

**Perbandingan Hasil Analisis *Cluster* Dengan Menggunakan
Metode *Average Linkage* Dan Metode Ward
(Studi Kasus : Kemiskinan Di Provinsi Kalimantan Timur Tahun 2018)**

***Comparison Of Cluster Analysis Results Using Average Linkage Method And Ward Method
(Case Study: Poverty in East Kalimantan Province in 2018)***

Imasdiani¹, Ika Purnamasari², dan Fidia Deny Tisna Amijaya¹

¹Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

²Laboratorium Statistika Ekonomi dan Bisnis FMIPA Universitas Mulawarman

E-mail: ¹18imasdiani@gmail.com

ABSTRACT

Hierarchical cluster analysis is an analysis used to classify data based on its characteristics. The average linkage method and the Ward method are methods of hierarchical cluster analysis. Grouping data from various aspects, one of which is poverty. This study uses poverty indicator data in East Kalimantan in 2018. The average linkage method is based on the average distance size, while the Ward method is based on the size of the distance between clusters by minimizing the number of squares. The purpose of this study was to determine the best method based on the average value of the standard deviation ratio. The results of the study using the average linkage method obtained two clusters, both the average linkage method and the Ward method both obtained two clusters. Where in the average linkage method, the first cluster consists of 7 districts / cities and the second cluster consists of 3 districts / cities. Whereas in the Ward method, the first cluster consists of 6 districts / cities and the second cluster consists of 4 districts / cities. For the best method based on the average standard deviation ratio in groups (S_w) and the standard deviation between groups (S_b), it is found that the ratio in the Ward method is smaller than the average linkage method, which is 2,681 which indicates that the average linkage method is the best method.

Keywords: *average linkage, Ward, Hierarchical cluster analysis, poverty indicators*

Pendahuluan

Analisis *cluster* adalah analisis multivariat yang digunakan untuk mengelompokkan objek menjadi beberapa *cluster* berdasarkan kemiripan variabel yang diamati, sehingga diperoleh kemiripan objek dalam *cluster* yang sama. Analisis *cluster* dibagi menjadi dua metode yaitu metode hirarki dan metode non-hirarki.

Kelebihan penggunaan metode hirarki dalam analisis *cluster* adalah mempercepat pengolahan dan menghemat waktu karena data yang diinputkan akan membentuk dendogram atau tingkatan sendiri sehingga mempermudah dalam penafsiran (Hardius dan Nurdin 2013).

Supranto (2004) menjelaskan bahwa analisis *cluster* jenis metode non-hirarki yaitu metode K-means (*C-means*), sedangkan jenis pada metode hirarki di antaranya yaitu metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata (*average linkage*), dan metode Ward (*Ward's method*).

Average Linkage adalah proses pengelompokan yang didasarkan pada jarak rata-rata antar objeknya. Berbeda halnya dengan *average linkage*, pada metode Ward proses pengelompokan menggunakan pendekatan analisis varians untuk menghitung jarak antar *cluster* (Kuncoro, 2003). Ukuran kemiripan yang digunakan pada metode *average linkage* dan

metode Ward adalah ukuran jarak *Euclid* yang paling sering digunakan pada penelitian. Jarak *Euclidean* merupakan suatu metode yang mencari kedekatan jarak dari 2 buah variabel (Johnson & Winchern, 2002).

Penelitian terkait *clustering* pada metode hirarki sudah banyak dilakukan, diantaranya tahun 2014 Laeli melakukan menggunakan penelitian analisis *cluster* dengan metode *average linkage* dan metode Ward untuk data responden nasabah asuransi jiwa unit link. Hasil yang di dapatkan pada penelitian Laeli (2014) bahwa metode *average linkage* memiliki kinerja lebih baik daripada metode Ward karena memiliki nilai rasio lebih kecil. Selanjutnya Nafisah (2017), melakukan penelitian analisis *cluster average linkage* berdasarkan faktor-faktor kemiskinan di Provinsi Jawa Timur, di dapatkan bahwa metode *average linkage* menghasilkan 3 kelompok yaitu kelompok tingkat rendah, kelompok tingkat sedang dan kelompok tingkat tinggi. Selain itu Salwa (2018) juga pernah melakukan penelitian penggunaan metode Ward pengelompokan Kabupaten/Kota di Provinsi Aceh berdasarkan indikator tanaman pangan dan perkebunan, di dapatkan hasil akhir yaitu 6 *cluster* dengan nilai rasio paling kecil.

Berdasarkan penelitian yang pernah dilakukan sebelumnya, metode *clustering* dapat

diaplikasikan dalam berbagai bidang, salah satunya yaitu bidang kemiskinan sebagaimana yang telah dilakukan oleh Lisda tahun 2018. Kemiskinan merupakan persoalan mendasar dan menjadi perhatian serius dari pemerintah. Salah satu alasan menggunakan metode *average linkage* dan Ward karena metode ini belum banyak dibahas pada penelitian. Berdasarkan uraian yang telah dijabarkan sebelumnya, maka penulis tertarik untuk melakukan penelitian terkait perbandingan hasil analisis *cluster* dengan menggunakan metode *average linkage* dan metode Ward pada data kemiskinan di setiap Kabupaten/Kota Provinsi Kalimantan Timur berdasarkan beberapa faktor yang mempengaruhinya, dalam bentuk skripsi atau tugas akhir dengan judul yaitu "Perbandingan Hasil Analisis *Cluster* dengan Menggunakan Metode *Average Linkage* dan Metode Ward (Studi Kasus : Kemiskinan di Provinsi Kalimantan Timur Tahun 2018)".

Analisis Multivariat

Menurut Sarwono (2012), analisis multivariat merupakan teknik analisis yang digunakan untuk menganalisis data yang menggunakan banyak variabel sebagai pengamatan yang diukur. Sedangkan menurut (Iriawan & Astuti, 2006), analisis multivariat adalah suatu metode analisis data statistik yang dilakukan secara serentak dengan memperhitungkan korelasi antar variabel. Data multivariat adalah data yang tidak hanya terdiri atas satu variabel yang digunakan untuk mengukur karakteristik tertentu. Analisis multivariat terbagi menjadi dua kategori utama yaitu:

1. Metode Ketergantungan (*Dependensi Methods*).

Analisis ketergantungan digunakan apabila tujuan dari analisis adalah untuk menjelaskan atau memprediksi variabel terikat berdasarkan dua atau lebih variabel bebas. Metode ini terdiri dari 4 macam yaitu Analisis Regresi Berganda (*Multiple Regression Analysis*), Analisis Multivariat Varians (*Multivariate Analysis of Variance*), dan Korelasi Kanonikal (*Canonical Correlation Analysis*).

2. Metode Saling Ketergantungan (*Interdependensi Methods*).

Metode yang digunakan untuk menjelaskan seperangkat variabel atau mengelompokkan berdasarkan variabel-variabel tertentu. Metode ini dikelompokkan menjadi tiga yaitu Analisis Faktor (*Factor Analysis*), Analisis *Cluster*, dan Skala Multidimensional (*Multidimensional Scaling*).

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan

pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dari database besar (Turban, dkk., 2005).

Tan (2016) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari basis data yang besar. Teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model. Setelah menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. *Data mining* dapat digunakan untuk memprediksi sekaligus pengelompokan data. Tujuan adanya metode *data mining* adalah agar kita dapat mengetahui pola universal data yang ada. *Data Mining* terbagi menjadi beberapa kelompok yaitu deskripsi (*Description*), estimasi (*Estimation*), prediksi (*Prediction*), klasifikasi (*classification*), pengklusteran (*clustering*), asosiasi (*association*).

Analisis Cluster

Analisis *cluster* merupakan suatu teknik analisis statistika yang bertujuan untuk mengelompokkan objek menjadi beberapa *cluster* berdasarkan kemiripan variabel yang diamati. Sehingga diperoleh kemiripan objek dalam *cluster* yang sama. Analisis *cluster* memiliki tujuan yaitu mengklasifikasikan objek seperti manusia, produk (barang), toko, perusahaan dan lain sebagainya ke dalam kelompok yang relatif homogen berdasarkan suatu variabel yang dipertimbangkan untuk diteliti (Simamora, 2005).

Menurut Supranto (2004) konsep dasar dalam analisis *cluster* yaitu pengelompokan dilakukan berdasarkan kemiripan (*similarity*) antar objek. Kemiripan diperoleh dengan meminimalkan jarak antar objek dalam *cluster* (*within-cluster*) dan memaksimalkan jarak antar *cluster* (*between-cluster*).

Asumsi Dalam Analisis Cluster

Sebelum melakukan pengelompokan pada analisis *cluster* harus memenuhi beberapa asumsi sebagaimana pada Santoso (2014) yaitu:

1. Sampel yang diambil harus mewakili populasi agar proses *cluster* dapat berjalan dengan benar.
2. Antar pengamatan memiliki korelasi. Dalam hal ini, perhitungan koefisien korelasi dapat membantu mendeteksi multikolinearitas pada data. Menurut Clave dkk (2005), rumus yang dapat digunakan dalam menghitung koefisien korelasi adalah:

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

Keterangan:

r = koefisien korelasi variabel X dengan variabel Y , dimana variabel Y merupakan variabel X yang ke-2, 3 dan seterusnya.

n = jumlah pengamatan penelitian.

Ukuran Jarak Kemiripan

Sartono dkk (2003), sesuai prinsip *cluster* yaitu mengelompokkan pengamatan yang mempunyai kemiripan, maka proses pertama pada analisis *cluster* adalah mengukur jarak antar pengamatan. Perkembangan model *cluster* dilakukan untuk mengetahui dua pengamatan tertentu lebih mirip dibandingkan dengan pengamatan lainnya, untuk mempermudah proses dalam pengklasifikasian.

Simamora (2005), jarak yang paling umum digunakan adalah jarak *Euclidean*. Misalkan ada dua pengamatan yaitu A dengan koordinat (x_1, y_1) dan B dengan koordinat (x_2, y_2) maka jarak antar kedua objek tersebut dapat diukur sebagaimana pada persamaan (2):

$$AB = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Selanjutnya, ukuran jarak atau ketidaksamaan antar pengamatan ke- i dengan pengamatan ke- j , disimbolkan dengan d_{ij} . Nilai d_{ij} diperoleh melalui perhitungan jarak *Euclidean* sebagaimana pada persamaan (3):

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

dimana:

p = banyaknya variabel.

d_{ij} = jarak *Euclidean* antar pengamatan ke- i dengan pengamatan ke- j .

x_{ik} = nilai pengamatan ke- i pada variabel ke- k .

x_{jk} = nilai pengamatan ke- j pada variabel ke- k .

$i, j = 1, 2, 3, \dots, n$.

Metode Hirarki

Metode hirarki adalah suatu metode pada analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon dimana terdapat tingkatan (Hirarki) yang jelas antara objek, dari yang paling mirip hingga yang paling tidak mirip, karena proses *cluster* dilakukan secara bertahap dan bertingkat.

Menurut Prasetyo (2012) strategi pengelompokkan hirarki umumnya terbagi menjadi dua jenis yaitu aglomeratif dan divisif. Pengelompokkan secara aglomeratif dimulai dari masing-masing pengamatan sebagai satu buah kelompok, kemudian kelompok yang memiliki tingkat kemiripan yang tinggi dengan kelompok

lain dipasangkan untuk membentuk satu kelompok yang lebih besar. Pengelompokkan secara

aglomeratif dimulai dari masing-masing pengamatan sebagai satu buah kelompok, kemudian kelompok yang memiliki tingkat kemiripan yang tinggi dengan kelompok lain dipasangkan untuk membentuk satu kelompok yang lebih besar. Metode hirarki juga memiliki beberapa metode diantaranya:

1. Metode pautan tunggal (*single linkage*)

Metode *single linkage* adalah suatu metode hirarki yang digunakan untuk membentuk kelompok dari gabungan individu berdasarkan jarak terpendek atau pengelompokkan berdasarkan tingkat kemiripan yang paling tinggi.

2. Metode pautan lengkap (*complete linkage*)

Metode *complete linkage* adalah suatu metode hirarki yang digunakan untuk membentuk kelompok dari gabungan individu berdasarkan jarak terjauh atau pengelompokkan berdasarkan tingkat kemiripan yang paling tinggi.

3. Metode pautan rata-rata (*average linkage*)

Metode *average linkage* adalah proses pengelompokkan yang didasarkan pada jarak rata-rata antar pengamatan.

4. Metode Ward

Metode Ward adalah proses pengelompokkan menggunakan pendekatan analisis varians untuk menghitung jarak antar *cluster* dengan meminimumkan jumlah kuadrat.

Metode pautan rata-rata (*average linkage*)

Metode *average linkage* adalah proses pengelompokkan yang didasarkan pada jarak rata-rata antar pengamatan. Prosedur ini hampir sama dengan *single linkage* maupun *complete linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu *cluster* dengan jarak seluruh individu dalam *cluster* yang lain. Untuk langkah dari algoritma diatas jarak-jarak antara *cluster* (ij) dan *cluster* k yang lain di tentukan oleh:

$$d_{(i,j)k} = \frac{\sum_i \sum_k x_{ik}}{2} \quad (4)$$

dimana:

$d_{(i,j)k}$ = jarak antar pengamatan ke- i dengan pengamatan ke- j .

Metode Ward

Metode Ward adalah proses pengelompokkan menggunakan pendekatan analisis varians untuk menghitung jarak antar *cluster* dengan meminimumkan jumlah kuadrat. Metode Ward merupakan bagian dari metode pengelompokkan yang mengelompokkan sebuah objek menjadi satu

cluster, dengan banyaknya cluster tidak diketahui. Metode Ward didasarkan pada kriteria *sum square error* (SSE) dengan ukuran kehomogenan antara dua pengamatan berdasarkan jumlah kuadrat yang paling minimal. SSE hanya dapat dihitung jika cluster memiliki elemen lebih dari satu objek. Metode Ward ini dihitung berdasarkan persamaan berikut:

$$SSE = \sum_{i=1}^n (X_i - \bar{X}) - (X_I - \bar{X}) \quad (5)$$

dimana:

X_i = data ke- i dari variabel X

\bar{X} = Vektor kolom yang berisi rata-rata nilai pengamatan dalam cluster

n = Banyaknya pengamatan.

Total jarak terdekat dihitung dengan rumus:

$$I = I_1 + I_2 + \dots + I_n \quad (6)$$

dimana:

I = Total jarak terdekat

Jarak antara pengamatan c_1, c_2 dan pengamatan $c_{(1,2)}$ dengan metode Ward yaitu sebagai berikut:

$$I_{(c_1 c_2) c_{1,2}} = \frac{n_{c_1} + n_{c_2}}{n_{c_1 c_2} + n_{c_{(1,2)}}} I_{c_1 c_{(1,2)}} + \frac{n_{c_2} + n_{c_{(1,2)}}}{n_{c_1 c_2} + n_{c_{(1,2)}}} I_{c_2 c_{(1,2)}}$$

dimana:

$I_{(c_1 c_2) c_{(1,2)}}$ = Jarak antar cluster 1,2 dan (1,2)

$I_{c_1 c_{(1,2)}}$ = Jarak antar cluster 1 dan (1,2)

$I_{c_2 c_{(1,2)}}$ = Jarak antar cluster 2 dan (1,2)

$I_{c_1 c_2}$ = Jarak antar cluster 1 dan 2

$n_{c_1} n_{c_2} n_{c_{(1,2)}}$ = Banyak pengamatan pada cluster ke 1, 2, (1,2)

Metode Ward juga memiliki kelebihan yaitu lebih efisien dan cenderung menciptakan cluster sampai ukuran terkecil.

Pemilihan Metode Terbaik dengan Simpangan Baku

Menurut Nuritha, dkk (2013), sebuah metode pengelompokan dikatakan baik jika mempunyai nilai simpangan baku dalam kelompok (S_w) yang minimum dan nilai simpangan baku antar kelompok (S_b) yang maksimum. Dengan rumus sebagaimana pada Persamaan (7):

$$S_w = \frac{1}{c} \sum_{k=1}^p S_k \quad (7)$$

Simpangan baku dalam cluster didefinisikan sebagaimana pada Persamaan (8):

$$S_k = \sqrt{\frac{\sum_{i=1}^n (x_{i(k)} - \bar{x}_{(k)})^2}{n-1}} \quad (8)$$

$$\bar{x}_k = \sqrt{\frac{1}{c-1} \sum_{k=1}^p (x_{(k)} - \bar{x})^2} \quad (9)$$

Simpangan baku antar cluster dapat dirumuskan sebagaimana pada Persamaan (10):

$$S_b = \sqrt{\frac{1}{c-1} \sum_{k=1}^p (\bar{x}_{(k)} - \bar{x})^2} \quad (10)$$

$$\bar{x} = \frac{1}{c} \sum_{k=1}^p \bar{x}_k \quad (11)$$

dengan :

c = banyaknya cluster yang terbentuk

$x_{i(k)}$ = data pengamatan ke- i variabel ke- k

S_k = simpangan baku cluster ke- k

$\bar{x}_{(k)}$ = rata-rata dari cluster ke- k

\bar{x} = rata-rata keseluruhan pengamatan dalam cluster.

Pengelompokan yang baik akan memiliki nilai S_w minimum dan S_b maksimum atau dalam hal ini metode terbaik menghasilkan nilai rasio simpangan baku minimum S_w terhadap S_b dengan rumus sebagaimana pada persamaan (12):

$$R = \frac{S_w}{S_b} \quad (12)$$

Pengertian Kemiskinan

Kemiskinan seringkali ditandai dengan tingginya tingkat pengangguran dan keterbelakangan. Masyarakat miskin umumnya lemah dalam kemampuan berusaha dan terbatas aksesnya terhadap kegiatan ekonomi sehingga akan tertinggal jauh dari masyarakat lainnya yang mempunyai potensi lebih tinggi. Ukuran kemiskinan dilihat dari tingkat pendapatan dapat dikelompokkan menjadi kemiskinan absolut dan kemiskinan relatif (Ginandjar, 1996).

Metode Penelitian

Variabel yang digunakan dalam penelitian ini adalah jumlah sepuluh indikator kemiskinan yang ada di Kalimantan Timur pada tahun 2018. Adapun langkah-langkah yang dilakukan dalam analisis data adalah sebagai berikut:

1. Pengumpulan data dari Badan Pusat Statistik Provinsi Kalimantan Timur.
2. Input data dengan bantuan Ms. Excel 2007.
3. Menentukan objek dan variabel penelitian.
4. Mendeskripsikan data dengan statistika deskriptif yaitu melihat nilai mean.
5. Melakukan pengujian asumsi dalam analisis cluster dengan menghitung koefisien korelasi mengacu pada persamaan (1).
6. Menghitung ukuran jarak kemiripan dengan menggunakan perhitungan jarak *Euclidean* mengacu pada persamaan (3).
7. Membentuk cluster dengan metode *average linkage* sebagai berikut:
 - a. Membuat matriks jarak antar data dengan menggunakan persamaan (3).
 - b. Mencari jarak terkecil dengan menggunakan d_{ij} .
 - c. Mencari jarak rata-rata dengan menggunakan persamaan (4).
 - d. Menghitung kelompok (ij) dan k menjadi kelompok baru (ijk). Menghitung

- kembali matriks jarak dengan cara menghapus baris dan kolom kelompok (ij) dan k, dan menambahkan satu baris dan kolom untuk kelompok (ij)k.
- e. Mengulangi langkah (c) dan (e) sebanyak n-1 setiap iterasi.
 - f. Menginterpretasikan hasil cluster.

Hasil dan Pembahasan

1. Statistika Deskriptif

Tahap awal yang dilakukan dalam penelitian ini adalah analisis deskriptif, bertujuan untuk menggambarkan karakteristik data indikator kemiskinan di Kabupaten/kota yang berada di Kalimantan Timur pada tahun 2018. Karakteristik yang digambarkan pada analisis deskriptif adalah ekonomi, sumber daya manusia, dan kesehatan.

2. Pengujian Analisis Cluster

Pengujian asumsi analisis cluster dilakukan agar mengetahui apakah terjadi multikolinearitas pada data tersebut dan untuk mengetahui kuat tidaknya korelasi yang terjadi antar dua variabel. Semakin besar nilai korelasinya maka variabel tersebut semakin kuat hubungannya. Diperoleh bahwa nilai dari koefisien korelasi antar variabel penelitian tertinggi adalah sebesar 1,000 yang bernilai diatas 0,8 yang artinya terjadi multikolinearitas antar variabel dalam penelitian. Semakin besar nilai korelasinya maka variabel tersebut semakin kuat hubungannya. Diperoleh bahwa rata-rata variabel memiliki nilai korelasi yang kuat terhadap variabel lain. Dengan demikian semakin besar kemungkinan kedua variabel tersebut bergabung dalam satu cluster.

3. Menentukan Ukuran Jarak Kemiripan

Perhitungan matriks jarak Euclidean dengan menggunakan Persamaan (3). Contoh perhitungan jarak antara X₁ dan X₂ adalah sebagai berikut:

$$d_{12} = \sqrt{\sum_{k=1}^{10} (x_{ik} - x_{2k})^2} = 31,667$$

Dengan menggunakan perhitungan yang sama, diperoleh juga jarak antara Kota 1 dan Kota 3 dan seterusnya, Semakin kecil nilai jarak antar dua pengamatan, maka semakin mirip kedua pengamatan tersebut.

4. Metode Average Linkage

Proses klasifikasi dengan metode average linkage adalah sebagai berikut:

1. Proses klasifikasi dengan menggunakan matriks jarak kemiripan yang telah diperoleh, sehingga diperoleh hasil seperti pada Tabel 4.3. Pada proses awal ada sepuluh cluster. Tahap pertama yang dilakukan adalah mencari jarak yang terdekat antara dua pengamatan dari sepuluh pengamatan yang ada. Jarak antara Kota 2 dan Kota 6 merupakan jarak terdekat di

antara yang lainnya yaitu sebesar 12,399, sehingga kedua pengamatan tersebut menjadi satu cluster. Sekarang tersisa sembilan cluster.

2. Kemudian dilakukan perbaikan matriks jarak menggunakan metode average linkage dengan Persamaan (4). Selanjutnya dilakukan perhitungan jarak tiap cluster terhadap cluster baru. Contoh perhitungan untuk perbaikan matriks menggunakan Persamaan (4) sebagai berikut:

$$d_{(2,6)1} = \frac{x_{2,1} + x_{6,1}}{2} = 33,379$$

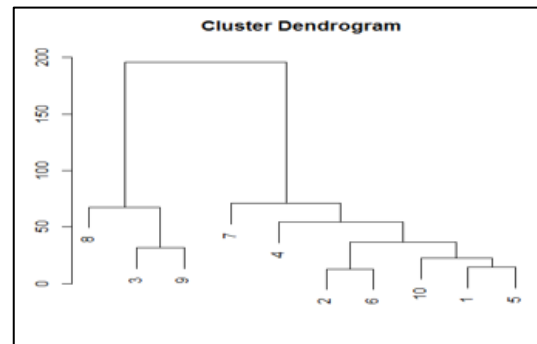
$$d_{(2,6)3} = \frac{x_{2,3} + x_{6,3}}{2} = 178,741$$

Selanjutnya dilakukan perhitungan jarak tiap cluster terhadap cluster baru, Sehingga terbentuk matriks jarak yang baru.

Tabel 2 Matriks Jarak Dua Cluster

d	(2,6,1,5,10,4,7)	(3,9,8)
(2,6,1,5,10,4,7)	0	218,606
(3,9,8)	218,606	0

Berdasarkan matriks jarak pada Tabel 2 didapatkan jarak terdekat adalah Kota 2, 6, 1, 5, 10, 4, dan 7 dengan Kota 3, 9, dan 8 sebesar 28,606 maka proses pengklasifikasian terhenti. Obyek-obyek tersebut menjadi satu cluster. Untuk memperjelas proses penggabungan satu demi satu dapat digambarkan dalam bentuk dendrogram berikut ini:



Gambar 1. Hasil Dendrogram untuk Metode Average Linkage

Dapat dilihat pada Gambar 1 hasil dendrogram untuk metode average linkage pada data indikator kemiskinan di Kalimantan Timur, terdapat dua cluster yaitu cluster pertama adalah Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Berau, Kabupaten Penajam Paser Utara, Kota Bontang, Kabupaten Kutai Timur, dan Kabupaten Mahakam Ulu. Cluster kedua adalah Kabupaten Kutai Kartanegara, Kota Balikpapan dan Kota Samarinda.

Setelah cluster terbentuk, tahap selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan dari masing-masing cluster yang terbentuk, yaitu dengan nilai rata-

rata dari banyaknya indikator kemiskinan tiap masing-masing Kabupaten/Kota. Selanjutnya ditentukan Kabupaten/Kota manakah yang masuk dalam *cluster* tersebut. Termasuk sepuluh wilayah yang ada di Kalimantan Timur. Wilayah terendah pada sepuluh indikator kemiskinan berada di Kabupaten Paser, Kutai Barat dan begitu seterusnya. Hasil perhitungan rata-rata disetiap Kabupaten/Kota dapat dilihat pada Tabel 4. Contoh Perhitungan rata-rata pada Kabupaten Paser dan Kabupaten Kutai Barat yang termasuk kedalam anggota *cluster* pertama dengan menggunakan sepuluh indikator kemiskinan di Provinsi Kalimantan Timur sebagai berikut:

Kabupaten Paser

$$d_{(2,6)1} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{9,03+67,94+48,175+5+\dots+3,25}{10} = 17,759$$

Untuk data yang lainnya juga mengikuti hal yang sama sesuai dengan anggota *cluster* pada masing-masing Kabupaten/kota di Kalimantan Timur, dapat dilihat bahwa proses *pengclusteran* sudah terbentuk yaitu sebagai berikut:

Tabel 3. Anggota *Cluster* pada Metode *Average Linkage*

<i>Cluster</i>	Kabupaten/Kota	Rata-rata
1	Paser	17,759
	Kutai Barat	14,706
	Berau	15,576
	Penajam Paser Utara	15,587
	Bontang	18,497
	Kutai Timur	21,097
	Mahakam Ulu	15,544
2	Kutai Kartanegara	31,467
	Balikpapan	38,137
	Samarinda	32,721

5. Menghitung Kemiripan atau Ketidak Miripan antara Dua Objek dengan Jarak Kuadrat *Euclidean*

Dalam menghitung kemiripan tiap objek (Kabupaten/Kota) dihitung dengan menggunakan perhitungan jarak kuadrat *Euclidean* dengan rumus:

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

$$d_{ij}^2 = (X_{1(1)} - X_{2(1)})^2 + \dots + (X_{1(10)} - X_{2(10)})^2$$

Berikut adalah perhitungan menggunakan rumus jarak kuadrat *Euclidean* tersebut dengan menggunakan data pada Lampiran 1. Misalkan akan dihitung kemiripan antara Kabupaten Paser dan Kabupaten Kutai Barat (objek 1 dan 2):

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

$$d_{ij}^2 = (X_{1(1)} - X_{2(1)})^2 + \dots + (X_{1(10)} - X_{2(10)})^2$$

$$d_{ij}^2 = (9,03 - 9,15)^2 + \dots + (3,25 - 0)^2$$

$$d_{ij}^2 = 1004,4197$$

Selanjutnya dengan menggunakan perhitungan yang sama, diperoleh jarak antara Kabupaten Paser dan Kabupaten Kutai Kartanegara (objek 1 dan 3) sebagai berikut:

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

$$d_{ij}^2 = (X_{1(1)} - X_{2(1)})^2 + \dots + (X_{1(10)} - X_{2(10)})^2$$

$$d_{ij}^2 = (9,03 - 7,41)^2 + \dots + (3,25 - 7,16)^2$$

$$d_{ij}^2 = 21407,35$$

Dengan menggunakan perhitungan yang sama, diperoleh juga jarak antara objek 1 dan 4 dan seterusnya. Semakin kecil nilai jarak antar dua objek, maka semakin mirip kedua objek tersebut. Berdasarkan hasil perhitungan jarak kuadrat *Euclidean* antara Kabupaten Kutai Barat dan Kabupaten Penajam Paser Utara memiliki jarak terdekat diantara kabupaten lainnya dengan jarak sebesar 153,76. Hal ini menunjukkan bahwa Kabupaten Kutai Barat dan Kabupaten Penajam Paser Utara memiliki kemiripan karakteristik dari sudut pandang ekonomi. Sedangkan Kabupaten Mahakam Ulu dengan Kota Balikpapan memiliki jarak terjauh diantara kabupaten lainnya dengan jarak sebesar 86.106,85. Demikian pula untuk penafsiran objek yang lainnya, semakin kecil jarak antara kedua objek maka akan semakin mirip karakteristik dari kedua objek tersebut.

6. Metode Ward

Proses pengklasifikasian dengan metode Ward sama seperti metode *average linkage* dengan menggunakan matriks jarak kemiripan yang telah dilakukan, karena proses dimulai dari dua objek yang terdekat, dari banyaknya kombinasi jarak sepuluh objek yang ada yaitu Kabupaten Kutai Barat dan Kabupaten Penajam Paser Utara dengan jarak sebesar 153,76 yang diperoleh dari perhitungan jarak kuadrat *Euclidean* dengan menggunakan Persamaan (3) sebagai berikut:

$$I_{(i,j)k} = \frac{1}{2}(x_{ik} - x_{jk})(x_{ik} - x_{jk})$$

$$= \frac{1}{2}d^2(x_{ik}, x_{jk}) = \frac{1}{2}\sum_{k=1}^p (x_{ik} - x_{jk})^2$$

$$= \frac{1}{2}(153,76) = 76,88$$

Proses di atas hanya untuk menghitung antar 2 kelompok yang terbentuk, dengan masing-masing kelompok terdiri dari 1 objek. Maka berikutnya menghitung jarak suatu *cluster* dengan *cluster* baru, ada dua *cluster* yang terbentuk, *cluster* 1 misalkan c_1 , dengan anggota kelompok

yaitu Kabupaten/Kota Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Balikpapan dan Samarinda. Cluster 2 misalkan c_2 , dengan anggota kelompok yaitu Kabupaten/Kota Paser, Kutai Barat, Kutai Kartanegara dan Bontang, maka:

a. Pembentukan Cluster 1

Untuk nilai SSE objek pada cluster 1 adalah c_1 (SSE_{c1}) maka:

$$\begin{aligned} SSE_{c1} &= \sum_{i=1}^{n_{c1}} (X_i - \bar{X}_{c1})(X_i - \bar{X}_{c1}) \\ &= (X_1 - \bar{X}_{c1})^2 + \dots + (X_{60} - \bar{X}_{c1})^2 \\ &= 156.035,512 \end{aligned}$$

Jadi, nilai untuk SSE cluster 1 sebesar 156.035,512 menandakan bahwa dari enam kabupaten/kota memiliki kemiripan dari variabel-variabel indikator kemiskinan.

b. Pembentukan Cluster 2

Untuk nilai SSE objek pada cluster 2 adalah c_2 (SSE_{c2}) maka:

$$\begin{aligned} SSE_{c2} &= \sum_{i=1}^{n_{c2}} (X_i - \bar{X}_{c2})(X_i - \bar{X}_{c2}) \\ &= (X_1 - \bar{X}_{c2})^2 + \dots + (X_{40} - \bar{X}_{c2})^2 \\ &= 156.035,512 \end{aligned}$$

Jadi, nilai untuk SSE cluster 2 sebesar 55.455,526 menandakan bahwa dari empat kabupaten/kota memiliki kemiripan dari variabel-variabel indikator kemiskinan.

Berdasarkan hasil analisis diketahui bahwa cluster 2 memiliki kemiripan karakteristik dari variabel-variabel indikator kemiskinan dimana jarak terdekat sebesar 55.455,526. Sedangkan cluster 1 memiliki jarak terjauh dengan jarak sebesar 156.035,512.

Nilai SSE objek pada cluster 1 dan cluster 2 (SSE_{c1c2}) sebagai berikut:

$$\begin{aligned} SSE_{c1c2} &= \sum_{i=1}^{n_{c1c2}} (X_i - \bar{X}_{c1c2})(X_i - \bar{X}_{c1c2}) \\ &= (X_1 - \bar{X}_{c1c2})^2 + \dots + (X_{40} - \bar{X}_{c1c2})^2 \\ &= 156.035,512 + 55.455,526 \\ &= 211.491,038 \end{aligned}$$

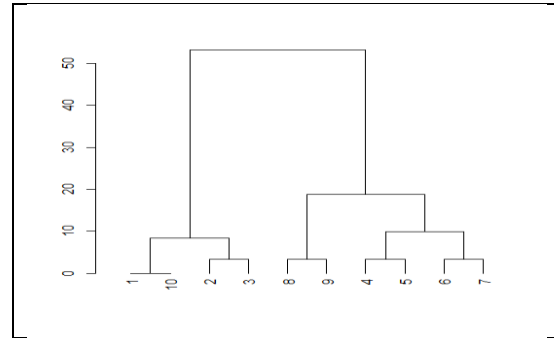
Sehingga terbentuk matriks jarak sebagai berikut:

Tabel 4 Matriks Jarak Dua Cluster

d	(4,5,6,7,8,9)	(1,2,3,10)
(4,5,6,7,8,9)	0	211.491,038
(1,2,3,10)	211.491,038	0

Jadi, nilai untuk SSE cluster 1 dan cluster 2 sebesar 211.491,038 menandakan bahwa dari sepuluh kabupaten/kota memiliki kemiripan dari variabel-variabel indikator kemiskinan.

Untuk hasil Dendrogram dapat dilihat pada Gambar 2 sebagai berikut:



Gambar 2. Hasil Dendrogram untuk Metode Ward

Dapat dilihat pada Gambar 2 diperoleh hasil dendrogram untuk metode Ward pada data indikator kemiskinan di Kalimantan Timur tahun 2018 yaitu 2 cluster. Dengan cluster pertama adalah Kabupaten Kutai Timur, Kabupaten Berau, Kabupaten Penajam Paser Utara, Kabupaten Mahakam Ulu, Kota Balikpapan dan Kota Samarinda. Cluster kedua adalah Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Kutai Kartanegara, dan Kota Bontang.

7. Evaluasi Hasil Pengklasifikasian

Untuk melihat kinerja kedua metode tersebut digunakan kriteria nilai simpangan baku, yaitu simpangan baku dalam cluster (s_w) dan simpangan baku antar cluster (s_b).

1. Perhitungan simpangan baku dalam cluster (s_w) dan simpangan baku antar cluster (s_b) metode Average Linkage

Contoh perhitungan nilai rata-rata cluster 1

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^{10} x_i}{n_1} \\ \bar{x}_{(2)1} &= \frac{9,03+9,15+\dots+4,67}{7} = 8,019 \end{aligned}$$

Perhitungan simpangan untuk setiap cluster (s_k)

$$s_k = \sqrt{\frac{\sum_{i=1}^n (x_{i(k)} - \bar{x}_{(k)})^2}{n_1 - 1}} = \sqrt{\frac{37,184}{6}} = 2,489$$

Berdasarkan hasil perhitungan dapat dilihat nilai rata-rata cluster pertama adalah 8,019 dengan nilai simpangan baku di dalam cluster sebesar 2,489. Berdasarkan hasil yang telah diperoleh maka perhitungan nilai s_w menggunakan Persamaan (7) adalah sebagai berikut:

$$s_w = \frac{1}{c} \sum_{k=1}^p = \frac{1}{2} (2,489) = 1,245$$

Berdasarkan nilai rata-rata cluster 1 yang telah didapat maka, perhitungan simpangan baku antar cluster menggunakan Persamaan (2.19) sebagai berikut:

$$s_b = \sqrt{\frac{1}{c-1} \sum_{k=1}^p (\bar{x}_{(k)} - \bar{x})^2}$$

$$s_b = \sqrt{\frac{1}{1}(16,072)} = 4,009$$

Berdasarkan hasil perhitungan s_b dapat dilihat bahwa nilai simpangan baku antar cluster untuk variabel x_1 cluster 1 adalah sebesar 4,009. Berdasarkan hasil perhitungan nilai s_w dan s_b maka nilai rasio rata-rata simpangan baku dalam cluster dan simpangan baku antar cluster untuk cluster 1 adalah sebagai berikut:

$$\text{Nilai Rasio} = \frac{s_w}{s_b} = \frac{1,245}{4,009} = 0,311$$

Perhitungan nilai rata-rata cluster 2

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{n_2}$$

$$\bar{x}_{(2)1} = \frac{7,41+2,64+\dots+4,59}{3} = 4,880$$

Perhitungan simpangan untuk setiap cluster (s_k)

$$s_k = \sqrt{\frac{\sum_{i=1}^n (x_{i(k)} - \bar{x}_{(k)})^2}{n_2 - 1}} = \sqrt{\frac{6.507,642}{2}} = 57,042$$

Berdasarkan hasil perhitungan dapat dilihat nilai rata-rata cluster kedua adalah 4,880 dengan nilai simpangan baku di dalam cluster sebesar 57,042. Berdasarkan hasil yang telah diperoleh maka perhitungan nilai s_w menggunakan Persamaan (7) adalah sebagai berikut:

$$s_w = \frac{1}{c} \sum_{k=1}^p = \frac{1}{2}(57,042) = 28,251$$

Berdasarkan nilai rata-rata cluster 2 yang telah didapat maka, perhitungan simpangan baku antar cluster menggunakan Persamaan (10) sebagai berikut:

$$s_b = \sqrt{\frac{1}{2-1}(4,880 - 2,440)^2} = 1,562$$

Berdasarkan hasil perhitungan s_b dapat dilihat bahwa nilai simpangan baku antar cluster untuk variabel x_1 cluster 2 adalah sebesar 1,562. Berdasarkan hasil perhitungan nilai s_w dan s_b maka nilai rasio rata-rata simpangan baku dalam cluster dan simpangan baku antar cluster adalah sebagai berikut:

$$\text{Nilai Rasio} = \frac{s_w}{s_b} = \frac{1,245}{4,009} = 0,311$$

Tabel 5 Perbandingan Nilai Rata-rata Rasio Simpangan Baku Kedua Metode

No	Metode	Rata-rata Rasio Simpangan Baku (S_w/S_b)
1	Average Linkage	9,198
2	Ward	2,681

Berdasarkan Tabel 5, nilai rata-rata rasio simpangan baku (R) sebesar 2,681 menunjukkan bahwa metode Ward memiliki nilai rata-rata rasio simpangan baku yang lebih kecil di bandingkan

dengan nilai rata-rata rasio simpangan baku pada metode *average linkage*. Oleh karena itu, kinerja yang paling baik dari kedua metode yang diteliti adalah metode Ward.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, dapat disimpulkan bahwa hasil perbandingan antara metode Ward dan metode *average linkage* berdasarkan rata-rata nilai simpangan baku pada pengelompokan data kemiskinan disetiap Kabupaten/Kota Provinsi Kalimantan Timur tahun 2018. Diperoleh bahwa rata-rata nilai rasio simpangan baku pada metode Ward memiliki nilai rasio simpangan baku dalam kelompok (s_w) dan nilai rasio simpangan baku antar kelompok (s_b) lebih kecil dari metode *average linkage* yaitu sebesar 2,681.

Daftar Pustaka

- Clave, J. T., P. G. Benson., & T.Sincich. (2005). *Statistik untuk Bisnis dan Ekonomi Jilid 2*. Jakarta: Erlangga.
- Hardius, U., & Nurdin, S. (2013). *Aplikasi Teknik Multivariate Riset Pemasaran*. Jakarta: Raja Grafindo Persada.
- Iriawan, Nur & Septian Puji Astuti. (2006). *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*. Yogyakarta: Andi Offset.
- Johnson, R. A., & D. W. Wichern. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Kartasamita, Ginandjar. (1996). *Pembangunan Untuk Rakyat; Memadukan Pertumbuhan dan Pemerataan*. Jakarta. CIDES.
- Kuncoro, M. 2003. *Metode Riset untuk Bisnis & Ekonomi*. Jakarta: Erlangga.
- Laeli, Sofya. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Lin*. Yogyakarta.
- Nafisah, Qonitatin. (2017). *Analisis Cluster Average Linkage Berdasarkan Faktor-faktor Kemiskinan di Provinsi Jawa Timur*, Lamongan.
- Nuritha, I., S. Bukhori., & W. E. Y. Retnani. (2013). Identifikasi Pengaruh Lokasi Usaha Terhadap Tingkat Keberhasilan Usaha Minimarket Waralaba di Kabupaten Jember dengan Sistem Infomasi Geografis. *Jurnal Sainstek UNEJ*, 1(1), 825-835.
- Prasetyo, E. (2012). *Data Mining dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi
- Santoso, S. (2014). *Statistik Multivariat Edisi Revisi Konsep dan Aplikasi dengan SPSS*. Jakarta: Gramedia.

Sartono, B., Farid M. Affendi, Utami Dyah Syafitri, I. Made Sumertajaya. & Yenni Angraeni. (2003). *Analisis Peubah Ganda*. Bogor: IPB.

Sarwono, Jonathan. (2012). *Statistik ultivariat Aplikasi untk Riset Skripsi*. Yogyakarta: Andi Offset.

Simamora, B. (2005). *Analisis Multivariat Pemasaran EdisiPertama*. Jakarta: PT. Gramedia Pustaka Utama.

