

ANALISIS GEROMBOL MENGUNAKAN METODE *TWO STEP CLUSTER* (Studi kasus : data Potensi Desa Sensus Ekonomi 2003 wilayah Jawa Barat)

I Made Sumertajaya dan Erfiani
Departemen Statistika FMIPA IPB

Windy D.Y. Putri
Business Planning & Analytics - Consumer Mass Market
PT. Bank Danamon Indonesia, Tbk

Abstrak

Penggerombolan adalah proses mengelompokkan objek ke dalam kelompok-kelompok yang memiliki kemiripan. Beberapa masalah yang sering dijumpai dalam analisis gerombol yaitu skala pengukuran peubah tidak sama dan jumlah objek besar serta jumlah gerombol tidak diketahui. Salah satu pendekatan untuk menangani masalah ini tanpa mentransformasi peubah-peubah tersebut adalah dengan menggunakan metode Two Step Cluster. Penelitian ini bertujuan sebagai penerapan metode Two Step Cluster dengan menggerombolkan desa/kelurahan yang berada di Jawa Barat.

Hasil penggerombolan dengan metode TwoStep Cluster, gerombol awal yang dihasilkan pada tahap pertama adalah sebanyak delapan gerombol, sedangkan gerombol optimal yang dihasilkan pada tahap dua adalah sebanyak tiga gerombol. Gerombol satu tidak dapat dikatakan sebagai suatu gerombol, karena anggota-anggota didalamnya merupakan objek-objek yang memencil ekstrim dan tidak dapat dimasukkan ke dalam gerombol lainnya. Desa/kelurahan yang termasuk gerombol dua memiliki karakteristik pedesaan. Desa/kelurahan tersebut memiliki lahan terluas, jumlah rumah tangga pertanian terbanyak, namun belum berkembang dalam bidang industri serta komunikasi dan informasi. Sehingga untuk meningkatkan potensi desa pada gerombol ini, yang harus diperhatikan adalah peubah-peubah yang tingkat perkembangannya masih rendah. Gerombol tiga memiliki karakteristik desa yang berstatus perkotaan. Desa/kelurahan pada gerombol ini memiliki jarak terdekat ke pusat kota, cukup maju dalam bidang industri, komunikasi dan informasi, namun memiliki angka pengangguran tertinggi.

Kata Kunci : Analisis gerombol, two step cluster

PENDAHULUAN

Latar Belakang

Penggerombolan adalah suatu proses mengelompokkan objek ke dalam kelompok-kelompok yang memiliki kemiripan. Saat ini, metode penggerombolan yang digunakan secara luas adalah metode penggerombolan berhirarki dan metode penggerombolan tak berhirarki. Kedua metode ini berbasis pada data yang berskala interval atau rasio.

Beberapa analisis yang sering dijumpai dalam analisis gerombol yaitu skala pengukuran peubah tidak sama dan jumlah objek besar serta jumlah gerombol tidak diketahui. Salah satu pendekatan untuk menangani masalah ini adalah dengan memilih tipe skala pengukuran tertentu, dan mentransformasikan peubah-peubahnya sehingga memiliki tipe skala pengukuran yang sama (Anderberg 1973).

Metode *Two Step cluster* merupakan suatu metode penggerombolan yang dapat mengatasi masalah skala pengukuran, khususnya untuk data berukuran besar dengan peubah yang memiliki tipe data kategorik dan kontinu (Bacher. 2004), serta mengetahui gerombol optimal yang terbentuk. Gerombol optimal memiliki

jarak antar gerombol yang paling jauh, dan jarak antar obyek yang paling dekat.

Tujuan

Tujuan penelitian ini adalah penerapan metode *Two Step cluster* untuk menggerombolkan desa/kelurahan di wilayah Jawa Barat.

TINJAUAN PUSTAKA

Potensi Desa

Potensi Desa (Podes) adalah kemampuan yang memiliki kemungkinan untuk dikembangkan dalam wilayah otonomi desa. Data podes merupakan satu-satunya data yang berurusan dengan wilayah atau tata ruang dengan basis desa atau kelurahan (BPS, 2003).

Salah satu tujuan pengumpulan data Podes ini adalah tersedianya data potensi atau keadaan pembangunan di desa dan perkembangannya meliputi keadaan sosial, ekonomi, sarana dan prasarana, serta potensi yang ada di desa/kelurahan.

Analisis Gerombol

Analisis gerombol merupakan suatu metode peubah ganda untuk mengelompokkan n objek ke dalam m gerombol ($m \leq n$) berdasarkan karakter-karakternya (Johnsons & Winchern, 2002).

Tujuan dari penggerombolan ini adalah menemukan gerombol alamiah dari sekumpulan unit pengamatan. Ada dua metode dalam analisis gerombol satu tahapan, yaitu :

1. Metode berhirarki.

Metode penggerombolan berhirarki digunakan apabila banyak gerombol yang akan dibentuk belum diketahui sebelumnya. Metode ini dapat dibedakan atas dua metode yaitu metode penggabungan (*agglomerative*) dan metode pembagian (*divisive*). Dalam metode berhirarki terdapat beberapa ukuran jarak antar gerombol, antara lain metode pautan tunggal (*single linkage*), pautan lengkap (*complete linkage*), pautan rata-rata (*average linkage*), metode *ward*, dan metode *centroid*. Fungsi jarak yang sering digunakan diantaranya jarak Euclidean dan jarak Mahalanobis.

2. Metode tak berhirarki

Metode ini digunakan apabila banyak gerombol yang akan dibentuk sudah diketahui terlebih dahulu. Salah satu contohnya adalah metode *K-means*. Pada metode ini harus ditentukan terlebih dahulu besarnya k , yaitu banyaknya gerombol. Pemilihan k dapat ditentukan secara subjektif berdasarkan latar belakang bidang masing-masing. Jarak yang biasanya digunakan adalah jarak Euclidean.

Metode penggerombolan dapat dilakukan hanya dengan satu tahapan, tetapi untuk data yang berukuran besar menggunakan penggerombolan dengan dua tahapan atau lebih. Penggerombolan secara bertahap dilakukan untuk mempermudah dalam menganalisa data berukuran besar.

Metode penggerombolan yang dilakukan dengan dua tahapan diantaranya adalah :

1. Analisis gerombol Hibrid (*Hybrid Clustering*)

Metode Hibrid merupakan penggabungan dari metode gerombol berhirarki dengan tak berhirarki. Tahap pertama yaitu menggerombolkan objek-objek ke dalam k gerombol menggunakan metode *K-means*. Kedua, menggerombolkan gerombol yang diperoleh pada tahapan pertama dengan menggunakan metode pautan tunggal (*single-linkage*) sebagai ukuran jarak antar gerombol, dan jarak Euclidean sebagai ukuran jarak antar obyeknya. Metode ini dapat digunakan pada data dengan jumlah amatan yang besar dan jumlah gerombol yang diinginkan tidak diketahui. Pada metode ini, peubah yang digunakan bertipe numerik.

2. BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*)

BIRCH merupakan salah satu penerapan dari metode gerombol berhirarki untuk peubah yang berskala pengukuran interval dan rasio. Teknik yang dilakukan pada tahap awal sama dengan yang dilakukan metode *Two Step Cluster* pada tahap pertama, yaitu pembentukkan *CF Tree*. Tahapan

selanjutnya adalah pengambilan contoh secara acak dari data tersebut. Jarak antar gerombol yang dapat digunakan pada metode ini adalah jarak Euclidean dan jarak Manhattan.

3. Analisis gerombol dua langkah (*Two Step Cluster Analysis*)

Metode ini dapat mengatasi masalah skala pengukuran yang tidak sama, dalam hal ini bertipe kontinu dan kategorik, serta memiliki jumlah objek amatan relatif besar. Metode ini masih memiliki kelemahan yaitu sensitif terhadap data yang berupa urutan atau tingkatan, sehingga masih tidak mampu dalam menangani data ordinal. Apabila terdapat peubah yang bertipe ordinal, maka sebelum dianalisis peubah tersebut harus ditransformasi terlebih dahulu.

Jarak antara dua gerombol didefinisikan sebagai jarak antar pusat dari masing-masing gerombol tersebut. Pusat dari suatu gerombol adalah vektor dari rata-rata masing-masing peubahnya. Jarak yang digunakan dalam metode *Two Step Cluster* adalah jarak Log-Likelihood dan jarak Euclidean.

Prosedur penggerombolan objek dalam metode *Two Step Cluster* dilakukan melalui dua tahapan (SPSS, 2004; Bacher *et al.*, 2004), yaitu :

• Tahap 1 : Pembentukan gerombol awal

Tahapan ini bertujuan untuk meminimalisasi jumlah amatan yang relatif besar. Yang dilakukan pada tahapan ini adalah pembentukan *Cluster Feature (CF) Tree*. *CF Tree* terdiri dari beberapa *Cluster Feature (CF)*. Pada *CF Tree*, data diamati satu persatu secara random. Berdasarkan kriteria jarak, data tersebut ditentukan apakah digabungkan dengan gerombol sebelumnya atau membentuk gerombol baru.

Pencilan merupakan data yang tidak sesuai apabila dimasukkan ke dalam gerombol yang tersedia. Setelah dibentuk *CF-tree*, diperiksa kembali apakah pencilan dapat dimasukkan ke dalam gerombol yang sudah ada tanpa harus membuat CF baru.

Setelah itu gerombol yang ada diamati dan dilakukan perhitungan jarak log-likelihood untuk gerombol yang terdapat pencilan dan gerombol tanpa pencilan. Gerombol yang memiliki jarak terbesar dikatakan memiliki pencilan jika jarak antara gerombol tersebut lebih besar dari titik kritis C , dengan rumusnya sebagai berikut :

$$C = \log(V)$$

dimana :

$$V = \prod_k R_k \prod_m L_m$$

R_k = range dari peubah kontinu ke- k

L_m = jumlah kategori untuk peubah kategori ke- m

Pada jarak Euclidean, data yang memuat pencilan memiliki prosedur yang sama dengan jarak Log-Likelihood. Dikatakan pencilan jika jarak Euclidean terbesar antara gerombol tersebut lebih besar dari titik kritis C , dengan rumus C sebagai berikut :

$$C = 2 \left(\sum_{i=1}^{K^A} \frac{\hat{\sigma}_{kl}^2}{K^A} \right)^{\frac{1}{2}}$$

dimana :

R_k = range dari peubah kontinu ke- k

K^A = jumlah total peubah kontinu

$\hat{\sigma}_{kl}^2$ = ragam dugaan untuk peubah kontinu ke- l dalam gerombol k .

• Tahap 2 : Pembentukan gerombol optimal

Suatu gerombol dikatakan optimal apabila memiliki jarak antar gerombolnya paling jauh dan jarak antar objek dalam gerombol tersebut paling dekat. Semakin dekat jarak antar objek maka semakin besar kemiripan antar obyek dalam satu gerombol. Pada tahapan ini, hasil dari tahap pertama yaitu *Cluster Feature (CF)* digerombolkan menggunakan analisis gerombol berhirarki dengan metode penggabungan.

Dalam penentuan jumlah gerombol optimal, ada dua langkah yang harus dilakukan. Langkah yang pertama yaitu menghitung BIC (*Bayesian Information Criterion*) atau AIC (*Akaike's Information Criterion*) untuk tiap-tiap gerombol. Kemudian hasil perhitungan tersebut digunakan untuk menduga jumlah gerombol. Langkah yang kedua yaitu mencari peningkatan jarak terbesar antara dua gerombol terdekat pada masing-masing tahapan penggerombolan.

Rumus BIC dan AIC untuk Gerombol j adalah sebagai berikut :

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N)$$

$$AIC(J) = -2 \sum_{j=1}^J \xi_j + m_j$$

dimana :

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_K - 1) \right\}$$

K^A = jumlah total peubah kontinu

K^B = jumlah total peubah kategorik

L_K = jumlah kategori untuk peubah kategorik ke- k

N = jumlah total data

Solusi gerombol yang terbaik memiliki BIC terkecil, tetapi ada beberapa kasus dalam penggerombolan dimana BIC akan terus menurun nilainya bila jumlah gerombol semakin meningkat. Maka dalam situasi tersebut, *ratio BIC Changes* (rasio perubahan BIC) dan *ratio of Distance Measure Changes* (rasio perubahan jarak) mengidentifikasi solusi gerombol terbaik. Solusi akan memiliki rasio *BIC changes* dan rasio *distance measure* yg besar.

Jumlah gerombol yang terbentuk dapat diketahui dengan menggunakan perbandingan antar jarak untuk k gerombol, dengan rumus perbandingannya sebagai berikut :

$$R(k) = d_{k-1} / d_k$$

$$d_k = l_{k-1} - l_k$$

dimana :

$$l_v = (r_v \log n - BIC_v) / 2 \text{ atau}$$

$$l_v = (2r_v - AIC_v) / 2$$

$$v = k, k-1$$

d_{k-1} = jarak jika k gerombol digabungkan dengan $k-1$ gerombol

Ukuran Jarak

Ukuran kemiripan dan ketakmiripan yang digunakan dalam analisis gerombol adalah jarak antar objek dan jarak antar gerombol.

Fungsi jarak yang sering digunakan diantaranya :

1. Jarak Euclidean

Jarak Euclidean paling sering digunakan hampir di berbagai metode analisis gerombol, tetapi hanya dapat digunakan apabila semua peubah yang digunakan adalah kontinu.

Jarak Euclidean antara gerombol ke- i dan gerombol ke- j dari p peubah didefinisikan :

$$d(i, j) = \left[\sum_{i=1}^p (\bar{X}_i - \bar{X}_j)^2 \right]^{\frac{1}{2}}$$

dimana :

$d(i, j)$ = jarak antara objek i ke objek j

\bar{X}_i = nilai tengah pada gerombol ke- i

\bar{X}_j = nilai tengah pada gerombol ke- j

p = banyaknya peubah yang diamati

2. Jarak Manhattan

Ukuran jarak ini dapat dikatakan sebagai bentuk yang lebih bersifat umum dari jarak Euclidean. Fungsi jaraknya adalah sebagai berikut :

$$d(i, j) = \left[\sum_{i=1}^p (\bar{X}_i - \bar{X}_j)^k \right]^{\frac{1}{k}}$$

dimana :

$d(i, j)$ = jarak antara objek i ke objek j

\bar{X}_i = nilai tengah pada gerombol ke- i

\bar{X}_j = nilai tengah pada gerombol ke- j

p = banyaknya peubah yang diamati

Jika dipilih $k=2$, maka akan menghasilkan jarak Euclidean.

3. Jarak Mahalanobis

Jarak Mahalanobis sangat berguna dalam menghilangkan atau mengurangi perbedaan skala pada masing-masing komponen. Pada permasalahan tertentu, pada saat menentukan jarak, perlu juga dipertimbangkan ragam dan peragam. Jarak Mahalanobis didefinisikan sebagai berikut :

$$d(i, j) = \left[(\bar{X}_i - \bar{X}_j)' S^{-1} (\bar{X}_i - \bar{X}_j) \right]^{\frac{1}{2}}$$

dimana :

$d(i, j)$ = jarak antara objek i ke objek j

\bar{X}_i = nilai tengah pada gerombol ke- i

\bar{X}_j = nilai tengah pada gerombol ke- j

S^{-1} = matriks ragam-peragam gabungan antara \bar{X}_i

dan \bar{X}_j .

4. Jarak Log-Likelihood

Jarak Log-Likelihood dapat digunakan untuk peubah kontinu maupun kategorik. Jarak antara gerombol j dan s didefinisikan sebagai berikut:

$$d(j, s) = \xi_j + \xi_s - \xi_{\langle j, s \rangle}$$

dimana:

$$\xi_j = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) + \sum_{k=1}^{K^B} \hat{E}_{jk} \right) N \dots =$$

jumlah total data.

N_{jkl} = jumlah data di gerombol j untuk peubah kategorik ke- k dengan kategori ke- l

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke- k dalam gerombol j .

K^A = jumlah total peubah kontinu

K^B = jumlah total peubah kategorik

L_m = jumlah kategori untuk kategori ke- k

$d(j, s)$ = jarak antara gerombol j dan s .

$\langle j, s \rangle$ = indeks kombinasi gerombol j dan s .

BAHAN DAN METODE

Data yang digunakan sebagai studi kasus adalah data Podes Sensus Ekonomi tahun 2003 untuk wilayah Jawa Barat.

Satuan pengamatan adalah populasi desa/kelurahan dengan jumlah amatan sebanyak 5758 desa.. Peubah-peubah yang digunakan yaitu mengenai profil tentang desa atau kelurahan, kependudukan, ekonomi, komunikasi dan informasi dan penggunaan lahan.

Dari data tersebut, dilakukan analisis gerombol menggunakan metode *Two Step Cluster*. Perangkat lunak yang digunakan adalah *SPSS 13 for Windows* dan *Ms Excel*.

Beberapa tahapan yang dilakukan adalah :

1. Tahap sebelum analisis.

Yang pertama kali dilakukan adalah mengkategorikan peubah-peubah berdasarkan skala pengukurannya. Peubah bertipe kontinu distandarisasi terlebih dahulu lalu dilakukan analisis *Two Step Cluster*.

2. Tahap analisis.

Dalam analisis, untuk mengetahui karakteristik penting pada masing-masing gerombol dilakukan dengan uji *Chi-Square* untuk peubah kategorik dan uji *t-Student* untuk peubah kontinu. Pada uji *Chi-Square*, yang digunakan adalah proporsi individu/objek antar kategorik dalam satu gerombol. Sedangkan pada uji *t-Student* adalah rataan dari karakteristik didalam gerombol terhadap rataan umum.

3. Tahap setelah analisis.

Setelah dianalisis, dilakukan tahap verifikasi Kabupaten Bogor pada masing-masing gerombolnya dengan cara mengunjungi kantor kelurahan daerah tersebut.

HASIL DAN PEMBAHASAN

Penggerombolan Desa/Kelurahan dengan Metode Two Step Cluster

Sebelum dilakukan penggerombolan, peubah yang bertipe kontinu perlu dilakukan standarisasi dari data awal ke bentuk baku (Z) untuk mengurangi keragaman yang disebabkan oleh satuan pengukuran yang tidak sama. Pada kasus ini, ada 18 peubah yang distandarisasi. Peubah-peubah tersebut adalah $X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19}, X_{20}, X_{21}$.

Ukuran jarak yang digunakan adalah jarak Log Likelihood, karena data yang digunakan bertipe kategorik dan kontinu. Dalam penentuan jumlah gerombol, yang digunakan adalah perhitungan BIC. Hal ini ditentukan secara subjektif, karena perhitungan BIC dan AIC memberikan hasil tidak terlalu berbeda.

Kriteria gerombol dilihat dari masing-masing BIC yang didapat. Semakin kecil nilai BIC maka semakin bagus. Gerombol yang dihasilkan pada tahap pertama sebanyak 8 gerombol. Hal ini dilihat dari rasio BIC_k/BIC_1 pada gerombol 8, yaitu sebesar 0.04, bernilai sama dengan konstanta c_1 ($c_1=0.04$, didapat dari simulasi penelitian yang dilakukan oleh penulis dari SPSS Two Step Clustering pada tahun 2004). Jumlah gerombol didapat dari solusi yang memiliki *ratio change* terbesar. *Ratio change* merupakan perbandingan antara dua $R(k)$ yang terbesar. Dari simulasi penelitian yang dilakukan sebelumnya juga didapat suatu batas $c_2 = 1.15$. Jadi, jika hasil perbandingannya lebih besar dari batas c_2 , maka jumlah gerombol tersebut merupakan solusi akhir. Dari hasil yang didapat, dua nilai $R(k)$ terbesar adalah untuk tiga gerombol ($R(k) = 2.446$) dan lima gerombol ($R(k) = 1.938$). Rasio antara kedua nilai tersebut adalah 1.262 dan lebih besar dari batas nilai konstanta 1.15. Oleh karena itu, dalam kasus ini tiga gerombol merupakan solusi optimal.

Distribusi anggota dari masing-masing gerombol yang terbentuk untuk lebih jelasnya dapat dilihat pada Tabel 3. Dari 5758 desa/kelurahan yang berada di Jawa Barat, sebanyak 185 (3.2%) berasal dari gerombol 1, 3820 (66.3%) dari gerombol 2, dan 1753 (30.4%) berasal dari gerombol 3.

Karakteristik Gerombol Desa/Kelurahan

Berdasarkan hasil-hasil gerombol yang didapat, gerombol satu tidak dapat dikatakan sebagai suatu gerombol, karena anggota-anggota didalamnya merupakan objek-objek yang memencil ekstrim dan tidak dapat dimasukkan ke dalam gerombol lainnya. Anggota dari gerombol ini hanya sebanyak 3.2 % dari keseluruhan populasi.

Gerombol yang dihasilkan dapat dibedakan menjadi dua jenis gerombol. Jenis pertama yaitu gerombol yang terdiri dari desa/kelurahan yang memiliki karakteristik ekstrim, sedangkan jenis kedua yaitu gerombol yang anggota-anggota di tiap gerombol memiliki kemiripan karakteristik.

Tabel BIC (Bayesian Information Criterion)

Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change ^a	Ratio of BIC Changes ^b	Ratio of Distance Measures ^c
1	73625.428			
2	53402.884	-20222.544	1.000	1.887
3	42826.244	-10576.640	.523	2.446
4	38681.605	-4144.639	.205	1.505
5	36028.812	-2652.793	.131	1.934
6	34803.230	-1225.582	.061	1.088
7	33701.870	-1101.361	.054	1.246
8	32878.212	-823.658	.041	1.243
9	32275.170	-603.042	.030	1.062
10	31724.961	-550.210	.027	1.060
11	31222.785	-502.175	.025	1.010
12	30728.800	-493.986	.024	1.289
13	30413.401	-315.398	.016	1.083
14	30145.320	-268.081	.013	1.115
15	29935.978	-209.342	.010	1.122

- a. The changes are from the previous number of clusters in the table.
- b. The ratios of changes are relative to the change for the two cluster solution.
- c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Gerombol desa/kelurahan ekstrim

Gerombol yang termasuk ke dalam gerombol desa/kelurahan ekstrim adalah gerombol satu. Dilihat dari status desa, desa/kelurahan ekstrim yang berstatus pedesaan sebanyak 3.5 % dan yang berstatus perkotaan sebanyak 2.6 %. Berdasarkan lokasi, yang berasal dari daerah non pesisir sebanyak 3.2 % dan dari daerah pesisir sebanyak 4.1 %. Dari status keterpencilan, terdapat sebanyak 3,2 % yang berasal dari daerah non terpencil dan 2.9 % berasal dari daerah terpencil. Desa/kelurahan yang termasuk desa terpencil adalah desa Sedari di kabupaten Karawang.

Peubah kategorik menggunakan uji *Chi-Square* dan peubah kontinu menggunakan uji *t-Student* sebagai statistik uji untuk mengetahui karakteristik penting pada masing-masing gerombol. Apabila terdapat peubah-peubah yang memiliki statistik uji lebih besar dari nilai kritis, maka dapat disimpulkan bahwa peubah-peubah itu berpengaruh pada pembentukan gerombol tersebut.

Pada desa/kelurahan yang memiliki karakteristik ekstrim, semua peubah kategorik tidak signifikan, karena anggotanya memiliki karakteristik yang berbeda-beda. Sehingga dapat disimpulkan bahwa dalam pembentukan gerombol satu tidak berdasarkan pada peubah-peubah kategorik yang digunakan. Sedangkan untuk peubah kontinu, yang memiliki statistik uji lebih besar dari titik kritis hanya peubah X_6 , X_7 , X_{11} , X_{16} , X_{17} .

Sebagai ilustrasi, desa/kelurahan yang ekstrim di bidang industri adalah desa Dramaga di Kabupaten Bogor Kecamatan Dramaga. Desa ini memiliki tingkat industri yang jauh lebih maju bila dibandingkan desa lainnya.

Gerombol desa/kelurahan non ekstrim

Gerombol dua dan gerombol tiga termasuk ke dalam gerombol desa/kelurahan non ekstrim. Berdasarkan status desa, sebagian besar gerombol dua berstatus pedesaan, yaitu sebesar 96.5 %. Desa yang berada di gerombol tiga sebanyak 95.9 % berstatus perkotaan. Pada gerombol tiga, desa yang berstatus pedesaan hanya sebanyak satu desa, yaitu desa Wadas Kabupaten Karawang.

Desa/kelurahan di wilayah Jawa Barat mayoritas berlokasi di daerah non pesisir, pada gerombol dua sebanyak 65.5 % berlokasi di daerah non pesisir dan 87.7 % anggotanya berada di daerah pesisir. Pada gerombol tiga, sebanyak 31.3 % berlokasi di daerah non pesisir dan hanya 8.2% % anggotanya berada di daerah pesisir.

Berdasarkan status keterpencilan, sebanyak 66.2 % desa/kelurahan yang termasuk pada gerombol dua merupakan daerah non terpencil. Pada gerombol tiga, seluruh anggotanya merupakan daerah non terpencil.

Desa/kelurahan yang merupakan gerombol tiga memiliki jarak yang paling dekat dengan kecamatan dan kota terdekat dibanding gerombol lainnya. Gerombol dua memiliki persentase rumah tangga pertanian serta persentase rumah tangga sejahtera I dan prasejahtera terbanyak, tetapi memiliki persentase terkecil untuk rumah tangga yang menggunakan listrik, televisi dan berlangganan telepon. Tingkat penduduk yang menganggur paling tinggi berada di gerombol tiga. Gerombol dua mempunyai pendapatan asli daerah dan pengeluaran anggaran terendah. Desa/kelurahan yang merupakan gerombol tiga memiliki luas lahan sawah, ladang, perkebunan dan hutan yang paling kecil. Hal ini disebabkan oleh anggota gerombol tiga yang mayoritas berstatus perkotaan. Gerombol dua memiliki jumlah industri yang paling sedikit.

Peubah kategorik yang berpengaruh pada pembentukan gerombol dua dan gerombol tiga adalah status desa. Peubah lokasi dan keterpencilan memiliki nilai statistik uji yang kurang dari nilai kritis. Hal ini menyebabkan hasil gerombol dua dan gerombol tiga cenderung terpisah menurut status desa. Desa yang merupakan anggota gerombol dua mayoritas berstatus pedesaan dan gerombol tiga mayoritas berstatus perkotaan.

Peubah kontinu yang tidak berpengaruh pada pembentukan gerombol dua adalah luas perkebunan, luas hutan rakyat dan persentase penduduk menganggur. Pada gerombol tiga, peubah kontinu yang tidak berpengaruh adalah persentase penduduk menganggur, pengeluaran anggaran rutin dan pendapatan asli desa.

Verifikasi Gerombol Desa/Kelurahan di Wilayah Kabupaten Bogor

Tahapan verifikasi dilakukan untuk memastikan bahwa desa/kelurahan yang berada pada masing-masing gerombol yang terbentuk memang memiliki karakteristik gerombol tersebut. Pada tahapan ini, diamati kecamatan Dramaga yang berada di wilayah kabupaten Bogor yang memiliki 10 desa, dengan 1 desa berada pada gerombol pertama, 5 desa pada gerombol kedua dan 4 desa di gerombol ketiga. Pemilihan kecamatan ini ditentukan secara subjektif.

Desa Dramaga yang berstatus perkotaan merupakan anggota gerombol satu. Desa ini termasuk desa yang memiliki tingkat industri yang berkembang. Industri yang berkembang pada desa ini adalah industri kerajinan, industri pakaian, industri rumah tangga, industri bahan bangunan, industri alat pertanian dan perdagangan. Tingkat perkembangan industri desa Dramaga jauh lebih tinggi dibanding desa lainnya. Ini merupakan faktor yang menyebabkan desa tersebut masuk ke gerombol ini.

Desa yang tergabung pada gerombol dua berada jauh dari kecamatan dan kota terdekat dan berstatus pedesaan. Desa di gerombol ini memiliki luas lahan sawah yang luas, sehingga dapat ditingkatkan potensinya dengan pemanfaatan lahan tersebut. Desa/kelurahan tersebut memiliki kelemahan di bidang komunikasi dan informasi.

Desa yang termasuk pada gerombol tiga merupakan desa yang berstatus perkotaan. Diantara desa/kelurahan yang berada di Kecamatan Dramaga, desa/kelurahan pada gerombol tiga cukup maju dalam komunikasi dan informasi.

KESIMPULAN DAN SARAN

Kesimpulan

Dari hasil penggerombolan dengan metode *Two Step Cluster*, gerombol awal yang dihasilkan pada

tahap pertama adalah sebanyak delapan gerombol, sedangkan gerombol optimal yang dihasilkan pada tahap dua adalah sebanyak tiga gerombol.

Gerombol satu merupakan gerombol khusus yang memiliki karakteristik ekstrim. Anggota dari gerombol ini merupakan kumpulan dari desa/kelurahan yang memiliki karakteristik yang sangat menonjol atau ekstrim, sehingga tidak dapat dilihat karakteristik spesifik mengenai gerombol ini.

Desa/kelurahan yang termasuk gerombol dua memiliki karakteristik pedesaan. Desa/kelurahan tersebut memiliki lahan terluas, jumlah rumah tangga pertanian terbanyak namun belum berkembang dalam bidang industri serta komunikasi dan informasi. Sehingga untuk meningkatkan potensi desa pada gerombol ini, yang harus diperhatikan adalah peubah-peubah yang tingkat perkembangannya masih rendah.

Sebaliknya, gerombol tiga memiliki karakteristik desa yang berstatus perkotaan. Desa/kelurahan pada gerombol ini memiliki jarak terdekat ke pusat kota, cukup maju dalam bidang industri, komunikasi dan informasi, namun memiliki angka pengangguran tertinggi.

Saran

Saran untuk penelitian selanjutnya adalah membandingkan antara metode *Two Step Cluster* dengan pendekatan analisis gerombol lain dalam masalah skala pengukuran yang berbeda.

DAFTAR PUSTAKA

- Anderberg, M.R. 1973. *Cluster Analysis for Application*. New York: Academic Press.
- Bacher, J., K.Wenzig, M. Vogler. 2004. SPSS Two Step Cluster - A First Evaluation. [terhubung berkala]. <http://www.statisticalinnovations.com/products/TwoStep.pdf>. [10 Juni 2005].
- [BPS] Badan Pusat Statistik. 2003. *Jawa Barat dalam Angka 2003*. BPS, Jakarta.
- Johnson, R.A., D.W. Wichern. 2002. *Applied Multivariate Statistical Analysis*. Ed ke-5. New Jersey: Prentice-Hall.
- Muenchen, B. 2002. Discover Depth and Flexibility with Two Step Cluster Analysis. [terhubung berkala]. <http://www.spss.com/pdfs/S115ad8-1202A.pdf>. [20 Mei 2005].
- Sartono, B. *et al.* 2003. *Modul Teori Analisis Peubah Ganda*. Bogor: Departemen Statistika FMIPA IPB.
- Zhang, T., R. Ramakrishnan, M. Livny. 1996. BIRCH: An Efficient Data Clustering Method for Very Large Databases. Di dalam: *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*; Montreal, 4- 6 Jun 1996. Canada: ACM Press. hlm 103-114.