

OPTIMALISASI JUMLAH KELOMPOK PADA METODE *SINGLE LINKAGE* DAN *COMPLETE LINKAGE* MENGGUNAKAN INDEKS *DAVIES BOULDIN*

OPTIMIZING NUMBER OF CLUSTERS ON SINGLE LINKAGE AND COMPLETE LINKAGE METHODS USING DAVIES BOULDIN INDEX

Oleh: Siti Sakdiyah ¹⁾ Retno Subekti, M. Sc.²⁾

¹⁾ Mahasiswa Pendidikan Matematika Universitas Negeri Yogyakarta

²⁾ Dosen Pendidikan Matematika Universitas Negeri Yogyakarta

¹⁾ sitisakdiyah27@gmail.com ²⁾ retnosubekti@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk mengoptimalkan jumlah kelompok menggunakan metode *single linkage* dan *complete linkage* dengan indeks *Davies Bouldin* yang diterapkan pada kasus data indikator pendidikan di Indonesia tahun 2016. Langkah-langkah yang dilakukan dalam metode *single linkage* dan *complete linkage* terdiri atas menentukan variabel yang akan digunakan sebagai data penelitian, mendeteksi adanya *outliers*, memilih ukuran kesamaan (jarak) yang digunakan, melakukan standarisasi data jika diperlukan, melakukan pengujian terhadap asumsi-asumsi dalam analisis kelompok, memilih prosedur pengelompokan, menentukan jumlah kelompok yang akan dibentuk, melakukan interpretasi terhadap kelompok yang terbentuk, melakukan uji validasi terhadap hasil pengelompokan untuk menentukan jumlah kelompok optimal, dan membuat profil kelompok yang terbentuk. Hasil penelitian menunjukkan bahwa pada metode *single linkage* dan *complete linkage* diperoleh kelompok optimal yang sama, yaitu sebanyak 2 kelompok dengan indeks *Davies Bouldin* sebesar 0,0759. Diperoleh pengelompokan provinsi berdasarkan tingkat partisipasi pendidikan, yaitu salah satu kelompok mempunyai tingkat partisipasi pendidikan rendah yang beranggotakan Provinsi Papua, dan kelompok lain memiliki tingkat partisipasi pendidikan tinggi.

Kata kunci : Analisis Kelompok, *Single Linkage*, *Complete Linkage*, dan Indeks *Davies Bouldin*

Abstract

This research aims to optimize the number of clusters using the single linkage and complete linkage method with Davies Bouldin index applied in the case of education indicator data in Indonesia on 2016. This methods consist of determining variables to be used as research data, detecting outliers, choosing the size of the similarity (distance) used, standardizing the data if necessary, testing the assumptions in the analysis cluster, selecting the grouping procedure, determining the number of clusters to be formed, interpreting the cluster formed, validating to determine the number of groups optimal, and creating profiles of cluster that are formed. The results showed the single linkage and complete linkage method, the same optimal cluster was obtained, that is 2 clusters with Davies Bouldin index about 0.0759. The clustering can be obtained based on educational participation rank, that are low educational participation cluster that consist of Papua Province, and high educational participation cluster.

Keywords : Cluster Analysis, *Single Linkage*, *Complete Linkage*, and *Davies Bouldin Index*

PENDAHULUAN

Cluster analysis (analisis kelompok) merupakan metode analisis untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok, sehingga akan diperoleh kelompok dimana objek-objek dalam satu kelompok memiliki banyak persamaan sedangkan dengan anggota kelompok lain memiliki banyak perbedaan (Johnson &

Winchern, 2007:671). Dalam analisis kelompok dikenal dua metode pengelompokan, yaitu metode hierarki dan metode nonhierarki. Metode hierarki terbagi atas dua bagian, yaitu *agglomerative method* (metode penyatuan) dan *divisive method* (metode pemisahan). Dalam metode penyatuan dikenal beberapa metode, yaitu metode *linkage*, metode *variance*, dan metode *centroid*, dimana metode *linkage* terdiri dari

metode *single linkage*, *complete linkage*, dan *average linkage*. Pada metode *variance* terdiri dari metode *Ward*. Setelah dilakukan pemilihan metode yang digunakan untuk membentuk kelompok dan diperoleh hasil pengelompokan, selanjutnya dilakukan perhitungan untuk memperoleh jumlah kelompok optimal dan metode terbaik dengan menggunakan indeks validitas.

Penelitian mengenai analisis kelompok dan indeks validitas telah banyak dilakukan. Seperti penelitian yang menggunakan metode *K-means* dan indeks validitas *Davies Bouldin* yang diimplementasikan pada data sekuens DNA untuk mengenali suatu spesies dan membedakan spesies yang satu dengan yang lainnya berdasarkan kesamaan ciri yang dimiliki (Sinurat, 2014). Hasil penelitian tersebut menunjukkan nilai pengelompokan terbaik dicapai oleh pengelompokan dengan tiga kelompok yang ditunjukkan dari nilai indeks *Davies Bouldin* terkecil. Terdapat juga penelitian mengenai pengelompokan desa/kelurahan di Kota Denpasar menurut indikator pendidikan (Aprilia A.P., Srinadi, & Sari, 2016). Penelitian ini menggunakan indikator pendidikan sebagai variabel yang digunakan. Indikator-indikator pendidikan dalam penelitian tersebut meliputi banyak sekolah TK, SD, SMP, SM sederajat/SMK; banyak siswa TK, SD, SMP, SM sederajat/SMK; dan banyak guru TK, SD, SMP, SM sederajat/SMK. Penelitian tersebut menggunakan metode *linkage* dan indeks *Cluster Tightness Measure* (CTM) dengan jarak euclid dan pearson. Hasil penelitian menunjukkan bahwa berdasarkan nilai terkecil CTM, metode terbaik menggunakan *average linkage* dengan jarak pearson.

Selain dua penelitian di atas, analisis kelompok juga dapat digunakan untuk mengelompokkan provinsi di Indonesia. Pengelompokan provinsi salah satunya berguna dalam penyetaraan capaian pendidikan semua wilayah di Indonesia. Hal tersebut dilakukan agar pemerintah mengetahui tingkat partisipasi

pendidikan di setiap provinsi di Indonesia. Penjaminan kepastian, penyediaan, dan peningkatan layanan pendidikan merupakan salah satu isu strategis di bidang pendidikan yang tercantum dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2015-2019. Isu strategis tersebut bertujuan untuk tercapainya perluasan dan pemerataan akses pendidikan yang berkualitas, berorientasi pada pembentukan karakter, dan relevan dengan kebutuhan masyarakat dan pembangunan daerah. Isu strategis tersebut diwujudkan dalam program Indonesia Pintar yang direalisasikan melalui pelaksanaan Wajib Belajar (Wajar) 12 Tahun. Salah satu sasaran yang ingin dicapai dalam Program Indonesia Pintar melalui pelaksanaan Wajar 12 tahun pada RPJMN 2015-2019 adalah meningkatnya angka partisipasi pendidikan dasar dan menengah. Angka partisipasi pendidikan tersebut dapat dilihat dari nilai Angka Partisipasi Kasar (APK) dan Angka Partisipasi Murni (APM).

Dengan mengacu pada penelitian-penelitian sebelumnya dan kondisi pendidikan di Indonesia pada saat ini, penulis tertarik melakukan penelitian dengan judul, "Optimalisasi Jumlah Kelompok pada Metode *Single Linkage* dan *Complete Linkage* Menggunakan Indeks *Davies Bouldin* (Studi Kasus : Indikator Pendidikan di Indonesia Tahun 2016). Pengelompokan dilakukan berdasarkan kemiripan sifat yang dimiliki, yaitu capaian partisipasi pendidikan, sehingga diperoleh jumlah kelompok optimal yang sesuai kondisi pendidikan Indonesia saat ini.

TINJAUAN PUSTAKA

Cluster Analysis (Analisis Kelompok)

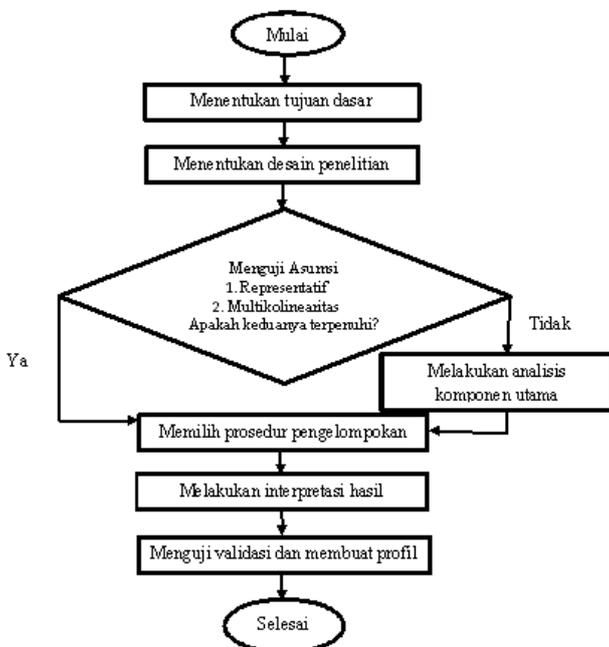
Analisis kelompok adalah salah satu teknik multivariat yang memiliki tujuan utama untuk mengklasifikasi objek-objek berdasarkan karakteristik yang dimiliki. Analisis kelompok mengklasifikasi objek-objek ke dalam kelompok-kelompok sehingga objek-objek yang berada dalam satu kelompok serupa sedang objek yang

berada dalam kelompok yang lain berbeda. Analisis tersebut bertujuan untuk memaksimalkan homogenitas objek dalam kelompok sementara itu juga memaksimalkan heterogenitas antarkelompok (Hair, et. al, 2009:478).

Analisis kelompok termasuk ke dalam analisis statistik multivariat metode interdependen, sehingga tujuannya tidak menghubungkan atau membedakan sampel/variabel lain. Analisis kelompok berguna untuk meringkas data dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang akan diteliti. Kelompok yang baik mempunyai ciri-ciri sebagai berikut (Ary, 2015):

- memiliki homogenitas yang tinggi antar anggota dalam satu kelompok (*within cluster*)
- memiliki heterogenitas yang tinggi antarkelompok yang satu dengan kelompok yang lain (*between cluster*).

Tahapan-tahapan dalam proses analisis kelompok dapat direpresentasikan seperti pada Gambar 1.



Gambar 1. Diagram Alir Analisis Kelompok

Deteksi terhadap univariat *outlier* dapat dilakukan dengan menentukan nilai batas yang akan dikategorikan sebagai data *outlier*, yaitu

dengan cara mengkonversi nilai data ke dalam skor standardized atau yang biasa disebut *z-score*, yang memiliki nilai *means* (rata-rata) nol dan standar deviasi satu. Jika sebuah data memiliki *z-score* lebih besar dari 3 maka dipertimbangkan sebagai *outliers* (Prasetyo, 2012:200). Jika data diketahui terdapat satu atau lebih data *outliers*, pada *outliers* tersebut bisa dilakukan beberapa penanganan. Data dihilangkan, jika data *outliers* tersebut didapat karena kesalahan pengambilan data, kesalahan pemasukan data pada komputer, dan sebagainya. Data *outlier* tetap dipertahankan dan tidak perlu dihilangkan, jika tidak terdapat kesalahan pada proses sampling maupun pemasukan data (Yulianto, 2014). Nilai *z-score* diperoleh menggunakan rumus persamaan berikut (Prasetyo, 2012:200):

$$z = \frac{x_i - \bar{x}}{s} \tag{1}$$

dimana :

- z = nilai standar/*z-score*
- x_i = nilai observasi/data
- \bar{x} = mean data
- s = deviasi standar sampel

Jarak yang dipakai sebagai ukuran kesamaan menjelaskan seberapa mirip atau seberapa berbeda objek-objek yang dikelompokkan. Misalkan terdapat p variabel. Pengukuran variabel V_1, V_2, \dots, V_p dilakukan terhadap setiap individu anggota sejumlah n dengan a_1, a_2, \dots, a_n sebagai individu anggota. Jarak antar individu anggota dapat dinyatakan dalam bentuk matriks seperti pada Persamaan (2).

$$D_{n \times n} = \begin{bmatrix} D_{11} & D_{12} & \dots & D_{1j} & \dots & D_{1n} \\ D_{21} & D_{22} & \dots & D_{2j} & \dots & D_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{i1} & D_{i2} & \dots & D_{ij} & \dots & D_{in} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \dots & D_{nj} & \dots & D_{nn} \end{bmatrix} \tag{2}$$

(Yulianto, 2014)

Pada Persamaan (2), matriks D adalah matriks jarak yang terdiri atas kumpulan jarak antar objek sejumlah n dengan D_{ij} merupakan jarak antara individu ke- i dan ke- j (untuk setiap

$i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, n$). Semakin kecil nilai D , maka semakin besar kemiripan antara kedua pengamatan tersebut. Sebaliknya bila D besar, semakin besar ketidakmiripan dari pengamatan tersebut (Yulianto & Hidayatullah, 2014). Salah satu ukuran jarak yang digunakan adalah *euclidean distance* (jarak euclid) (Johnson & Wichern, 2007:673).

$$D_{ij} = \sqrt{\sum_{k=1}^p (a_{ik} - a_{jk})^2} \quad (3)$$

dengan

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

a_{ik} = nilai pengamatan objek ke- i variabel ke- k

a_{jk} = nilai pengamatan objek ke- j variabel ke- k

Standarisasi data dilakukan jika ditemukan adanya perbedaan skala yang besar di antara variabel-variabel yang digunakan dalam penelitian (Simamora, 2005:232). Untuk melakukan standarisasi data menggunakan rumus seperti pada Persamaan (1).

Dalam analisis kelompok, masalah seberapa besar sampel mewakili populasi (representatif) dan adanya multikolinearitas harus lebih diperhatikan (Santoso, 2010:116).

Pengujian multikolinieritas digunakan untuk mengetahui ada tidaknya variabel independen yang memiliki kemiripan dengan variabel independen lainnya (Hidayat, 2012). Jenis uji parametrik untuk pengujian hubungan, yaitu korelasi *product moment pearson*. Pengujian ini digunakan untuk menguji dua variabel apakah ada hubungan atau tidak, dengan jenis data keduanya sama yaitu rasio atau interval. Rumus sebagai berikut (Suparlan, 1983:74-75):

$$r_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[(n \sum x^2) - (\sum x)^2][(n \sum y^2) - (\sum y)^2]}} \quad (4)$$

dengan

r_{xy} = koefisien korelasi r pearson

n = ukuran sampel/observasi

x = variabel pertama

y = variabel kedua

Jika ada nilai korelasi antarvariabel bebas lebih besar dari 0,8 maka dapat diindikasikan adanya multikolinearitas (korelasi sangat kuat) (Hidayat, 2012). Data yang mengandung multikolinearitas maka dilakukan proses analisis komponen utama. Analisis Komponen Utama (*Principal Component Analysis*) adalah analisis multivariat yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi lagi (Rufaidah & Effindi, 2017). Variabel baru tersebut disebut sebagai komponen utama (*principal component/KU*). Secara aljabar linier, KU adalah kombinasi-kombinasi linier tertentu dari p variabel acak V_1, V_2, \dots, V_p dimana koefisiennya adalah vektor ciri (*eigen vector*). KU tergantung pada matriks kovarian (Σ) dan matriks korelasi (ρ) dari V_1, V_2, \dots, V_p dikarenakan vektor ciri dihasilkan dari akar ciri (*eigen value*) matriks kovarian atau dapat juga dihasilkan dari matriks korelasi.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \text{ dan}$$

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & \rho_{pp} \end{bmatrix}$$

Melalui persamaan karakteristik matriks kovarian diperoleh akar ciri-akar cirinya, yaitu $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ dan vektor-vektor cirinya, yaitu e_1, e_2, \dots, e_p . Menyusutkan dimensi variabel asal V menjadi KU Y , karena itu sering disebut vektor pembobot. KU pertama adalah kombinasi linier terbobot dari variabel asal yang mempunyai varian terbesar. KU kedua juga merupakan kombinasi linier terbobot dari variabel asal dengan varian terbesar kedua dan antara kedua KU tersebut tidak saling berkorelasi demikian seterusnya. Secara umum, KU ke- i dapat dituliskan sebagai,

$$Y_i = e_{1i}V_1 + e_{2i}V_2 + e_{3i}V_3 + \dots + e_{pi}V_p$$

$$= e_i'V; i = 1, 2, 3, \dots, p \quad (5)$$

dimana e_i adalah vektor ciri (*eigen vector*) ke- i yang berpadanan dengan akar cirinya. Matriks kovarian Σ digunakan jika variabel yang diamati ukurannya pada skala dengan perbedaan tidak besar atau jika satuan ukurannya sama. Bila variabel yang diamati ukurannya pada skala dengan perbedaan sangat besar, atau ukurannya tidak sama, maka variabel tersebut perlu dibakukan sehingga KU ditentukan dari variabel baku. Variabel baku (Z) diperoleh dari transformasi terhadap variabel asal dalam matriks sebagai berikut:

$$Z = (V^{\frac{1}{2}})^{-1}(V - \mu) \quad (6)$$

dengan $V^{\frac{1}{2}}$ adalah matriks simpangan baku dengan unsur diagonal utama $(\sigma_{ii})^{1/2}$ dengan unsur lainnya adalah 0. Nilai harapan $E(Z) = 0$ dan keragamannya $Cov(Z) = (V^{\frac{1}{2}})^{-1}$ dan $S(V^{\frac{1}{2}})^{-1} = r$. Dengan demikian KU dari Z dapat ditentukan dari vektor ciri yang diperoleh melalui matriks korelasi variabel asal r .

Untuk mencari akar ciri dan menentukan vektor pembobotnya sama seperti pada matriks S . Akar ciri yang didapat dari matriks korelasi akan memiliki *trace* (jumlah seluruh akar ciri) sama dengan jumlah variabel yang dipakai.

$$\sum_{i=1}^p \lambda_i = \lambda_1 + \lambda_2 + \dots + \lambda_p = p \quad (7)$$

(Yulianto, 2013:30-39)

Pada tahap interpretasi dilakukan pengujian terhadap masing-masing kelompok untuk menamai atau memberikan keterangan secara tepat sebagai profil sifat dari kelompok.

Single Linkage

Single linkage (jarak terdekat) atau tautan tunggal merupakan metode hierarki yang mengelompokkan data berdasarkan tetangga terdekat (*nearest neighbour*) atau similaritas maksimum antara dua data (Amponsah, et al, 2013). Ukuran ini mendefinisikan jarak antara dua kelompok sebagai jarak minimum yang ditemukan antara satu kasus dari kelompok pertama dan satu kasus dari kelompok kedua (Odilia, 2015). Berikut rumus yang digunakan

untuk menentukan kelompok dalam metode *single linkage*.

$$D_{(uv)w} = \min(D_{uw}, D_{vw}) \quad (8)$$

dimana, D_{uw}, D_{vw} adalah jarak antara tetangga terdekat dari kelompok (U dan W) juga kelompok V dan W , begitu sebaliknya (Maylana, 2014).

Complete Linkage

Juga disebut sebagai tetangga terjauh atau metode maksimal. Ukuran ini serupa dengan ukuran *single linkage* yang dijelaskan di atas, namun *complete linkage* menggunakan jarak terjauh antara pasangan kasus (Odilia, 2015). Berikut adalah rumus yang digunakan dalam metode *complete linkage*.

$$D_{(uv)w} = \max(D_{uw}, D_{vw}) \quad (9)$$

dimana, D_{uw}, D_{vw} adalah jarak antara tetangga terjauh dari kelompok U dan W juga kelompok V dan W , begitu sebaliknya (Maylana, 2014).

Davies Bouldin Index (Indeks Davies Bouldin)

Indeks *Davies Bouldin* adalah pengukuran yang memaksimalkan jarak inter-kelompok antara kelompok C_k dan C_l dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah kelompok. Andai $Sc(k)$ adalah jarak intra-kelompok dalam kelompok k . Rumus untuk menghitung $Sc(k)$ ialah (Sinurat, 2014):

$$Sc(k) = \frac{\sum_i^{N_k} \|X_i - C_k\|^2}{N_k} \quad (10)$$

dengan

N_k = banyak titik dalam kelompok k

C_k = *centroid* dari kelompok k

X_i = data ke- i dalam kelompok k

Prosedur perlu dijabarkan menurut tipe penelitiannya. Bagaimana penelitian dilakukan dan data akan diperoleh, perlu diuraikan dalam bagian ini.

Jarak inter-kelompok didefinisikan:

$$d_{kl} = \|C_k - C_l\| \quad (11)$$

dengan

d_{kl} = jarak antara kelompok k dan kelompok l

C_k = centroid kelompok k

C_l = centroid kelompok l

Indeks *Davies Bouldin* dihitung dengan menggunakan rumus:

$$DB(nc) = \frac{1}{nc} \sum_{k=1}^{nc} \max \left(\frac{sc(k)+sc(l)}{d_{kl}} \right) \quad (12)$$

dengan

nc = banyak kelompok

Skema pengelompokan yang optimal menurut Indeks *Davies Bouldin* ialah yang memiliki nilai Indeks *Davies Bouldin* minimal.

HASIL PENELITIAN DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari buku saku Ikhtisar Data Pendidikan Tahun 2016/2017. Penelitian ini terbatas pada persoalan peningkatan angka partisipasi pendidikan dasar dan menengah sehingga akan digunakan enam indikator pendidikan, yaitu APK SD (V_1), APK SMP (V_2), APK SM sederajat (V_3), APM SD (V_4), APM SMP (V_5), dan APM SM sederajat (V_6) sebagai variabel untuk pengolahan data lebih lanjut. Data yang digunakan adalah data nilai indikator pendidikan setiap provinsi di Indonesia tahun 2016. Analisis data dilakukan dengan bantuan RStudio versi terbaru yang dirilis pada tanggal 31 Maret 2017.

Proses analisis data dimulai dengan melakukan pendeteksian *outliers* dengan melihat nilai *z-score*. Berdasarkan hasil perhitungan, diperoleh bahwa nilai *z-score* untuk semua data kurang dari 3, maka disimpulkan bahwa data tidak mengandung *outliers*. Sehingga semua data tetap diperhitungkan untuk pengolahan data selanjutnya.

Kriteria ukuran kesamaan yang digunakan dalam penelitian ini adalah ukuran jarak. Ukuran jarak euclid dipilih sebagai ukuran kesamaan dalam pengelompokan data,

kemudian dilakukan perhitungan dan diperoleh matriks jarak.

Data yang digunakan dalam penelitian ini (nilai APK dan APM) disajikan dalam bentuk persen sehingga tidak terdapat perbedaan skala. APK adalah proporsi anak sekolah pada suatu jenjang tertentu terhadap penduduk pada kelompok usia tertentu. Sedangkan APM merupakan persentase jumlah anak yang sedang bersekolah pada jenjang pendidikan yang sesuai dengan usianya terhadap jumlah seluruh anak pada kelompok usia sekolah yang bersangkutan. Oleh karena itu, standarisasi data tidak dilakukan dalam penelitian ini.

Dalam skripsi ini terdapat 204 data dengan 34 objek penelitian dan 6 variabel. Tiga puluh empat objek penelitian adalah representasi dari 34 provinsi yang ada di Indonesia. Nilai korelasi antarvariabel yang digunakan dapat dilihat selengkapnya pada Tabel 1.

Tabel 1. Nilai Korelasi Antar Variabel

Variabel	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	0,802	0,469	0,289	0,102	0,069
V_2	0,802	1	0,648	0,699	0,361	0,447
V_3	0,469	0,648	1	0,884	0,798	0,767
V_4	0,289	0,699	0,884	1	0,741	0,843
V_5	0,102	0,361	0,798	0,741	1	0,947
V_6	0,069	0,447	0,767	0,843	0,947	1

Dari Tabel 1 diketahui bahwa terdapat 4 pasangan variabel yang mempunyai korelasi yang cukup besar. Yaitu, variabel V_1 dengan V_2 sebesar 0,802; variabel V_3 dengan V_4 sebesar 0,884; variabel V_4 dengan V_6 sebesar 0,843; dan variabel V_4 dengan V_5 sebesar 0,947. Karena data mengandung korelasi maka dilakukan proses analisis komponen utama. Karena data mengandung skala yang sama, maka digunakan matriks kovarian. Nilai kovarian seperti pada Tabel 2.

Tabel 2. Nilai Kovarian Antar Variabel

Variabel	V_1	V_2	V_3	V_4	V_5	V_6
V_1	24,2349	16,9473	17,4111	9,8172	4,8188	2,6105
V_2	16,9473	18,4263	20,9696	20,9567	14,8762	14,7798

V_3	17,4111	20,9696	56,7919	46,5546	57,7085	44,5265
V_4	9,8172	20,9567	46,5546	48,7802	49,6884	45,3548
V_5	4,8188	14,8762	57,7085	49,6884	92,0902	70,0065
V_6	2,6105	14,7798	44,5265	45,3548	70,0065	59,3911

V_6	0,1572
-------	--------

Selanjutnya, dicari akar ciri (*eigen value*) matriks kovarian dan diperoleh akar ciri matriks kovarian seperti pada Tabel 3.

Tabel 3. Akar Ciri Matriks Kovarian

Variabel	Akar Ciri
V_1	232,9353
V_2	43,7279
V_3	13,4614
V_4	8,4579
V_5	0,9749

Kemudian dihitung vektor ciri yang dihasilkan dari akar ciri. Sehingga diperoleh fungsi komponen utamanya seperti berikut.

$$\begin{aligned}
 Y_1 &= -0,0908V_1 + 0,6503V_2 + 0,4971V_3 - 0,2025V_4 + 0,3537V_5 + 0,3945V_6 \\
 Y_2 &= 0,1667V_1 + 0,4688V_2 + 0,0779V_3 + 0,4804V_4 + 0,5294V_5 + 0,4851V_6 \\
 Y_3 &= -0,4549V_1 + 0,3076V_2 + 0,0377V_3 + 0,7198V_4 + 0,1266V_5 - 0,4035V_6 \\
 Y_4 &= -0,4172V_1 + 0,1984V_2 - 0,6921V_3 + 0,0112V_4 - 0,1008V_5 + 0,5452V_6 \\
 Y_5 &= -0,5932V_1 - 0,3866V_2 + 0,5017V_3 - 0,0231V_4 - 0,4322V_5 + 0,2442V_6 \\
 Y_6 &= 0,4806V_1 + 0,2717V_2 + 0,1210V_3 + 0,4576V_4 + 0,6178V_5 + 0,2991V_6
 \end{aligned}$$

Dengan memasukkan nilai data awal ke dalam persamaan komponen utama, maka diperoleh nilai komponen utama seperti pada Tabel 4. Nilai inilah yang akan diolah pada tahap analisis kelompok selanjutnya untuk dihasilkan kelompok-kelompok.

Tabel 4. Komponen Utama Data APK dan APM setiap Provinsi di Indonesia tahun 2016

Nama Provinsi	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
DKI Jakarta	-202,3013	105,6825	29,6984	-23,5640	-6,5967	-0,7946
Jawa Barat	-176,7464	114,9952	26,6486	-22,9369	-6,0132	-1,0949
Banten	-175,4294	115,3448	29,7594	-23,3854	-4,9680	-1,1068
Jawa Tengah	-175,8904	110,6215	28,5990	-21,7815	-4,2912	-0,9749
Di Yogyakarta	-199,6823	95,0446	30,4221	-25,8850	-3,6875	-1,4860
Jawa Timur	-187,7647	110,4653	26,7199	-23,4124	-3,6791	-1,4045
Aceh	-186,1534	107,8804	30,0960	-20,2811	-4,3285	-0,4850
Sumatera Utara	-193,4741	108,1094	36,7310	-24,8649	-4,3485	-0,1892
Sumatera Barat	-192,3674	108,9800	32,9482	-25,3991	-2,8671	-1,0949
Riau	-174,0557	113,1393	31,8148	-26,8353	-4,7143	-0,6531
Kepulauan Riau	-188,3093	94,9952	31,5600	-24,6553	-4,5152	-0,0634
Jambi	-179,1165	113,6457	34,0869	-24,7176	-4,5843	-0,6166
Sumatera Selatan	-176,4131	113,1925	37,8242	-24,6475	-4,5325	-0,5452
Bangka Belitung	-165,2231	115,6183	33,9524	-23,4650	-5,7863	-1,0614
Bengkulu	-184,6057	115,0348	33,4606	-24,0753	-2,9596	-1,2636
Lampung	-176,3243	111,9447	29,5801	-22,1236	-3,3861	-1,3225
Kalimantan Barat	-173,6291	120,0523	33,0207	-19,5816	-3,2824	-0,1312
Kalimantan Tengah	-171,1329	119,4940	31,0623	-22,4354	-5,0477	-1,1201
Kalimantan Selatan	-170,4509	113,0779	29,9960	-22,7733	-4,3802	-1,4495
Kalimantan Timur	-183,6414	110,3987	33,5658	-24,8335	-5,6113	-0,9223
Kalimantan Utara	-172,0342	108,5704	33,3255	-26,8792	-4,9399	-0,9958
Sulawesi Utara	-187,0945	106,3878	38,6338	-16,8056	-6,0672	-1,2213
Gorontalo	-180,5769	114,6094	32,5889	-26,1865	-3,6144	-0,0721
Sulawesi Tengah	-186,9896	116,3701	28,1275	-18,6926	-3,6762	-0,3519
Sulawesi Selatan	-183,3339	109,2504	32,0489	-21,0733	-4,8943	-0,8552
Sulawesi Barat	-181,0573	111,6864	27,3767	-17,4725	-3,7765	-0,6802
Sulawesi Tenggara	-193,4238	105,2611	34,3441	-17,8050	-3,3804	-0,7709
Maluku	-190,1998	100,6265	36,9895	-16,5260	-4,7133	-1,0699

Maluku Utara	-192,4234	110,7192	35,9573	-19,5487	-4,3093	-1,0882
Bali	-199,2386	103,8896	32,4030	-26,7634	-4,2507	-0,8908
Nusa Tenggara Barat	-194,2076	107,0669	31,4006	-23,2789	-2,3913	-1,1776
Nusa Tenggara Timur	-174,4891	116,0580	36,9080	-22,8201	-2,8379	-1,1860
Papua	-112,7608	91,9941	31,9912	-21,8402	-3,7591	-0,7948
Papua Barat	-174,8985	114,5059	43,6415	-24,3014	-4,0301	-1,2707

Dalam penelitian ini dipilih dua metode, yaitu *single linkage* dan *complete linkage*. Kedua metode dipilih karena memiliki ukuran yang berbeda dalam pembentukan kelompok. Untuk mengelompokkan data menggunakan metode AHC pada Rstudio dengan perintah `hclust`. Selanjutnya, akan dilihat perbedaan pembentukan kelompok pada kedua metode.

```
> cluster_single <-hclust
  (matriks_jarak, method="single")
> cluster_complete <-hclust
  (matriks_jarak, method="complete")
```

Secara visual hal ini sering digambarkan dalam sebuah diagram pohon (dendrogram) dengan perintah `plot` seperti yang terlihat pada Gambar 2 dan Gambar 3.

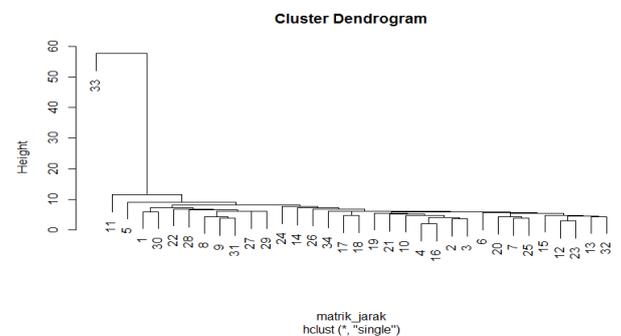
```
> plot(cluster_single)
> plot(cluster_complete)
```

Dendogram merupakan visualisasi kelompok. Sumbu vertikal menunjukkan jarak (*distance*) dimana kelompok digabung dan sumbu horizontal menunjukkan nomer identitas data. Selain digunakan untuk menunjukkan anggota kelompok yang ada, dendogram juga dapat digunakan untuk menentukan jumlah kelompok yang terbentuk. Dari Gambar 2 jika ingin membuat 2 kelompok, maka kelompok I beranggotakan data 33, sementara kelompok II beranggotakan data 1 sampai data 32 dan data 34. Jika ingin membuat 3 kelompok, maka kelompok I beranggotakan data 33, kelompok II beranggotakan data 11, dan kelompok III beranggotakan data 1 sampai data 32, kecuali data 11, dan data 34. Jika ingin membuat 4 kelompok, maka kelompok I beranggotakan data 33, kelompok II beranggotakan data 11, kelompok III beranggotakan data 5, dan kelompok IV beranggotakan data selain data 5, data 11, dan data 33. Dan jika ingin membuat 5 kelompok, maka kelompok I beranggotakan data

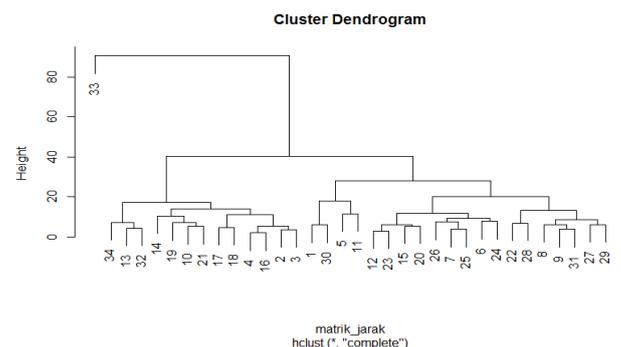
33, kelompok II beranggotakan data 11, kelompok III beranggotakan data 5, kelompok IV beranggotakan data 1, data 8, data 9, data 22, data 27, data 28, data 29, data 30, data 31. Sementara kelompok V beranggotakan data selain yang telah disebutkan.

Pembentukan kelompok pada metode *complete linkage* analog dengan pembentukan kelompok pada *single linkage*. Pembentukan kelompok tersebut dengan melihat dendrogram pada Gambar 3.

Setelah hasil pengelompokan menggunakan metode *single linkage* dan *complete linkage* diperoleh, kemudian dihitung nilai Indeks *Davies Bouldin* dengan menggunakan Persamaan (10), (11), dan (12). Selanjutnya diperoleh nilai Indeks *Davies Bouldin* seperti tampak pada Tabel 5.



Gambar 2. Dendrogram Hierarchical Clustering menggunakan Single Linkage



Gambar 3. Dendrogram Hierarchical Clustering menggunakan Complete Linkage

Tabel 5. Nilai Indeks Davies Bouldin pada Metode Single Linkage dan Complete Linkage

Jumlah Kelompok	Nilai Indeks <i>Davies Bouldin</i>	
	Metode <i>Single Linkage</i>	Metode <i>Complete Linkage</i>
2	0,0759	0,0759
3	0,2133	0,3014
4	0,1474	0,2517
5	0,1481	0,2381

Hasil pada Tabel 5 menunjukkan bahwa dari kedua metode pengelompokan hierarki yang digunakan dalam penelitian ini diperoleh jumlah kelompok optimal adalah 2 kelompok dengan nilai indeks Davies Bouldin sebesar 0,0759. Nilai tersebut sama baik pada metode single linkage maupun metode complete linkage. Nilai indeks Davies Bouldin yang sama tersebut menunjukkan bahwa kedua metode memiliki tingkat kinerja yang sama dan tidak ada perbedaan hasil dalam penerapannya. Profil setiap kelompok dapat dilihat pada Tabel 6.

Tabel 6. Profil setiap Kelompok

Jenis Kelompok	Rata-rata					
	V_1	V_2	V_3	V_4	V_5	V_6
Kelompok I	88,75	72,3	64,93	42,86	49,31	33,24
Kelompok II	107,34	93,37	101,95	75,45	84,47	62,07

Dengan melihat hasil pada Tabel 6, diperoleh pengelompokan provinsi berdasarkan tingkat partisipasi pendidikan, yaitu kelompok dengan tingkat partisipasi pendidikan rendah dan kelompok dengan tingkat partisipasi pendidikan tinggi. Pengelompokan ini diilustrasikan oleh Gambar 4.



Gambar 4. Pengelompokan Provinsi di Indonesia dengan Kelompok Optimal

Daerah yang berwarna kuning adalah daerah yang termasuk dalam kelompok I. Sementara daerah berwarna biru adalah daerah-daerah-daerah yang termasuk ke dalam kelompok II. Kelompok I adalah kelompok dengan tingkat partisipasi pendidikan rendah. Nilai APK SD, APM SD, APK SMP, APM SMP, APK SM sederajat, dan APM SM sederajat kelompok I tidak memenuhi sasaran yang ditargetkan pemerintah pada tahun 2016. Nilai-nilai tersebut memiliki rentang yang sangat jauh dengan nilai yang ditargetkan. Kelompok I hanya beranggotakan Provinsi Papua atau mewakili 2,94% wilayah Indonesia. Wilayah tersebut sudah seharusnya menjadi prioritas utama pemerintah dalam upaya penyetaraan capaian pendidikan. Kelompok II adalah kelompok dengan tingkat partisipasi pendidikan tinggi. Nilai APK SD, APM SD, APM SMP, APK SM sederajat, dan APM SM sederajat kelompok II hampir memenuhi sasaran yang ditargetkan pemerintah pada tahun 2016. Bahkan nilai APK SMP melampaui nilai yang ditargetkan. Meskipun demikian, dari nilai APK dan APM yang dicapai dapat diketahui bahwa tingkat partisipasi pendidikan di Indonesia sudah baik karena 97,06% wilayah di Indonesia termasuk dalam kelompok II dimana wilayah tersebut mampu menampung penduduk usia SD, SMP, dan SM sederajat sesuai target yang diinginkan pemerintah daerah tersebut dan sudah banyak penduduk usia SD, SMP, dan SM sederajat yang memanfaatkan fasilitas yang telah disediakan pemerintah. Pemerintah telah menyediakan fasilitas pendidikan sebagai sarana meningkatkan kualitas sumber daya manusia Indonesia dan sudah dimanfaatkan dengan baik di hampir seluruh wilayah di Indonesia.

SIMPULAN DAN SARAN

Simpulan

Setelah dilakukan perhitungan dalam penerapan optimalisasi jumlah kelompok menggunakan metode *single linkage* dan *complete linkage* dengan indeks *Davies Bouldin* yang diterapkan pada kasus data indikator

pendidikan di Indonesia tahun 2016, maka dapat disimpulkan bahwa pada metode *single linkage* dan *complete linkage* diperoleh kelompok optimal yang sama, yaitu sebanyak 2 kelompok dengan indeks *Davies Bouldin* sebesar 0,0759. Dari kelompok yang terbentuk diperoleh tingkat partisipasi pendidikan, yaitu kelompok dengan tingkat partisipasi pendidikan rendah (kelompok I) dan kelompok dengan tingkat partisipasi pendidikan tinggi (kelompok II). Nilai indeks *Davies Bouldin* yang sama yang ditunjukkan oleh metode *single linkage* dan *complete linkage* menunjukkan bahwa kedua metode memiliki tingkat kinerja yang sama dan tidak ada perbedaan hasil dalam penerapannya.

Saran

Bagi peneliti yang juga ingin melakukan pengelompokan data dengan jumlah kelompok optimal maka dapat dilakukan penelitian pada metode-metode analisis kelompok dan indeks-indeks validitas yang lain, mengingat cakupan metode analisis kelompok dan indeks validitas yang cukup banyak serta dapat dikembangkan dengan mengaplikasikan pada bidang ilmu yang lain.

DAFTAR PUSTAKA

- Amponsah, J., Adamtey, N., Elegba, W., & et al. (2017). *In situ morphometric characterization of aframomum melegueta accessions in ghana*. Diambil dari www.aobplants.oxfordjournals.org pada tanggal 16 Maret 2017.
- Aprilia A.P, N. W., Srinadi, I. G., & Sari, K. (2016). *Pengelompokan desa/keurahan di Kota Denpasar menurut indikator pendidikan*. Diambil dari <https://ojs.unud.ac.id/index.php/mtk/article/view/21296> pada tanggal 30 Maret 2017.
- Ary, M. 2015. *Pengklasifikasian karakteristik mahasiswa baru dalam memilih program studi menggunakan analisis cluster*. Diambil dari <https://www.researchgate.net/publication/290315357> pada tanggal 14 Maret 2017.
- Badan Pengawasan Keuangan dan Pembangunan. (2015). *Buku I RPJMN 2015-2019*. Diambil dari: <http://www.bpkp.go.id> pada tanggal 7 Juni 2017.
- Badan Pengawasan Keuangan dan Pembangunan. (2015). *Buku II RPJMN 2015-2019*. Diambil dari: <http://www.bpkp.go.id> pada tanggal 7 Juni 2017.
- Badan Pusat Statistik. (2017). *Angka partisipasi kasar (APK)*. Diambil dari <https://sirusa.bps.go.id/index.php?r=indikator/view&id=8> pada tanggal 7 April 2017.
- Badan Pusat Statistik. (2017). *Angka partisipasi murni (APM)*. Diambil dari <https://sirusa.bps.go.id/index.php?r=indikator/view&id=9> pada tanggal 7 April 2017.
- Badan Pusat Statistik. (2017). *Angka partisipasi sekolah (APS)*. Diambil dari <https://sirusa.bps.go.id/index.php?r=indikator/view&id=10> pada tanggal 7 April 2017.
- Gudono. (2011). *Analisis data multivariat (edisi pertama)*. Yogyakarta: BPEE.
- Hair, J. F., Black, W. C., Babin, B. J., & et al. (2009). *Multivariate data analysis (7th ed.)*. Upper Saddle River: Prentice Hall.
- Hidayat, A. (2012, Juli 1). *Uji pearson product moment dan asumsi klasik*. Dipetik Juli 2017, 11, dari <https://www.statistikian.com/2012/07/pearson-dan-asumsi-klasik.html>
- Hidayat, A. (2016, 11 11). *Pengertian multikolinearitas dan dampaknya*. Diambil dari <https://www.statistikian.com/2016/11/multikolinearitas.html> pada tanggal 14 Juni 2017.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis (6th Ed)*. New Jersey: Prentice International Inc.
- Kementerian Pendidikan dan Kebudayaan. (2017). *Ikhtisar data pendidikan tahun 2016/2017*. (2016). Diambil dari http://publikasi.data.kemdikbud.go.id/upload/Dir/isi_FC1DCA36-A9D8-4688-8E5F-0FB5ED1DE869_.pdf pada tanggal 28 Maret 2017.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., et al. (2017). *Package 'cluster'*.

- Diambil dari <https://cran.r-project.org> pada tanggal 15 April 2017.
- Maylana, R. L. (2014). *Beberapa metode pautan pada analisis kelompok menggunakan jarak euclidean dan square euclidean*. Diambil dari <http://statistik.studentjournal.ub.ac.id> pada tanggal 30 Maret 2017.
- Odilia, Yim, & Kylee T. Ramdeen. 2015. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *The Quantitative Methods for Psychology*, 11:8-21.
- Prasetyo, E. 2012. *Data mining : konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.
- Rufaidah, A., & Effindi, M. A. (2017). Analisis komponen utama pada penerapan aplikasi pembelajaran metode glenn doman. *Jurnal Ilmiah Edutic*, 3:107-112.
- Santoso, S. (2010). *Statistika multivariat*. Jakarta: PT Alex Media Komputindo .
- Simamora, B. (2005). *Analisis multivariat pemasaran (edisi pertama)*. Jakarta: PT Gramedia Pustaka.
- Sinurat, B. (2014). *Pengelompokan sekuens DNA menggunakan metode k-means dan fitur n-mers frequency*. Diambil dari <http://e-repository.ipb.ac.id/handle/123456789/69384> pada tanggal 29 Maret 2017
- Suparlan, I. A. 1983. *Statistika Sosial*. Jakarta : CV. Rajawali.
- Yulianto, S., & Hidayatullah, K. H. (2014). Analisis klaster untuk pengelompokan kabupaten/kota di Provinsi Jawa Tengah Berdasarkan indikator kesejahteraan rakyat. *Statistika*, 2:56-63.
- Yulianto, S., & Putriana, U. (Mei 2013). Analisis komponen utama (AKU) untuk pengelompokan area pelayanan dan jaringan (APJ) Daerah Jawa Tengah dan D. I. Yogyakarta. *Buletin Median*, 5:30-39.