

**Penerapan Metode Complete Linkage dan Metode Hierarchical
Clustering Multiscale Bootstrap
(Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016)**

**Application of Complete Linkage Method and Hierarchical Clustering
Multiscale Bootstrap Method
(Case Study: Poverty in East Kalimantan Year 2016)**

Lisda Ramadhani¹, Ika Purnamasari², dan Fidia Deny Tisna Amijaya³

^{1,2}Laboratorium Ekonomi Dan Bisnis FMIPA Universitas Mulawarman

³Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

¹E – mail: ramadhanilisda@gmail.com

Abstract

Cluster analysis is an analysis that has a purpose to grouping the data (object). The multiscale bootstrap method in cluster analysis is used as a manner for looking at the validity from the result of cluster analysis. The working process of multiscale bootstrap in cluster analysis is taking a sample that has been bootstrapped and then take the one of bootstrap resampling result that has been reputed to represent the distribution in East Kalimantan 2016. The purpose of this research is looking at the result of data agglomeration in poverty indicator in East Kalimantan 2016 in using a multiscale bootstrap method that produces four cluster types. The first cluster consists of two regencies/cities who has the low percentage of poverty indicator 49,32%. Additionally, the second cluster contains of five regencies/cities with the high percentage of poverty indicator 53,39%. In addition, the third cluster involves of two regencies/cities with the percentage of poverty indicator in high sufficient 51,46%. Finally, the fourth cluster consists of a regency/city which has a percentage of poverty indicator low adequate 51,02%.

Keywords: cluster analysis, multiscale bootstrap, poverty.

Pendahuluan

Analisis *cluster* adalah salah satu analisis statistika yang bertujuan mengelompokkan obyek-obyek berdasarkan kesamaan karakteristik di antara obyek-obyek tersebut. Obyek tersebut akan dikelompokkan ke dalam satu atau beberapa kelompok (*cluster*) sehingga obyek-obyek yang berada dalam satu kelompok akan mempunyai kemiripan satu dengan yang lain (Kuncoro, 2003).

Terdapat dua metode umum dalam algoritma *cluster*, yaitu hierarki dan non-hierarki. Analisis *cluster* dengan metode hierarki 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-hierarki terdapat metode *K-means*. Pada metode hierarki, *cluster* terbentuk dengan sendirinya secara alami dan digambarkan dalam bentuk dendrogram serta tidak mudah diterapkan pada kasus dengan jumlah obyek yang sangat besar, sedangkan pada metode non-hierarki jumlah *cluster* yang terbentuk harus ditentukan terlebih dahulu oleh peneliti dan dapat diterapkan pada kasus dengan obyek yang sangat besar yaitu dengan menggunakan metode *K-means*.

Metode hierarki dibagi berdasarkan teknik pengukuran jarak antar kelompok yang digunakan untuk mengelompokkan obyek. Pengukuran jarak dalam metode hierarki antara

lain single linkage, proses pengelompokkan diawali dengan menemukan dua obyek yang mempunyai jarak minimum. Selanjutnya obyek tersebut menjadi satu kelompok, sedangkan pada metode complete linkage menggunakan jarak maksimum antar kelompok dan proses pengelompokkan sama seperti pada metode single linkage. Penggunaan metode hierarki dengan teknik pengukuran jarak hanya memberikan satu solusi dalam penyelesaiannya, yaitu didasarkan pada ukuran kemiripan pada teknik jarak yang digunakan. Sehingga ukuran akurasi dalam pengelompokkan tidak dapat ditemukan dengan metode klasik ini. Hal ini dapat diatasi dengan menggunakan *multiscale bootstrap*. Metode ini bekerja dengan pendekatan *bootstrap* resampling untuk setiap *cluster*. *Multiscale bootstrap resampling* digunakan untuk menghitung nilai *probability bootstrap* dan *approximately unbiased*.

Penelitian terdahulu yang pernah dilakukan adalah perbandingan hasil analisis *cluster* dengan menggunakan metode *single linkage* dan metode *k-means* (perusahaan di Kutai Barat berdasarkan tingkat pencemaran udara) yang dilakukan oleh Goreti tahun (2015), analisis *cluster* non-hierarki dengan menggunakan metode *k-modes* yang dilakukan oleh Amah tahun (2016), perbandingan kinerja metode *complete linkage* dan *average linkage* dalam menentukan hasil *cluster* (data produksi palawija Provinsi Kalimantan Timur) yang dilakukan oleh Ningsih

tahun (2016). Perbedaan penelitian dengan penelitian sebelumnya yaitu terletak pada perbandingan metode yang digunakan.

Analisis Multivariat

Analisis multivariat dilihat dari akar kata merupakan pengembangan lanjutan dari analisis univariat maupun bivariat. Analisis multivariat berasal dari akar kata *multi* dan *variate*. Dari dua akar kata ini maka analisis multivariat ini merujuk kepada teknik statistika tertentu yang menganalisis banyak variabel secara simultan. Variabel di dalam analisis multivariat dibagi menjadi variabel dependen (*dependent variable*) dan variabel independen (*independent variable*). Sebagai sebuah alat analisis statistika yang bersifat general, analisis multivariat terdiri dari beberapa jenis metode. Metode dalam multivariat dapat dikelompokkan menjadi metode dependen dan metode independen. Metode dependen yaitu ketika variabel dependen dipengaruhi oleh variabel independen. Sedangkan metode independen merupakan metode dimana semua variabel saling berpengaruh. Dengan kata lain semua adalah variabel independen.

Menurut Widarjono (2002), Data variabel dependen maupun independen di dalam analisis multivariat dikelompokkan menjadi dua jenis data yaitu kuantitatif dan kualitatif. Variabel kuantitatif adalah variabel yang disajikan dalam bentuk angka atau metrik. Sehingga variabel ini seringkali disebut dengan variabel yang mempunyai data metrik. Contoh variabel kuantitatif seperti harga mobil, keuntungan perusahaan, harga saham, berat badan seseorang, dan lain sebagainya. Sedangkan variabel kualitatif adalah variabel yang disajikan dengan tidak dalam bentuk angka atau nonmetrik sehingga variabel ini sering kali disebut dengan variabel yang mempunyai data nonmetrik. Data variabel kualitatif dinyatakan dalam bentuk atribut atau merupakan karakteristik dari variabel. Variabel ini seringkali disebut dengan variabel yang mempunyai data nonmetrik. Contoh variabel kualitatif misalnya jenis kelamin, tingkat pendidikan, dan lain sebagainya.

Metode Dependen

Menurut Widarjono (2002), Bila dalam menganalisis hubungan antara variabel di dalam analisis multivariat dapat dibedakan antara variabel dependen dan independen maka metode ini disebut dengan metode dependen. Ada beberapa jenis analisis metode dependen di dalam analisis multivariat. Pengelompokkannya didasarkan oleh dua hal yaitu jumlah variabel dependen dan jenis pengukuran data terhadap variabel baik dependen maupun independen.

Berdasarkan jumlah variabel dependen, maka analisis multivariat dikelompokkan apakah mempunyai satu variabel dependen, dua variabel dependen, atau beberapa variabel dependen. Selanjutnya, setelah diketahui jumlah variabel dependen maka pengelompokkan selanjutnya berdasarkan jenis data variabel dependen maupun variabel independennya.

Tabel 1. Jenis Metode Multivariat Dependen

Jenis Analisis Multivariat	Variabel dependen		Jenis Variabel independen
	Jumlah Variabel	Jenis Variabel	
Regresi			Metrik/
Regresi Logistik	1	Metrik	Nonmetrik
Analisis Diskriminan	1	Nonmetrik	Metrik/
Analisis Konjoin	1	Nonmetrik	Nonmetrik
Analisis Konjoin	>1	Metrik	Nonmetrik
Analisis MANOVA	>1	Metrik	Metrik
			Nonmetrik

Metode Independen

Menurut Widarjono (2002), Tujuan dari analisis independen adalah menganalisis mengapa dan bagaimana variabel-variabel yang ada saling berhubungan. Karena tidak dapat memisahkan antara variabel dependen dan independen, pembagian metode independen ini didasarkan pada jenis variabel yang ada yaitu apakah bersifat metrik atau nonmetrik. Jika jenis variabel adalah metrik maka ada tiga jenis analisis yaitu: analisis faktor, analisis *cluster*, dan *multidimensional scalling*. Sedangkan bila jenis variabel adalah variabel nonmetrik maka ada satu analisis yaitu analisis koresponden.

Tabel 2. Jenis Metode Multivariat Independen

Jenis Analisis	Jenis Variabel
Analisis Faktor	Metrik
Analisis <i>Cluster</i>	Metrik
Analisis Korespondensi	Nonmetrik
Analisis <i>Multidimensional Scalling</i>	Metrik

Analisis Cluster

Menurut Prasetyo (2012), Analisis *cluster* merupakan analisis yang bertujuan untuk mengelompokkan data (obyek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan obyek tersebut dan hubungan di antaranya. Tujuannya adalah agar obyek-obyek yang bergabung dalam sebuah kelompok merupakan obyek-obyek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan obyek dalam kelompok yang lain. Lebih besar kemiripannya (homogenitas) dalam kelompok dan lebih besar perbedaannya di antara kelompok lainnya.

Menurut Prasetyo (2012), Analisis *cluster* akan membagi sejumlah data pada satu atau

beberapa *cluster* tertentu. Pertanyaan yang kemudian timbul adalah ‘apa yang menjadi batas bahwa sejumlah data dapat disebut sebagai satu *cluster*?’. Secara logika, sebuah *cluster* yang baik adalah *cluster* yang mempunyai:

1. Homogenitas (kesamaan) yang tinggi antara anggota dalam satu *cluster*. sebagai contoh, *cluster* konsumen rumah yang peduli lingkungan tentu terdiri atas orang yang mengutamakan kebersihan dan kenyamanan lingkungan rumahnya; mereka yang mengutamakan harga murah tentu tidak dapat digabungkan menjadi anggota *cluster*.
2. Heterogenitas (perbedaan) yang tinggi antara *cluster* yang satu dengan *cluster* lainnya. Dalam contoh di atas, anggota dari *cluster* konsumen rumah yang peduli lingkungan tentu mempunyai pendapat yang jelas berbeda dengan anggota-anggota *cluster* konsumen rumah yang mementingkan harga murah.

Menurut Supranto (2004), Berikut ini merupakan langkah-langkah dalam melakukan proses analisis *cluster*:

1. Merumuskan masalah, hal yang paling penting di dalam analisis *cluster* adalah pemilihan variabel-variabel yang akan dipergunakan
2. Memilih ukuran kedekatan, Pada analisis *cluster* terdapat tiga ukuran kedekatan untuk mengukur kesamaan jarak antara pasangan obyek, yaitu:
 - a. Jarak *Euclidean* mengukur jumlah kuadrat perbedaan nilai masing-masing variabel. Dirumuskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$
 (1)
 - b. Jarak *Squared Euclidean* yang merupakan variasi dari jarak *Euclidean*. Jika pada jarak *Euclidean* diakarkan, maka pada jarak *Squared Euclidean* tidak diakarkan. Dirumuskan sebagai berikut:

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$
 (2)
 - c. *Cityblock* atau biasa disebut dengan jarak *Manhattan*, jarak antar dua obyek merupakan jumlah perbedaan mutlak di dalam nilai untuk setiap variabel. Dirumuskan sebagai berikut:

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$
 (3)
3. Memilih prosedur proses analisis *cluster*, Terdapat dua metode umum dalam algoritma *cluster*, yaitu hierarki dan non-hierarki. Analisis *cluster* dengan metode hierarki ada beberapa metode yaitu metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata

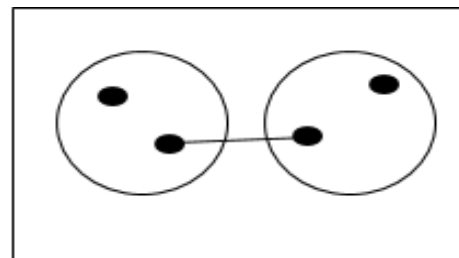
(*average linkage*), dan metode Ward (*Ward’s method*). Pada metode non-hierarki terdapat metode K-means.

4. Menentukan banyaknya *cluster*, Menurut Sudjana (2002), Isi utama dalam analisis *cluster* adalah menentukan berapa banyaknya *cluster*. Sebenarnya tidak ada aturan yang baku menentukan berapa banyaknya jumlah *cluster*, namun demikian ada beberapa petunjuk yang bisa dipergunakan yaitu:
 - a. Pertimbangan teoritis, konseptual, praktis, mungkin dapat diusulkan untuk menentukan berapa banyaknya *cluster* yang sebenarnya.
 - b. Di dalam pengelompokkan hierarki, jarak dimana *cluster* digabung dapat digunakan sebagai kriteria.
 - c. Penentuan banyaknya *cluster* memperhatikan aspek kegunaan
5. Menginterpretasi hasil *cluster*, Menurut Simamora (2005), interpretasi dibutuhkan untuk memberi nama spesifik yang dapat menggambarkan isi *cluster* dan untuk mengetahui profil dari setiap *cluster* yang dibentuk dari rata-rata setiap *cluster* pada setiap obyek.

Metode Hierarki

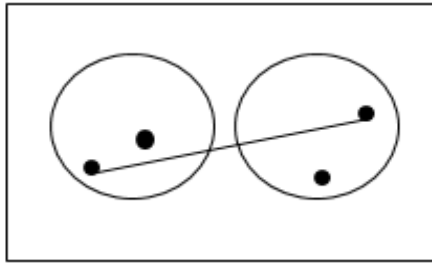
Metode hierarki adalah suatu metode pada analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon karena proses *cluster* dilakukan secara bertahap dan bertingkat. Hasil dari metode hierarki dapat disajikan dalam bentuk dendrogram. Dendrogram adalah representatif visual dari langkah-langkah dalam analisis *cluster* yang menunjukkan bagaimana *cluster* terbentuk dan nilai koefisien jarak pada setiap langkah. Ada empat metode aglomeratif dalam pembentukan *cluster* menurut Simamora (2005) yaitu:

1. *Single Linkage*, metode ini mengelompokkan dua obyek yang mempunyai jarak yang terdekat terlebih dahulu.



Gambar 1. Pautan tunggal antara dua obyek

2. *Complete Linkage*, metode pautan lengkap didasarkan pada jarak maksimum. Jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan obyek yang mempunyai jarak terjauh. Berikut contoh berdasarkan jarak maksimum di antara *cluster*:



Gambar 2. Pautan lengkap antara dua obyek

3. *Average Linkage*, merupakan variasi dari algoritma *single linkage* dan *complete linkage* yaitu menghitung jarak antara dua *cluster* yang disebut sebagai jarak rata-rata dimana jarak tersebut dihitung pada masing-masing *cluster* dengan meminimumkan rata-rata jarak antara pasangan *cluster* yang digabungkan.
4. *Ward's Method*, Metode varian bertujuan untuk memperoleh *cluster* yang memiliki varians internal *cluster* yang sekecil mungkin. Metode varians yang umum digunakan adalah metode ward dimana rata-rata untuk setiap *cluster* dihitung, lalu dihitung jarak antara setiap obyek dan nilai rata-rata kemudian jarak tersebut dihitung semua. Pada setiap tahap, dua *cluster* yang memiliki kenaikan *sum of square* dalam *cluster* yang terkecil digabungkan.

Statistika Deskriptif

Analisis statistika deskriptif adalah bentuk analisis statistik yang bertujuan untuk mendeskripsikan terhadap obyek yang diteliti melalui data sampel atau populasi sebagaimana adanya, tanpa melakukan analisis dan membuat kesimpulan yang berlaku secara umum. Tujuan dari analisis ini adalah memaparkan data secara sederhana sehingga dapat dibaca dan dianalisis secara sederhana. Bentuk penyajian hasil dari analisis deskriptif tergantung dari jenis atau skala data dari variabel yang sedang dianalisis.

Menurut Supranto (2002), Salah satu bagian dari statistika deskriptif yaitu ukuran pemusatan. Ukuran pemusatan terbagi menjadi nilai rata-rata, nilai maksimum, nilai minimum, dan standar deviasi. Nilai maksimum adalah nilai terbesar dari suatu data. Nilai minimum adalah nilai terkecil dari data. Nilai rata-rata adalah nilai yang mewakili himpunan atau kelompok data. Nilai rata-rata umumnya cenderung terletak di tengah suatu kelompok data yang disusun menurut besar atau kecilnya nilai. Berikut ini merupakan cara menghitung rata-rata:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n (x_1 + x_2 + \dots + x_n) \tag{4}$$

Metode Bootstrap

Bootstrap merupakan salah satu metode resampling (pengambilan sampel ulang dari data

yang ada). Efron (1979) menggabungkan metode-metode resampling menjadi suatu metode resampling baru yaitu *bootstrap* nonparametrik sederhana. Pada metode ini resampling data dilakukan dengan pengembalian dari sampel hasil observasi dengan replikasi B kali. Prinsip metode *bootstrap* adalah untuk memperkirakan parameter masing-masing sampel *bootstrap* sebanyak B yang merupakan sampel acak berukuran n yang diambil dengan pengembalian dari populasi n pengamatan.

Berdasarkan hasil penelitian Efron dan Tibshirani (1998), berikut dua poin penting dalam menentukan besaran B:

1. Menggunakan replikasi *bootstrap* yang sedikit, katakanlah B= 25, biasanya sudah cukup informatif. B=50 sudah cukup untuk memberikan estimasi simpangan baku dengan baik. Dalam analisis *cluster* jumlah replikasi yang dibutuhkan adalah B=100 sampai dengan 500.
2. Penggunaan replikasi B=200 ke atas jarang diperlukan untuk estimasi simpangan baku. Akan tetapi, penggunaan B yang lebih besar diperlukan untuk membuat interval kepercayaan *bootstrap*.

Multiscale Bootstrap

Menurut Efron (1998), Metode *multiscale bootstrap* pada analisis *cluster* digunakan sebagai cara untuk melihat kevalidan dari hasil analisis *cluster* sebelumnya. Cara kerja dari *multiscale bootstrap* pada analisis *cluster* adalah mengambil sampel yang telah *bootstrap* kemudian memilih salah satu dari hasil resampling *bootstrap* yang dianggap telah mewakili distribusi data asli. Pada pendekatan *multiscale bootstrap* terdapat nilai *probability bootstrap* dan *approximately unbiased p-values*. Nilai *probability bootstrap* adalah nilai peluang sampel data asli yang muncul saat proses *resampling bootstrap* dilakukan. Berikut ini merupakan ilustrasi penerapan *multiscale bootstrap* pada *hierarchical clustering*:

1. Data \bar{D} didefinisikan sebagai data sampel berukuran n yang terdiri dari $\bar{D}_i = \bar{D}_1, \bar{D}_2, \dots, \bar{D}_n$ dimana \bar{D}_i merupakan vektor data pengamatan.
2. Sampel data diambil secara acak dan diperoleh data sampel baru yang didefinisikan sebagai \bar{D}^* .
3. Langkah ke-2 dilakukan secara berulang sebanyak B kali sehingga didapatkan himpunan data *bootstrap* $(\bar{D}^{*1}, \bar{D}^{*2}, \dots, \bar{D}^{*B})$ dimana setiap sampel *bootstrap* merupakan sampel acak yang independen.
4. Cara menentukan nilai N' adalah sebagai berikut:

- a. N' yang pertama = $\frac{N}{2}$ (5)
 - b. N' yang kedua = N' yang pertama + $\left(\frac{N}{p}\right)$ (6)
 - c. Dan seterusnya sampai N' yang kesepuluh
5. Nilai v dan c diperoleh hasil perhitungan $z(N')$. Nilai $z(N')$ adalah sebagai berikut:

$$z(N') = v\sqrt{N'/N} + c\sqrt{N/N'} \quad (7)$$

6. Menurut Efron (1998), Nilai *Approximately Unbiased* (AU) dirumuskan sebagai berikut

$$AU = \Phi(-v + c) \quad (8)$$

$$BP = \Phi|v + c| \quad (9)$$

Menurut Efron (1998), Nilai *Approximately Unbiased* (AU) memiliki fungsi yaitu menunjukkan kevalidan suatu *cluster*. *Cluster* yang valid harus memiliki nilai $AU \geq 0,95$. Sedangkan nilai *Bootstrap Probability* (BP) memiliki fungsi untuk melihat peluang suatu *cluster* untuk menjadi *cluster* yang valid.

Kemiskinan

Kemiskinan adalah kondisi kehidupan yang serba kekurangan yang dialami seorang atau rumah tangga sehingga tidak mampu memenuhi kebutuhan minimal atau yang layak bagi kehidupannya. Kebutuhan dasar minimal yang dimaksud adalah yang berkaitan dengan kebutuhan pangan, sandang, perumahan dan kebutuhan sosial yang diperlukan oleh penduduk atau rumah tangga untuk memenuhi kebutuhan hidupnya secara layak (Suharto, 2009).

Adapun dimensi kemiskinan menurut Suharto (2009) menyangkut beberapa aspek-aspek sebagai berikut :

1. Aspek ekonomi, secara ekonomi, kemiskinan dapat di definisikan sebagai kekurangan sumber daya yang dapat di gunakan untuk memenuhi kebutuhan hidup dan meningkatkan kesejahteraan sekelompok orang.
2. Aspek Politik, kemiskinan dapat dilihat dari tingkat akses terhadap kekuasaan. Kekuasaan dalam pengertian ini mencakup tatanan sistem politik yang dapat menentukan kemampuan sekelompok orang dalam menjangkau dan menggunakan sumberdaya.

3. Aspek sosial-psikologi, kemiskinan secara sosial-psikologis menunjukkan pada kekurangan jaringan dan struktur sosial yang mendukung dalam mendapatkan kesempatan-kesempatan peningkatan produktivitas.

Hasil dan Pembahasan

Data

Data yang digunakan dalam penelitian adalah data sepuluh indikator kemiskinan yang berada pada sektor ekonomi, sumber daya manusia, pendidikan, dan kesehatan yang terdapat pada kabupaten/kota di Kalimantan Timur tahun 2016. Data yang diperoleh dapat dilihat pada Tabel 3.

Keterangan:

- i= Obyek pengamatan
- 1= Kota Balikpapan
- 2= Kabupaten Kutai Kartanegara
- 3= Kota Samarinda
- 4= Kabupaten Penajam Paser Utara
- 5= Kabupaten Paser
- 6= Kabupaten Kutai Barat
- 7= Kabupaten Kutai Timur
- 8= Kabupaten Berau
- 9= Kota Bontang
- 10= Kabupaten Mahakam Ulu

- X_1 = Persentase jumlah penduduk miskin usia 15 tahun ke atas yang tidak bekerja (satuan dalam %).
- X_2 = Persentase jumlah rumah tangga yang pernah membeli beras raskin(satuan dalam %).
- X_3 = Persentase jumlah pengeluaran perkapita untuk non makanan (satuan dalam %).
- X_4 = Persentase jumlah penduduk miskin usia 15 tahun ke atas yang tidak tamat SD (satuan dalam %).
- X_5 = Persentase angka Melek huruf penduduk miskin usia 15-55 tahun (satuan dalam %).
- X_6 = Persentase jumlah pengguna alat KB di rumah tangga miskin (satuan dalam %).
- X_7 = Persentase jumlah rumah tangga miskin dengan luas lantai perkapita $\leq 8 \text{ m}^2$ (satuan dalam %).
- X_8 = Persentase jumlah rumah tangga miskin yang menggunakan air bersih.(satuan dalam %).
- X_9 = Persentase jumlah rumah tangga miskin yang mendapatkan pelayanan jaminan kesehatan (satuan dalam %).

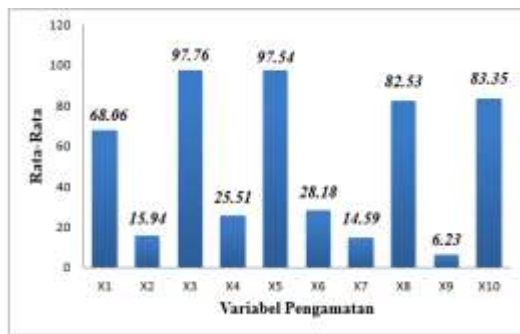
Tabel 3. Data Sembilan Indikator Kemiskinan di Kalimantan Timur

i	Variabel Penelitian									
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
1	70,59	4,99	99,18	25,36	89,08	21,93	11,17	98,75	4,9	99,05
2	67,7	7,86	98,44	27,22	98,44	17,4	12,15	91,61	4,37	94,51
3	70,96	9,51	99,13	23,39	79,13	30,14	15,07	98,85	1,95	36,04
4	66,9	32,27	99,48	24,31	89,38	31,69	13,34	84,04	12,89	96,6
5	68,13	12,46	95,78	25,43	90,63	20,59	17,4	73,96	6,54	86,01
6	67,46	27,1	98,29	26,46	99,46	34,61	15,37	70,86	6,06	73,05
7	68,6	13,44	95,67	29,63	85,66	24,85	10,85	84,72	2,86	69,51
8	67,02	3,63	97,15	30,1	84,15	36,17	15,33	79,9	3,08	87,95
9	68,79	12,66	97,94	20,35	65,94	43,91	16,23	95,27	9,66	91,07
10	64,42	35,45	96,54	22,87	67,62	20,48	19,07	47,33	9,97	99,75

X₁₀= Persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama (satuan dalam %).

Statistika Deskriptif

Grafik rata-rata dari persentase indikator kemiskinan dapat dilihat pada Gambar 3.



Gambar 3. Grafik rata-rata persentase indikator kemiskinan.

Metode Complete Linkage

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

1. Menghitung matriks jarak antar data dengan menggunakan perhitungan jarak *Euclidean*. Contoh perhitungan matriks jarak antara *cluster* 1 dan *cluster* 2 adalah sebagai berikut:

$$d_{12} = \sqrt{\sum_{k=1}^{10} (x_{1k} - x_{2k})^2} = 10,7$$

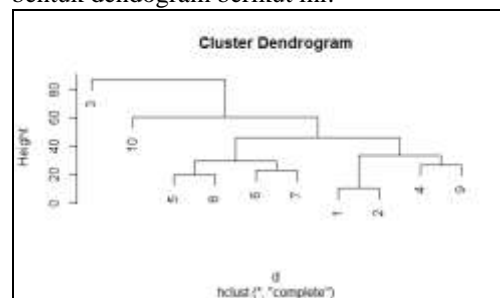
Dengan menggunakan perhitungan yang sama, matriks jarak *cluster* 1 dan *cluster* 3, *cluster* 1 dan *cluster* 4, dan seterusnya

2. Menentukan jarak terkecil atau terdekat dari matriks jarak.
3. Menghitung jarak *cluster* gabungan dengan *cluster* lainnya. Contoh perhitungan untuk menghitung jarak *cluster* gabungan dengan *cluster* lainnya sebagai berikut:

$$d_{(12)3} = \max\{d_{13}, d_{23}\} = \max\{(63,92), (60,64)\} = 63,92$$

4. Membuat matriks jarak terbaru berdasarkan perhitungan sebelumnya.

5. Mengulangi langkah (2) sampai langkah (4) hingga terbentuk empat *cluster*.
6. Berdasarkan matriks jarak telah didapatkan empat *cluster* maka proses pengelompokan terhenti. Karena dalam penentuan jumlah *cluster* yang terbentuk dalam penelitian ini adalah empat *cluster*. Empat *cluster* yang terbentuk mengacu kepada kategori perhitungan kemiskinan pada Badan Pusat Statistik Provinsi Kalimantan Timur (2017). Untuk memperjelas proses penggabungan satu demi satu dapat digambarkan dalam bentuk dendrogram berikut ini:



Gambar 4. Dendrogram metode complete linkage

Setelah *cluster* terbentuk, tahap selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan dari masing-masing *cluster* yang terbentuk yaitu dengan melihat rata-rata persentase indikator kemiskinan tiap masing kabupaten/kota. Selanjutnya ditentukan kabupaten/kota manakah yang termasuk dalam *cluster* wilayah dengan persentase indikator kemiskinan tertinggi, *cluster* wilayah dengan persentase indikator kemiskinan cukup tinggi, *cluster* wilayah dengan persentase indikator kemiskinan cukup rendah, dan *cluster* wilayah dengan persentase indikator kemiskinan terendah. Contoh perhitungan pada rata-rata persentase wilayah sebagai berikut:

Balokpapan

$$= \frac{\sum_{j=1}^{10} x_{1j}}{n} = \frac{70,59+4,99+\dots+99,05}{10} = 53,5$$

Tabel 4. Anggota *cluster* pada metode complete linkage

Cluster	Kabupaten/kota	Rata-rata persentase
1	Balikpapan	53,51
	Kutai Kartanegara	51,97
	Penajam Paser Utara	56,1
	Bontang	55,38
2	Samarinda	48,42
3	Paser	50,21
	Kutai barat	51,75
	Kutai timur	51,76
4	Berau	49,58
	Mahulu	51,02

Berdasarkan Tabel 4 dapat diketahui dari keempat *cluster* yang terbentuk adalah sebagai berikut:

i. *Cluster* pertama

$$= \frac{53,51+51,97+56,1+55,38}{4} = 54,29.$$

Memiliki rata-rata persentase yaitu 54,29% yang berarti bahwa *cluster* pertama termasuk wilayah yang persentase indikator kemiskinan tertinggi.

ii. *Cluster* kedua = $\frac{48,42}{1} = 48,42.$

Memiliki rata-rata persentase yaitu 48,42% yang berarti bahwa *cluster* kedua termasuk wilayah yang persentase indikator kemiskinan terendah.

iii. *Cluster* ketiga = $\frac{50,21+51,75+51,76+49,58}{4} = 50,83.$

Memiliki rata-rata persentase yaitu 50,83% yang berarti bahwa *cluster* ketiga termasuk wilayah yang persentase indikator kemiskinan cukup rendah.

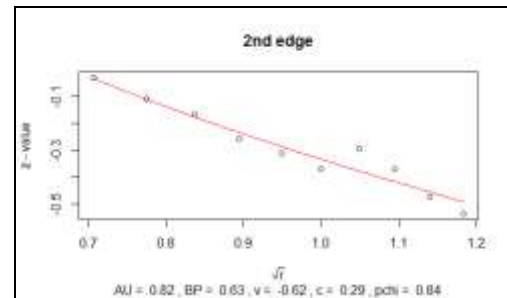
iv. *Cluster* keempat = $\frac{51,02}{1} = 51,02.$

Memiliki rata-rata persentase yaitu 51,02% yang berarti bahwa *cluster* keempat termasuk wilayah yang persentase indikator kemiskinan cukup tinggi.

Metode Multiscale Bootstrap

Metode *multiscale bootstrap* pada analisis *cluster* digunakan sebagai cara untuk melihat kevalidan dari hasil analisis *cluster* sebelumnya. Perhitungan menggunakan metode *multiscale bootstrap* untuk *cluster* kedua dapat dilihat sebagai berikut:

- Melakukan *resampling* sebanyak $b=500$ dan menghitung nilai N' berdasarkan persamaan (4) dan persamaan (7).
- Menghitung Nilai $z(N')$, nilai v dan c . Berikut ini merupakan grafik untuk *cluster* pada iterasi kedua:



Gambar 5. *z-value* untuk iterasi kedua

Berdasarkan Gambar 5 maka diperoleh nilai $z(N')$, nilai \sqrt{r} , nilai AU, nilai BP, nilai v , dan nilai c . Langkah selanjutnya adalah menghitung nilai v dan nilai c yang mengacu pada persamaan (7). Langkah selanjutnya adalah menghitung nilai v dengan mengeliminasi nilai c .

Berdasarkan Tabel 5. diperoleh perhitungan sebagai berikut:

$$-0,034 = v\sqrt{2.598,45/5.196,9} + c\sqrt{5.196,9/2.598,45}$$

$$-0,111 = v\sqrt{3.118,14/5.196,9} + c\sqrt{5.196,9/3.118,14}$$

Diperoleh hasil sebagai berikut:

$$-0,034 = 0,707 v + 1,414 c$$

$$-0,111 = 0,775 v + 1,291 c$$

Langkah selanjutnya adalah menghitung nilai v dengan mengeliminasi nilai c :

$$-0,034 = 0,707 v + 1,414 c | \times 1,291 |$$

$$-0,111 = 0,775 v + 1,291 c | \times 1,414 |$$

Kemudian diperoleh:

$$-0,04389 = 0,912737 v + 1,825474 c$$

$$-0,15695 = 1,09585 v + 1,825474 c$$

$$0,11306 = -0,18311 v$$

$$v = \frac{0,11306}{-0,18311}$$

$$v = -0,61743$$

$$v \approx -0,62$$

Selanjutnya nilai $v = -0,61743$ dimasukkan dalam persamaan dibawah ini:

$$-0,034 = 0,707 v + 1,414 c$$

$$-0,034 = 0,707 (-0,61743) + 1,414 c$$

$$-0,034 = -0,43653 + 1,414 c$$

$$0,40253 = 1,414 c$$

$$c = 0,284675$$

$$c \approx 0,29$$

Berdasarkan perhitungan diatas diperoleh nilai $v = -0,62$ dan nilai $c = 0,29$. Dengan menggunakan persamaan (7), dapat dilihat pada Tabel 5.

- Menghitung nilai AU berdasarkan persamaan (8) dan menghitung nilai BP berdasarkan persamaan (9).

Perhitungan AU dapat dilihat sebagai berikut:

$$AU = \Phi(-(-0,62) + 0,29)$$

$$AU = \Phi(0,91)$$

$$AU = 0,82$$

Nilai AU sebesar 0,82 menyatakan bahwa *cluster* pada iterasi kedua tidak valid, hal tersebut disebabkan karena nilai AU pada iterasi kedua $\leq 0,95$.

Perhitungan BP dapat dilihat sebagai berikut:

$$BP = \Phi |(-0,62 + 0,29)|$$

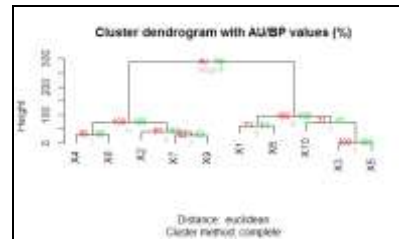
$$BP = \Phi |-0,33|$$

$$BP = 0,63$$

Nilai BP sebesar 0,63 menyatakan bahwa peluang *cluster* pada iterasi kedua untuk menjadi *cluster* yang valid hanya sebesar 0,63.

Berdasarkan hasil yang didapat diketahui pada awal pembentukan *cluster* terdapat sembilan *cluster*. Dari kesembilan *cluster* tersebut

dibentuk kembali secara berulang sampai hanya menjadi empat *cluster* saja. Untuk memperjelas proses penggabungan satu demi satu dapat digambarkan dalam bentuk dendrogram berikut ini:



Gambar 6. Dendrogram multiscale *bootstrap*

Tabel 5. Perhitungan v dan c .

No	N^*	$z(N^*)$	$z(N^*) = v\sqrt{N^*/N} + c\sqrt{N/N^*}$
1	2.598,45	-0,034	$-0,034 = v\sqrt{2.598,45/5.196,9} + c\sqrt{5.196,9/2.598,45}$
2	3.118,14	-0,111	$-0,111 = v\sqrt{3.118,14/5.196,9} + c\sqrt{5.196,9/3.118,14}$
3	3.637,83	-0,177	$-0,177 = v\sqrt{3.637,83/5.196,9} + c\sqrt{5.196,9/3.637,83}$
4	4.157,52	-0,235	$-0,235 = v\sqrt{4.157,52/5.196,9} + c\sqrt{5.196,9/4.157,52}$
5	4.677,21	-0,287	$-0,287 = v\sqrt{4.677,21/5.196,9} + c\sqrt{5.196,9/4.677,21}$
6	5.196,9	-0,334	$-0,334 = v\sqrt{5.196,9/5.196,9} + c\sqrt{5.196,9/5.196,9}$
7	5.716,59	-0,378	$-0,378 = v\sqrt{5.716,59/5.196,9} + c\sqrt{5.196,9/5.716,59}$
8	6.236,28	-0,418	$-0,418 = v\sqrt{6.236,28/5.196,9} + c\sqrt{5.196,9/6.236,28}$
9	6.755,97	-0,456	$-0,456 = v\sqrt{6.755,97/5.196,9} + c\sqrt{5.196,9/6.755,97}$
10	7.275,66	-0,492	$-0,492 = v\sqrt{7.275,66/5.196,9} + c\sqrt{5.196,9/7.275,66}$

Berdasarkan Gambar 6 dapat diketahui terdapat empat *cluster* yang memiliki nilai AU $\geq 0,95$ yaitu *cluster* pada iterasi pertama, iterasi ketiga, iterasi keenam, dan terakhir pada iterasi kedelapan.

Setelah *cluster* terbentuk, tahap selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan dari masing-masing *cluster* yang terbentuk yaitu dengan melihat rata-rata persentase indikator kemiskinan tiap masing kabupaten/kota. Selanjutnya ditentukan kabupaten/kota manakah yang termasuk dalam *cluster* wilayah dengan persentase indikator kemiskinan tertinggi, *cluster* wilayah dengan persentase indikator kemiskinan cukup tinggi, *cluster* wilayah dengan persentase indikator kemiskinan cukup rendah, dan *cluster* wilayah dengan persentase indikator kemiskinan terendah. Rata-rata persentase wilayah pada setiap *cluster* dapat dilihat pada Tabel 6.

Tabel 6. Anggota *cluster* pada multiscale *bootstrap*

Cluster	Kabupaten/kota	Rata-rata persentase
1	Samarinda	48,42
	Paser	50,21
	Penajam Paser Utara	56,1
2	Kutai Barat	51,75
	Kutai Kartanegara	51,97
	Kutai Timur	51,76
3	Bontang	55,38
	Balikpapan	53,51
4	Berau	49,58
	Mahakam Ulu	51,02

Berdasarkan Tabel 6 dapat diketahui dari kedua *cluster* yang terbentuk adalah sebagai berikut:

- i. *Cluster* pertama = $\frac{48,42+50,21}{2} = 49,32$.
Memiliki rata-rata persentase yaitu 49,32% yang berarti bahwa *cluster* pertama termasuk wilayah yang persentase indikator kemiskinan terendah.
- ii. *Cluster* kedua = $\frac{56,1+51,75+51,97+51,76+55,38}{5} = 53,39$.

Memiliki rata-rata persentase yaitu 53,39% yang berarti bahwa *cluster* kedua termasuk wilayah yang persentase indikator kemiskinan tertinggi.

$$\text{iii. Cluster ketiga} = \frac{53,51+49,58}{2} = 51,46.$$

Memiliki rata-rata persentase yaitu 51,46% yang berarti bahwa *cluster* ketiga termasuk wilayah yang persentase indikator kemiskinan cukup tinggi.

$$\text{iv. Cluster keempat} = \frac{51,02}{1} = 51,02.$$

Memiliki rata-rata persentase yaitu 51,02% yang berarti bahwa *cluster* keempat termasuk wilayah yang persentase indikator kemiskinan cukup rendah.

Kesimpulan

Berdasarkan hasil analisis pada penelitian ini, maka dapat diperoleh kesimpulan sebagai berikut:

1. Terdapat empat *cluster* yang terbentuk dengan menggunakan metode *complete linkage*. Keempat *cluster* tersebut adalah *Cluster* pertama memiliki rata-rata persentase yaitu 54,29% yang berarti bahwa *cluster* pertama termasuk wilayah yang persentase indikator kemiskinan tertinggi. Anggota *cluster* pertama adalah Kota Balikpapan, Kabupaten Kutai Kartanegara, Kabupaten Penajam Paser Utara, Kota Bontang. *Cluster* kedua memiliki rata-rata persentase yaitu 48,42% yang berarti bahwa *cluster* kedua termasuk wilayah yang persentase indikator kemiskinan terendah. Anggota *cluster* kedua adalah Kota Samarinda. *Cluster* ketiga memiliki rata-rata persentase yaitu 50,83% yang berarti bahwa *cluster* ketiga termasuk wilayah yang persentase indikator kemiskinan cukup rendah. Anggota *cluster* ketiga adalah Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Kutai Timur, Kabupaten Berau. *Cluster* keempat memiliki rata-rata persentase yaitu 51,02% yang berarti bahwa *cluster* keempat termasuk wilayah yang persentase indikator kemiskinan cukup tinggi. Anggota *cluster* keempat adalah Kabupaten Mahakam Ulu.
2. Terdapat empat *cluster* yang terbentuk dengan menggunakan metode *multscale bootstrap*. Keempat *cluster* tersebut adalah *Cluster* pertama memiliki rata-rata persentase yaitu 49,32% yang berarti bahwa *cluster* pertama termasuk wilayah yang persentase indikator kemiskinan terendah. Anggota *cluster* pertama adalah Kota Samarinda dan Kabupaten Paser. *Cluster* kedua memiliki rata-rata persentase yaitu 53,39% yang berarti bahwa *cluster* kedua termasuk wilayah yang persentase indikator kemiskinan tertinggi.

Anggota *cluster* kedua adalah Kabupaten Penajam Paser Utara, Kabupaten Kutai Barat, Kabupaten Kutai Kartanegara, Kota Bontang, Kabupaten Kutai Timur. *Cluster* ketiga memiliki rata-rata persentase yaitu 51,46% yang berarti bahwa *cluster* ketiga termasuk wilayah yang persentase indikator kemiskinan cukup tinggi. Anggota *cluster* ketiga adalah Kota Balikpapan dan Kabupaten Berau. *Cluster* keempat memiliki rata-rata persentase yaitu 51,02% yang berarti bahwa *cluster* keempat termasuk wilayah yang persentase indikator kemiskinan cukup rendah. Anggota *cluster* keempat adalah Kabupaten Mahakam Ulu.

Daftar Pustaka

- Amah, N. (2016). Analisis *Cluster* Non-Hierarki Dengan Menggunakan Metode K-Modes. *Jurnal: FMIPA Universitas Mulawarman Samarinda*.
- Badan Pusat Statistik Kalimantan Timur, (2017). *Kalimantan Timur Dalam Angka 2017*. Samarinda.
- Clave, J.T.,P. G. Benson., dan Sincich, T. (2005). *Statistik Untuk Bisnis dan Ekonomi Jilid 2*. Jakarta: Erlangga.
- Efron, B dan Tibshirani, R.J. (1998). *An Introduction to the Bootsrap. 2nd ed.* United States of America: Chapman & Hall.
- Goreti, M. (2015). Perbandingan Hasil Analisis *Cluster* Dengan Menggunakan Metode Single Linkage dan Metode C-Means. *Jurnal: FMIPA Universitas Mulawarman Samarinda*.
- Gujarati, D. (1999). *Ekonometrika Dasar*. Bandung: Erlangga.
- Kuncoro, M. (2003). *Metode Riset Untuk Bisnis Dan Ekonomi*. Jakarta: Erlanga.
- Ningsih, S. (2016). Perbandingan Kinerja Metode Complete Linkage dan Average Linkage Dalam Menentukan Hasil *Cluster*. *Jurnal: FMIPA Universitas Mulawarman Samarinda*.
- Prasetyo, E. (2012). *Data Mining Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Santoso, S. (2015). *Menguasai Statistik Multivariat*. Jakarta: Gramedia.
- Sarwako. (2005). *Dasar-Dasar Ekonometrika*. Yogyakarta: Andi.
- Sekaran, U. (2006). *Metode Penelitian Bisnis*. Jakarta: Salemba Empat.
- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka.
- Sudjana. (2002). *Metode Statistika*. Bandung: Tarsito.
- Sugiyono. (2005). *Statistik Untuk Penelitian*. Bandung: Alfabeta.

- Suharto, E. (2009). *Membangun Masyarakat Memberdayakan Rakyat, Kajian Strategis Pembangunan Kesejahteraan Sosial Dan Pekerjaan Sosial*. Bandung: PT Rafika Aditama.
- Supranto, J. (2000). *Teknik Sampling Untuk Survei Dan Eksperimen*. Bandung: PT. Rineka Cipta.
- Supranto, J. (2004). *Analisis Multivariat Arti dan Interpretasi*. Bandung: PT. Rineka Cipta.
- Suzuki, R. (2007). An Application Of Multiscale the *Bootstrap* Resampling To Hierarchical Clustering Of Microarray: How Accurate Are These Clusters, In *Prosceedings By the Fifteenth International Conference On Genome Informatics*.
- Triton. (2006). *SPSS 13.0 Terapan: Riset Statistika Parametrik*. Jakarta: Andi Offset.
- Widarjono, A. (2002). *Analisis Multivariat Terapan*. Jakarta: UPP STIM YKPN.