

## PENCARIAN *CLUSTER OPTIMUM* PADA *SINGLE LINKAGE*, *COMPLETE LINKAGE* DAN *AVERAGE LINKAGE*

Nur Asiska, Neva Satyahadewi, Hendra Perdana

### INTISARI

*Analisis cluster merupakan teknik multivariat yang digunakan untuk mengelompokkan objek/kasus (responden) menjadi kelompok-kelompok yang lebih kecil dimana setiap kelompok berisi objek/kasus yang mirip satu sama lain. Dalam analisis cluster dua prosedur yang digunakan untuk pengelompokan yaitu analisis cluster hierarki dan non-hierarki. Penentuan jumlah cluster optimum yang tepat untuk digunakan diperoleh melalui identifikasi pola pergerakan varian pada cluster yang mencapai global optimum. Penemuan posisi cluster yang mencapai global optimum pada pola pergerakan varian diperoleh melalui penerapan metode valley-tracing. Pada penelitian, digunakan penerapan analisis cluster hierarki untuk mengelompokkan kabupaten/kota di Kalimantan Barat berdasarkan indikator IPM. Dari hasil analisis pembentukan cluster optimum pada metode single linkage diperoleh cluster optimum sebanyak 4 cluster. Pada metode complete linkage diperoleh cluster optimum sebanyak 5 cluster. Metode average linkage menghasilkan cluster optimum sebanyak 5 cluster*

**Kata Kunci** : *Analisis Multivariat, Analisis Cluster, Cluster Optimum*

### PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan tolak ukur keberhasilan pembangunan manusia yang mencakup tiga dimensi kehidupan manusia, yaitu umur panjang dan hidup sehat, pengetahuan serta standar hidup layak. Umur panjang dan hidup sehat diukur melalui Angka Harapan Hidup (AHH). Pengetahuan diukur melalui Rata-Rata Lama Sekolah (RLS) dan Harapan Lama Sekolah (HLS). Sedangkan standar hidup layak diukur melalui Pengeluaran Per kapita (PP). IPM merupakan indikator yang digunakan untuk melihat perkembangan pembangunan dalam jangka panjang. Terdapat dua aspek yang perlu diperhatikan untuk melihat kemajuan pembangunan manusia, yaitu kecepatan dan status pencapaian. Pembangunan manusia di Kalimantan Barat terus mengalami kemajuan selama periode 2010 hingga 2017. Hal ini ditandai dengan meningkatnya nilai IPM Kalimantan Barat dari 61,97% pada tahun 2010 menjadi 66,26% pada tahun 2017 [1].

Salah satu prasyarat keberhasilan program-program pembangunan sangat tergantung pada ketepatan pengidentifikasian target grup dan target area. Oleh karena itu, sangat penting mempertimbangkan pengelompokan dari 14 kabupaten/kota di Kalimantan Barat berdasarkan indikator IPM. Analisis *cluster* merupakan analisis yang tepat digunakan untuk penelitian ini. Analisis *cluster* merupakan teknik multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik-karakteristik yang dimilikinya. Secara umum terdapat dua prosedur pengelompokan yang digunakan dalam analisis *cluster* yaitu metode hierarki dan non hierarki. Metode hierarki terdiri atas *single linkage*, *complete linkage*, *average linkage*, *centroid linkage* serta *ward's*. Sedangkan pada metode non hierarki yaitu *k-means*. Dalam analisis *cluster*, belum ada dasar yang kuat mengenai banyaknya *cluster* yang tepat untuk digunakan. Banyaknya *cluster* optimum diperoleh ketika *cluster* mencapai global optimum melalui pola pergerakan varian. Metode yang digunakan untuk menemukan global optimum pada pergerakan varian yaitu *valley-tracing* [2].

Penelitian ini bertujuan untuk membentuk *cluster* pada *single linkage*, *complete linkage* dan *average linkage*. Identifikasi pola pergerakan varian dari *single linkage*, *complete linkage* dan *average linkage* melalui penerapan *valley-tracing* untuk mencari *cluster* optimum pada kabupaten/kota di Kalimantan Barat berdasarkan indikator IPM tahun 2017.

Data yang digunakan merupakan data indikator IPM di Kalimantan Barat tahun 2017 yaitu AHH, RRLS, HLS, PP yang diperoleh dari Badan Pusat Statistik (BPS). Langkah pertama yang dilakukan adalah standarisasi data menggunakan *z-score* dan menghitung jarak *euclidean*. Selanjutnya analisis *cluster* hierarki yaitu *single linkage*, *complete linkage*, *average linkage*. Langkah berikutnya analisis kepadatan *cluster* pada *single linkage*, *complete linkage*, *average linkage*, menentukan *cluster* yang mencapai global optimum melalui identifikasi pola pergerakan varian pada *single linkage*, *complete linkage*, *average linkage* dengan penerapan metode *valley-tracing*. Kemudian mencari perbedaan nilai tinggi yang maksimum pada *cluster* yang mencapai global optimum.

## ANALISIS CLUSTER

Analisis *cluster* merupakan teknik multivariat yang digunakan untuk mengelompokkan objek menjadi kelompok-kelompok yang lebih kecil dimana setiap kelompok berisi objek yang mirip satu sama lain [3].

### 1. Standarisasi Data

Proses standarisasi dilakukan apabila diantara variabel-variabel yang diteliti terdapat variabilitas satuan. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis *cluster* menjadi tidak valid. Proses standarisasi dilakukan dengan transformasi pada data asli sebelum dianalisis lebih lanjut. Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk *z-score* [3].

$$z = \frac{x - \bar{x}}{s} \quad (1)$$

Dengan  $x$  : Data

$\bar{x}$  : Rata-rata data

$s$  : Standar deviasi data

### 2. Mengukur Kemiripan Antar Objek

Sesuai dengan prinsip *cluster* yaitu mengelompokkan objek yang memiliki kemiripan, maka dilakukan pengukuran seberapa jauh ada kesamaan kesamaan antar objek [3]. Terdapat beberapa macam ukuran jarak yang biasa dipakai dalam analisis *cluster*, diantaranya yaitu jarak *euclidean*. Jarak *euclidean* adalah jarak antara dua objek dari  $p$  dimensi pengamatan. Terdapat objek pertama yang diamati adalah  $X = (x_1, x_2, \dots, x_p)$  dan  $Y = (y_1, y_2, \dots, y_p)$ , jarak *euclidean* dapat di hitung dengan rumus:

$$d_{(XY)} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (2)$$

Dengan  $d_{(XY)}$  : Jarak *euclidean* antara objek  $X$  dengan objek  $Y$

$p$  : Banyaknya variabel yang diamati

$x_j$  : Nilai  $j$  pada objek  $X$

$y_j$  : Nilai  $j$  pada objek  $Y$

### 3. Memilih Suatu Prosedur Analisis Cluster

Prosedur yang digunakan untuk melakukan pengelompokan pada analisis *cluster* terbagi menjadi dua yaitu metode hierarki dan non-hierarki [3]. Secara umum langkah-langkah dalam metode *cluster* hierarki tipe *agglomerative* untuk membentuk kelompok dari  $N$  objek adalah sebagai berikut:

- a) Dimulai dengan  $n$  *cluster*, dimana masing-masing memuat satu kesatuan. Jika terdapat matriks  $n \times n$  dengan jarak  $D = \{d_{(XY)}\}$ .

- b) Mencari matriks jarak untuk pasangan *cluster* terdekat. Misalkan pasangan *cluster* paling mirip objek  $U$  dan  $V$  maka  $D = \{d_{(UV)}\}$ , sehingga  $U$  dan  $V$  dipilih.
- c) Menggabungkan *cluster*  $U$  dan  $V$  menjadi *cluster* baru ( $UV$ ) perbarui *entry* matriks dengan cara:
- 1) Menghapus baris dan kolom sesuai dengan *cluster*  $U$  dan  $V$
  - 2) Menambahkan baris dan kolom dengan memberikan nilai jarak antara *cluster* baru ( $UV$ ) dan semua sisa *cluster*.
- d) Mengulangi langkah (b) dan (c) sebanyak  $(n - 1)$  kali. (Semua objek akan berada dalam *cluster* tunggal pada berakhirnya algoritma).

Metode *cluster* hierarki tipe *agglomerative* yang digunakan dalam penelitian ini adalah sebagai berikut:

a). *Single Linkage*

Metode *single linkage* menggunakan prinsip jarak minimum. Pengelompokan dimulai dengan mencari nilai minimum dari  $D = \{d_{(XY)}\}$  dan menggabungkan objek-objek yang saling berdekatan, misal  $U$  dan  $V$  untuk mendapatkan *cluster* ( $UV$ ). Selanjutnya pada langkah (c) dari algoritma yang dijelaskan sebelumnya, jarak antar *cluster* ( $UV$ ) dengan objek lain, misal  $W$  adalah:

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (3)$$

b). *Complete Linkage*

Metode *complete linkage* merupakan kebalikan dari pendekatan yang digunakan pada *single linkage*. Prinsip jarak yang digunakan adalah jarak terjauh antar objek. Pengelompokan dimulai dengan mencari nilai minimum dari  $D = \{d_{(XY)}\}$  dan menggabungkan objek-objek yang saling berdekatan, misal  $U$  dan  $V$  untuk mendapatkan *cluster* ( $UV$ ). Selanjutnya pada langkah (c) dari algoritma yang dijelaskan sebelumnya, jarak antar *cluster* ( $UV$ ) dengan objek lain, misal  $W$  adalah:

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (4)$$

c). *Average Linkage*

Metode *average linkage* mengikuti prosedur yang sama dengan kedua metode sebelumnya. Prinsip ukuran jarak yang digunakan adalah jarak rata-rata antar tiap pasangan objek yang mungkin. Pengelompokan dimulai dengan mencari nilai minimum dari  $D = \{d_{(XY)}\}$  dan menggabungkan objek-objek yang saling berdekatan, misal  $U$  dan  $V$  untuk mendapatkan *cluster* ( $UV$ ). Selanjutnya pada langkah (c) dari algoritma yang dijelaskan sebelumnya, jarak antar *cluster* ( $UV$ ) dengan objek lain, misal  $W$  adalah:

$$d_{(UV)W} = \text{average}(d_{UW}, d_{VW}) \quad (5)$$

## ANALISIS CLUSTER OPTIMUM

Penentuan *cluster* yang tepat untuk digunakan pada analisis *cluster* tidak ada aturan yang baku. *Cluster* optimum diperoleh ketika *cluster* mencapai global optimum [2].

### 1. Analisis Kepadatan Cluster

Kepadatan suatu *cluster* ditentukan melalui nilai *variance within cluster* dan *variance between cluster*. *Cluster* ideal mempunyai nilai *variance within cluster* minimum dan nilai *variance between cluster* maksimum [2]. Jika ada beberapa *cluster*  $c_i$ , dimana  $i = 1, 2, \dots, k$  dan masing-masing *cluster* memiliki anggota  $x_i$ , dimana  $i = 1, 2, \dots, n$  dan  $n$  adalah jumlah anggota dari masing-masing *cluster*, dan  $\bar{x}_p$  adalah rata-rata dari *cluster*  $p$ . Kemudian Varian dari tiap *cluster*  $p$  ( $\delta_p^2$ ) dapat dihitung menggunakan rumus:

$$\delta_p^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_p)^2 \quad (6)$$

Jika  $N$  adalah jumlah anggota di semua *cluster*, nilai *variance within cluster* ( $V_w^2$ ) dapat dicari dengan rumus:

$$V_w^2 = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) \delta_i^2 \quad (7)$$

Kemudian, jika terdapat nilai rata-rata dalam setiap *cluster* ke- $i$  ( $\bar{x}$ ) maka komponen dari setiap *cluster* berbeda maka *variance between cluster* ( $V_b^2$ ) dapat dihitung menggunakan rumus:

$$V_b^2 = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 \quad (8)$$

Karena *cluster* memiliki  $V_w^2$  minimum  $V_b^2$  maksimum, itu artinya *cluster* ideal memiliki  $V$  minimum, dimana:

$$V = \frac{V_w^2}{V_b^2} \times 100\% \quad (9)$$

## 2. Penentuan Cluster Optimum

Penentuan banyaknya *cluster* optimum ditentukan melalui pola pergerakan varian. Identifikasi pergerakan varian merupakan metode untuk memperoleh *cluster* mencapai yang global optimum [2]. Posisi yang mungkin untuk menemukan global optimum menggunakan penerapan metode *valley-tracing*. Pada *valley-tracing* didefinisikan bahwa kemungkinan mencapai global optimum terletak pada tahap ke- $i$ , jika memenuhi syarat:

$$(V_{i-1} \geq V_i) \cap (V_{i+1} > V_i) \quad (10)$$

Dimana  $V_i$  adalah varian pada *cluster* ke- $i$  hasil perbandingan  $V_w^2$  dan  $V_b^2$ ,  $i : 1, 2, \dots, n$  tahap pembentuk *cluster*.

Untuk menentukan *cluster* optimum pada beberapa *cluster* yang mencapai global optimum berasal dari nilai maksimum dari nilai tinggi ( $\partial$ ). Rumus nilai tinggi ( $\partial$ ) pada setiap tahap sebagai berikut:

$$\partial = (V_{i+1} + V_{i-1}) - (2 \cdot V_i) \quad (11)$$

Keakuratan suatu metode hierarki menggunakan *valley-tracing* digunakan rumus:

$$\varphi = \frac{\max(\partial)}{\text{nilai\_terdekat\_ke\_max}(\partial)} \quad (12)$$

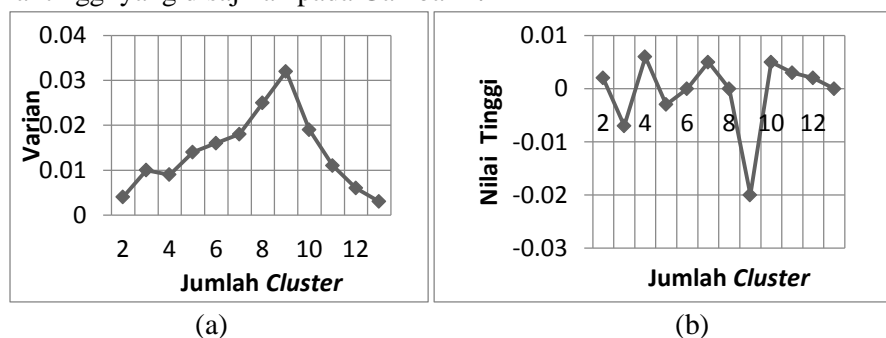
Dimana *nilai\_terdekat\_ke\_max*( $\partial$ ) adalah Nilai ( $\partial$ ) sebelum  $\max(\partial)$ . Nilai  $\varphi \geq 2$  menunjukkan *cluster* yang terbentuk merupakan *cluster* yang *well-separated* (terpisah dengan baik). Jika nilai  $\varphi$  bertanda negatif berarti global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik [4].

## HASIL DAN PEMBAHASAN

Data sekunder yang digunakan dalam penelitian ini adalah data indikator IPM di Kalimantan barat tahun 2017 yang diperoleh dari Badan Pusat Statistik Kalimantan Barat. Hasil *cluster* optimum dari metode *single linkage*, *complete linkage* dan *average linkage* berdasarkan pola pergerakan varian dan nilai tinggi adalah sebagai berikut:

### 1. Cluster Optimum Pada Single Linkage

Penentuan *cluster* optimum pada *single linkage* melalui identifikasi pola pergerakan varian dan maksimum nilai tinggi yang disajikan pada Gambar 1.

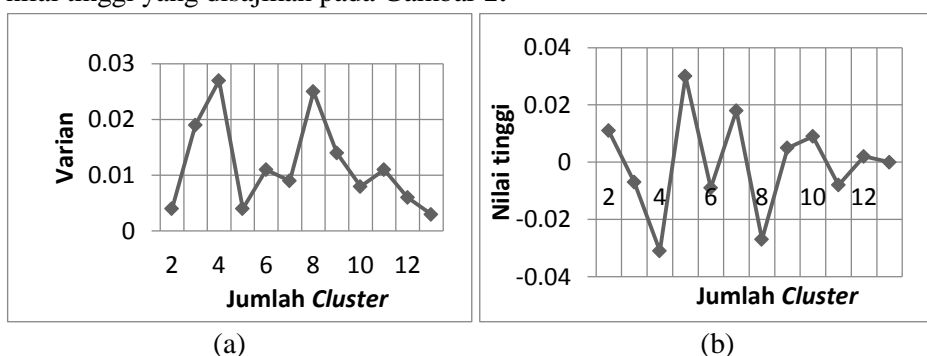


Gambar 1. Pola Pergerakan Varian (a) dan Nilai Tinggi *Single Linkage* (b)

Pada Gambar 1 dapat dilihat bahwa pola pergerakan varian pada metode *single linkage cluster* yang mencapai global optimum terletak pada *cluster* 4. Perbedaan nilai tinggi yang maksimum yaitu ketika jumlah *cluster* sebanyak 4 dengan perbedaan nilai tinggi sebesar 0,006. Hal ini berarti bahwa jumlah *cluster* yang optimum pada *single linkage* berdasarkan indikator IPM pada kabupaten/kota di Kalimantan Barat tahun 2017 adalah sebanyak 4 *cluster*. Keakuratan dari suatu metode pembentukan *cluster* optimum pada *single linkage* menggunakan Persamaan 12, diperoleh nilai  $\varphi$  yaitu  $-0,857$ . Nilai  $\varphi$  bertanda negatif artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

## 2. Cluster Optimum Pada Complete Linkage

Penentuan *cluster* optimum pada *complete linkage* melalui identifikasi pola pergerakan varian dan maksimum nilai tinggi yang disajikan pada Gambar 2.

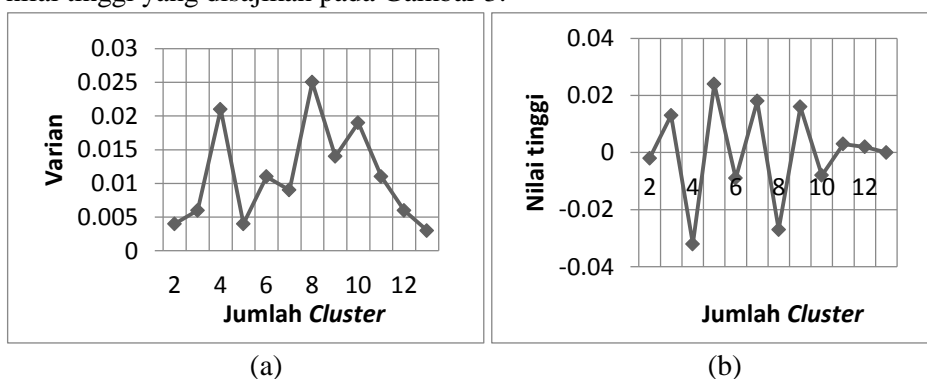


Gambar 2. Pola Pergerakan Varian (a) dan Nilai Tinggi *Complete Linkage* (b)

Pada Gambar 2 dapat dilihat bahwa pola pergerakan varian pada metode *complete linkage cluster* yang mencapai global optimum terletak pada *cluster* 5,7 dan 10. Perbedaan nilai tinggi yang maksimum yaitu ketika jumlah *cluster* sebanyak 5 dengan perbedaan nilai tinggi sebesar 0,003. Hal ini berarti bahwa jumlah *cluster* yang optimum pada *complete linkage* berdasarkan indikator IPM pada kabupaten/kota di Kalimantan Barat tahun 2017 adalah sebanyak 5 *cluster*. Keakuratan dari suatu metode pembentukan *cluster* optimum pada *complete linkage* menggunakan Persamaan 12, diperoleh nilai  $\varphi$  yaitu  $-0,967$ . Nilai  $\varphi$  bertanda negatif artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

## 3. Cluster Optimum Pada Average Linkage

Penentuan *cluster* optimum pada *average linkage* melalui identifikasi pola pergerakan varian dan maksimum nilai tinggi yang disajikan pada Gambar 3.



Gambar 3. Pola Pergerakan Varian (a) dan Nilai Tinggi *Average Linkage* (b)

Pada Gambar 3 dapat dilihat bahwa pola pergerakan varian pada metode *average linkage cluster* yang mencapai global optimum terletak pada *cluster* 5,7 dan 9. Perbedaan nilai tinggi yang maksimum yaitu ketika jumlah *cluster* sebanyak 5 dengan perbedaan nilai tinggi sebesar 0,024. Hal ini berarti

bahwa jumlah *cluster* yang optimum pada *average linkage* berdasarkan indikator IPM pada kabupaten/kota di Kalimantan Barat tahun 2017 adalah sebanyak 5 *cluster*. Keakuratan dari suatu metode pembentukan *cluster* optimum pada *average linkage* menggunakan Persamaan 12, diperoleh nilai  $\varphi$  yaitu  $-0,75$ . Nilai  $\varphi$  bertanda negatif artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

### KESIMPULAN

Berdasarkan hasil analisis pembentukan *cluster* optimum pada metode *single linkage*, *complete linkage* dan *average linkage* berdasarkan indikator IPM di Kalimantan Barat tahun 2017 diperoleh:

1. *Cluster* optimum yang terbentuk pada metode *single linkage* adalah sebanyak 4 *cluster* dengan nilai keakuratan sebesar  $-0,857$ , artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.
2. *Cluster* optimum yang terbentuk pada metode *complete linkage* adalah sebanyak 5 *cluster* dengan nilai keakuratan sebesar  $-0,976$ , artinya global optimum yang ditentukan telah benar dan *cluster* terpisah dengan baik.
3. *Cluster* optimum yang terbentuk pada metode *average linkage* adalah sebanyak 5 *cluster* dengan nilai keakuratan sebesar  $-0,75$ , artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik Kalimantan Barat. *Indeks pembangunan manusia Kalimantan Barat*. Pontianak: BPS;2017.
- [2] Arai, K., and Barakbah, A.R. Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering. *IES Proc. Pp.* 2004; 409-413.
- [3] Gunawan, Imam. *Pengantar Statistika Inferensial*, Jakarta: Rajawali Pers, 2016.
- [4] Arai, K., and Barakbah, A.R. Cluster Construction Method Based on Global Optimum Cluster Determination with The Newly Defined Moving Variance, *Reports of the Faculty of Science and Engineering Saga University*. 2007; 36(1), hal. 9-15.

NUR ASISKA : Jurusan Matematika FMIPA UNTAN, Pontianak  
nur.asiska06@gmail.com

NEVA SATYAHADEWI : Jurusan Matematika FMIPA UNTAN, Pontianak  
neva.satya@math.untan.ac.id

HENDRA PERDANA : Jurusan Matematika FMIPA UNTAN, Pontianak  
hendra.perdana@math.untan.ac.id

---