

**PENERAPAN METODE *HIERARCHICAL CLUSTERING*
MULTISCALE BOOTSTRAP
(STUDI KASUS: INDEKS PEMBANGUNAN MANUSIA (IPM)
DI KALIMANTAN TIMUR TAHUN 2017)**

Yulia Safarina^{1*}, Memi Nor Hayati¹, dan Yuki Novia Nasution¹

¹Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Mulawarman, Indonesia

Corresponding author: yuliatat14@gmail.com

Abstrak. Analisis *cluster* hirarki merupakan analisis yang digunakan untuk mengelompokkan data (objek) berdasarkan karakteristiknya. Metode *complete linkage* merupakan salah satu metode pada analisis *cluster* hirarki yang digunakan untuk mengelompokkan data di berbagai aspek, salah satunya yaitu IPM. Penelitian ini menggunakan data indikator IPM di Kalimantan Timur Tahun 2017. Metode *complete linkage* hanya memberikan satu solusi penyelesaian, yaitu didasarkan pada ukuran kedekatan jarak, sehingga penelitian ini menggunakan metode *multiscale bootstrap* untuk memperoleh validitas dari *cluster* yang terbentuk. Hasil penelitian yang diperoleh yaitu terdapat empat *cluster* yang terbentuk. Kota Balikpapan, Kota Samarinda, dan Kota Bontang termasuk dalam *cluster* pertama, Kabupaten Kutai Kartanegara termasuk dalam *cluster* kedua, Kabupaten Penajam Paser Utara, Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Kutai Timur, dan Kabupaten Berau termasuk dalam *cluster* ketiga, Kabupaten Mahakam Ulu termasuk dalam *cluster* keempat. Dari keempat *cluster* yang terbentuk, terdapat tiga *cluster* yang *valid* yaitu *cluster* kedua, *cluster* ketiga, dan *cluster* keempat.

Kata Kunci: Analisis *cluster* hirarki, *complete linkage*, IPM, *multiscale bootstrap*.

1 PENDAHULUAN

Analisis *cluster* bertujuan untuk mengelompokkan objek-objek pengamatan berdasarkan karakteristik yang dimiliki. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki kesamaan karakteristik [1]. Terdapat dua metode umum dalam algoritma *cluster*, yaitu hirarki dan non-hirarki. Analisis *cluster* dengan metode hirarki ada beberapa metode, yaitu metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata (*average linkage*), dan metode *Ward* (*Ward's method*). Pada metode non-hirarki terdapat metode *K-Means* [2].

Analisis *cluster* memiliki beberapa kelemahan, salah satunya yaitu tidak dapat melihat validitas dari hasil analisis *cluster*, sehingga perlu menggunakan metode lain untuk melihat validitas dari hasil analisis *cluster* sebelumnya. Salah satu metode yang dapat digunakan yaitu metode *multiscale bootstrap*. Metode *multiscale bootstrap* bekerja dengan pendekatan *bootstrap resampling* untuk setiap *cluster* yang terbentuk [3].

Analisis *cluster* dapat digunakan untuk mengelompokkan data di berbagai macam kasus, salah satunya yaitu IPM. IPM didefinisikan sebagai suatu tolak ukur maju atau tidaknya suatu wilayah ataupun daerah [4]. Program yang telah dibuat pemerintah baik pusat maupun daerah perlu didukung oleh strategi perencanaan yang matang untuk meningkatkan IPM. Karena strategi perencanaan program untuk tiap-tiap daerah itu tidak harus sama, maka perlu dilakukan pengelompokkan karakteristik IPM pada setiap kabupaten/kota di Kalimantan Timur.

Pada penelitian, analisis *cluster* yang digunakan yaitu analisis *cluster* hirarki dengan metode *complete linkage*. Kemudian, data yang digunakan yaitu 7 indikator IPM yang berada pada dimensi kehidupan yang layak, dimensi pendidikan, dan dimensi umur panjang dan kesehatan yang terdapat pada kabupaten/kota di Kalimantan Timur tahun 2017. Lalu, jumlah replikasi (B) yang digunakan sebanyak 500.

2 TUJUAN PENELITIAN

Tujuan penelitian ini yaitu:

- a. Untuk mengetahui hasil *cluster* yang terbentuk berdasarkan IPM pada setiap kabupaten/kota di Kalimantan Timur.
- b. Untuk mengetahui validitas dari hasil *cluster* yang terbentuk

3 METODOLOGI

Pada penelitian ini, analisis data dilakukan dengan langkah-langkah berikut:

1. Melakukan deteksi Non-Multikolinieritas antar variabel menggunakan korelasi *Spearman*. Dua variabel atau lebih dikatakan multikolinieritas apabila nilai korelasinya lebih dari 0,8. Jika setelah dideteksi ternyata diketahui bahwa terdapat multikolinieritas antar variabel, adanya multikolinieritas dapat diatasi dengan mereduksi variabel [5].
2. Melakukan analisis *hierarchical clustering multiscale bootstrap* metode *complete linkage* untuk mengelompokkan kabupaten/kota di Kalimantan Timur berdasarkan IPM. Dirumuskan sebagai berikut [6]:

$$d_{UV(W)} = \max \{d_{UW}, d_{VW}\} \quad (1)$$

Pada penelitian ini, ukuran kedekatan yang digunakan yaitu Jarak *Euclidean*. Dirumuskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (2)$$

dimana,

d_{ij} = jarak antar objek ke- i dengan objek ke- j

p = banyaknya variabel

x_{il} = nilai objek ke- i pada variabel ke- l

x_{jl} = nilai objek ke- j pada variabel ke- l

$l = 1, 2, \dots, p$

- Melihat validitas dari hasil analisis *hierarchical clustering* metode *complete linkage* menggunakan metode *multiscale bootstrap*. Dirumuskan sebagai berikut:

$$AU = z(-v + c) \quad (3)$$

Suatu *cluster* dikatakan *valid* jika nilai *Approximately Unbiased* (AU) $\geq 0,95$ [7]. Proses validasi menjamin bahwa solusi *cluster* yang dihasilkan dalam analisis *cluster* dapat menggambarkan populasi sebenarnya [8].

4 HASIL DAN PEMBAHASAN

4.1 Data

Data yang digunakan dalam penelitian ini adalah data 7 indikator IPM yang berada pada dimensi kehidupan yang layak, dimensi pendidikan, dan dimensi umur panjang yang terdapat pada kabupaten/kota di Kalimantan Timur tahun 2017. Data dapat dilihat pada Tabel 1.

Keterangan:

X_1 = Persentase Tingkat Partisipasi Angkatan Kerja (TPAK).

X_2 = Persentase Tingkat Pengangguran Terbuka (TPT).

X_3 = Persentase jumlah pengeluaran perkapita.

X_4 = Persentase Angka Partisipasi Sekolah (APS).

X_5 = Persentase harapan lama sekolah.

X_6 = Persentase angka buta huruf penduduk.

X_7 = Persentase jumlah kematian bayi.

Tabel 1: Data Penelitian

No.	Objek Pengamatan	Variabel Pengamatan						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	Kota Balikpapan	63,38	10,39	58,19	25,09	13,75	0,85	12,11
2	Kab. Kutai Kartanegara	62,95	5,72	50,21	27,02	13,56	1,15	26,09
3	Kota Samarinda	63,05	6,19	58,29	28,23	14,64	0,73	5,12
4	Kab. PPU	61,00	4,75	50,78	27,32	12,53	1,79	6,99
5	Kab. Paser	62,15	5,54	44,78	23,07	12,98	1,51	10,25
6	Kab. Kutai Barat	66,01	5,87	44,28	23,35	12,82	1,51	8,54
7	Kab. Kutai Timur	66,75	4,61	48,56	25,49	12,48	1,03	11,18

No.	Objek Pengamatan	Variabel Pengamatan						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7
8	Kota Berau	66,00	6,45	51,95	26,35	13,29	1,06	11,80
9	Kota Bontang	64,52	12,44	57,46	28,55	12,88	0,59	6,52
10	Kab. Mahakam Ulu	77,12	9,05	46,76	21,35	12,47	2,67	1,40

4.2 Pendeteksian Multikolinearitas

Hasil pendeteksian multikolinearitas dari data penelitian dapat dilihat pada tabel 2.

Tabel 2: Matriks Korelasi

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1,00	0,30	0,26	0,39	0,48	0,01	0,20
X_2	0,30	1,00	0,47	0,09	0,28	0,40	0,21
X_3	0,26	0,47	1,00	0,71	0,64	0,73	0,01
X_4	0,39	0,09	0,71	1,00	0,37	0,62	0,07
X_5	0,48	0,28	0,64	0,37	1,00	0,73	0,39
X_6	0,01	0,40	0,73	0,62	0,58	1,00	0,15
X_7	0,20	0,21	0,01	0,07	0,39	0,15	1,00

Berdasarkan Tabel 2, dapat dilihat bahwa tidak terdapat nilai korelasi yang lebih dari 0,80. Sehingga dapat disimpulkan bahwa tidak terjadi multikolinearitas antar variabel.

4.3 Metode *Complete Linkage*

Langkah-langkah dari metode *complete linkage* adalah sebagai berikut:

- a. Menghitung matriks jarak antar data. Matriks jarak yang dibentuk berdasarkan Persamaan 2 dapat dilihat pada Tabel 3.

Tabel 3: Matriks Jarak *Euclidean*

Cluster	1	2	3	4	5	6	7	8	9	10
1	0,00	16,88	8,79	11,22	14,61	15,42	11,84	7,95	7,08	21,33
2	16,88	0,00	22,54	19,27	17,24	19,15	15,63	14,75	22,05	29,48
3	8,79	22,54	0,00	8,52	15,49	15,64	12,64	9,95	6,86	20,22
4	11,22	19,27	8,52	0,00	8,18	9,32	7,71	7,38	10,93	19,03
5	14,61	17,24	15,49	8,18	0,00	4,28	6,60	8,97	16,09	17,98
6	15,42	19,15	15,64	9,32	4,28	0,00	5,68	8,90	15,85	14,00
7	11,84	15,63	12,64	7,71	6,60	5,68	0,00	4,15	13,30	15,68
8	7,95	14,75	9,95	7,38	8,97	8,90	4,15	0,00	10,08	17,14
9	7,08	22,05	6,86	10,93	16,09	15,85	13,30	10,08	0,00	19,16
10	21,33	29,48	20,22	19,03	17,98	14,00	15,68	17,14	19,16	0,00

- b. Menentukan jarak terdekat dari matriks jarak yang diperoleh. Berdasarkan Tabel 3, dapat dilihat bahwa *cluster 7* dan *cluster 8* merupakan *cluster* dengan jarak terdekat yaitu sebesar 4,15, sehingga kedua *cluster* digabungkan menjadi satu *cluster*.

- c. Menghitung jarak cluster gabungan dengan cluster lain menggunakan Persamaan (1).
- d. Membuat matriks jarak terbaru yang dibentuk berdasarkan perhitungan pada langkah (c).
- e. Mengulangi langkah (b) sampai dengan langkah (d) hingga terbentuk jumlah cluster yang diinginkan. Pada penelitian ini, peneliti ingin membentuk empat cluster berdasarkan status pembangunan manusia suatu negara atau wilayah di UNDP. Matriks jarak terbaru empat cluster dapat dilihat pada Tabel 4.

Tabel 4: Matriks Jarak Terbaru Empat Cluster

Cluster	1,3,9	2	4,5,6,7,8	10
1,3,9	0,00	22,54	16,09	21,33
2	22,54	0,00	19,27	29,48
4,5,6,7,8	16,09	19,27	0,00	19,03
10	21,33	29,48	19,03	0,00

Tahap selanjutnya yaitu memberi ciri spasifik untuk menggambarkan keanggotaan dari masing-masing cluster dengan melihat rata-rata persentase indikator IPM untuk setiap kabupaten/kota di Prov. Kalimantan Timur. Hasil perhitungan rata-rata persentase wilayah dapat dilihat pada Tabel 5.

Tabel 5: Anggota Cluster yang Terbentuk

Cluster	Kabupaten /Kota	Rata-Rata Persentase (%)
1	Balikpapan	26,25
	Bontang	25,18
	Samarinda	26,14
2	Kutai Kartanegara	26,67
3	PPU	23,59
	Paser	22,90
	Kutai Barat	23,20
	Kutai Timur	24,30
	Berau	25,27
4	Mahakam Ulu	24,40

Berdasarkan Tabel 5, dapat dilihat dari keempat cluster yang terbentuk adalah sebagai berikut:

- Cluster pertama

$$\text{Cluster pertama} = \frac{26,25 + 25,18 + 26,14}{3} = 25,86$$

Memiliki rata-rata persentase sebesar 25,86% yang menandakan bahwa cluster pertama termasuk wilayah yang memiliki persentase indikator IPM tinggi.

- Cluster kedua

$$\text{Cluster kedua} = 26,67$$

Memiliki rata-rata persentase sebesar 26,67% yang menandakan bahwa cluster kedua termasuk wilayah yang memiliki persentase indikator IPM sangat tinggi.

– Cluster ketiga

$$\text{Cluster ketiga} = \frac{23,59 + 22,90 + 23,20 + 24,30 + 25,27}{5} = 23,85$$

Memiliki rata-rata persentase sebesar 23,85% yang menandakan bahwa *cluster* ketiga termasuk wilayah yang memiliki persentase indikator IPM rendah.

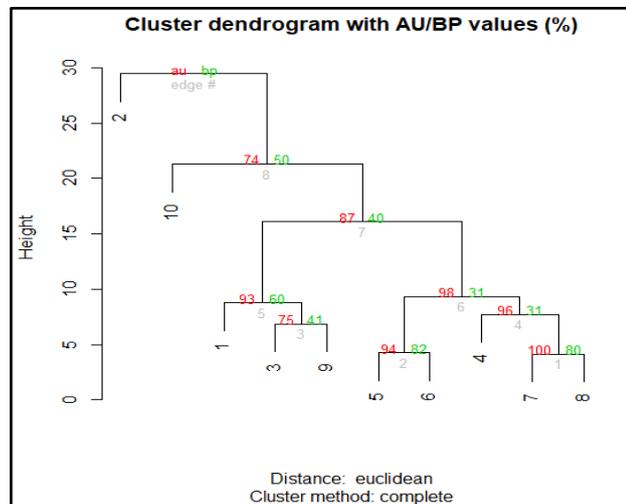
– Cluster keempat

$$\text{Cluster keempat} = 24,40$$

Memiliki rata-rata persentase sebesar 24,40% yang menandakan bahwa *cluster* keempat termasuk wilayah yang memiliki persentase indikator IPM sedang.

4.4 Metode *Multiscale Bootstrap*

Hasil penelitian dengan metode *multiscale bootstrap* menggunakan *software R* dapat dilihat pada Gambar 1.



Gambar 1: Dendrogram *multiscale bootstrap*

Pada *cluster* pertama, nilai AU diperoleh sebesar 0,93. Karena nilai AU (0,93) ≤ 0,95, maka dapat disimpulkan bahwa *cluster* pertama tidak *valid*. Kemudian pada *cluster* kedua, hanya terdapat satu anggota yaitu Kabupaten Kutai Kartanegara dengan nilai AU sebesar 1,00. Karena nilai AU (1,00) ≥ 0,95, maka dapat disimpulkan bahwa *cluster* kedua *valid*. Kemudian pada *cluster* ketiga, nilai AU diperoleh sebesar 0,98. Karena nilai AU (0,98) ≥ 0,95, maka dapat disimpulkan bahwa *cluster* ketiga *valid*. Kemudian pada *cluster* keempat, hanya terdapat satu anggota yaitu Kabupaten Mahakam Ulu dengan AU sebesar 1,00. Karena nilai AU (1,00) ≥ 0,95, maka dapat disimpulkan bahwa *cluster* keempat *valid*.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan, didapatkan kesimpulan sebagai berikut:

1. Pada penelitian ini, *cluster* yang terbentuk menggunakan metode *complete linkage* ada empat. *Cluster* pertama merupakan *cluster* wilayah yang memiliki persentase indikator IPM tinggi dengan anggota *cluster* pertama yaitu Kota Balikpapan, Kota Samarinda, dan Kota Bontang. Kemudian *cluster* kedua

merupakan *cluster* wilayah yang memiliki persentase indikator IPM sangat tinggi dengan anggota *cluster* kedua yaitu Kabupaten Kutai Kartanegara. Kemudian *cluster* ketiga merupakan *cluster* wilayah yang memiliki persentase indikator IPM rendah dengan anggota *cluster* ketiga yaitu Kabupaten Penajam Paser Utara, Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Kutai Timur, dan Kabupaten Berau. Sedangkan cluster keempat merupakan *cluster* wilayah yang memiliki persentase indikator IPM sedang dengan anggota *cluster* keempat yaitu Kabupaten Mahakam Ulu.

2. Pada penelitian ini, terdapat tiga *cluster* yang *valid*, yaitu *cluster* kedua, *cluster* ketiga, dan *cluster* keempat. Hal ini menandakan bahwa ketiga *cluster* tersebut sudah akurat sehingga dapat menggambarkan populasi yang sebenarnya.

DAFTAR PUSTAKA

- [1] Yulianto, S. dan Hidayatullah, K. H. (2014). Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Statistika*. 2. 56-63. Diakses dari <https://jurnal.unimus.ac.id/>.
- [2] Supranto, J. (2004). *Analisis Multivariat Arti & Interpretasi*. Jakarta: Penerbit Rineka Cipta.
- [3] Efron, B. dan Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman Hall.
- [4] UNDP. (1995). *Human Development Report*. New York: UNDP.
- [5] Nugroho, S. (2008). *Statistika Multivariat Terapan edisi pertama*. Bengkulu: UNIB Press.
- [6] Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka.
- [7] Anuraga, G. (2015). *Hierarchical Clustering Multiscale Bootstrap* untuk Pengelompokan Kemiskinan di Jawa Timur. *Jurnal Statistika Universitas Muhammadiyah Semarang*. 3. 27-33. Diakses dari <https://jurnal.unimus.ac.id/>.
- [8] Yatskiv, I. dan Lada, G. (2005). *The Methods of Cluster Analysis Resulty Validation*. *Proceeding of International Conference RelStat'04*. 6. 75-80. Diakses dari <https://www.researchgate.net>