

PENGELOMPOKAN KABUPATEN/KOTA BERDASARKAN FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN DI JAWA TENGAH MENGGUNAKAN METODE WARD DAN AVERAGE LINKAGE

REGENCIES/CITIES CLASSIFICATION BASED ON FACTORS RELATED TO POVERTY IN CENTRAL JAVA USING WARD METHOD AND AVERAGE LINKAGE

Oleh: Meilia Wulan Puspitasari¹⁾, Mathilda Susanti²⁾
Program Studi Matematika, Jurusan Pendidikan Matematika FMIPA UNY
meilia_wulan@yahoo.com¹⁾

Abstrak

Analisis *cluster* merupakan analisis multivariat yang bertujuan mengelompokkan objek-objek pengamatan berdasarkan kesamaan karakteristik. Penelitian ini bertujuan untuk mengelompokkan 35 kabupaten/kota di Jawa Tengah, sehingga pemerintah dapat menjalankan program yang sesuai pada tiap *cluster*. Faktor yang mempengaruhi kemiskinan yaitu Produk Domestik Regional Bruto (PDRB), pengeluaran perkapita, upah minimum, gizi buruk balita, rata-rata lama sekolah, rasio ketergantungan, dan pengangguran tahun 2014. Metode yang digunakan adalah *Ward* dan *Average Linkage*. *Ward* menghasilkan 4 *cluster*: *cluster* 1 sebanyak 8 kabupaten, *cluster* 2 sebanyak 21 kabupaten, *cluster* 3 sebanyak 5 kota, *cluster* 4 sebanyak 1 kota. *Average Linkage* menghasilkan 4 *cluster*: *cluster* 1 sebanyak 4 kabupaten, *cluster* 2 sebanyak 25 kabupaten, *cluster* 3 sebanyak 5 kota, *cluster* 4 sebanyak 1 kota. Variabel yang mempengaruhi tiap *cluster* untuk *Ward* dan *Average Linkage* sama: *cluster* 1 pengangguran, *cluster* 2 rasio ketergantungan, *cluster* 3 rata-rata lama sekolah, *cluster* 4 PDRB. Digunakan rasio simpangan baku dalam *cluster* (s_w) dan antar *cluster* (s_b) untuk mengetahui metode terbaik. Metode yang memiliki rasio paling kecil merupakan metode terbaik. Rasio *Ward* sebesar 0,203165, sedangkan *Average Linkage* sebesar 0,218543. *Ward* adalah metode terbaik dalam mengelompokkan 35 kabupaten/kota di Jawa Tengah berdasarkan faktor yang mempengaruhi kemiskinan tahun 2014.

Kata kunci: *Ward*, *Average Linkage*, simpangan baku dalam dan antar *cluster*.

Abstract

Cluster analysis is a multivariate analysis aiming to classify observed objects based on possessed characteristics. This study aims to classify 35 regencies/cities in Central Java, thus government will be able to execute programs according to each cluster. Factors related to poverty were Gross Regional Domestic Product (GRDP), expenditure per capita, minimum wage, toddler malnutrition, average school duration, dependence ratio and unemployment in 2014. Method was used Ward and Average Linkage. Ward resulted 4 clusters: cluster 1 consisted of 8 regencies, cluster 2 consisted of 21 regencies, cluster 3 consisted of 5 cities, cluster 4 consisted of 1 city. Average Linkage resulted 4 clusters: cluster 1 consisted of 4 regencies, cluster 2 consisted of 25 regencies, cluster 3 consisted of 5 cities, cluster 4 consisted of 1 city. Variables that affect each cluster for Ward and Average Linkage are same: cluster 1 unemployment, cluster 2 dependence ratio, cluster 3 average school duration, cluster 4 GRDP. Used ratio standard deviation values of within cluster (s_w) and between cluster (s_b) for knowing the best method. Method which possess the least ratio was then considered the best. Ratio of Ward was 0.203165, while the ratio for Average Linkage was 0.218543. Ward method was the best method in classifying 35 regencies/cities in Central Java based on factors related to poverty in 2014.

Keywords: *Ward*, *Average Linkage*, within and between cluster standard deviation.

PENDAHULUAN

Analisis *cluster* bertujuan untuk mengelompokkan objek-objek pengamatan berdasarkan kesamaan karakteristik yang dimiliki. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Dua objek yang

memiliki jarak paling dekat akan bergabung menjadi satu *cluster*. Kedekatan jarak yang dimiliki menunjukkan kedua objek tersebut memiliki kesamaan karakteristik. Ciri-ciri pengelompokan yang baik yaitu (Yulianto dan Hidayatullah, 2014: 57):

- a) mempunyai kesamaan (homogenitas) yang tinggi antar anggota dalam satu *cluster* (*within cluster*), dan
- b) mempunyai ketaksamaan (heterogenitas) yang tinggi antar *cluster* (*between cluster*).

Secara garis besar analisis *cluster* terdiri dari dua metode, yaitu metode hirarki dan non hirarki. Metode hirarki terbagi ke dalam dua metode yaitu metode *agglomerative* dan *divisive*. Metode *agglomerative* adalah metode yang menganggap setiap objek sebagai sebuah *cluster* tersendiri. Dua *cluster* yang mempunyai jarak paling dekat akan bergabung menjadi satu *cluster* sampai seluruh *cluster* bergabung menjadi sebuah *cluster* besar (Johnson dan Winchern, 2007: 680-681). Metode *agglomerative* terdiri dari beberapa metode, dua diantaranya yaitu: *ward* dan *average linkage*.

Pada analisis *cluster* terdapat asumsi-asumsi yang harus dipenuhi, yaitu sampel yang diambil harus mewakili populasi dan multikolinearitas. Multikolinearitas adalah adanya hubungan linear atau korelasi yang tinggi antar variabel. Sebaiknya dalam analisis *cluster* tidak terjadi multikolinearitas. Setiap variabel diberi bobot yang sama dalam perhitungan jarak (Gudono, 2014: 284). Jika terjadi multikolinearitas, maka diatasi menggunakan *Principal Component Analysis (PCA)*.

Principal Component Analysis (PCA) bertujuan untuk menyederhanakan variabel-variabel yang diamati dengan mereduksi dimensinya tanpa kehilangan banyak informasi dari variabel asal/aslinya. Hal ini dilakukan dengan cara menghilangkan korelasi antar variabel bebas melalui transformasi variabel bebas asal ke variabel bebas baru yang tidak

berkorelasi sama sekali atau sering disebut dengan *principal component* (Johnson dan Wichern, 2007: 430).

Dalam kehidupan sehari-hari terdapat permasalahan yang dapat diatasi menggunakan analisis *cluster*, salah satunya masalah kemiskinan. Berdasarkan Berita Resmi Statistik 4 Januari 2016, sebagian besar penduduk miskin Indonesia berada di Pulau Jawa yaitu sebesar 15,31 juta jiwa. Dari tabel 1, terlihat bahwa Provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat mempunyai jumlah penduduk miskin tinggi yaitu lebih dari empat juta jiwa. Dari keenam provinsi yang ada di Pulau Jawa, Provinsi Jawa Tengah mempunyai persentase penduduk miskin paling tinggi yaitu sebesar 13,32% atau sebesar 4505,78 ribu jiwa.

Tabel 1. Jumlah dan Persentase Penduduk Miskin di Pulau Jawa Periode September 2015

Provinsi	Jumlah Penduduk Miskin (ribu jiwa)	Persentase Penduduk Miskin (%)
DKI Jakarta	368,67	3,61
Jawa Barat	4485,66	9,57
Jawa Tengah	4505,78	13,32
DI Yogyakarta	485,56	13,16
Jawa Timur	4775,97	12,28
Banten	690,66	5,75
Total	15312,3	10,52

Faktor-faktor yang mempengaruhi kemiskinan di provinsi Jawa Tengah, antara lain (Prastyo, 2010): pertumbuhan ekonomi, upah minimum, tingkat pendidikan dan pengangguran. Berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas) BPS, penduduk yang tidak bekerja di Jawa Tengah sebesar 34,73%. Angka tersebut sudah termasuk pengangguran dan bukan angkatan kerja. Rendahnya tingkat pendidikan masyarakat juga sangat mempengaruhi kemiskinan. Semakin rendah tingkat pendidikan

seseorang, maka peluang untuk bersaing dalam dunia kerja semakin kecil.

Penerapan analisis *cluster* dapat digunakan untuk mengelompokkan 35 kabupaten/kota Provinsi Jawa Tengah berdasarkan faktor-faktor yang mempengaruhi kemiskinan tahun 2014. Pengelompokan dilakukan untuk mengetahui kesamaan karakteristik antara 35 kabupaten/kota sehingga pemerintah dapat menjalankan program-program yang sesuai pada tiap *cluster*. Pengelompokan dilakukan menggunakan metode *Ward* dan *Average Linkage*. Pemilihan metode *Ward* karena metode ini menghasilkan *cluster* yang tidak terlalu besar dan merupakan metode hirarki terbaik. Sebagai pembanding digunakan metode *Average Linkage* karena metode ini dianggap mempunyai ketelitian yang lebih baik daripada metode hirarki yang lain. Ukuran kemiripan yang digunakan pada kedua metode tersebut adalah ukuran jarak Euclid kuadrat (Supranto, 2010: 154). Apabila terjadi multikolinearitas antar variable diatasi menggunakan *Principal Component Analysis (PCA)*.

HASIL PENELITIAN DAN PEMBAHASAN

Analisis *cluster* merupakan analisis yang bertujuan untuk mengelompokkan objek-objek pengamatan berdasarkan karakteristik yang dimiliki. Asumsi-asumsi dalam analisis *cluster* yaitu:

1. Sampel yang diambil harus mewakili populasi

Untuk mengetahui apakah sampel yang digunakan telah mewakili populasi, dapat dilihat dari nilai *Kaiser Meyer Olkin (KMO)*. *KMO* adalah indeks perbandingan nilai

koefisien korelasi terhadap korelasi parsial (Yamin dan Kurniawan, 2014: 181).

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2}$$

$$a_{ij} = \frac{-r_{ij}}{\sqrt{r_{ij} \cdot r_{ij}}}$$

Jika nilai *KMO* < 0,5, maka sampel tidak mewakili populasi, sedangkan jika nilai *KMO* > 0,5, maka sampel mewakili populasi sehingga layak untuk dilakukan analisis *cluster* (Yamin, Rachmach, dan Kurniawan, 2011: 122).

2. Multikolinearitas

Multikolinearitas adalah adanya hubungan linear atau korelasi yang tinggi antar variabel. Untuk mengetahui ada tidaknya multikolinearitas, dapat dilihat dari nilai-nilai korelasi pada matriks korelasi. Koefisien korelasi dapat diperoleh dari persamaan (Johnson dan Wichern, 2007: 8):

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} \\ = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Dua variabel atau lebih dikatakan multikolinearitas apabila $r_{xy} > 0,70$ (Yamin dan Kurniawan, 2014: 70). Apabila terjadi multikolinearitas diatasi menggunakan *Principal Component Analysis (PCA)*.

Principal Component Analysis (PCA) bertujuan untuk menyederhanakan variabel yang diamati dengan cara mereduksi dimensinya. Prinsip utama dari *PCA* adalah terdapatnya korelasi antar variabel sehingga diduga bahwa variabel-variabel tersebut dapat direduksi. Hal ini dilakukan dengan cara menghilangkan korelasi antar variabel bebas melalui transformasi variabel

bebas asal ke variabel bebas baru yang tidak berkorelasi sama sekali atau sering disebut dengan *principal component* (Johnson dan Wichern, 2007: 430).

Principal component (PC) merupakan suatu kombinasi linear dari variabel-variabel asal. Penelitian ini menggunakan matriks korelasi sebagai dasar pembentukan *PC* karena variabel yang digunakan memiliki satuan berbeda. Sebelum dilakukan *PCA*, terlebih dahulu dilakukan beberapa pengujian sebagai berikut:

1. Uji *Kaiser Meyer Olkin (KMO)*

Uji *KMO* telah dilakukan pada asumsi analisis *cluster*.

2. Uji Bartlett

Uji Bartlett digunakan untuk mengetahui apakah matriks korelasi hubungan antara variabel adalah matriks identitas. Hal ini digunakan untuk menguji kecukupan hubungan antara variabel, dimana matriks identitas mempunyai diagonal matriks bernilai 1 dan yang lainnya adalah 0.

Hipotesis:

H_0 : matriks korelasi = matriks identitas

H_1 : matriks korelasi \neq matriks identitas

Statistik uji :

$$\chi^2_{obs} = -\ln|R| \left[(n - 1) - \frac{(2p + 5)}{6} \right]$$

Kriteria keputusan:

Tolak H_0 jika $\chi^2_{obs} > \chi^2_{\alpha, p(p-1)/2}$ atau $p - value \leq \alpha$

Principal component hasil dari *PCA* akan digunakan sebagai variabel bebas baru untuk dianalisis *cluster* menggunakan metode *Ward* dan *Average Linkage*. Sebelum penyusunan *cluster*, terlebih dahulu dilakukan pemilihan ukuran jarak. Ukuran jarak yang digunakan adalah jarak Euclid

kuadrat. Adapun persamaan Euclid kuadrat adalah sebagai berikut:

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

Metode *Ward* bertujuan untuk memperoleh *cluster* yang memiliki varian dalam *cluster (within cluster)* sekecil mungkin. Ukuran kesamaan yang digunakan adalah jarak Euclid kuadrat (Supranto, 2010: 154). Metode *Ward* mengelompokkan objek didasarkan pada kenaikan *sum square error (SSE)*. Pada tiap tahap, dua *cluster* yang memiliki kenaikan *SSE* paling kecil digabungkan (Simamora, 2005: 218). *Sum Square Error (SSE)* dapat dihitung menggunakan persamaan (Gudono, 2014: 294):

$$SSE = \sum_{i=1}^k (x_i - \bar{x})^2$$

Metode *average linkage* (pautan rata-rata) adalah metode pengelompokan yang jarak antara dua *cluster* didefinisikan sebagai rata-rata jarak seluruh pasangan yang berada dalam satu *cluster* dengan *cluster* yang lain. Untuk menghitung rata-rata jarak pada metode *average linkage* digunakan persamaan (Johnson dan Wichern, 2007: 690):

$$d_{(IJ)K} = \frac{\sum_a \sum_b d_{ab}}{N_I N_K}$$

Untuk menentukan metode yang menghasilkan *cluster* terbaik, digunakan nilai-nilai simpangan baku dalam *cluster (within cluster)* dan antar *cluster (between cluster)*. Persamaan-persamaan simpangan baku dalam dan antar *cluster* sebagai berikut:

1. Simpangan baku dalam *cluster (within cluster)*

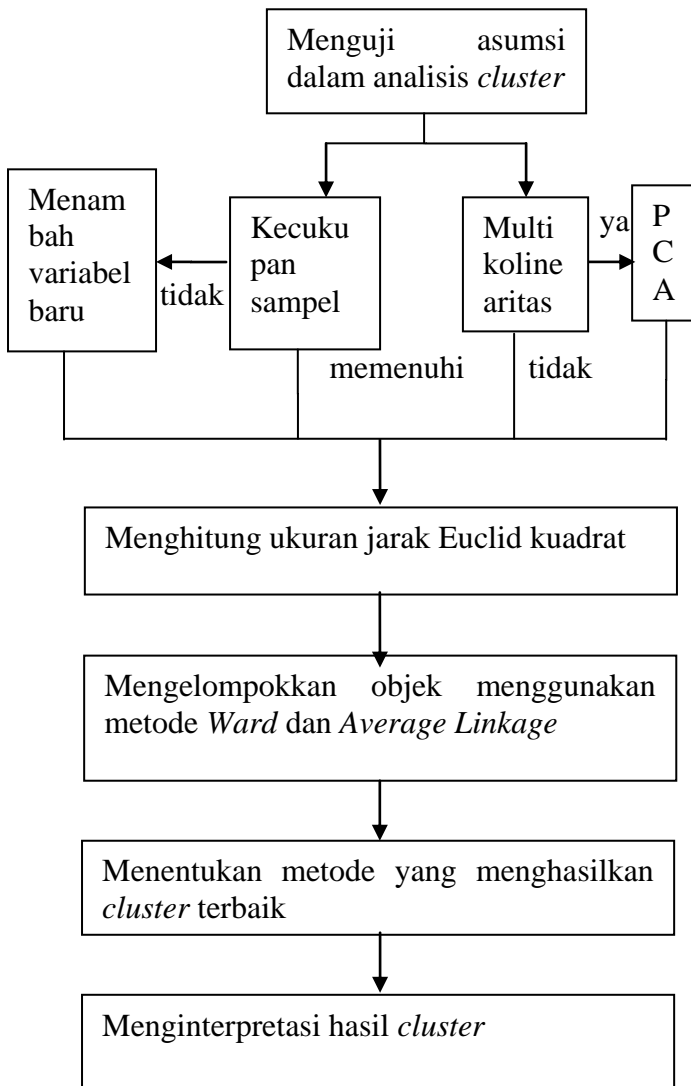
$$s_w = K^{-1} \sum_{k=1}^K s_k$$

2. Simpangan baku antar cluster (between cluster)

$$s_b = \left[(K - 1)^{-1} \sum_{k=1}^K (\bar{x}_k - \bar{x})^2 \right]^{1/2}$$

Metode yang mempunyai rasio terkecil merupakan metode terbaik (Bunkers dkk, 1996: 136). Semakin kecil nilai s_w dan semakin besar nilai s_b , maka metode yang digunakan mempunyai kinerja yang baik, artinya homogenitasnya tinggi.

Dari uraian diatas, langkah-langkah yang dilakukan dalam analisis cluster disajikan dalam bagan sebagai berikut:



Gambar 1. Langkah-langkah Analisis Cluster

Data yang digunakan merupakan data yang diperoleh dari publikasi BPS berupa buku yang berjudul “Jawa Tengah Dalam Angka 2015”, “Tinjauan PDRB Kabupaten/Kota Se Jawa Tengah 2014”, dan “Data dan Informasi Kemiskinan Kabupaten/Kota Tahun 2014”. Adapun variabel yang digunakan antara lain: Produk Domestik Regional Bruto (PDRB) (X_1), Pengeluaran perkapita (X_2), Upah minimum (X_3), Gizi buruk balita (X_4), Rata-rata lama sekolah (X_5), Rasio Ketergantungan (X_6), dan Pengangguran (X_7).

Hasil uji asumsi analisis cluster adalah sebagai berikut:

1. Sampel yang diambil harus mewakili populasi
Berdasarkan perhitungan menggunakan SPSS 20, diperoleh nilai KMO sebesar 0,713. Nilai tersebut lebih dari 0,5, maka sampel telah mewakili populasi sehingga layak untuk dilakukan analisis cluster.

2. Multikolinearitas
Berdasarkan perhitungan menggunakan SPSS 20, nilai-nilai korelasi antar variabel sebagai berikut:

Tabel 2. Korelasi Antar Variabel

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1,00						
X_2	0,250	1,00					
X_3	0,435	0,466	1,00				
X_4	0,502	-0,403	-0,036	1,00			
X_5	0,247	0,777	0,476	-0,425	1,00		
X_6	-0,269	-0,646	-0,586	0,441	-0,683	1,00	
X_7	0,516	-0,234	0,147	0,862	-0,281	0,207	1,00

Tabel 2 menunjukkan bahwa variabel X_2 dan X_5 , X_4 dan X_7 mengalami multikolinearitas karena nilai korelasi lebih dari 0,7. Karena terjadi multikolinearitas antar

variabel, maka perlu dilakukan penanganan multikolinearitas guna memenuhi asumsi dalam analisis *cluster*. Penanganan masalah multikolinearitas menggunakan *Principal Component Analysis (PCA)*.

Sebelum melakukan *PCA*, terlebih dahulu melakukan uji *KMO* dan Bartlett. Uji *KMO* telah dilakukan untuk memenuhi asumsi dalam analisis *cluster* dan diperoleh nilai *KMO* sebesar 0,713. Hal ini menunjukkan bahwa data layak untuk dilakukan *PCA*. Uji Bartlett digunakan untuk mengetahui apakah matriks korelasi hubungan antara variabel adalah matriks identitas, sehingga diketahui korelasi antar variabel. Berdasarkan perhitungan menggunakan SPSS 20 diperoleh hasil nilai $p - value = 0,000 < \alpha = 0,05$. Hal ini menunjukkan bahwa terdapat korelasi antar variabel sehingga perlu dilakukan *PCA*.

Pembentukan *Principal Component (PC)* untuk ketujuh variabel yang digunakan berdasarkan matriks korelasi. Pemilihan penggunaan matriks korelasi dikarenakan ketujuh variabel mempunyai satuan berbeda, maka harus dilakukan standarisasi data terlebih dahulu. Hasil standarisasi data sebagai berikut:

$$Z = \begin{bmatrix} 2,800178 & -0,34771 & -0,47511 & \dots & 0,930991 \\ 0,35037 & -0,0462 & -0,62906 & \dots & 0,798469 \\ -0,42869 & -0,68838 & -0,42078 & \dots & -0,27904 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -0,67798 & 1,149711 & -0,23061 & \dots & -1,04064 \end{bmatrix}$$

Matriks korelasi (*R*) adalah perkalian antara matriks *Z'* dengan *Z*. Matriks *Z'* sebagai berikut:

$$Z' = \begin{bmatrix} 2,800178 & 0,35037 & -0,42869 & \dots & -0,67798 \\ -0,34771 & -0,0462 & -0,68838 & \dots & 1,149711 \\ -0,47511 & -0,62906 & -0,42078 & \dots & -0,23061 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0,930991 & 0,798469 & -0,27904 & \dots & -1,04064 \end{bmatrix}$$

Sehingga matriks korelasi (*R*) adalah sebagai berikut:

$$R = \begin{bmatrix} 1 & 0,249574 & 0,435267 & 0,501512 & 0,24728 & -0,26956 & 0,516328 \\ 0,249574 & 1 & 0,465551 & -0,40295 & 0,77744 & -0,6462 & -0,23416 \\ 0,435267 & 0,465551 & 1 & -0,03639 & 0,476429 & -0,58586 & 0,14671 \\ 0,501512 & -0,40295 & -0,03639 & 1 & -0,42504 & 0,441316 & 0,861983 \\ 0,247284 & 0,77744 & 0,476429 & -0,42504 & 1 & -0,68314 & -0,28088 \\ -0,26956 & -0,6462 & -0,58586 & 0,441316 & -0,68314 & 1 & 0,206817 \\ 0,516328 & -0,23416 & 0,14671 & 0,861983 & -0,28088 & 0,206817 & 1 \end{bmatrix}$$

Langkah selanjutnya yaitu mencari nilai eigen dari matriks korelasi (*R*). Dengan perhitungan menggunakan SPSS 20 diperoleh hasil sebagai berikut:

Tabel 3. Nilai Eigen, Total Varian, Total Kumulatif Varian dari Matriks Korelasi (*R*)

Komponen	Nilai Eigen	Total Varian (%)	Total Kumulatif Varian (%)
1	3,243	46,323	46,323
2	2,295	32,781	79,103
3	0,534	7,630	86,733
4	0,343	4,893	91,626
5	0,287	4,103	95,729
6	0,218	3,118	98,848
7	0,081	1,152	100,000

Kriteria minimal yang digunakan untuk menentukan *Principal Component (PC)* yaitu nilai eigen lebih dari 1 ($\lambda > 1$), maka komponen tersebut terpilih sebagai *PC* (Gudono, 2014: 210). Dari tabel 3, terbentuk dua *PC* yaitu *PC*₁ dan *PC*₂. *PC*₁ terdiri dari variabel pengeluaran perkapita (*X*₂), upah minimum (*X*₃), rata-rata lama sekolah (*X*₅), dan rasio ketergantungan (*X*₆), sedangkan *PC*₂ terdiri dari variabel PDRB (*X*₁), gizi buruk balita (*X*₄), dan pengangguran (*X*₇). Kedua *PC* telah menjelaskan varian dari ketujuh variabel sebesar 79,103%. Kedua *PC* inilah yang digunakan untuk mengelompokkan 35

kabupaten/kota di Jawa Tengah menggunakan metode *Ward* dan *Average Linkage*.

Sebelum melakukan pengelompokan kabupaten/kota di Jawa Tengah menggunakan metode *Ward* dan *Average Linkage*, terlebih dahulu dilakukan penghitungan jarak antar objek. Jarak inilah yang digunakan sebagai ukuran kesamaan antar objek. Ukuran kesamaan antar objek yang digunakan adalah jarak Euclid kuadrat. Semakin besar jarak menunjukkan ketaksamaan antar objek, sebaliknya semakin kecil jarak menunjukkan kesamaan antar objek.

Hasil pengelompokan 35 kabupaten/kota di Jawa Tengah menggunakan metode *Ward* sebagai berikut:

Tabel 4. Pengelompokan Kabupaten/Kota di Jawa Tengah menggunakan Metode *Ward*

Cluster	Kabupaten/Kota
1	Kab. Cilacap, Kab. Banyumas, Kab. Magelang, Kab. Grobogan, Kab. Pati, Kab. Pemalang, Kab. Tegal, Kab. Brebes
2	Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Blora, Kab. Rembang, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Temanggung, Kab. Kendal, Kab. Batang, Kab. Pekalongan
3	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan, Kota Tegal
4	Kota Semarang

Hasil pengelompokan 35 kabupaten/kota di Jawa Tengah menggunakan metode *Average Linkage* sebagai berikut:

Tabel 5. Pengelompokan Kabupaten/Kota di Jawa Tengah menggunakan Metode *Average Linkage*

Cluster	Kabupaten/Kota
1	Kab. Cilacap, Kab. Pemalang, Kab. Tegal, Kab. Brebes
2	Kab. Banyumas, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Temanggung, Kab. Kendal, Kab. Batang, Kab. Pekalongan
3	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan, Kota Tegal
4	Kota Semarang

Setelah terbentuk 4 *cluster* dari masing-masing metode, dilakukan penentuan metode yang menghasilkan hasil *cluster* terbaik. Pengecekan dilakukan menggunakan simpangan baku dalam *cluster* (*within cluster*) dan antar *cluster* (*between cluster*). Berikut nilai-nilai simpangan baku dalam dan antar *cluster*:

Tabel 6. Simpangan baku dalam dan antar *cluster* untuk Metode *Ward* dan *Average Linkage*

Simpangan Baku	Metode <i>Ward</i>	Metode <i>Average Linkage</i>
simpangan baku dalam <i>cluster</i> (s_w)	0,319176	0,338324
simpangan baku antar <i>cluster</i> (s_b)	1,571015	1,548091

Simpangan baku dalam *cluster* (s_w) metode *Ward* lebih kecil daripada metode *average linkage*, sedangkan simpangan baku antar *cluster* (s_b) metode *Ward* lebih besar daripada metode *average linkage*. Perbandingan (rasio) antara s_w dan s_b kedua metode sebagai berikut:

Tabel 7. Perbandingan (rasio) s_w dan s_b untuk

Metode Ward dan Average Linkage

Rasio	Metode Ward	Metode Average Linkage
$\frac{s_w}{s_b}$	0,203165	0,218543

Dari tabel 7 terlihat bahwa rasio metode Ward sebesar 0,203165 lebih kecil daripada rasio metode average linkage sebesar 0,218543. Metode yang mempunyai rasio terkecil merupakan metode dengan kinerja terbaik. Dengan kata lain, untuk mengelompokkan 35 kabupaten/kota Provinsi Jawa Tengah berdasarkan faktor-faktor yang mempengaruhi kemiskinan tahun 2014 lebih baik menggunakan metode Ward.

Untuk menginterpretasi dan memprofilkan hasil cluster digunakan nilai-nilai centroid tiap cluster. Centroid adalah rata-rata nilai objek yang terdapat dalam cluster pada tiap variabel. Berikut nilai-nilai centroid tiap cluster untuk masing-masing metode:

Tabel 8. Rata-rata variabel pada setiap cluster untuk Metode Ward

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1 (PDRB)	0,299439	-0,17125	-0,56249	4,01327
X_2 (PP)	-0,6581	-0,18554	1,443998	1,94109
X_3 (UP)	-0,42188	-0,08408	0,386978	3,20597
X_4 (GB)	1,189608	-0,15182	-1,55689	1,45573
X_5 (RLS)	-0,66338	-0,23398	1,58368	2,30212
X_6 (RK)	0,557673	0,198812	-1,28834	-2,19474
X_7 (PG)	1,226821	-0,30629	-1,15905	2,41286

Dari tabel 8 dapat diinterpretasikan masing-masing cluster sebagai berikut:

1) Cluster 1 mempunyai nilai centroid tinggi untuk variabel X_7 (pengangguran). Hal ini berarti bahwa pada cluster 1, penduduk yang tidak bekerja cukup banyak. Oleh karena itu, pemerintah dapat

memprioritaskan bantuan tambahan lapangan pekerjaan pada daerah di cluster 1.

- 2) Cluster 2 mempunyai nilai centroid tinggi untuk variabel X_6 (rasio ketergantungan). Hal ini berarti bahwa pada cluster 2, penduduk usia tidak produktif yang harus ditanggung penduduk usia produktif cukup banyak. Oleh karena itu, pemerintah dapat memberikan bantuan dengan membuat program-program usaha mandiri untuk penduduk usia tidak produktif pada daerah di cluster 2.
- 3) Cluster 3 mempunyai nilai centroid tinggi untuk variabel X_5 (rata-rata lama sekolah). Hal ini berarti bahwa pada cluster 3, tingkat pendidikan penduduknya cukup baik.
- 4) Cluster 4 mempunyai nilai centroid tinggi untuk variabel X_1 (PDRB). Hal ini berarti bahwa pada cluster 4, pendapatan daerahnya sangat tinggi. Cluster 4 hanya terdiri dari satu kota, yaitu Kota Semarang.

Tabel 9. Rata-rata variabel pada setiap cluster untuk Metode Average Linkage

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1 (PDRB)	0,615893	-0,14658	-0,56249	4,01327
X_2 (PP)	-0,89179	-0,22376	1,443998	1,94109
X_3 (UP)	-0,44116	-0,13505	0,386978	3,20597
X_4 (GB)	1,60621	-0,00385	-1,55689	1,45573
X_5 (RLS)	-0,94256	-0,25801	1,58368	2,30212
X_6 (RK)	0,913723	0,199262	-1,28834	-2,19474
X_7 (PG)	1,692533	-0,13551	-1,15905	2,41286

Dari tabel 9 dapat diinterpretasikan masing-masing cluster sebagai berikut:

1) Cluster 1 mempunyai nilai centroid tinggi untuk variabel X_7 (pengangguran). Hal ini berarti bahwa pada cluster 1, penduduk

yang tidak bekerja cukup banyak. Oleh karena itu, pemerintah dapat memprioritaskan bantuan tambahan lapangan pekerjaan pada daerah di *cluster* 1.

- 2) *Cluster* 2 mempunyai nilai *centroid* tinggi untuk variabel X_6 (rasio ketergantungan). Hal ini berarti bahwa pada *cluster* 2, penduduk usia tidak produktif yang harus ditanggung penduduk usia produktif cukup banyak. Oleh karena itu, pemerintah dapat memberikan bantuan dengan membuat program-program usaha mandiri untuk penduduk usia tidak produktif pada daerah di *cluster* 2.
- 3) *Cluster* 3 mempunyai nilai *centroid* tinggi untuk variabel X_5 (rata-rata lama sekolah). Hal ini berarti bahwa pada *cluster* 3, tingkat pendidikan penduduknya cukup baik.
- 4) *Cluster* 4 mempunyai nilai *centroid* tinggi untuk variabel X_1 (PDRB). Hal ini berarti bahwa pada *cluster* 4, pendapatan daerahnya sangat tinggi. *Cluster* 4 hanya terdiri dari satu kota, yaitu Kota Semarang.

SIMPULAN DAN SARAN

Simpulan

Metode *Ward* dalam mengelompokkan objek berdasarkan pada kenaikan *SSE* paling kecil. Hasil dari pengelompokan yaitu terbentuk 4 *cluster*: *cluster* 1 sebanyak 8 kabupaten, *cluster* 2 sebanyak 21 kabupaten, *cluster* 3 sebanyak 5 kota, *cluster* 4 sebanyak 1 kota.

Metode *Average Linkage* mengelompokkan objek dengan mendefinisikan jarak antara dua *cluster* sebagai rata-rata jarak

seluruh pasangan yang berada dalam satu *cluster* dengan *cluster* yang lain. Hasil dari pengelompokan yaitu terbentuk 4 *cluster*: *cluster* 1 sebanyak 4 kabupaten, *cluster* 2 sebanyak 25 kabupaten, *cluster* 3 sebanyak 5 kota, *cluster* 4 sebanyak 1 kota.

Variabel yang mempengaruhi tiap *cluster* untuk *Ward* dan *Average Linkage* sama, yaitu: *cluster* 1 pengangguran, *cluster* 2 rasio ketergantungan, *cluster* 3 rata-rata lama sekolah, *cluster* 4 PDRB. Metode *Ward* merupakan metode dengan kinerja terbaik karena mempunyai rasio simpangan baku dalam *cluster* (s_w) dan antar *cluster* (s_b) lebih kecil daripada metode *average linkage*.

Saran

Untuk penelitian selanjutnya, dapat membandingkan kinerja antara metode *Ward* atau *Average Linkage* dengan metode non hirarki *K-means*. Apabila terjadi masalah multikolinearitas dapat diatasi menggunakan ukuran jarak Mahalonobis, metode Ridge.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2015a). *Data dan Informasi Kemiskinan Kabupaten/Kota Tahun 2014*. Jakarta: Badan Pusat Statistik (CV. Nario Sari).
- . (2015b). *Jawa Tengah Dalam Angka Jawa Tengah in Figures 2015*. Semarang: BPS Provinsi Jawa Tengah.
- . (2015c). *Tinjauan PDRB Kabupaten/Kota Se Jawa Tengah 2014*. Semarang: BPS dan BAPPEDA Provinsi Jawa Tengah.
- . (2016). *Profil Kemiskinan di Indonesia September 2015 Persentase Penduduk Miskin September 2015*

- mencapai 11,13 persen. Berita Resmi Statistik BPS, No. 05/01/Th. XIX, 4 Januari 2016. Diakses dari https://www.bps.go.id/website/brs_ind/brsInd-20160104121812.pdf pada tanggal 26 Juni 2016, pukul 12.00 WIB.
- Bunkers, M. J. et al. (1996). *Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique*. *Journal of Climate*, Vol. 9, Hlm. 130-146.
- Gudono. (2014). *Analisis Data Multivariat Edisi Tiga*. Yogyakarta: BPFE Yogyakarta.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis Six Edition*. New Jersey: Pearson Prentice Hall.
- Prastyo, A. A. (2010). *Analisis Faktor-faktor yang Mempengaruhi Tingkat Kemiskinan (Studi Kasus 35 Kabupaten/Kota di Jawa Tengah Tahun 2003-2007)*. Skripsi. Undip.
- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Jakarta: Gramedia Pustaka Utama.
- Supranto, J. (2010). *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Yamin, S & Kurniawan, H. (2014). *SPSS Complete: Teknik Analisis Terlengkap dengan Software SPSS*. Jakarta: Salemba Infotek.
- Yamin, S., Rachmah, L. A., & Kurniawan, H. (2011). *Regresi dan Korelasi dalam Genggaman Anda: Aplikasi dengan Software SPSS, Eviews, MINITAB, dan STATGRAPHICS*. Jakarta: Salemba Empat.
- Yulianto, S., & Hidayatullah, K. H. (2014). *Analisis Kluster untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat*. *Statistika*, Vol. 2, No. 1, Hlm. 56-63.