

MONOTHETIC DIVISIVE CLUSTERING (MONA) PADA DATA INTERVAL DAN HISTOGRAM (Studi Kasus : Data Indeks Komponen IPM pada Provinsi-provinsi di Indonesia Wilayah Timur Tahun 2012)

Anisa Bella Fathia, Dewi Rachmatin, Jarnawi Afgani Dahlan,

Departemen Pendidikan Matematika FPMIPA Universitas Pendidikan Indonesia

ABSTRAK: *Monothetic Divisive Clustering* (MONA) merupakan salah satu metode pengklasteran dalam analisis kluster hierarki, lebih tepatnya metode *divisive* atau pemecahan. Metode ini mula-mulanya menghimpun seluruh objek dalam satu kluster besar, kemudian kluster tersebut dipecah ke dalam dua sub-kluster berdasarkan pertanyaan biner, lalu kedua sub-kluster tersebut dipecah lagi ke dalam dua sub-kluster, begitu seterusnya hingga dalam satu kluster terdiri dari objek tunggal (*singleton*) atau tidak ada lagi variabel pemisah yang tersisa. Ketika menghadapi suatu data histogram atau data berdistribusi peluang biasanya sangat sulit untuk dilakukan penganalisisan khususnya pengelompokkan untuk mengidentifikasi masing-masing objek. Penggunaan metode *Monothetic Divisive Clustering* (MONA) pada data interval dan histogram sangat tepat dan efisien untuk digunakan.

Kata kunci : *monothetic divisive clustering*, analisis kluster hierarki, *divisive methods*, data interval, data histogram, *symbolic data*.

PENDAHULUAN

Monothetic analysis (MONA) termasuk ke dalam metode hirarki divisif. Metode ini berturut-turut memisahkan satu kluster ke dalam dua sub kluster. Pada setiap tahap, salah satu variabel digunakan untuk memisahkan data dengan cara memisahkan objek-objek yang variabelnya diambil dari variabel biner, sehingga meminimumkan keragaman antar kelompok. Karena setiap pemisahan dilakukan menggunakan satu variabel, maka disebut *monothetic*.

Sejalan dengan tujuan analisis kluster yakni untuk mengidentifikasi kelompok-kelompok berdasarkan kemiripannya, maka penggunaan analisis kluster ini banyak sekali ditemukan dalam bidang sosial, politik, ekonomi, biologi, industri, kesehatan dan lain-lain. Dalam hal mengelompokkan data berdistribusi peluang atau data histogram, metode *monothetic divisive clustering* (MONA) dapat dijadikan salah satu alternatif untuk menangani permasalahan tersebut.

LANDASAN TEORI

Variabel Interval

Sebuah variabel interval didefinisikan oleh $Y_j: \Omega \rightarrow B$ sehingga $Y_j(\omega_i) = [l_i, u_i], l_i \leq u_i$, di mana B adalah himpunan interval dari sebuah himpunan dasar $\mathcal{O} \subseteq \mathbb{R}$. Diberikan I sebuah matriks $n \times p$ yang mewakili nilai-nilai dari variabel interval p pada Ω . Setiap $\omega_i \in \Omega$ diwakili oleh p -tuple dari interval, $I_i = (I_{i1}, \dots, I_{ip}), i = 1, \dots, n$ dengan $I_{ij} = [l_{ij}, u_{ij}], j = 1, \dots, p$.

Variabel Histogram

Sebuah variabel histogram Y_j didefinisikan oleh $Y_j: \Omega \rightarrow B$, di mana B adalah himpunan peluang atau distribusi frekuensi dalam sub-interval $\{I_{i1}, \dots, I_{ik_i}\}$ (k_i adalah jumlah interval dalam $Y_j(\omega_i)$) dan $Y_j(\omega_i) = (I_{i1}, p_{i1}; \dots; I_{ik_i}, p_{ik_i})$, dengan $p_{i\ell}$ adalah peluang atau frekuensi yang berasosiasi dengan sub-interval $I_{i\ell} = [l_{i\ell}, \bar{l}_{i\ell})$ di mana $\ell = 1, \dots, k_i$ dan $p_{i1} + \dots + p_{ik_i} = 1$. Oleh karena itu, $Y_j(\omega_i)$ dapat diwakili oleh histogram berdasarkan Bock & Diday (2000) dalam Brito & Chavent (2012):

$$H_{Y_j(\omega_i)} = ([l_{i1}, \bar{l}_{i1}), p_{i1}; \dots; [l_{ik_i}, \bar{l}_{ik_i}), p_{ik_i})$$

di mana

$$i \in \{1, 2, \dots, n\}, l_{i\ell} \leq \bar{l}_{i\ell} \leq l_{i(\ell+1)}$$

Diasumsikan bahwa di dalam setiap sub-interval $[l_{i\ell}, \bar{l}_{i\ell})$, nilai dari variabel Y_j pada observasi atau pengamatan ω_i berdistribusi seragam. Untuk setiap variabel Y_j jumlah dan panjang dari sub-interval dalam $Y_j(\omega_i), i = 1, \dots, n$ pada kenyataannya adalah berbeda. Untuk menggunakan metode pengklasteran, semua observasi dari setiap variabel bernilai histogram harus dibentuk menggunakan partisi yang sama, sehingga dapat dibandingkan dengan tepat. Untuk setiap variabel, kita bentuk kembali setiap histogram yang diobservasi menggunakan titik potong dari partisi atau interval yang diberikan.

Untuk memperoleh interval yang sama, maka interval $I_{i\ell}$ yang diberikan di partisi ke dalam interval I_{ij} sehingga akan terdapat $k \in Z(\mathcal{g})$ dan menghasilkan nilai frekuensi observasi yang berasosiasi dengan interval I_{ij} , yakni (Billard & Diday, 2003:478):

$$p_{ij} = \sum_{k \in Z(\mathcal{g})} \frac{\|Z(k; \omega) \cap I_{ij}\|}{Z(k; \omega)} p_{i\ell} ; \ell = 1, \dots, k_i$$

Ukuran Similaritas

Oleh karena tujuan pengklasteran ialah untuk mengelompokkan objek yang mirip dalam kluster yang sama, maka semua algoritma pengklasteran memerlukan beberapa ukuran untuk mengakses seberapa mirip atau berbeda objek-objek tersebut. Salah satu ukuran tersebut yakni ukuran jarak. Dengan menghitung jarak dapat diketahui bahwa dua objek tertentu lebih mirip dibandingkan dengan objek lainnya.

Dalam metode *divisive clustering*, khususnya MONA, Kuadrat jarak Euclid digunakan untuk menghitung disimilaritas antar objek. Diberikan $Y_j(\omega_i) = H_{Y_j}(\omega_i) = ([\underline{l}_{j1}, \bar{l}_{j1}], p_{l_{j1}}; \dots; [\underline{l}_{jK_j}, \bar{l}_{jK_j}], p_{l_{jK_j}})$. Kuadrat jarak Euclid adalah sebagai berikut (Brito & Chavent, 2012:3):

$$d^2_E(x_i, x_{i'}) = \sum_{j=1}^{K_j} (p_{l_{ij}} - p_{l_{i'j}})^2$$

dengan $d^2_E(x_i, x_{i'}) =$ kuadrat jarak Euclid antar objek ke- i dengan objek ke- i'

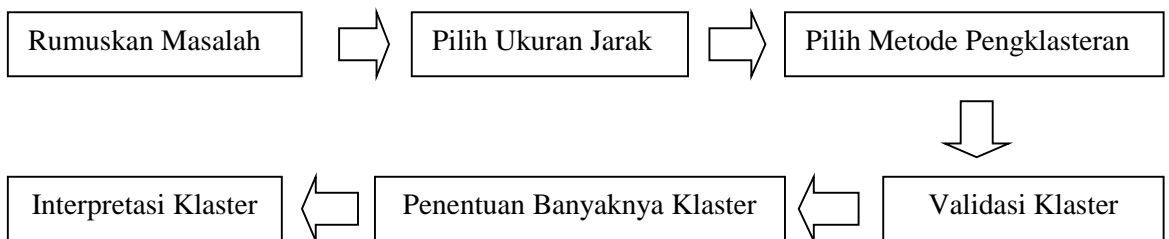
$K_j =$ jumlah interval untuk variabel $j, j = 1, 2, \dots, p$

$p_{l_{ij}} =$ peluang yang berasosiasi dengan interval $l_{ij} = [\underline{l}_{ij}, \bar{l}_{ij})$

$p_{l_{i'j}} =$ peluang yang berasosiasi dengan interval $l_{i'j} = [\underline{l}_{i'j}, \bar{l}_{i'j})$.

Proses Analisis Kluster

Untuk melakukan pengklasteran, harus melalui sebuah proses sebelum diperoleh hasil analisis. Dalam proses analisis kluster ada beberapa tahap yang harus dilakukan, yakni:



Kriteria Inersia

Kualitas dari sebuah partisi yang diberikan $P_m = \{C_1^{(m)}, C_2^{(m)}, \dots, C_m^{(m)}\}$ diukur oleh sebuah kriteria $Q(m)$, jumlah simpangan intra-kluster untuk setiap kluster (Brito & Chavent, 2012:3) yakni:

$$Q(m) = \sum_{\alpha=1}^K I(C_\alpha) = \sum_{\alpha=1}^K \sum_{\omega_i, \omega_i'} c_{\alpha}^{(m)} D^2(\omega_i, \omega_i')$$

dengan

$$D^2(\omega_i, \omega_i') = \sum_{j=1}^p d^2(x_{ij}, x_{ij}')$$

di mana d^2 adalah jarak kuadrat antar objek sehingga untuk setiap kluster, simpangan intra-kluster didefinisikan sebagai jumlah D^2 dari semua pasangan jarak kuadrat antar elemen kluster. Pada setiap tahap, satu kluster dipilih untuk dipecah ke dalam dua sub-kluster, sehingga $Q(m+1)$ minimum, atau secara ekuivalen $Q(m) - Q(m+1)$ maksimum.

Pertanyaan Biner

Bipartisi dilakukan pada setiap tahap yang didefinisikan oleh satu variabel tunggal. Dengan mempertimbangkan kondisi sebagai berikut (Brito & Chavent, 2012:4):

$$R_j := Y_j \quad I_j, g = 1, \dots, K_j - 1, j = 1, \dots, p$$

Kita anggap batas atas I_j dari semua interval (kecuali untuk yang terakhir) bersesuaian dengan setiap variabel.

Setiap kondisi R_j menuntun sebuah bipartisi dari sebuah kluster. Sub-kluster 1 menghimpun elemen yang memenuhi kondisi dan sub-kluster 2 menghimpun elemen yang tidak memenuhi kondisi. Menurut Chavent (2000) dalam (Brito & Chavent, 2012:4), sebuah elemen ω_i memenuhi kondisi $R_j = Y_j \quad I_j$ jika:

$$\sum_{\alpha=1}^g p_{i\alpha} \geq 0,5$$

Pada setiap tahap, kluster $C^{(m)}$ dan kondisi pemecahan R_j dipilih, sehingga resultan partisi P_{m+1} dalam $m+1$ kluster meminimumkan $Q(m+1)$.

Algoritma MONA

Diberikan $P_m = \{C_1^{(m)}, \dots, C_m^{(m)}\}$ yang menjadi partisi pada tahap m .

1. Inisialisasi: $P_1 = \{C_1^{(1)}\}$.

2. Pada tahap m , tentukan kluster $C_M^{(m)}$ dan pertanyaan biner $R_j := Y_j - I_j, g = 1, \dots, K_j - 1, j = 1, \dots, p$, sehingga jumlah partisi baru $P_{m+1} = \{C_1^{(m+1)}, \dots, C_{m+1}^{(m+1)}\}$ dalam $m + 1$ kluster meminimumkan simpangan intra-kluster yang diberikan oleh $Q(m) = \sum_{l=1}^m \omega_l \omega_l' C^{(m)} \sum_{j=1}^p d^2(x_l, x'_j)$ diantara partisi dalam $m + 1$ kluster yang diperoleh dari pemecahan sebuah kluster $P(m)$ ke dalam dua kluster.
3. Perhatikan bahwa untuk meminimumkan $Q(m)$ adalah ekuivalen dengan memaksimumkan $Q = I(C_M^{(m)}) - I(C_1^{(m+1)} + I(C_2^{(m+1)}))$.
4. Ketika diinginkan dan jumlah kluster telah dicapai atau P memiliki n kluster dengan elemen tunggal, maka algoritma berhenti.

Data yang digunakan adalah data indeks komponen IPM pada provinsi Maluku Utara, Maluku, Papua dan Papua Barat tahun 2012 yang diperoleh dari *website* resmi BPS berdasarkan provinsi. Data tersebut dibentuk ke dalam data histogram sedemikian sehingga nilai-nilai yang berada dalam interval pada setiap variabel yakni, IHH (Indeks Harapan Hidup), IP (Indeks Pendidikan) dan IDB (Indeks Daya Beli) pada setiap observasi berdistribusi seragam. Data dapat dilihat pada Tabel 1.

TABEL.1. Komponen IPM

Provinsi	ω	IHH (Y_1)	IP (Y_2)	IDB (Y_3)
Maluku Utara	1	([63.33,67.41),0.22; [67.41,68.13),0.22; [68.13,68.74),0.22; [68.74,70.98),0.22; [70.98,77.11),0.12)	([72.61,81.16),0.22; [81.16,81.39),0.22; [81.39,83),0.22; [83,86.7),0.22; [86.7,91.15],0.12)	([57.48,69.06),0.22; [69.06,70.43),0.22; [70.43,70.67),0.22; [70.67,78.86),0.22; [78.86,83.4],0.12)
Maluku	2	([63.33,67.41),0.18; [67.41,68.74),0.18; [68.74,70.98),0.18; [70.98,72),0.18; [72,77.11),0.18; [77.11,80.28],0.10)	([72.61,81.16),0.18; [81.16,83),0.18; [83,84.23),0.18; [84.23,85.89),0.18; [85.89,91.15),0.18; [91.15,91.34],0.10)	([57.48,69.06),0.18; [69.06,69.67),0.10; [69.67,70.67),0.18; [70.67,73.72),0.18; [73.72,75.63),0.18; [75.63,83.4],0.18)
Papua	3	([63.33,68.74),0.17; [68.74,69.38),0.17; [69.38,69.79),0.17; [69.79,70.74),0.17; [70.74,72),0.17; [72,77.11],0.15)	([26.6,28.35),0.17; [28.35,31),0.17; [31,46.99),0.17; [46.99,72.61),0.17; [72.61,81.16),0.17; [81.16,91.15),0.15)	([57.48,64.16),0.17; [64.16,67.21),0.17; [67.21,69.06),0.17; [69.06,71.49),0.17; [71.49,75.63),0.17; [75.63,83.4],0.15)

Provinsi	ω	IHH (Y_1)	IP (Y_2)	IDB (Y_3)
Papua Barat	4	([63.33,69.68),0.18; [69.68,70.98),0.18; [70.98,72),0.10; [72,72.51),0.18; [72.51,75.18),0.18; [75.18,80.28],0.18)	([46.99,72.61),0.18; [72.61,76.92),0.18; [76.92,79.23),0.18; [79.23,80.01),0.18; [80.01,86.7),0.18; [86.7,91.15],0.10)	([33.79,64.16),0.18; [64.16,67.21),0.18; [67.21,69.06),0.18; [69.06,69.67),0.18; (69.67,70.67),0.18; (70.67,83.4],0.10)

Pembentukan Data histogram setelah di Partisi

Untuk mengaplikasikan metode pengelompokan, semua observasi pada variabel histogram harus dibentuk menggunakan partisi yang sama, sehingga dapat dibandingkan dengan tepat. Untuk setiap variabel, dibentuk kembali histogram observasi menggunakan titik potong dari partisi yang diberikan. Nilai frekuensi atau peluang dari setiap interval diperoleh menggunakan rumus:

$$p_{k_l} = \sum_{k \in Z(g)} \frac{\|Z(k; \omega) \cap I_{k_l}\|}{|Z(k; \omega)|} p_{k_l} ; l = 1, \dots, k_l$$

Data yang akan dianalisis menggunakan metode *Monothetic Divisive Clustering* (MONA) pada data interval dan histogram dapat dilihat pada Tabel 2.

TABEL.2

Provinsi	ω	Indeks Harapan Hidup (Y_1)	Indeks Pendidikan (Y_2)	Indeks Daya Beli (Y_3)
Maluku Utara	1	([63.33,67.41),0.22; [67.41,68.13),0.22; [68.13,68.74),0.22; [68.74,69.38),0.06; [69.38,69.68),0.03; [69.68,69.79),0.01; [69.79,70.74),0.10; [70.74,70.98),0.02; [70.98,72),0.02; [72,72.51),0.01; [72.51,75.18),0.05; [75.18,77.11),0.03; [77.11,80.28],0.01)	([26.6,28.35),0; [28.35,31),0; [31,46.99),0; [46.99,72.61),0; [72.61,76.92),0.11; [76.92,79.23),0.06; [79.23,80.01),0.03; [80.01,81.16),0.02; [81.16,81.39),0.22; [81.39,83),0.22; [83,84.23),0.08; [84.23,85.89),0.10; [85.89,86.7),0.04; [86.7,91.15),0.12 (91.15,91.34],0.00)	([33.79,57.48),0; [57.48,64.16),0.15; [64.16,67.21),0.03; [67.21,69.06),0.04; [69.06,69.67),0.10; [69.67,70.43),0.12; [70.43,70.67),0.22; [70.67,71.49),0.03; [71.49,73.72),0.06; [73.72,75.63),0.05; [75.63,78.86),0.08; [78.86,83.4],0.12)

Provinsi	ω	Indeks Harapan Hidup (Y_1)	Indeks Pendidikan (Y_2)	Indeks Daya Beli (Y_3)
Maluku	2	([63.33,67.41),0.18; [67.41,68.13),0.10; [68.13,68.74),0.08; [68.74,69.38),0.05; [69.38,69.68),0.02; [69.68,69.79),0.01; [69.79,70.74),0.07; [70.74,70.98),0.03; [70.98,72),0.18; [72,72.51),0.05; [72.51,75.18),0.08; [75.18,77.11),0.05; [77.11,80.28],0.10)	([26.6,28.35),0; [28.35,31),0; [31,46.99),0; [46.99,72.61),0; [72.61,76.92),0.09; [76.92,79.23),0.05; [79.23,80.01),0.02; [80.01,81.16),0.02; [81.16,81.39),0.03; [81.39,83),0.15; [83,84.23),0.18; [84.23,85.89),0.18 [85.89,86.7),0.03; [86.7,91.15),0.15; (91.15,91.34],0.10)	([33.79,57.48),0; [57.48,64.16),0.12; [64.16,67.21),0.03; [67.21,69.06),0.03; [69.06,69.67),0.10; [69.67,70.43),0.13; [70.43,70.67),0.05; [70.67,71.49),0.05; [71.49,73.72),0.13; [73.72,75.63),0.18; [75.63,78.86),0.07; [78.86,83.4],0.11)
Papua	3	([63.33,67.41),0.13; [67.41,68.13),0.02 ; [68.13,68.74),0.02; [68.74,69.38),0.17; [69.38,69.68),0.12; [69.68,69.79),0.05; [69.79,70.74),0.17; [70.74,70.98),0.02; [70.98,72),0.15; [72,72.51),0.02; [72.51,75.18),0.06; [75.18,77.11),0.07; [77.11,80.28],0.00)	([26.6,28.35),0.17; [28.35,31),0.17; [31,46.99),0.17; [46.99,72.61),0.17; [72.61,76.92),0.09; [76.92,79.23),0.04; [79.23,80.01),0.02; [80.01,81.16),0.02 [81.16,81.39),0.00; [81.39,83),0.03; [83,84.23),0.01; [84.23,85.89),0.03; [85.89,86.7),0.01; [86.7,91.15),0.07; (91.15,91.34]0.00)	([33.79,57.48),0; [57.48,64.16),0.17; [64.16,67.21),0.17; [67.21,69.06),0.17; [69.06,69.67),0.04; [69.67,70.43),0.05; [70.43,70.67),0.03; [70.67,71.49),0.05; [71.49,73.72),0.09; [73.72,75.63),0.08; [75.63,78.86),0.07; [78.86,83.4],0.08)

Provinsi	ω	Indeks Harapan Hidup (Y_1)	Indeks Pendidikan (Y_2)	Indeks Daya Beli (Y_3)
Papua Barat	4	([63.33,67.41),0.11; [67.41,68.13),0.01; [68.13,68.74),0.02; [68.74,69.38),0.02; [69.38,69.68),0.02; [69.68,69.79),0.02; [69.79,70.74),0.13; [70.74,70.98),0.03; [70.98,72),0.10; [72,72.51),0.18; [72.51,75.18),0.18; [75.18,77.11),0.07; [77.11,80.28],0.11)	([26.6,28.35),0; [28.35,31),0; [31,46.99),0; [46.99,72.61),0.18; [72.61,76.92),0.18; [76.92,79.23),0.18; [79.23,80.01),0.18; [80.01,81.16),0.03; [81.16,81.39),0.01; [81.39,83),0.04; [83,84.23),0.02; [84.23,85.89),0.04; [85.89,86.7),0.04; [86.7,91.15),0.09; (91.15,91.34],0.01)	([33.79,57.48),0.14; [57.48,64.16),0.04; [64.16,67.21),0.18; [67.21,69.06),0.18; [69.06,69.67),0.18; [69.67,70.43),0.14; [70.43,70.67),0.04; [70.67,71.49),0.01; [71.49,73.72),0.03; [73.72,75.63),0.02; [75.63,78.86),0.02; [78.86,83.4],0.02)

Ukuran Jarak

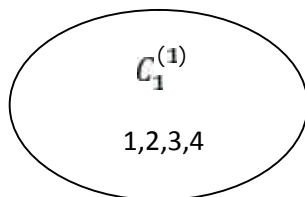
Ukuran jarak yang dipakai dalam analisis ini adalah jarak kuadrat Euclid. Dengan menggunakan rumus $d^2_E(x_i, x_{i'}) = \sum_{j=1}^{K_j} (p_{ij} - p_{i'j})^2$, maka diperoleh matriks kuadrat jarak Euclid 4×4 seperti berikut:

$$\begin{pmatrix} 0 & 0,19 & 0,44 & 0,45 \\ 0,19 & 0 & 0,33 & 0,34 \\ 0,44 & 0,33 & 0 & 0,31 \\ 0,45 & 0,34 & 0,31 & 0 \end{pmatrix}$$

Algoritma Metode MONA

1. Inisialisasi partisi pada tahap pertama $P_1 = \{C_1^{(1)}\}$

P_1



2. Menentukan pertanyaan biner R_f atau variabel pemisah yang paling baik untuk memisahkan $C_1^{(1)}$ ke dalam kluster-kluster $C_1^{(2)}$ dan $C_2^{(2)}$ yang dapat meminimumkan jumlah simpangan intra kluster diantara partisi dalam $m + 1$ kluster. Oleh karena terdapat tiga buah variabel deskriptif, maka terdapat 6 kemungkinan untuk mendapatkan urutan variabel pemisah yang paling baik. Kemungkinan-kemungkinannya adalah sebagai berikut:

$$\begin{array}{ll}
 Y_1, Y_2, Y_3 & Y_2, Y_1, Y_3, Y_3, Y_1, Y_2 \\
 Y_1, Y_3, Y_2 & Y_2, Y_3, Y_1, Y_3, Y_2, Y_1
 \end{array}$$

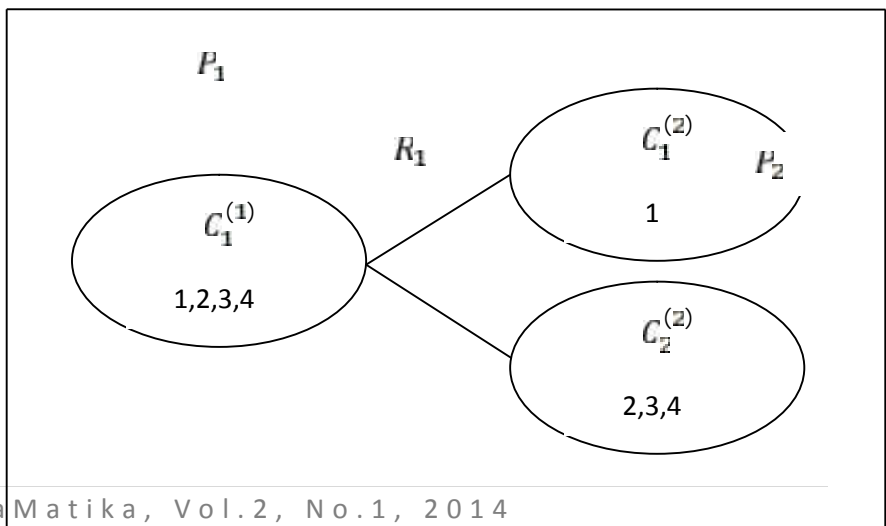
Y_1, Y_2, Y_3

- Berdasarkan pertanyaan biner $R_1 := Y_1 \quad I_1 = 68,74$, akan dicari anggota ω_i yang memenuhi $\sum_{i=1}^m P_i \geq 0,5$, yaitu

ω	Indeks Harapan Hidup (Y_1)	
1	([63.33,67.41),0.22; [67.41,68.13),0.22; [68.13,68.74),0.22; [68.74,69.38),0.06; [69.38,69.68),0.03; [69.68,69.79),0.01; [69.79,70.74),0.10; [70.74,70.98),0.02; [70.98,72),0.02; [72,72.51),0.01; [72.51,75.18),0.05; [75.18,77.11),0.03; [77.11,80.28],0.01)	$\rightarrow 0,22+0,22+0,22=0,66 \geq 0,5$
2	([63.33,67.41),0.18; [67.41,68.13),0.10; [68.13,68.74),0.08; [68.74,69.38),0.05; [69.38,69.68),0.02; [69.68,69.79),0.01; [69.79,70.74),0.07; [70.74,70.98),0.03; [70.98,72),0.18; [72,72.51),0.05; [72.51,75.18),0.08; [75.18,77.11),0.05; [77.11,80.28],0.10)	$\rightarrow 0,18+0,10+0,08=0,36 < 0,5$

ω	Indeks Harapan Hidup (Y_1)	
3	([63.33,67.41],0.13; [67.41,68.13],0.02; [68.13,68.74],0.02; [68.74,69.38],0.17; [69.38,69.68],0.12; [69.68,69.79],0.05; [69.79,70.74],0.17; [70.74,70.98],0.02; [70.98,72),0.15; [72,72.51],0.02; [72.51,75.18],0.06; [75.18,77.11],0.07; [77.11,80.28],0.00)	→ 0,13+0,02+0,02=0,17<0,5
4	([63.33,67.41],0.11; [67.41,68.13],0.01; [68.13,68.74],0.02; [68.74,69.38],0.02; [69.38,69.68],0.02; [69.68,69.79],0.02; [69.79,70.74],0.13; [70.74,70.98],0.03; [70.98,72),0.10; [72,72.51],0.18; [72.51,75.18],0.18; [75.18,77.11],0.07; [77.11,80.28],0.11)	→ 0,11+0,01+0,02=0,13<0,5

Dari perhitungan tersebut, maka dapat dengan mudah dibentuk sub-kluster $C_1^{(2)}$ dan $C_2^{(2)}$. $C_1^{(2)}$ berisi anggota ω_1 yang memenuhi kondisi dan $C_2^{(2)}$ berisi anggota ω_1 yang tidak memenuhi kondisi.



Selanjutnya, hitung jumlah simpangan intra-kluster pada partisi 2 menggunakan persamaan

$$Q(m) = \sum_{\alpha=1}^K I(C_{\alpha}) = \sum_{\alpha=1}^K \sum_{\omega_i, \omega_i'} C_{\alpha}^{(m)} \sum_{j=1}^P d^{zz}(x_i, x_i' j)$$

Maka, $F_2 = \{C_1^{(z)}, C_2^{(z)}\}$ memiliki jumlah simpangan intra-kluster sebesar 0,98.

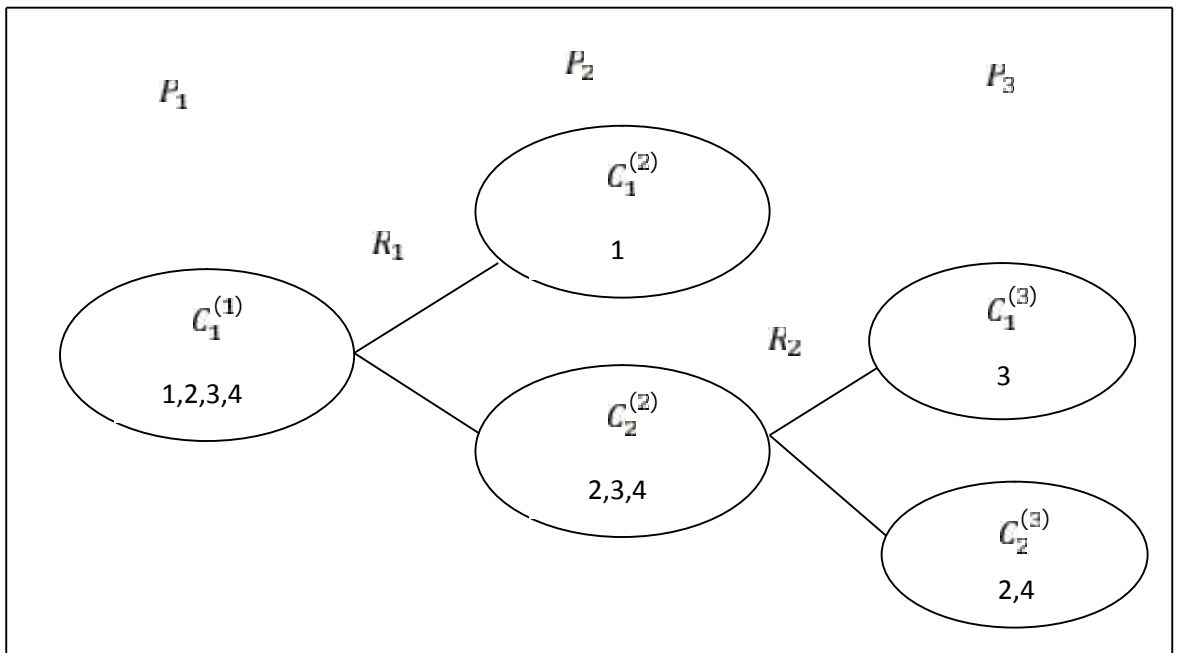
- Pada tahap ketiga, pecah lagi sub-kluster $C_2^{(z)}$ ke dalam $C_1^{(s)}$ dan $C_2^{(s)}$ dengan cara yang sama. Berdasarkan pertanyaan biner $R_2 := Y_2 \quad I_2 = 46,99$

ω	Indeks Pendidikan (Y_2)
2	<p>([26.6,28.35),0; [28.35,31.00),0; [31.00,46.99),0; [46.99,72.61),0; [72.61,76.92),0.09; [76.92,79.23),0.05; [79.23,80.01),0.02; [80.01,81.16),0.02; [81.16,81.39),0.03; [81.39,83),0.15; [83,84.23),0.18; [84.23,85.89),0.18 [85.89,86.7),0.03; [86.7,91.15),0.15; (91.15,91.34],0.10)</p>
	<p>→ 0+0+0=0<0,5</p>
3	<p>([26.6,28.35),0.17; [28.35,31.00),0.17; [31.00,46.99),0.17; [46.99,72.61),0.17; [72.61,76.92),0.09; [76.92,79.23),0.04; [79.23,80.01),0.02; [80.01,81.16),0.02 [81.16,81.39),0.00; [81.39,83),0.03; [83,84.23),0.01; [84.23,85.89),0.03; [85.89,86.7),0.01; [86.7,91.15),0.07; (91.15,91.34]0.00)</p>
	<p>→ 0,17+0,17+0,17=0,51≥0,5</p>

ω	Indeks Pendidikan (Y_2)
4	([26.6,28.35) 0; [28.35,31.00) 0; [31.00,46.99) 0; [46.99,72.61) 0.18; [72.61,76.92) 0.18; [76.92,79.23) 0.18; [79.23,80.01) 0.18; [80.01,81.16) 0.03; [81.16,81.39) 0.01; [81.39,83) 0.04; [83,84.23) 0.02; [84.23,85.89) 0.04; [85.89,86.7) 0.04; [86.7,91.15) 0.09; (91.15,91.34 0.01)

→ 0+0+0=0<0,5

Dari perhitungan tersebut, maka di dapat sub-kuster $C_1^{(2)}$ dan $C_2^{(2)}$ sebagai berikut:



Nilai $Q(m)$ pada partisi P_3 yaitu sebesar 0,34 sehingga $P_3 = \{C_1^{(3)}, C_2^{(3)}, C_3^{(3)} = C_1^{(2)}\}$ memiliki jumlah simpangan intra-kuster sebesar 0,34 yang lebih minimum dari jumlah simpangan intra-kuster pada P_2 .

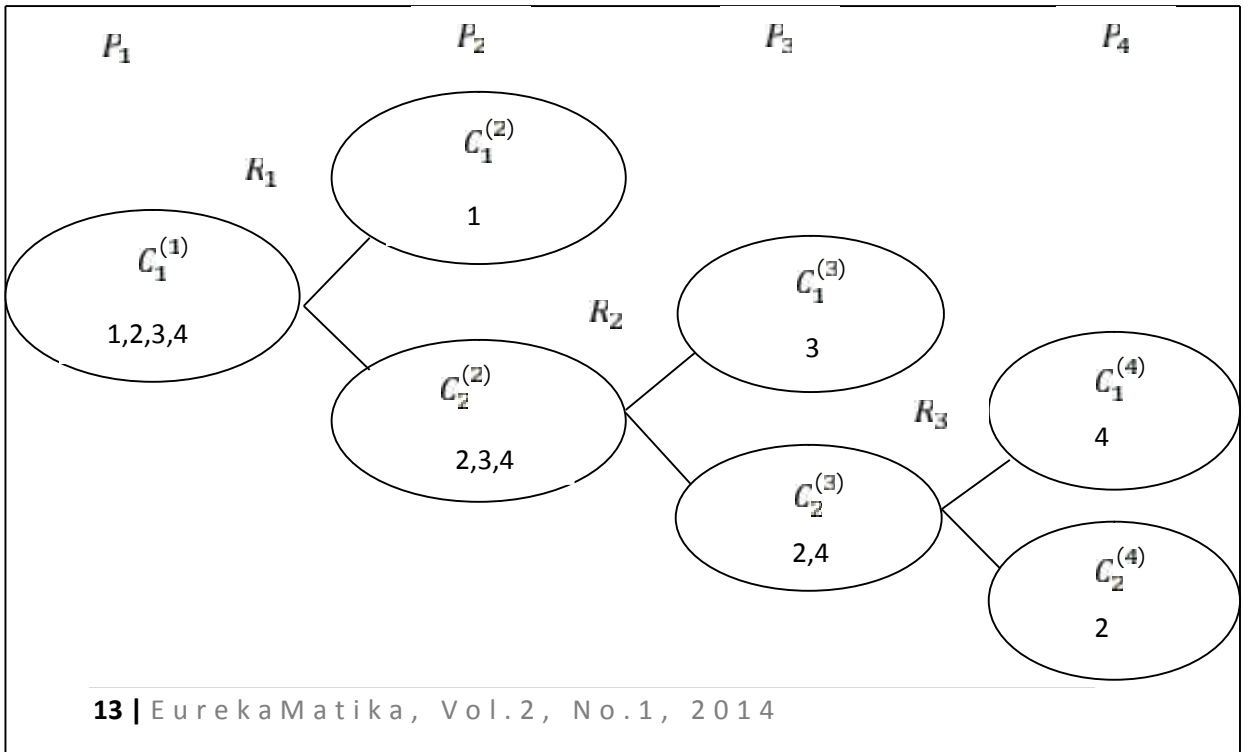
- Pada tahap keempat, sub-kuster $C_2^{(3)}$ dipecah ke dalam $C_1^{(4)}$ dan $C_2^{(4)}$ berdasarkan pertanyaan biner $R_3 := Y_3 \quad I_3 = 69,06$

ω	Indeks Daya Beli (Y_3)
2	$([33.79,57.48],0;$ $[57.48,64.16),0.12;$ $[64.16,67.21),0.03;$ $[67.21,69.06),0.03;$ $[69.06,69.67),0.10;$ $[69.67,70.43),0.13;$ $[70.43,70.67),0.05;$ $[70.67,71.49),0.05;$ $[71.49,73.72),0.13;$ $[73.72,75.63),0.18;$ $[75.63,78.86),0.07;$ $[78.86,83.4],0.11)$
4	$([33.79,57.48],0.14;$ $[57.48,64.16),0.04;$ $[64.16,67.21),0.18;$ $[67.21,69.06),0.18;$ $[69.06,69.67),0.18;$ $[69.67,70.43),0.14;$ $[70.43,70.67),0.04;$ $[70.67,71.49),0.01;$ $[71.49,73.72),0.03;$ $[73.72,75.63),0.02;$ $[75.63,78.86),0.02;$ $[78.86,83.4],0.02)$

$$0+0,12+0,03+0,03=0,18 < 0,5$$

$$0,14+0,04+0,18+0,18=0,54 \geq 0,5$$

- Dari perhitungan tersebut, maka di dapat $C_1^{(4)}$ dan $C_2^{(4)}$ sebagai berikut:



Nilai $Q(m)$ pada partisi P_4 yaitu 0 sehingga $P_4 = \{C_1^{(4)}, C_2^{(4)}, C_3^{(4)} = C_1^{(4)}, C_4^{(4)} = C_1^{(4)}\}$ memiliki jumlah simpangan intra-klaster adalah 0 yang lebih minimum dari jumlah simpangan dalam klaster pada P_3 .

- Oleh karena sub-klaster terakhir sudah berisi objek tunggal dan tidak ada lagi variabel pemisah sehingga tidak bisa dipecah lagi, maka algoritma berhenti

Dengan perhitungan yang sama dengan sebelumnya, maka didapat tabel urutan variabel pemisah terbaik seperti berikut:

Urutan Variabel Pemisah	$Q(m)$ pada Partisi P_2	$Q(m)$ pada Partisi P_3	$Q(m)$ pada Partisi P_4
Y_1, Y_2, Y_3	0,98	0,34	0
Y_1, Y_3, Y_2	0,94	0,31	0
Y_2, Y_1, Y_3	0,98	0,34	0
Y_2, Y_3, Y_1	0,98	0,19	0
Y_3, Y_1, Y_2	0,5	0	-
Y_3, Y_2, Y_1	0,5	0	-

Dari hasil tabel tersebut, maka dipilih Y_2, Y_3, Y_1 sebagai urutan variabel pemisah terbaik, hal itu karena urutan tersebut paling optimum untuk meminimumkan jumlah simpangan intra klaster $Q(m)$ pada $m + 1$ klaster, dilihat dari selisih partisi P_2 dan partisi P_3 yaitu 0,79. Selengkapnya diagram hasil pengklasteran *monothetic divisive clustering* (MONA) pada data indeks komponen IPM pada provinsi-provinsi di Indonesia wilayah timur tahun 2012 dapat dilihat pada lampiran.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa:

1. langkah-langkah metode *monothetic divisive clustering* (MONA) dalam mengelompokkan data interval dan histogram adalah data yang sudah berbentuk histogram harus disusun ke dalam data histogram dengan sub-sub interval yang sama pada setiap kelas interval. Setelah itu, seluruh objek pengamatan digabungkan ke dalam satu klaster besar, kemudian pisahkan ke dalam dua sub klaster berdasarkan pertanyaan biner atau variabel pemisah pada persamaan $R_j := Y_j \quad I_j, g = 1, \dots, K_j - 1, j = 1, \dots, p$. Objek-objek yang

memenuhi persamaan tersebut dimasukkan ke dalam sub-kluster kesatu dan yang tidak memenuhi dimasukkan pada sub-kluster kedua, kemudian pisahkan lagi sub-kluster tersebut ke dalam dua sub-kluster yang lebih kecil begitu seterusnya hingga satu kluster berisi objek tunggal atau tidak ada lagi variabel sisa yang bisa memisahkan sub-kluster tersebut. Setiap tahap pemisahan harus meminimumkan simpangan intra kluster pada tahap pemisahan selanjutnya berdasarkan kriteria pada persamaan

$$Q(m) = \sum_{\alpha=1}^K I(C_{\alpha}) = \sum_{\alpha=1}^K \sum_{\omega_i, \omega_i'} C_{\alpha}^{(m)} D^2(\omega_i, \omega_i')$$

2. Hasil Pengklasteran dalam kasus data indeks komponen IPM pada provinsi-provinsi di Indonesia wilayah timur tahun 2012 menghasilkan empat kluster dengan urutan variabel pemisah terbaik adalah Y_2, Y_3, Y_1 . Hal itu karena urutan tersebut paling optimum untuk meminimumkan jumlah simpangan intra kluster $Q(m)$ pada $m + 1$ kluster, dilihat dari selisih partisi P_2 dan partisi P_3 yaitu 0,79.

Dari hasil urutan variabel pemisah terbaik yaitu Y_2, Y_3, Y_1 maka objek-objek dapat dikelompokkan sebagai berikut:

- Berdasarkan variabel Indeks Pendidikan $> 72,61$, maka keanggotaan klasternya adalah Maluku Utara, Maluku, dan Papua Barat.
- Berdasarkan variabel Indeks Pendidikan $72,61$, maka keanggotaan klasternya adalah Papua.
- Berdasarkan variabel Indeks Daya Beli $> 69,06$, maka keanggotaan klasternya adalah Maluku Utara dan Maluku.
- Berdasarkan variabel Indeks Daya Beli $69,06$, maka keanggotaan klasternya adalah Papua Barat.
- Berdasarkan variabel Indeks Harapan Hidup $> 68,74$, maka keanggotaan klasternya adalah Maluku Utara.
- Berdasarkan variabel Indeks Harapan Hidup $68,74$, maka keanggotaan klasternya adalah Maluku.

Saran

1. Teoritis

Diharapkan terdapat beberapa kriteria atau rumusan dalam menentukan urutan variabel pemisah terbaik, karena pada kenyataannya untuk beberapa kasus, variabel deskriptif bisa banyak sekali dan tidak mungkin untuk dilakukan pengecekan seperti penelitian yang penulis kemukakan.

2. Praktis

Kepada peneliti berikutnya yang berminat untuk melakukan penelitian selanjutnya bisa mencoba menerapkan *polythetic divisive clustering*, yakni menggunakan lebih dari satu variabel untuk memisahkan kluster ke dalam sub-kluster, dan bisa juga dibandingkan metode mana yang lebih baik hasil pengelompokannya.

REFERENSI

- Akyuwen,R. (2010).*Potret Maluku di Mata Nasional:Disparitas Pembangunan Manusia*. [online]. Tersedia di <http://robertoakyuwen.blogspot.com/2010/09/potret-maluku-di-mata-nasional.html>. Diakses 12 Agustus 2014.
- Asumpta,E.(2010).*Metode Centroid dan Metode Median dalam Analisis Kluster*. (Skripsi). Universitas Pendidikan Indonesia,Bandung.
- Billard,L., & Diday,E.(2003).From the statistics of data to the statistics of knowledge:symbolic data analysis.*Journal of the American Statistical Association*,98 (462),hlm.471-478.
- BPS.(2013).*Indeks Pembangunan Manusia Kabupaten Halmahera Utara 2012*. [online]. Tersedia di:http://halutkab.bps.go.id/publikasi/2013/ipm_8207/files/search/searchtext.xml. Diakses 20 Juli 2014.
- BPS.(2013).*Indeks Pembangunan Manusia Kabupaten Manokwari tahun 2012*. [online]. Tersedia di:http://manokwarikab.bps.go.id/data/publikasi/publikasi_3/publikasi/files/search/searchtext.xml. Diakses 20 Juli 2014.
- BPS.(2013).*Indeks Pembangunan Provinsi Maluku Utara 2012*. [online]. Tersedia di:<http://malut.bps.go.id/publikasi/flipbook/2013/ipm/files/search/searchtext.xml>. Diakses 20 Juli 2014.
- BPS.(2013).*Waropen dalam Angka 2013*. [online]. Tersedia di:<http://waropenkab.bps.go.id/yiiweb/download/DDA%202013/perbandingan.pdf>. Diakses 20 Juli 2014.
- Brito,P., & Chavent,M.(2012).Divisive monothetic clustering for interval and histogram-valued data.*ICPRAM 2012-1st International Conference on Pattern Recognition Applications and Methods*,hlm 1-5.

- Chavent,M.,Lechevallier,B.,& Briant,O.(2007).DIVCLUS-T:a monothetic divisive hierarchical clustering method.*Computational Statistics & Data Analysis*,hlm 6.
- Emiliana.(2010).*Centroid dan Median Method dalam Analisis Kluster*. (Skripsi).Universitas Pendidikan Indonesia,Bandung.
- Everitt,B.(2011). *Cluster Analysis 5th Edition*.UK:John Willey & Sons.
- Gunawan,C.(2010).*Pengklasteran Data dengan Menggunakan Divisive Analysis Method (DIANA)*. (Skripsi). Universitas Pendidikan Indonesia,Bandung.
- Herrhyanto,N. dan Giantini,T.(2009).*Pengantar Statistika Matematis*. Bandung:CV.Yrama Widya.
- Johnson,R.A., & Wichern,D.W.(2007).*Applied Multivariate Statistical Analysis*. New Jersey:Prentice Hall.
- Kaufman,L., & Rousseeuw, J.P.(2005).*Finding Groups in Data*.Canada: John Willey & Sons.
- Maqin,A.(2007).*Indeks Pembangunan Manusia:Tinjauan Teoritis dan Empiris di Jawa Barat*.Pelatihan program pendanaan kompetisi indeks pembangunan manusia,hlm 4-12.
- Rosalina,F.*Single Linkage,Complete Linkage,Average Linkage, dan Ward's Method pada Analisis Cluster*.(Skripsi).Universitas Pendidikan Indonesia,Bandung.
- Sharma,S.(1996).*Applied Multivariate Techniques*.Canada: John Willey & Sons.
- Sudjana.(2005).*Metode Statistika*.Bandung:Tarsito.
- Xiannong,M.(2002).*FrequencyTest*.[online].Tersedia di:
<http://www.eg.bucknell.edu/~xmeng/Course/CS6337/Note/master/node43.html>.Diakses 02 Juli 2014.