

HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP PADA CAPAIAN DIMENSI IPM TAHUN 2020 YANG TERDAMPAK COVID-19

Anjeli Dwi Laksmi, Dadan Kusnandar, Nurfitri Imro'ah

INTISARI

Analisis cluster hirarki adalah pengamatan yang dipakai guna melakukan pengelompokan data (objek) menurut kriterianya. Salah satu metode di analisis cluster yang dapat dipakai guna melakukan pengelompokan data pada berbagai aspek adalah metode average linkage. Penelitian ini menggunakan data dari variabel yang mempengaruhi capaian dimensi Indeks Pembangunan Manusia (IPM) ketika periode 2020 yang terdampak Covid-19, ialah dimensi umur panjang serta hidup sehat, dimensi pengetahuan, serta dimensi standar hidup layak. Pengkajian ini memiliki tujuan guna melakukan pengelompokan Kabupaten/Kota di Kalimantan Barat menurut data yang diperoleh dengan memakai metode hierarchical clustering multiscale bootstrap. Tahapan yang dilakukan dalam penelitian ini yaitu, metode hirarki menggunakan average linkage untuk membentuk cluster dilanjutkan dengan multiscale bootstrap untuk melihat validitas hasil cluster berdasarkan nilai approximately unbiased (AU) jika memiliki nilai $\geq 0,95$ maka valid. Dengan metode average linkage terbentuk sebanyak empat cluster yaitu cluster pertama terdiri dari Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kubu Raya, Kota Pontianak, dan Kota Singkawang. Cluster kedua terdiri dari Kabupaten Landak dan Kabupaten Kapuas Hulu. Cluster ketiga terdiri dari Kabupaten Sanggau, Kabupaten Ketapang, Kabupaten Sintang, Kabupaten Sekadau, dan Kabupaten Melawi. Kemudian yang terakhir cluster keempat terdiri dari Kabupaten Kayong Utara. Hasil cluster kemudian dilanjutkan dengan multiscale bootstrap dengan melihat nilai AU diperoleh hasilnya yaitu cluster pertama, cluster kedua, dan cluster ketiga valid sedangkan cluster keempat tidak valid.

Kata Kunci: Average Linkage, Multiscale Bootstrap, IPM

PENDAHULUAN

Analisis *cluster* merupakan suatu teknik dalam analisis multivariat yang dipakai guna pengelompokan objek. Analisis *cluster* bermaksud yaitu guna mengelompokkan objek-objek pengamatan menurut kriteria yang dipunyai. Objek-objek yang mempunyai jarak terdekat berkelompok sebagai satu *cluster*. Kedekatan jarak yang diperoleh membuktikan bahwasanya objek-objek itu mempunyai keserupaan kriteria. Ukuran kesamaan yang digunakan adalah ukuran jarak antar-objek [1].

Ada dua teknik umum pada analisis *cluster*, ialah non-hirarki serta hirarki. Pada teknik non-hirarki terdapat teknik *k-means*, teknik *k-harmonic means*, dan metode *k-modes*. Pada teknik hirarki terdapat beberapa teknik, ialah teknik *single linkage*, *complete linkage*, *average linkage*, serta *ward* [2]. Teknik hirarki merupakan sebuah teknik dalam analisis *cluster* yang membentuk tingkat terkhusus semisal struktural pohon dikarenakan tahapan *cluster* dilaksanakan dengan bertahap dan bertingkat. Teknik hirarki dibagi menurut metode proses ukur jarak tiap kelompok yang dipakai guna pengelompokan objek.

Validitas merupakan hal yang penting dalam analisis *cluster*. Validitas digunakan untuk melihat keakuratan hasil analisis sehingga dapat menggambarkan populasinya. *Multiscale bootstrap* merupakan metode yang dapat dipakai guna meninjau validitas sebuah *cluster*. Teknik tersebut dapat digunakan untuk melihat validitas dari analisis *cluster* sebelumnya. Teknik tersebut berproses melalui pendekatan *bootstrap resampling* guna tiap *cluster*. *Multiscale bootstrap resampling* dipakai guna melakukan perhitungan angka AU untuk setiap *cluster* yang terbentuk [3].

Pada awal tahun 2020 pandemi *coronavirus diseases* (Covid-19) mulai menginfeksi berbagai bangsa. Indonesia merupakan satu diantara bangsa yang mengalami dampak dari pandemi tersebut, kasus pertama kali yang muncul pada awal maret tahun 2020 hingga jumlah kasus terus meningkat dari waktu ke waktu. Dalam penyusunan IPM meliputi tiga dimensi, ialah dimensi usia panjang serta hidup sehat, pengetahuan, serta tolak ukur hidup baik. Berdasarkan publikasi Badan Pusat Statistik (BPS) tahun 2020 menyatakan bahwa gapaian dimensi usia panjang serta hidup sehat tumbuh lambat selama pandemi Covid-19. Hal ini dapat dipengaruhi oleh empat faktor tertentu yang bisa menyebabkan hal tersebut ialah faktor lingkungan, tindakan sehat, layanan kesehatan, serta keturunan. Kemudian ada capaian dimensi pengetahuan tumbuh melambat selama pandemi Covid-19 yang mana pada tahun 2010 hingga tahun 2020 memang mengalami peningkatan tetapi sejak tahun 2014 cenderung mengalami perlambatan. Dan yang terakhir dimensi standar hidup layak, mengalami perlambatan selama pandemi dikarenakan kegiatan perekonomian yang menurun sehingga menyebabkan menurunnya penyerapan karyawan serta Tingkat Pengangguran Terbuka (TPT) berdampak terhadap meningkatnya kemiskinan [4]. Dalam pengkajian ini analisis *cluster* dipakai guna pengelompokan data pada variabel yang mempengaruhi capaian dimensi penyusun IPM tahun 2020.

Pengkajian ini memiliki tujuan guna mengelompokkan Kabupaten/Kota di Kalimantan Barat berdasarkan variabel yang mempengaruhi capaian dimensi IPM tahun 2020 yang terdampak Covid-19 menggunakan metode *Hierarchical Clustering Multiscale Bootstrap*. Dengan langkah-langkah yang akan dilakukan yaitu menggunakan metode *average linkage* untuk menghasilkan jumlah *clusternya* kemudian dilanjutkan dengan *multiscale bootstrap* untuk melihat validitas dari hasil *clusternya*. Objek yang diteliti merupakan variabel yang mempengaruhi capaian dimensi IPM tahun 2020 yang terdampak Covid-19. Data yang dipakai pada pengkajian ini ialah data seluruh Kabupaten/Kota di Kalimantan Barat pada tahun 2020 berdasarkan dimensi penyusun IPM yang terdampak Covid-19.

ANALISIS CLUSTER

Analisis *cluster* suatu analisis guna melakukan pengelompokan objek-objek pada kelompok yang hampir serupa. Objek untuk masing-masing kelompok memiliki kesamaan satu dengan lainnya serta berbeda jauh terhadap objek dari *cluster* yang lain. Analisis *cluster* meneliti semua korelasi interdependensi, tidaklah terdapat ketidakserupaan variabel independen serta variabel dependen [2]. Analisis *cluster* membagi sejumlah objek-objek dalam satu ataupun beberapa *cluster*. *Cluster* yang benar ialah *cluster* yang mempunyai [5]:

1. Kesamaan yang besar antar-objek pada suatu *cluster*.
2. Perbedaan yang besar antar-*cluster* dengan *cluster* lain.

Cluster dalam dasarnya melakukan pencarian serta mengelompokkan objek-objek yang memiliki kesamaan satu dan yang lainnya, sehingga memiliki kesamaan adalah karakteristik dasar dari teknik *cluster*. Tahapan perubahan objek-objek hingga objek tersebut bisa dikategorikan pada satu ataupun beberapa *cluster* ialah [6]:

1. Menentukan ukuran jarak

Mengukur kesamaan antar-objek selaras dengan pedoman dasar *cluster* yang melakukan pengelompokkan objek yang memiliki kesamaan, sehingga tahapan awal ialah melakukan pengukuran seberapa jauh terdapat keserupaan tiap objek. Salah satu teknik yang dapat digunakan yaitu melakukan pengukuran jarak antara dua objek. Ada tiga ukuran kedekatan yang dapat digunakan untuk mengukur kesamaan jarak, yaitu:

- 1) Jarak *Euclidean*.

Digunakan untuk melakukan pengukuran total kuadrat perbedaan angka dari setiap variabel. Dibentuk rumus dibawah ini:

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

2) Jarak *Squared Euclidean*.

Berbeda dengan jarak *Euclidean* yang diakarkan, pada jarak *Square Euclidean* tidak diakarkan. Rumusnya ialah:

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

3) Jarak *Cityblock*.

Pada jarak *Cityblock*, jarak antar dua objek adalah total ketidakserupaan mutlak pada angka guna tiap variabel. Rumusnya sebagai berikut:

$$d_{ij} = \sum_{l=1}^p |x_{il} - x_{jl}| \quad (3)$$

dimana,

- d_{ij} : Jarak antar objek ke- i dengan objek ke- j
- p : Banyak variabel
- x_{il} : Nilai objek ke- i pada variabel ke- l
- x_{jl} : Nilai objek ke- j pada variabel ke- l
- l : $1, 2, \dots, p$

2. Standarisasi data.

Sesudah metode proses ukur jarak ditentukan, yang butuh ditinjau juga ialah apakah data yang digunakan memiliki ketidakserupaan yang signifikan. Contohnya, apabila variabel morbiditas memiliki satuan puluhan dan TPT memiliki satuan ratusan juta, maka perbedaan ini membuat hasil yang diperoleh menjadi tidak valid.

3. Proses *cluster*.

Setelah objek yang memiliki satuan yang berbeda distandarisasikan, tahapan setelahnya ialah membentuk *cluster*. Tahapan utaman *cluster* ialah mengelompokkan objek, yang dapat dilaksanakan melalui dua teknik:

1) Teknik non-hirarki.

Dalam teknik ini diawali melalui penentuan total *cluster* yang dikehendaki. Sesudah total *cluster* didapati, barulah tahapan *cluster* dilaksanakan dengan tidak mengiringi tingkatan.

2) Teknik hirarki.

Teknik ini mengawali pengelompokan melalui dua ataupun lebih objek yang memiliki keserupaan terdekat. Setelah itu, tahapan dilanjutkan kepada objek lainnya yang memiliki kesamaan kedua. Hingga *cluster* ada tingkat yang gamblang pada tiap objek, dari yang paling sama hingga paling tidak sama.

4. Memberi karakteristik pada *cluster*.

Menginterpretasi pada *cluster*, yang dalam intinya memberikan nama spesifik guna menceritakan isi *cluster* itu.

5. Validasi.

Cluster yang terbentuk setelah itu dilakukan validasi apakah peroleha itu valid atau tidak.

ASUMSI ANALISIS CLUSTER

Ada asumsi yang wajib diperhatikan pada analisis *cluster* yang wajib dipenuhi. Asumsi tersebut terdiri dari dua hal yaitu sebagai berikut [7]:

1. Sampel yang diperoleh untuk penelitian bisa mewakili populasi.
2. Tidak terjadi multikolinearitas (Non-Multikolinearitas).

Multikolinearitas adalah adanya suatu korelasi yang tinggi pada dua variabel atau lebih pada sebuah model regresi. Multikolinearitas terjadi ketika variabel yang digunakan untuk penelitian saling terkait. Untuk mengetahui terdapat ataupun tidak multikolinearitas, bisa dilaksanakan perhitungan melalui melihat dari nilai-nilai koefisien korelasi *pearson*. Koefisien korelasi *Pearson* dipakai guna melakukan pengukuran korelasi antar satu variabel terhadap variabel lainnya apakah terdapat korelasi. Dengan rumus sebagai berikut:

$$r_{XY} = \frac{n \sum_{i=1}^n X_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)(n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)}} \quad (4)$$

dimana,

- n : Banyak pasangan data
- X_i : Variabel X ke- i
- Y_i : Variabel Y ke- i
- X_i^2 : Kuadrat dari variabel X ke- i
- Y_i^2 : Kuadrat dari variabel Y ke- i
- i : $1, 2, 3, \dots, n$

METODE AVERAGE LINKAGE

Average linkage ialah variasi dari metode *single linkage* serta metode *complete linkage* ialah menghitung jarak antar dua *cluster* dimana jarak itu dilakukan perhitungan dalam setiap *cluster* melalui meminimalkan rerata jarak antar pasangan *cluster* yang dikelompokkan. Tahapan yang dilakukan adalah sebagai berikut [6]:

1. Diawali n *cluster*, dengan tiap-tiap berisikan satu kesatuan. Apabila ada matriks $n \times n$ berjarak $D = \{d_{(ij)}\}$.
2. Menghitung matriks D pada pasangan *cluster* terdekat. Semisal dipilih pasangan *cluster* yang memiliki kesamaan yaitu objek U dan V maka $D = \{d_{(UV)}\}$.
3. Menggabungkan *cluster* U dan V menjadi sebuah *cluster* yang baru (UV) melalui metode sebagai berikut:
 - 1) Menghilangkan baris serta kolom sesuai dengan *cluster* yang mirip yaitu *cluster* U serta V .
 - 2) Melakukan penambahan baris serta kolom baru melalui pemberian angka jarak antar *cluster* (UV) dengan menggunakan rumus sebagai berikut:

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

4. Diulangi tahapan (2) dan (3) sejumlah $n - 1$ kali, yang mana n ialah total objek.

METODE BOOTSTRAP

Bradley Efron memperkenalkan pertama kali metode *bootstrap* pada tahun 1979. Metode *Bootstrap* memiliki prinsip yaitu untuk memprediksi parameter pada tiap-tiap sampel *bootstrap* sebanyak B dimana adalah sampel acak yang memiliki ukuran n yang diperoleh dengan mengambil populasi n penganalisisan. Penganalisisan ke- i ($i = 1, 2, \dots, n$) pertama mungkin muncul beberapa kali dalam sampel *bootstrap* replikasi ke- r ($r = 1, 2, \dots, B$), adapun penganalisisan lainnya tidak akan muncul sama sekali. Kedudukan sampel asli dalam *bootstrap* dipandang sebagai populasi [2]. Ukuran sampel sebaiknya berjumlah lebih dari 30 dan kurang dari 500 dan *bootstrap* dilakukan karena jumlah ukuran sampel kurang dari 30. Ada dua hal krusial pada penentuan besaran B [8]:

1. Memakai replika *bootstrap* yang minim, misalnya $B = 25$, umumnya telah cukup informatif, $B = 50$ telah mencukupi guna memberi perkiraan persimpangan baku secara benar. Pada pengamatan *cluster* total replikan yang diperlukan ialah $B = 100$ hingga 500.
2. Pemakaian replika $B = 200$ ke atas jarang dibutuhkan guna perkiraan persimpangan baku. Namun, pemakaian B yang relatif besar dibutuhkan guna membentuk interval kepercayaan *bootstrap*.

MULTISCALE BOOTSTRAP

Dalam pengamatan *cluster* metode *multiscale bootstrap* dipakai guna meninjau validitas melalui perolehan pengamatan *cluster* sebelumnya. Metode kerjanya yaitu mengumpulkan sampel yang sudah di *bootstrap* lalu memilih satu diantara perolehan *resampling bootstrap* yang dirasa sudah mewakili data asli [3]. Pendekatan *multiscale bootstrap* memiliki angka *bootstrap probability* (BP) serta nilai *approximately unbiased* (AU). Angka *bootstrap probability* (BP) ialah angka kesempatan sampel data asli yang timbul ketika tahapan *resampling bootstrap* dilaksanakan. Sedangkan angka AU adalah nilai yang membuktikan validitas sebuah *cluster*. Suatu *cluster* dikatakan valid jika apabila $AU \geq 0,95$. Dibawah adalah gambaran pengimplementasian *multiscale bootstrap* dalam *hierarchical clustering* [9]:

1. Data \vec{D} didefinisikan menjadi data sampel dengan ukuran n yang meliputi $\vec{D}_i, i = 1, 2, \dots, n$ yang mana \vec{D}_i adalah vektor data pengamat.
2. Sampel data diperoleh dengan acak serta didapatkan data sampel baru yang diartikan menjadi \vec{D}^* .
3. Tahapan setelahnya ialah dilaksanakan dengan berulang sejumlah b kali hingga diperoleh himpunan data *bootstrap* $(\vec{D}^{*1}, \vec{D}^{*2}, \dots, \vec{D}^{*B})$ yang mana tiap sampel *bootstrap* adalah sampel acak yang independen.
4. Metode penentuan angka N_i' ialah:
 - 1) N_i' yang pertama = $\frac{N}{2}$
 - 2) N_i' yang kedua = N_i' yang pertama + $\left(\frac{N}{p}\right)$
5. Nilai v dan c diperoleh dari hasil perhitungan $z(N_i')$.

$$z(N_i') = v \sqrt{\frac{N_i'}{N}} + c \sqrt{\frac{N}{N_i'}} \quad (6)$$

dimana,

$z(N_i')$:	Nilai z
N	:	Total data pengamatan
N_i'	:	Jumlah pengamatan <i>bootstrap</i> ke - i
p	:	Banyak objek pengamatan
i	:	$1, 2, 3, \dots, p$
v dan c	:	Estimasi parameter dari setiap <i>cluster</i>

6. Nilai *approximately unbiased* (AU) dirumuskan sebagai berikut:

$$AU = \Phi(-v + c) \quad (7)$$

STUDI KASUS

Data yang dipakai dalam pengkajian ini merupakan data sekunder yang diperoleh melalui <https://kalbar.bps.go.id/>. Data yang dipakai adalah variabel yang mempengaruhi capaian dimensi penyusunan IPM pada tahun 2020 yang terdampak Covid-19 di Kabupaten/Kota Kalimantan Barat,

seluruh data memiliki satuan persen. Ada sebanyak sepuluh variabel yang dipakai dalam pengkajian ini, detailnya ada dalam Tabel 1.

Tabel 1. Variabel Penelitian

Dimensi Penyusun IPM	Variabel
Umur Panjang Dan Hidup Sehat	X_1 Rumah tangga dengan akses air minum layak
	X_2 Rumah tangga dengan yang tidak memiliki fasilitas buang air besar
	X_3 Wanita pernah kawin dengan umur ≤ 16 tahun
	X_4 Wanita pernah kawin yang proses melahirkan terakhirnya bukan ditolong oleh tenaga kesehatan
	X_5 Wanita pernah kawin yang proses melahirkan terakhirnya bukan di fasilitas kesehatan
	X_6 Morbiditas
Pengetahuan	X_7 APS 7 – 12 tahun
	X_8 APS 13 – 15 tahun
	X_9 APS 16 – 18 tahun
Standar Hidup Layak	X_{10} TPT

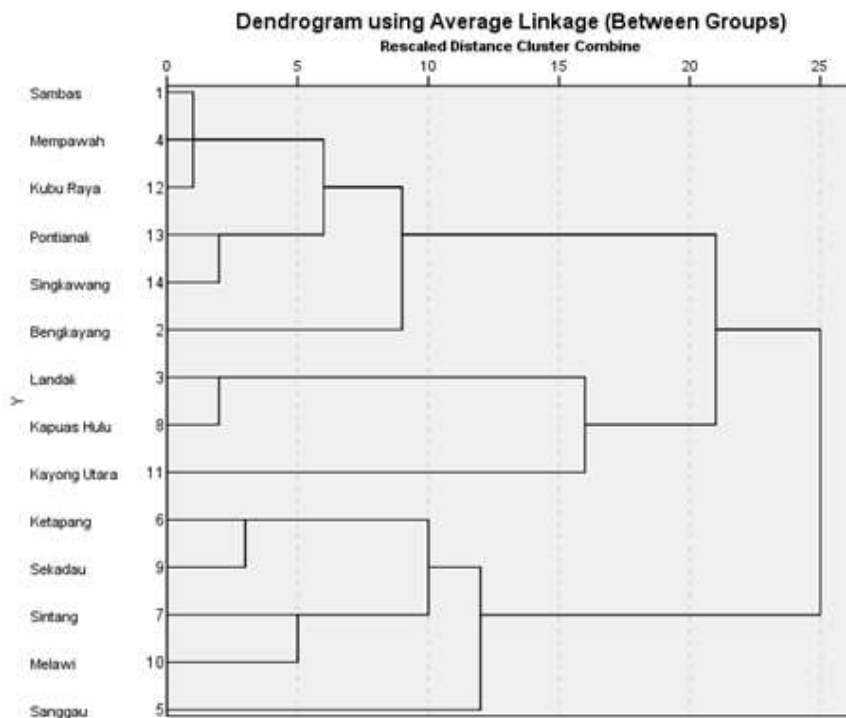
Hasil analisis statistika deskriptif dari data variabel yang didapatkan ada dalam Tabel 2.

Tabel 2. Analisis Statistika Deskriptif

Variabel	Min	Max	Mean	Sd
X_1	46,21	98,62	77,85	17,75
X_2	51,22	96,12	73,84	13,75
X_3	3,08	14,58	9,53	3,20
X_4	0,20	11,66	4,77	3,72
X_5	0,31	30,31	15,28	9,65
X_6	6,80	18,37	12,51	2,88
X_7	96,98	99,53	98,53	0,69
X_8	86,64	97,08	92,15	3,17
X_9	56,19	78,36	68,30	6,80
X_{10}	2,70	12,36	5,43	2,78

Pada Tabel 2 dapat diketahui bahwa untuk nilai minimal, nilai maksimal, dan nilai mean tertinggi terdapat pada variabel X_7 APS 7 – 12 tahun dan nilai terendah pada variabel X_4 wanita pernah kawin yang proses melahirkan terakhirnya bukan ditolong tenaga kesehatan. Nilai standar deviasi tertinggi terdapat pada variabel X_1 rumah tangga dengan akses air minum layak serta nilai terendah terdapat pada variabel X_7 APS 7 – 12.

Analisis *cluster* yang dipakai dalam pengkajian ini yaitu analisis *cluster* metode *average linkage* dengan mengambil sebanyak empat *cluster*. Sebelumnya dilakukan uji multikolinearitas untuk melihat data yang digunakan pada pengkajian ini sudahkah memenuhi syarat pada asumsi analisis *cluster* jika sudah maka dapat dilanjutkan dengan menentukan jarak yang digunakan pada analisis *cluster*. Terdapat tiga jarak yang dapat digunakan, pada pengkajian ini jarak yang digunakan yaitu menggunakan jarak *euclidean*. Hasil pengelompokan analisis *cluster* bisa ditinjau dalam Gambar 1.



Gambar 1. Dendrogram *Average Linkage*

Dapat diketahui bahwa dalam Gambar 1 terdapat sebanyak empat *cluster*, dimana *cluster* pertama memiliki anggota sebanyak enam, *cluster* kedua sebanyak dua anggota, *cluster* ketiga memiliki anggota sebanyak lima, dan anggota *cluster* keempat memiliki anggota sebanyak satu. Untuk lebih jelas pengelompokkan dari Gambar 1 bisa ditinjau dalam Tabel 3.

Tabel 3. Anggota *Cluster*

No	Cluster	Anggota Cluster
1	Pertama	Sambas
2		Bengkayang
3		Mempawah
4		Kubu Raya
5		Kota Pontianak
6		Kota Singkawang
7	Kedua	Landak
8		Kapuas Hulu
9	Ketiga	Sanggau
10		Ketapang
11		Sintang
12		Sekadau
13		Melawi
14	Keempat	Kayong Utara

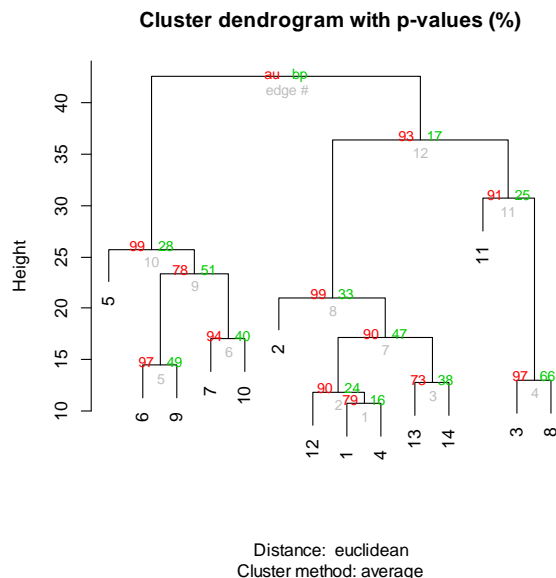
Berdasarkan perhitungan *cluster* yang telah dihasilkan, untuk memberikan ciri spesifik dari hasil *cluster* tersebut dapat dilihat pada angka rerata variabel untuk setiap *cluster*. Angka rerata dalam tiap variabel terdapat dalam Tabel 4.

Tabel 4. Nilai Rata-Rata Setiap *Cluster*

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1	91.07	76.38	58.42	98.62
X_2	85.42	52.98	70.42	63.19
X_3	7.58	13.73	10.07	10.13
X_5	6.97	10.72	24.79	26.66
X_6	12.14	12.83	11.67	18.37
X_7	98.81	99.08	98.04	98.15
X_8	94.17	90.24	90.17	93.72
X_9	72.60	71.96	60.82	72.63
X_{10}	7.24	3.70	4.28	3.71

Berdasarkan hasil Tabel 4 dapat diketahui bahwa untuk *cluster* pertama memiliki variabel yang paling berpengaruh dalam capaian dimensi penyusun IPM tahun 2020 yang terdampak Covid-19 yaitu variabel X_2 Rumah tangga dengan yang tidak memiliki fasilitas buang air besar, variabel X_8 APS 13 – 15 tahun, dan variabel X_{10} TPT. Pada *cluster* kedua memiliki nilai rata-rata tertinggi dari keseluruhan variabel terdapat pada variabel X_3 wanita pernah kawin dengan umur ≤ 16 tahun dan variabel X_7 APS 7 – 12 tahun, artinya pada *cluster* kedua variabel tersebut yang memberikan dampak paling besar. Untuk *cluster* ketiga tidak memiliki nilai rata-rata tertinggi di setiap variabelnya tetapi tetap mengalami dampak. *Cluster* keempat memiliki variabel yang berdampak yaitu variabel X_1 rumah tangga dengan akses air minum layak, variabel X_5 wanita pernah kawin yang proses melahirkan terakhirnya bukan di fasilitas kesehatan, variabel X_6 morbiditas, dan variabel X_9 APS 16 – 18 tahun.

Metode *multiscale bootstrap* digunakan untuk menghitung validitas melalui perolehan pengamatan *cluster* sebelumnya. Teknik ini dipakai memiliki tujuan ialah guna memberikan jaminan bahwasanya *cluster* yang dihasilkan bisa memberikan gambaran populasi sesungguhnya. Dengan syarat jika memiliki nilai $AU \geq 0,95$ maka *cluster* tersebut dapat dikatakan valid. Dalam analisis *cluster average linkage* telah terbentuk sebanyak empat *cluster* kemudian dari hasil tersebut dilanjutkan lagi dengan menggunakan metode *multiscale bootstrap* untuk menentukan keakuratan hasil analisis tersebut. Hasil perhitungan dapat dilihat pada Gambar 2.



Gambar 2. Dendrogram *Multiscale Bootstrap*

Cluster pertama dengan nilai AU yaitu sebesar 0,99 diperoleh pada iterasi kedelapan. *Cluster* kedua nilai AU diperoleh pada iterasi keempat dengan nilai sebesar 0,97. *Cluster* ketiga memiliki nilai AU sebesar 0,99 yang diperoleh dari iterasi kesepuluh, dan yang terakhir *cluster* keempat mempunyai nilai AU sebesar 0,91 yang diperoleh dari hasil iterasi ke-11. Berdasarkan hasil nilai AU yang telah diperoleh syarat untuk nilai AU valid adalah $\geq 0,95$. Pada *cluster* pertama dapat dikatakan valid karena memiliki nilai AU $(0,99) \geq 0,95$. *Cluster* kedua dengan nilai AU $(0,97) \geq 0,95$ sehingga *cluster* kedua juga dinyatakan valid. Kemudian *cluster* ketiga dinyatakan valid dengan nilai AU $(0,99) \geq 0,95$. Dan yang terakhir yaitu *cluster* keempat, tidak valid karena memiliki nilai AU $(0,91) \leq 0,95$. Jadi dari empat *cluster* yang terbentuk hanya *cluster* keempat saja yang dinyatakan tidak valid.

PENUTUP

Berdasarkan hasil perhitungan pada studi kasus yang sudah dilaksanakan, sehingga bisa disimpulkan. Ada empat *cluster* yang terbentuk dan signifikansi dengan nilai AU, *cluster* pertama meliputi Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kubu Raya, Kota Pontianak, dan Kota Singkawang. *Cluster* kedua terdiri dari Kabupaten Landak dan Kabupaten Kapuas Hulu. *Cluster* ketiga terdiri dari Kabupaten Sanggau, Kabupaten Ketapang, Kabupaten Sintang, Kabupaten Sekadau, dan Kabupaten Melawi. Kemudian yang terakhir *cluster* keempat hanya terdiri dari Kabupaten Kayong Utara. Variabel yang berpengaruh dalam penurunan capaian dimensi IPM tahun 2020 pada *cluster* pertama yaitu variabel X_2 rumah tangga dengan yang tidak memiliki fasilitas buang air besar, X_8 APS 13 – 15 tahun, dan X_{10} TPT serta nilai AU $(0,99) \geq 0,95$ sehingga dapat dikatakan bahwa *cluster* pertama valid. *Cluster* kedua variabel yang berpengaruh paling besar yaitu X_3 wanita pernah kawin dengan umur ≤ 16 tahun dan X_7 APS 7 – 12 tahun dengan nilai AU $(0,97) \geq 0,95$ sehingga dapat dikatakan bahwa *cluster* kedua valid. Untuk *cluster* ketiga tidak memiliki nilai rata-rata tertinggi di setiap variabel tetapi tetap mengalami dampak dan memiliki nilai AU sebesar $(0,99) \geq 0,95$ sehingga *cluster* ketiga dapat dikatakan valid. Pada *cluster* keempat variabel yang berdampak banyak X_1 rumah tangga dengan akses air minum layak, X_5 wanita pernah kawin yang proses melahirkan terakhirnya bukan di fasilitas kesehatan, X_6 morbiditas dan X_9 APS 16 – 18 tahun tetapi memiliki nilai AU $(0,91) \leq 0,95$ sehingga *cluster* keempat tidak valid.

DAFTAR PUSTAKA

- [1]. Yulianto S. Hidayatullah KH. Analisis Klaster untuk Pengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Statistika*. 2014; 2:56-63.
 - [2]. Supranto J. *Analisis Multivariat Arti & Interpretasi*. Penerbit Rineka Cipta: Jakarta; 2004.
 - [3]. Efron B. Tibshirani RJ. *An Introduction To The Bootstrap*. Chapman Hall: New York; 1993.
 - [4]. Badan Pusat Statistik. *Indeks Pembangunan Manusia 2020*. BPS: Jakarta; 2021
 - [5]. Prasetyo E. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Penerbit Andi: Yogyakarta; 2012.
 - [6]. Simamora B. *Analisis Multivariat Pemasaran*. PT. Gramedia Pustaka: Jakarta; 2005.
 - [7]. Santoso S. *Mahir Statistika Multivariat dengan SPSS*. PT Elex Media Komputindo: Jakarta; 2018.
 - [8]. Anugraha G. *Hierarchical Clustering Multiscale Bootstrap* untuk Pengelompokkan Kemiskinan di Jawa Timur. *Jurnal Statistika Universitas Muhammadiyah Semarang*. 2015; 3:27-33
 - [9]. Suzuki R. Shimodaira H. *An Application of Multiscale Bootstrap Resampling to Hierarchical Clustering Of Microarray: How Accurate Are These Cluster*. In *Proceedings By The Fifteenth International Conference On Genome Informatics*. 2004: hal.34
-

ANJELI DWI LAKSMI : Jurusan Matematika FMIPA Untan, Pontianak
anjelidwilakmi@student.untan.ac.id

DADAN KUSNANDAR : Jurusan Matematika FMIPA Untan, Pontianak
dkusnand@untan.ac.id

NURFITRI IMRO'AH : Jurusan Matematika FMIPA Untan, Pontianak
nurfitriimroah@math.untan.ac.id
