARD

TABEL 4. HASIL CLUSTER BESERTA ANGGOTANYA DENGAN METODE WARD

Kluster	Kabupaten/Kota
1	Kuningan, Purwakarta, Kota Cirebon, Kota Tasikmalaya, Pangandaran, Kota Banjar, Kota Sukabumi, Kota Bogor, dan Kota Cimahi
2	Cirebon, Ciamis, Subang, Bandung Barat, Majalengka, Indramayu, Karawang, Cianjur, Sumedang, dan Tasikmalaya
3	Bogor, Sukabumi, Garut, Bandung, dan Bekasi
4	Kota Bekasi, Kota Bandung, dan Kota Depok



GAMBAR 4. PETA ANALISIS GEROMBOL MENGGUNAKAN METODE WARD

Dalam menginterpretasikan dan memprofil kluster dengan menggunakan pengkajian mengenai *centroid* adalah rata-rata pada nilai objek yang terdapat dalam kluster untuk setiap variabel. Nilai *centroid* memungkinkan dalam menguraikan setiap kluster dengan cara memberikan label atau suatu nama. Berikut adalah rata-rata dari variabel pada setiap *cluster* yang telah terbentuk pada Tabel 5.

TABEL 5. RATA-RATA VARIABEL PADA SETIAP CLUSTER YANG TERBENTUK

Kluster	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$
1	105,908	92,016	72,022	18,674	99,252	84,16	54,124	15,496	99,682	95,044	62,61
2	103,882	93,585	71,673	12,338	98,127	81,578	55,861	9,603	99,362	94,413	66,833
3	103,399	95,695	85,632	21,03	96,478	81,317	63,861	16,850	99,4	95,757	72,476
4	104,183	89,667	100,546	37,553	98,71	82,136	72,016	28,933	99,32	97,753	81,853

Kluster	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$X_{16}$	$X_{17}$	$X_{18}$	$X_{19}$	$X_{20}$	$X_{21}$
1	21,07	1408,8	439	126,6	344450	66087,6	56389,4	14246	6499,2	2559,2
2	13,205	917,875	221	55,875	176906	26388,63	25472,5	8499,75	3788,375	1278,25
3	24,708	336,818	82,090	24,636	64904,91	10062,27	12981,27	3746,545	1656,727	696
4	37,26	515,666	272,666	105,333	200509	31161,33	45036,33	9748,667	4804	2447,667

**Keterangan:**

- Hijau bernilai rendah
- Kuning bernilai sedang
- Orange bernilai tinggi
- Merah bernilai sangat tinggi

Berdasarkan kajian menggunakan *Agglomerative Hierarchical Cluster*, terdapat beberapa variabel yang menarik untuk dikaji lebih lanjut. Pada variabel  $x_3$  dan  $x_7$  (APK dan APM SMA/MA/Paket C), terdapat perbedaan variansi yang cukup tinggi dengan rata-rata nilai tertinggi terdapat pada cluster 4. Hal ini menandakan bahwa pada Kota Bekasi, Kota Bandung, dan Kota Depok memiliki Angka Partisipasi Kasar bagi siswa SMA yang tinggi. Fenomena yang demikian adanya merupakan hal yang cukup wajar karena jika ditinjau dari aspek lain memang mengindikasikan proporsi yang tinggi untuk anak sekolah pada jenjang pendidikan terkait. Bukan menjadi rahasia umum lagi jika kualitas pendidikan di kota akan lebih bagus daripada wilayah kabupaten. Hal ini menyorot bahwa pendidikan akan lebih berkembang di kota dengan alasan pesatnya perkembangan IPTEK serta aksesibilitas yang tinggi [13]. Berdasarkan uraian tersebut,

dalam menambah efektivitas kurikulum merdeka belajar sekaligus memajukan pendidikan Indonesia, hal yang dapat digencarkan yaitu optimalisasi peran zonasi sehingga dapat terjadi pemerataan kualitas tiap klasternya.

Hal selanjutnya yang mengundang atensi penulis adalah variabel  $x_4$  dan  $x_8$  (APK dan APM Perguruan Tinggi). Terlihat adanya perbedaan rata-rata yang signifikan dengan klaster 4 yang menduduki posisi tertinggi. Dengan banyaknya keberadaan kampus unggul, baik negeri maupun swasta di klaster tersebut menjadi indikasi utama mengapa terjadi hal yang sedemikian ini, pemerataan pendidikan pada klaster tersebut terbilang baik. Sebut saja, tidak sedikit pelajar Indonesia yang berminat melanjutkan studinya di Perguruan Tinggi ternama. Menanggapi hal ini, rekomendasi yang dapat diupayakan guna meningkatkan pemerataan nilai rata-rata APK dan APM Perguruan Tinggi ialah dengan cara pengencaran beasiswa pada universitas di cluster 1, 2, dan 3; sehingga terjadi pemerataan pelajar yang mana diharapkan dapat menimbulkan efek domino yang baik terhadap mutu serta proporsi partisipasi untuk berkuliah. Lalu terdapat juga pada variabel  $x_{12}$  (APS pada perguruan tinggi yang ada pada klaster 4).

Seperti yang dijelaskan di atas tadi, klaster 4 memiliki keunggulan pada keberadaan dan pemerataan pendidikan yang terindikasi baik. Selanjutnya hal menarik lain yang menjadi atensi penulis adalah keterkaitan antara variabel:  $x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}$ , dan  $x_{21}$ . Dari variabel tersebut merupakan jumlah SD, SMP, dan SMA ditunjukkan pada variabel  $x_{13}$  sampai  $x_{15}$  pada klaster 1 mendapatkan skor yang cukup tinggi sehingga mengindikasikan bahwa jumlah SD, SMP, dan SMA pada klaster tersebut sangat banyak. Sehingga hal tersebut berdampak pada variabel  $x_{16}$  sampai  $x_{18}$  yang merupakan jumlah murid SD, SMP, dan SMA sangat komprehensif dibandingkan klaster lain pada variabel tersebut. Argumen tersebut didukung dengan adanya fasilitas dan kebijakan pemerintah mengenai program zonasi yang memiliki tujuan dalam menghadirkan pemerataan akses pada layanan pendidikan, serta pemerataan kualitas Pendidikan secara nasional.

Selanjutnya, variabel  $x_{19}$  hingga  $x_{21}$  menjalin hubungan yang sama seperti pada variabel sebelumnya hal ini dikarenakan efek domino atau efek beruntun yang menginvasi jumlah Guru SD, SMP, dan SMA. Selain itu, alasan lain yang melandasi tingginya angka jumlah guru pada klaster 1 yang memuat beberapa Kabupaten/Kota Kuningan, Purwakarta, Kota Cirebon, Kota Tasikmalaya, Pangandaran, Kota Banjar, Kota Sukabumi, Kota Bogor, dan Kota Cimahi berbanding linear pada fasilitas dan kebijakan pemerintah yang meningkatkan serta menghadirkan program peningkatan jumlah guru seperti PANRB PPK Guru, Guru Penggerak, Kebijakan Distribusi Guru, dan Beasiswa LPDP yaitu S1 Guru SMK. Program tersebutlah yang mendorong dalam menyuntikkan jumlah guru yang masif secara nasional.

Sebelumnya, sudah terdapat penelitian sebelumnya yang mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan variabel Angka Harapan Hidup, Angka Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Tingkat Pengeluaran per kapita. Purwandari dan Hidayat (2016) melakukan Pengelompokan Kabupaten dan Kota di Jawa Barat Menggunakan *Principal Component Analysis Biplot* [14]. Terdapat 3 kelompok kabupaten dan kota di Provinsi Jawa Barat dengan persentase keragaman data yang dihasilkan oleh Analisis *Biplot* sebesar 89,70%. Perbedaan jangkauan variabel yang digunakan menjadi alasan kuat mengapa terdapat perbedaan jumlah klaster yang terbentuk.

Terdapat pula penelitian sebelumnya yang melakukan pengelompokan kabupaten/kota di Provinsi Jawa Barat berdasarkan Analisis Klaster Hierarki. Sharon, dkk (2021) menggunakan analisis klaster hierarki untuk mengetahui karakteristik dari tiap klaster IPM di Jawa Barat tahun 2020 [15]. Berdasarkan 5 metode analisis klaster hierarki yang digunakan, diperoleh klaster optimum dengan nilai korelasi *cophenetic* terbesar terdapat pada metode *average linkage* sebesar 0,8262371. Pada penelitian ini, Provinsi Jawa Barat terbagi menjadi 4 klaster, berjumlah sama dengan hasil dalam penelitian ini.

#### IV. SIMPULAN DAN SARAN

Berdasarkan hasil penelitian Perbandingan Analisis Gerombol *Agglomerative Hierarchical Clustering* (AHC). Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat menggunakan komparasi metode *single linkage*, *complete linkage*, *average linkage*, dan *ward* dapat diambil simpulan bahwa perlu dilakukan peninjauan lebih lanjut mengenai pemerataan menjadi beberapa klaster. Nilai APK pada jenjang SMA

memiliki nilai standar deviasi yang cukup tinggi sekitar 14,5%, mengindikasikan pertumbuhan pendidikan formal jenjang Pendidikan SMA sederajat belum merata di setiap Kabupaten/Kota Provinsi Jawa Barat. Begitu juga pada kualitas pendidikan masyarakat usia 19 tahun keatas, hanya sekitar 7,7% penduduk Provinsi Jawa Barat yang memiliki ijazah sampai tamat perguruan tinggi.

Setelah dilakukan tinjauan komparasi korelasi *cophenetic*, Penggerombolan optimum didapatkan dengan menggunakan metode *Ward* dengan nilai korelasi *cophenetic* terbesar yaitu 0,952. Sedangkan untuk metode *single linkage*, *complete linkage*, dan *average linkage* memiliki nilai korelasi *cophenetic* sebesar 0,918; 0,886; dan 0,941 secara berurutan. Metode *Ward* membagi Provinsi Jawa Barat menjadi 4 cluster/kelompok yaitu : (a) Klaster 1 terdiri atas 9 anggota yaitu Kuningan, Purwakarta, Kota Cirebon, Kota Tasikmalaya, Pangandaran, Kota Banjar, Kota Sukabumi, Kota Bogor, dan Kota Cimahi; (b) Klaster 2 terdiri atas 10 anggota yaitu Cirebon, Ciamis, Bandung Barat, Majalengka, Indramayu, Karawang, Cianjur, Sumedang, dan Tasikmalaya; (c) Klaster 3 berisikan 5 anggota yaitu Bogor, Sukabumi, Garut, Bandung, dan Bekasi; serta (d) Klaster 4 berisi 3 anggota yaitu Kota Bekasi, Kota Bandung, dan Kota Depok.

Klaster 1 memiliki nilai rata-rata tertinggi pada variabel jumlah sekolah dan guru pada jenjang SD, SMP, SMA. Meninjau hal yang demikian adanya, guna mencapai pemerataan mutu pendidikan dan optimalisasi kualitas, pihak pemerintah dan atau yang berwenang disarankan untuk meninjau fasilitas dan kebijakan pemerintah yang meningkatkan serta menghadirkan program peningkatan jumlah guru seperti PANRB PPK Guru, Guru Penggerak, Kebijakan Distribusi Guru, dan Beasiswa LPDP yaitu S1 Guru SMK. Program tersebutlah yang mendorong dalam menyuntikkan jumlah guru yang masif secara nasional.

Klaster 4 memiliki nilai rata-rata tertinggi pada variabel partisipasi SMA sederajat dan Perguruan Tinggi jika dibandingkan pada klaster 1, 2, dan 3. Hal ini bermakna bahwa yang belum tercapainya pemerataan pendidikan khususnya pada jenjang sekolah menengah atas hingga perguruan tinggi di Provinsi Jawa Barat. Menanggapi hal tersebut perlu dipertimbangkan mengenai optimalisasi peran zonasi dan pengencaran beasiswa baik kip atau non-kip.

#### DAFTAR PUSTAKA

- [1] E. Muzdalifa, "Learning Loss Sebagai Dampak Pembelajaran Online Saat Kembali Tatap Muka Pasca Pandemi Covid 19," *GUAU: Jurnal Pendidikan Profesi Guru Agama Islam*, vol. 2, no. 1, pp. 187–192, Jun. 2022, Accessed: Nov. 28, 2022. [Online]. Available: <http://studentjournal.iaincurup.ac.id/index.php/guau/article/view/152>
- [2] "Estimates of COVID-19 Impacts on Learning and Earning in Indonesia: How to Turn the Tide," 2020. Accessed: Jan. 06, 2023. [Online]. Available: <https://www.worldbank.org/en/country/indonesia/publication/estimates-of-covid-19-impacts-on-learning-and-earning-in-indonesia-how-to-turn-the-tide>
- [3] Z. I. Alfianti, "Pengelompokan Wilayah Penyebaran COVID-19 di Kabupaten Karawang Menggunakan Algoritma K-Means," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 2, pp. 111–122, 2021, doi: 10.35760/IK.2021.V26I2.4155.
- [4] BPS, "Provinsi Jawa Barat dalam Angka 2021," <https://jabar.bps.go.id/publication/2021/02/26/4d3f7ec6c519dda0b9785d45/provinsi-jawa-barat-dalam-angka-2021.html>, 2021.
- [5] S. Sharma, J. Wiley, N. York, C. Brisbane, and T. Singapore, "Applied Multivariate Techniques."
- [6] B. Everitt and T. Hothorn, "An Introduction to Applied Multivariate Analysis with R (Use R)," 2011. [Online]. Available: <http://www.springer.com/series/6991>
- [7] A. Fadliana and F. Rozi, "Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 4, no. 1, pp. 35–40, Nov. 2015, doi: 10.18860/CA.V4I1.3172.
- [8] H. Hikmah, F. Fardinah, L. Qadrini, and E. Tande, "Analisis Klaster Pengelompokan Kecamatan di Sulawesi Barat Berdasarkan Indikator Pendidikan," *SAINTIFIK*, vol. 8, no. 2, pp. 188–196, Jul. 2022, doi: 10.31605/SAINTIFIK.V8I2.383.
- [9] R. Arnold. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*.
- [10] J. D. Miller, *Statistics for Data Science*. Birmingham: Packt Publishing, 2017.
- [11] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis)*, 1st ed. California: CreateSpace Independent Publishing Platform, 2017.
- [12] E. Widodo, N. N. Sari, I. Hidayati, F. Yubinas, M. Yuniarti, and R. D. Novyantika, "Analisis Cluster Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016," in *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) III 2018*, 2018. Accessed: Nov. 29, 2022. [Online]. Available: <http://publikasiilmiah.ums.ac.id/handle/11617/10140>
- [13] A. Y. Anas, A. W. Riana, and N. C. Apsari, "Desa dan Kota dalam Potret Pendidikan," *Prosiding Penelitian dan Pengabdian kepada Masyarakat*, vol. 2, no. 3, Nov. 2015, doi: 10.24198/JPPM.V2I3.13592.
- [14] T. Purwandari and Y. Hidayat, "Pengelompokan Kabupaten dan Kota di Jawa Barat Menggunakan Principal Component Analysis Biplot," *Sempoa*, May 2016, Accessed: Jan. 20, 2023. [Online]. Available: <http://publikasiilmiah.ums.ac.id/handle/11617/7624>

- [15] S. Syalomitha *et al.*, “Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Menurut Indikator Indeks Pembangunan Manusia Tahun 2020,” *E-Prosiding Seminar Nasional Statistika | Departemen Statistika FMIPA Universitas Padjadjaran*, vol. 10, pp. 108–123, Dec. 2021, doi: 10.1234/PNS.V10I.80.