PEMBENTUKAN *CLUSTER* OPTIMUM BERDASARKAN METODE HIERARKI *DIVISIVE*

Lidia Karnelia, Dadan Kusnandar, Shantika Martha

INTISARI

Analisis cluster merupakan teknik multivariat yang mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Prosedur pengelompokan yang digunakan dalam analisis cluster yaitu metode hierarki dan non hierarki. Metode hierarki terdiri dari divisive dan agglomerative. Pembentukan jumlah cluster optimum yang tepat untuk digunakan diperoleh melalui identifikasi pola pergerakan varians pada cluster yang mencapai global optimum. Penemuan posisi cluster yang mencapai global optimum pada pola pergerakan varians diperoleh melalui penerapan metode valley-tracing. Pada penelitian, digunakan penerapan analisis cluster hierarki divisive untuk mengelompokkan 13 desa di Kecamatan Nanga Taman berdasarkan kelompok pendidikan yang ditamatkan tahun 2019. Dari hasil analisis pembentukan cluster optimum pada metode hierarki divisive, diperoleh sebanyak empat cluster. Informasi tersebut akan mempermudah pemerintah Kecamatan Nanga Taman dalam menanggulangi permasalahan pendidikan karena tidak meratanya jumlah penduduk berdasarkan kelompok pendidikan yang ditamatkan. Anggota cluster pertama Desa Nanga Taman tidak memiliki kemiripan dengan desa lainnya, karena jumlah penduduk tamatan sarjana, diploma dan SMA terbanyak. Anggota cluster kedua Desa Nanga Mentukak dan Sungai Lawak, memiliki kemiripan berdasarkan jumlah penduduk tamatan diploma dan pada kelompok pendidikan lainnya tidak jauh berbeda. Anggota cluster ketiga Desa Nanga Koman dan Lubuk Tajau, memiliki kemiripan berdasarkan jumlah penduduk tamatan sarjana dan pada kelompok pendidikan lainnya tidak jauh berbeda. Anggota cluster keempat Desa Rirang Jati, Senangak, Nanga Kiungkang, Tapang Tingang, Nanga Mongko, Nanga Engkulun, Pantok dan Meragun, memiliki kemiripan berdasarkan kelompok pendidikan tidak/belum tamat SD dan tamatan SD.

Kata Kunci : Analisis Multivariat, Pendidikan, Global Optimum, Valley-Tracing.

PENDAHULUAN

Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Analisis *cluster* mengklasifikasi objek, sehingga setiap objek yang memiliki sifat yang mirip akan dikelompokkan kedalam satu *cluster* (kelompok) yang sama [1]. Secara umum terdapat dua prosedur pengelompokan yang digunakan dalam analisis *cluster* yaitu metode hierarki dan non hierarki. Metode hierarki terdiri dari dua yaitu *divisive* dan *agglomerative*. Dalam analisis *cluster*, belum ada dasar yang kuat mengenai banyaknya *cluster* yang tepat untuk digunakan. Banyaknya *cluster* terbaik diperoleh ketika *cluster* mencapai global optimum. Identifikasi pola pergerakan varians melalui penerapan metode *valley-tracing* untuk memperoleh banyaknya *cluster* optimum [2].

Pendidikan merupakan suatu proses menyiapkan individu untuk mampu menyesuaikan diri dengan perubahan lingkungan. Pendidikan mempunyai peran penting dalam pembangunan nasional karena pendidikan merupakan salah satu cara untuk membentuk sumber daya manusia yang berkualitas, terutama di Kecamatan Nanga Taman yang terletak di Kabupaten Sekadau Provinsi Kalimantan Barat.

Dalam penelitian kelompok pendidikan yang ditamatkan di Kecamatan Nanga Taman jumlah penduduk di 13 desa di Kecamatan Nanga Taman terdiri dari 28.321 jiwa pada tahun 2019. Kelompok pendidikan yang ditamatkan terdiri dari tidak atau belum sekolah sebanyak 7.500 jiwa. Tidak atau belum tamat SD sebanyak 5.723 jiwa. Tamat SD sebanyak 9.414 jiwa. SMP/ sederajat sebanyak 2.978 jiwa. SMA/sederajat sebanyak 2.215 jiwa. Diploma I-III sebanyak 240 jiwa dan sarjana sebanyak 251 jiwa [3].

Kelompok pendidikan yang ditamatkan dari setiap desa di Kecamatan Nanga Taman tahun 2019 dapat dikelompokkan berdasarkan kemiripan karakteristik yang dimiliki, oleh karena itu metode hierarki *divisive* dapat digunakan untuk memenuhi tujuan tersebut, *cluster* optimum diperoleh berdasarkan penerapan metode *valley-tracing*, setelah diperoleh *cluster* optimum maka akan mempermudah pemerintah Kecamatan Nanga Taman, dalam menanggulangi permasalahan dalam pendidikan dikarenakan tidak meratanya jumlah penduduk berdasarkan kelompok pendidikan yang ditamatkan di Kecamatan Nanga Taman.

ANALISIS CLUSTER OPTIMUM

Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki, dimana setiap kelompok berisi objek-objek yang mirip satu sama lain [4]. Beberapa proses yang harus dilakukan dalam analisis *cluster* diantaranya merumuskan masalah, mengukur kemiripan antar objek, memilih suatu prosedur pembentukan *cluster*, analisis kepadatan *cluster* dan penentuan *cluster* optimum.

Pemilihan variabel yang akan digunakan untuk membentuk *cluster* merupakan proses penting dalam analisis *cluster*. Pengamatan terhadap p variabel pada N objek dapat disajikan dalam bentuk matriks X sebagai berikut [5]:

$$X = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{Np} \end{bmatrix} = \{x_{mj}\}, m = 1, 2, \dots, N \text{ dan } j = 1, 2, \dots, p \quad (1)$$

Dimana z_{mj} = Elemen dari matriks X yang merupakan uraian dari m objek dan j variabel

N = Banyak objek yang diamati

p = Banyak variabel yang diamati

Tujuan analisis *cluster* mengelompokkan objek yang memiliki kemiripan kedalam *cluster* yang sama, maka dilakukan pengukuran seberapa jauh ada kesamaan antar objek. Terdapat beberapa macam ukuran jarak yang biasa dipakai dalam analisis *cluster*, diantaranya jarak euklidian. Jarak euklidian adalah jarak antara dua objek dari banyak variabel pengamatan. Jarak euklidian adalah akar dari jumlah kuadrat perbedaan nilai tiap variabel. Ukuran jarak objek ke- a dengan objek ke- b disimbolkan dengan $d_{(ab)}$ dan $j=1, 2, \dots, p$. Nilai $d_{(ab)}$ dapat dihitung dengan rumus[4]:

$$d_{(a,b)} = \sqrt{\sum_{j=1}^p (x_{aj} - x_{bj})^2} \quad (2)$$

Dimana $d_{(a,b)}$ = Jarak antar objek ke- a dan objek ke- b

p = Banyak variabel yang diamati

x_{aj} = Nilai dari objek ke- a pada variabel ke- j

x_{bj} = Nilai dari objek ke- b pada variabel ke- j

Memilih suatu prosedur pembentukan *cluster* dapat dilakukan dengan dua cara yaitu metode hierarki dan non hierarki [4]. Secara umum tahapan-tahapan dalam analisis pembentukan *cluster* hierarki *divisive* menggunakan *splinter average distance methods* adalah sebagai berikut [1]:

- Dimulai dengan mengelompokkan satu *cluster* besar yang berisikan seluruh objek.
- Mengitung jarak euklidian Persamaan (1).
- Mengitung jarak rata-rata objek terhadap objek lainnya, dapat dihitung dengan rumus [6]:

$$\bar{d}_m = \frac{1}{N-1} \sum_{m=1}^{N-1} d_m \quad (3)$$

Dimana $\bar{d}_m$ = Jarak rata-rata objek ke- m

d_m = Nilai-nilai pada objek m ($m = 1, 2, \dots, N$)

$N - 1$ = Jumlah semua objek dikurangi satu, karena jarak diantara objek yang sama bernilai nol.

- d. Memisahkan objek berdasarkan jarak rata-rata terjauh sehingga terbentuk dua grup yaitu grup utama dan *splinter group* ke- k .
- e. Membentuk anggota *splinter group* ke- k dengan syarat jarak rata-rata tanpa *splinter group* ke- $k \geq$ nilai objek *splinter group* ke- k .
- f. Mengulangi langkah (c), (d) dan (e) sedemikian sehingga semua objek terbentuk grup.

Kepadatan suatu *cluster* ditentukan melalui nilai *variance within cluster* dan *variance between cluster*. *Cluster* ideal mempunyai nilai *variance within cluster* minimum dan *variance between cluster* maksimum [2]. Nilai varians dari tiap *cluster* dapat dihitung menggunakan rumus:

$$\sigma_{c_i}^2 = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} (x_i - \bar{x}_i)^2 \quad (4)$$

Dimana c_i = *Cluster* i ($i = 1, 2, \dots, k$)

$\sigma_{c_i}^2$ = Varians pada *cluster* i ($i = 1, 2, \dots, k$)

n_i = Jumlah anggota pada *cluster* i ($i = 1, 2, \dots, k$)

x_i = Rata-rata variabel objek ke- m pada *cluster* ($i = 1, 2, \dots, k$)

$\bar{x}_i$ = Rata-rata dari x_i

Nilai *variance within cluster* (σ_w^2) dan *variance between cluster* (σ_b^2) dapat dihitung menggunakan rumus:

$$\sigma_w^2 = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) \sigma_{c_i}^2 \quad (5)$$

$$\sigma_b^2 = \frac{1}{k - 1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2 \quad (6)$$

Dimana N = Jumlah anggota di semua *cluster*

k = Jumlah *cluster*

$\bar{\bar{x}}$ = Rata-rata dari $\bar{x}_i$

Nilai batasan varians (V) dapat dihitung menggunakan rumus:

$$V = \frac{\sigma_w^2}{\sigma_b^2} \times 100\% \quad (7)$$

Penentuan banyaknya *cluster* optimum ditentukan melalui pola pergerakan varians. Identifikasi pergerakan varians merupakan metode untuk memperoleh *cluster* yang mencapai global optimum [2]. Pada *Valley-tracing* didefinisikan bahwa kemungkinan mencapai global optimum terletak pada tahap ke- i , jika memenuhi syarat:

$$(V_{i-1} \geq V_i) \cap (V_{i+1} > V_i) \quad (8)$$

Dimana V_i adalah batasan varians pada *cluster* ke- i ($i = 1, 2, \dots, k$). Menentukan *cluster* optimum pada beberapa *cluster* yang mencapai global optimum berasal dari nilai maksimum dari nilai tinggi (∂). Rumus nilai tinggi (∂) pada setiap tahap sebagai berikut:

$$\partial = (V_{i+1} + V_{i-1}) - (2.V_i) \quad (9)$$

Nilai keakuratan suatu metode hierarki berdasarkan penerapan metode *valley-tracing* dapat dihitung menggunakan rumus:

$$\varphi = \frac{\max(\partial)}{\text{nilai_terdekat_max}(\partial)} \quad (10)$$

Dimana $\max(\partial)$ adalah nilai tinggi terbesar *cluster* yang mencapai global optimum. Nilai terdekat ke $\max(\partial)$ adalah nilai sebelum atau nilai sesudah $\max(\partial)$. *Cluster* yang terbentuk merupakan *cluster* yang *well-separated* (terpisah dengan baik), jika nilai φ bertanda negatif berarti global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik [7].

STUDI KASUS

Data sekunder yang digunakan dalam penelitian ini adalah data kelompok pendidikan terakhir yang ditamatkan di Kecamatan Nanga Taman tahun 2019 yang diperoleh dari kantor Kecamatan Nanga Taman. Kelompok pendidikan terakhir yang ditamatkan diantaranya, tidak/belum sekolah, tidak/belum tamat sekolah dasar (SD), tamat Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP)/Sederajat, Sekolah Menengah Atas (SMA)/Sederajat, Diploma dan Sarjana. Objek dan variabel bebas dalam penelitian disajikan pada Tabel 1.

Tabel 1. Objek dan Variabel Bebas Dalam Penelitian

Objek	Variabel Kelompok Pendidikan
1. Nanga Taman(NT)	1. Tidak/belum sekolah (X_1)
2. Rirang Jati (RJ)	2. Tidak/ belum tamat SD (X_2)
3. Senangak (Se)	3. Tamat SD (X_3)
4. Nanga Kiungkan (NKi)	4. Tamat SMP (X_4)
5. Sungai Lawak (SL)	5. Tamat SMA (X_5)
6. Tapang Tingang (TT)	6. Tamat Diploma (X_6)
7. Nanga Mongko (NKO)	7. Tamat Sarjana (X_7)
8. Nanga Engkulun (NE)	
9. Nanga Koman (NKO)	
10. Lubuk Tajau (LT)	
11. Pantok (Pa)	
12. Meragun (Me)	
13. Nanga Mentukak (NMe)	

Hasil pembentukan *cluster* optimum pada kelompok pendidikan terakhir yang ditamatkan di Kecamatan Nanga Taman tahun 2019, berdasarkan metode hierarki *divisive* menggunakan *splinter average distance methods* dinyatakan pada Tabel 2.

Tabel 2. Pembentukan Cluster Dari Metode Hierarki Divisive

Cluster	Anggota Cluster
1	{NT,RJ,Se,NKi,SL,TT,NMo,NE,NKO,LT,Pa,Me,NMe}
2	{NT}{RJ,Se,NKi,SL,TT,NMo,NE,NKO,LT,Pa,Me,NMe}
3	{NT}{NMe,SL}{RJ,Se,NKi,TT,NMo,NE,NKO,LT,Pa,Me}
4	{NT}{NMe,SL}{NKo,LT}{RJ,Se,NKi,TT,NMo,NE,Pa,Me}
5	{NT}{NMe,SL}{NKo,LT}{Pa,NMo}{RJ,Se,NKi,TT,NE,Me}
6	{NT}{NMe,SL}{NKo,LT}{Pa,NMo}{Me,TT}{RJ,Se,NKi,NE}
7	{NT}{NMe,SL}{NKo,LT}{Pa,NMo}{Me,TT}{RJ,NKi}{Se,NE}
13	{NT}{NMe}{SL}{NKo}{LT}{Pa}{NMo}{Me}{TT}{RJ}{NK}{Se}{NE}

Pembentukan *cluster* berdasarkan metode hierarki *divisive*, pertama objek dinyatakan satu *cluster* besar, setelah proses *splinter average distance methods* diperoleh dua *cluster* sampai tujuh *cluster*, 13 *cluster* diperoleh berdasarkan metode hierarki *divisive*, dimana objek-objek yang dinyatakan satu *cluster* besar akan dipisahkan sampai setiap objek tersebut dinyatakan *cluster*.

Pembentukan *cluster* optimum berdasarkan metode hierarki *divisive*, pembentukan *cluster* optimum dengan penerapan metode *valley-tracing*, pada satu *cluster* dan 13 *cluster* tidak mempunyai batasan varians dikarenakan pada satu *cluster* tidak ada nilai *variance between cluster* dan 13 *cluster* tidak ada nilai *variance within cluster*. Pembentukan *cluster* optimum melalui identifikasi pola pergerakan varians dan nilai tinggi disajikan pada Gambar 1 dan Tabel 2.

Gambar 1. Pola Pergerakan Varians

Pada Gambar 1 dapat dilihat pola pergerakan varians yang memenuhi syarat Persamaan (8), *cluster* yang mencapai global optimum terletak pada *cluster* empat dan *cluster* enam, karena lebih dari satu *cluster* yang mencapai global optimum maka ditentukanlah nilai tinggi.

Tabel 2. Nilai Tinggi

<i>Cluster</i>	Nilai tinggi
2	-0,01884939
3	0,00857931
4	0,00084758
5	-0,00098883
6	0,00064556
7	-0,00029927

Berdasarkan identifikasi menggunakan Persamaan (9), nilai tinggi maksimum pada *cluster* empat dan *cluster* enam 0,00084758, artinya jumlah *cluster* optimum pada metode hierarki *divisive* berdasarkan kelompok pendidikan terakhir yang ditamatkan di Kecamatan Nanga Taman tahun 2019, adalah sebanyak empat *cluster*. Keakuratan dari suatu metode pembentukan *cluster* optimum pada metode hierarki *divisive* menggunakan persamaan (10), diperoleh nilai ϕ yaitu -0,857159003. Nilai ϕ bertanda negatif artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

KESIMPULAN

Berdasarkan hasil pembentukan *cluster* hierarki *divisive* diperoleh empat *cluster* optimum dengan anggota *cluster* kesatu terdiri dari Desa Nanga Taman tidak memiliki kemiripan dengan desa lainnya, karena jumlah penduduk tamatan sarjana, diploma dan SMA terbanyak. Anggota *cluster* kedua terdiri dari Desa Nanga Mentukak dan Sungai Lawak, memiliki kemiripan berdasarkan jumlah penduduk tamatan diploma dan pada kelompok pendidikan lainnya tidak jauh berbeda. Anggota *cluster* ketiga

terdiri dari Desa Nanga Koman dan Lubuk Tajau, memiliki kemiripan berdasarkan jumlah penduduk tamatan sarjana dan pada kelompok pendidikan lainnya tidak jauh berbeda. Anggota *cluster* keempat Desa pantok, Nanga Mongko, Meragun, Tapang Tingang, Rirang Jati, Senangak, Nanga Kiungkang dan Nanga Engkulun, memiliki kemiripan berdasarkan kelompok pendidikan tidak/belum tamat SD dan tamatan SD. Dengan penerapan metode *valley-tracing* dengan nilai keakuratan $-0,857159003$ artinya global optimum yang ditentukan telah benar dan *cluster* yang terbentuk terpisah dengan baik.

DAFTAR PUSTAKA

- [1]. Gunawan, Imam. *Pengantar Statistika Inferensial*, Jakarta: Rajawali Pers, 2016.
- [2]. Arai, K., and Barakbah, A.R. Reversed Pattern OF Moving Variance For Accelerating Automatic Clustering. *EEPIS Jounal.Proc.* 2004. Vol.2., No. 9.
- [3]. Data Agregat Kependudukan. *Data agregat kependudukan kelompok pendidikan yang ditamatkan*. Kantor Kecamatan Nanga Taman : Kantor Kecamatan Nanga Taman ; 2019.
- [4]. Supranto, J. *Analisis Multivariat Arti dan Interpretasi*, Jakarta: Rineka Cipta, 2010.
- [5]. Johnson, R.A and Whincern, D.W. *Applide Multivariate Statistical Analysis* 4nd Edition New Jersey: Pretice-Hall, Inc. 2002.
- [6]. Kaufman, L and Rousseeuw, P.J. *Fiding Groups in Data*. 9nd Edition. John Wiley& Sons. Ney Jersey. 2005.
- [7]. 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 Enineering, *Saga University*. 2007.Vol.36., No.1.

LIDIA KARNELIA	: Jurusan Matematika FMIPA UNTAN, Pontianak lidiakarnelia146@gmail.com
DADAN KUSNANDAR	: Jurusan Matematika FMIPA UNTAN, Pontianak dkusnand@untan.ac.id
SHANTIKA MARTHA	: Jurusan Matematika FMIPA UNTAN, Pontianak shantika.martha@math.untan.ac.id