

# Pengklasteran Provinsi Berdasarkan Persentase Penduduk yang Mempunyai Keluhan Kesehatan dalam Sebulan pada Tahun 2003

Lisa Noviyanti<sup>1)</sup>, Sigit Nugroho<sup>2)</sup>, Fachri Faisal<sup>2)</sup>

<sup>1)</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2)</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

## ABSTRAK

Strategi kebijakan pembangunan kesehatan baru yang dikenal dengan visi pembangunan kesehatan yang secara singkat dinyatakan sebagai “Indonesia Sehat 2010” bertujuan untuk mengoptimalkan derajat kesehatan. Derajat kesehatan yang optimal merupakan salah satu faktor untuk mensukseskan pembangunan nasional. Derajat kesehatan adalah tingkat kesehatan perorangan, kelompok atau masyarakat yang diukur dengan angka kesakitan, status gizi, angka kematian dan harapan hidup. Penelitian deskriptif eksploratif ini bertujuan untuk mengklasterkan provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003. Pengklasteran dilakukan dengan metode *k-means clustering* dan hirarki aglomeratif yaitu: *single linkage*, *complete linkage*, metode ward, median, dan sentroid. Pengklasteran ini menghasilkan 5 klaster yang ideal..

**Key words** : *cluster analysis, dendogram, aglomerative, k-means clustering*

## PENDAHULUAN

Pemerintah Republik Indonesia telah menyusun strategi kebijakan pembangunan kesehatan baru. Kebijakan ini didasarkan pada gerakan pembangunan berwawasan kesehatan sebagai Strategi Nasional menuju Indonesia Sehat 2010 yang dicanangkan oleh Presiden Republik Indonesia pada pembukaan Rakernas Departemen Kesehatan RI pada tanggal 1 Maret 1999.

Dengan rumusan ini diharapkan pada tahun 2010 bangsa Indonesia sudah hidup dalam lingkungan yang sehat, berperilaku hidup bersih dan sehat, dapat memilih, menjangkau dan memanfaatkan pelayanan kesehatan yang bermutu, merata dan berkeadilan sehingga memiliki derajat kesehatan yang optimal (Anonim, 2004).

Meningkatnya jumlah penduduk berarti upaya untuk meningkatkan derajat kesehatan masyarakat perlu diperhatikan. Hasil penelitian United Development Project (UNDP) menyatakan bahwa *Human Development Indeks* (indeks pembangunan manusia) Indonesia berada pada urutan ke 117 dari 175 negara. Oleh karena itu, masalah kesehatan tidak tepat lagi bila dipandang sebagai suatu konsumsi tapi sebagai investasi dalam upaya mencapai kesejahteraan masyarakat (Argadiredha, 2003).

Berbagai upaya untuk meningkatkan derajat kesehatan masyarakat telah dilakukan. Membuat dan melaksanakan program pembangunan di bidang kesehatan agar merata bagi segenap masyarakat Indonesia. Namun masih disadari bahwa masih terjadi ketimpangan sehingga derajat kesehatan masyarakat masih rendah dan sangat tertinggal jika dibandingkan dengan negara-negara tetangga (Anonim, 2002a).

Salah satu indikator yang digunakan untuk menentukan derajat kesehatan tersebut adalah angka kesakitan (Anonim, 2003). **Derajat Kesehatan** adalah tingkat kesehatan perorangan, kelompok atau masyarakat yang diukur dengan angka kesakitan, umur harapan hidup, status gizi dan angka kematian. Banyaknya orang yang sakit menggambarkan kondisi kesehatan suatu bangsa

(Anonim, 1991). Angka kesakitan ini diharapkan dapat menurun sesuai dengan sasaran program pokok upaya kesehatan baik untuk penyakit menular atau penyakit tidak menular.

Analisis klaster merupakan suatu teknik yang dapat digunakan untuk mengklasifikasikan obyek (provinsi-provinsi di Indonesia) ke dalam kelompok yang relatif homogen yang disebut klaster. Pengklasteran berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dapat memberikan informasi berupa profil sehingga dapat membantu mencari faktor penyebab timbulnya masalah kesehatan. Dengan demikian hasil pengklasteran dapat membantu pemerintah dalam mensukseskan program Indonesia Sehat 2010.

Berdasarkan hal tersebut maka penulis tertarik untuk mengklasterkan provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003.

Bagaimana membuat pengklasteran provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003 yang dapat memberikan profil yang menjelaskan karakteristik masing-masing provinsi di Indonesia.

Tujuan penelitian ini adalah mengklasterkan provinsi-provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan selama sebulan pada tahun 2003 sehingga provinsi-provinsi yang mempunyai kemiripan akan berada dalam satu klaster.

Penelitian ini dapat memberikan informasi berupa gambaran tentang kesehatan masyarakat dan profil yang menjelaskan karakteristik setiap klaster. Profil dapat digunakan sebagai sarana diagnosis dalam mencari sebab masalah kesehatan sehingga memudahkan pencarian pemecahan masalah. Dengan demikian hasil pengklasteran dapat membantu mengoptimalkan derajat kesehatan bangsa Indonesia dan mensukseskan program Indonesia Sehat 2010.

## TINJAUAN PUSTAKA

### **Indonesia Sehat 2010**

Tujuan pembangunan kesehatan (Undang-Undang Nomor 23 tahun 1992) adalah untuk meningkatkan kesadaran, kemauan dan kemampuan bagi tiap warga negara Indonesia agar berada pada kondisi kesehatan yang optimal. Hal senada juga tertuang dalam visi pembangunan kesehatan dan Garis-Garis Besar Haluan Negara (GBHN) 1999-2004 yang mengamanatkan bahwa pembangunan kesehatan diarahkan untuk dapat meningkatkan derajat kesehatan, meningkatkan mutu dan kemudahan pelayanan kesehatan agar terjangkau oleh seluruh lapisan masyarakat. Termasuk juga, meningkatkan keadaan gizi, membudayakan pola hidup bersih dan sehat (PHBS), didukung oleh pembangunan pemukiman yang layak yang dikelola melalui manajemen yang efisien dalam pemenuhan kebutuhan aspirasi masyarakat sasaran (Anonim, 2002a).

Derajat kesehatan atau proporsi penduduk dalam satu wilayah yang mengalami sakit selama periode tertentu adalah tingkat kesehatan perorangan, kelompok atau masyarakat. Dalam hal ini, keluhan kesehatan selama sebulan yang lalu sebelum survei. Keluhan kesehatan digambarkan lewat angka kesakitan macam-macam penyakit (persentase penduduk yang mengalami kesakitan selama periode tertentu) masing-masing jenis keluhan penduduk. Keluhan kesehatan adalah keadaan seseorang yang mengalami gangguan kesehatan dan kejiwaan baik karena penyakit akut, penyakit kronis, kecelakaan, kriminal atau hal lain (Anonim, 2003a).

### **Analisis Klaster**

Analisis klaster merupakan salah satu analisis multivariat yang termasuk dalam metode interdependensi yaitu variabel bebas  $x$  atau faktor penyebab tidak dibedakan dengan variabel terikat  $y$  atau respon (Supranto, 2004).

Istilah analisis klaster pertama kali digunakan oleh Tryon (1939) (Anonim, 2004b). Analisis klaster adalah suatu koleksi metode statistik yang mengidentifikasi kelompok sampel berdasarkan

karakteristik serupa (Sambamorthi, 2005). Analisis kluster mengelompokkan elemen mirip sebagai obyek penelitian yang mempunyai tingkat homogenitas yang tinggi antar obyek menjadi kluster yang berbeda dengan tingkat heterogenitas obyek yang tinggi antar kluster. Pengklasteran ini didasarkan pada gugus variabel yang diper timbangkan untuk diteliti.

Hasil pengklasteran diharapkan akan menyediakan beberapa pengertian yang mendalam untuk masing-masing provinsi di Indonesia. Pengklasteran mempunyai efek mengurangi banyak data dengan mengurangi banyaknya obyek (Anonim, 2005b). Analisis kluster harus memenuhi asumsi berikut (Santoso, 2002):

- a. Sampel yang diambil harus benar-benar bisa mewakili populasi
- b. Multikolinieritas yaitu korelasi antar obyek. Sebaiknya tidak ada, bila ada maka besar multikolinieritas tidaklah tinggi ( $< 0,5$ ).

Sebelum melakukan analisis kluster, data yang digunakan juga perlu diperhatikan. Apakah terdapat perbedaan nilai yang besar antar variabel. Perbedaan data yang besar akan menyebabkan perhitungan jarak menjadi tidak valid sehingga data harus ditransformasi. Transformasi dapat dilakukan dengan uji *z-score*.

Untuk menentukan kedua obyek dikatakan mirip, perlu didefinisikan ukuran kemiripan antar dua obyek. Hal ini dilakukan untuk memperoleh matrik *proximity* yaitu matrik persegi dan simetri dengan jumlah obyek yang sama pada baris dan kolom. Matrik ini menunjukkan kemiripan atau ketakmiripan antar obyek (Anonim, 2003c).

Adapun metode yang dapat digunakan untuk mengukur kesamaan antar obyek yaitu (Santoso, 2004)

1. Mengukur jarak antar dua obyek

Metode ini berbentuk matrik simetri  $n \times n$  yang berisi kemiripan atau ketakmiripan antar obyek sehingga jarak antar dua obyek bisa langsung diukur.

2. Mengukur korelasi antar sepasang obyek pada beberapa variabel

Pada metode ini datanya berbentuk matrik. Kesamaan antar obyek didapat dengan transformasi satu-satu sehingga indeks ketakmiripan bisa dikonversi menjadi indeks kemiripan. Salah satu yang jelas bisa menjadi ukuran ketakmiripan adalah fungsi jarak antara obyek  $a$  dan  $b$  di tulis  $d(a,b)$ . Fungsi ini harus memenuhi (Anonim (2003b), Portier (2003))

$$d(a,b) \geq 0$$

$$d(a,a) = 0$$

$$d(a,b) = d(b,a)$$

$$(a,b) \text{ meningkat seiring tidak miripnya } a \text{ dan } b$$

$$d(a,c) \leq d(a,b) + d(b,c)$$

Pengukuran jarak ada bermacam-macam namun yang paling sering digunakan adalah jarak *euclid* yaitu jarak geometris di ruang multidimensional (Anonim, 2004b). Beberapa cara dalam mengukur jarak yaitu

- a. Menggunakan jarak *euclid* yaitu akar jumlah kuadrat perbedaan nilai untuk tiap variabel (Wichern, 2002).

Jika  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  dan  $\mathbf{y}' = (y_1, y_2, \dots, y_p)$  maka

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

b. Menggunakan jarak kuadrat *euclid*

c. *The City Block or Manhattan Distance* antara dua obyek merupakan jumlah nilai perbedaan mutlak untuk tiap variabel. Jarak ini juga disebut jarak Minkowski (Dillon & Goldstein, 1984).

Jika  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ ;  $p$  adalah variabel

maka  $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  adalah kumpulan variabel pada obyek ke  $i$

$$d_{ij} = \left[ \sum_{k=1}^p |x_{ik} - y_{jk}|^r \right]^{\frac{1}{r}}$$

$d_{ij}$  adalah jarak antara obyek ke  $i$  dan obyek ke  $j$ .

d. *The Chebyshev Distance* antar dua obyek ialah nilai perbedaan mutlak yang maksimum pada tiap variabel (Supranto, 2004).

### 3. Mengukur asosiasi antar obyek

Pada metode ini, data berbentuk non matrik (nominal atau ordinal). Jika akan dibandingkan obyek  $r$  dan obyek  $s$  maka indeks kesamaan ditunjukkan pada tabel 1 (Portier, 2003).

Tabel 1 Kesamaan Antara Dua Obyek

Obyek r	obyek s		
	1	0	Jml
1	$a$	$b$	$a+b$
0	$c$	$d$	$c+d$
Jml	$a+c$	$b+d$	$a+b+c+d=p$

Asumsikan 1 jika karakteristik ada dan 0 jika karakteristik tidak ada. Koefisien kesamaan didefinisikan pada tabel 2.

Tabel 2 Koefisien Kesamaan Antara Dua Obyek

Nama Koefisien	Persamaan
Jaccard	$S_{rs} = \frac{a}{(a+b+c)}$
Dice	$S_{rs} = \frac{2a}{(2a+b+c)}$
Ochiai	$S_{rs} = \frac{a}{[(a+b)(a+c)]^{\frac{1}{2}}}$
Russell-Rao	$S_{rs} = \frac{a}{p}$
Yule	$S_{rs} = \frac{(ad-bc)}{ad+bc}$
Phi	$S_{rs} = \frac{(abc)}{[(a+b)(a+c)(b+d)(c+d)]^{\frac{1}{2}}}$

Konversi ketaksamaan untuk analisa  $d_{rs} = 1 - S_{rs}$

$d_{rs}$  = Koefisien ketaksamaan antara obyek  $r$  dan obyek  $s$

$S_{rs}$  = Koefisien kesamaan antara obyek  $r$  dan obyek  $s$

## Metode Pengklasteran

### Metode Hirarki

Metode hirarki bisa aglomeratif (*agglomeration*) atau devisif (*devisife*) (Anonim, 2005b). Metode Aglomeratif terdiri dari :

#### 1. Linkage Method (Metode Linkage)

Metode ini dibagi lagi menjadi tiga metode yaitu (Wichern, 2002):

##### i. Single linkage

*Single linkage* dan *complete linkage* kali pertama didiskusikan oleh Sneath (1957) lalu dibahas lagi oleh Sokal Sneath Classic (1963). Kemudian ke dua metode ini juga didiskusikan oleh Johnson (1967), (Dillon & Goldstein, 1984). Jika jarak antara kluster  $B_r$  dan  $B_s$  adalah  $h(B_r, B_s)$  didefinisikan sebagai berikut:

$$h(B_r, B_s) = \min \{ d(x_i, x_j); x_i \text{ anggota } B_r \text{ dan } x_j \text{ anggota } B_s \}$$

Kluster  $B_r$  dan  $B_s$  akan digabung jika  $h(B_r, B_s)$  adalah jarak yang terkecil sehingga metode ini juga disebut aturan tetangga dekat (Dillon & Goldstein, 1984).

##### ii. Complete linkage

Metode ini hampir sama dengan *single linkage* hanya saja pada metode ini menggunakan jarak yang paling jauh antara dua kluster yang berbeda  $B_r$  dan  $B_s$  yang didefinisikan sebagai berikut:

$$h(B_r, B_s) = \max \{ d(x_i, x_j); x_i \text{ anggota } B_r \text{ dan } x_j \text{ anggota } B_s \}$$

##### iii. Average linkage

Metode ini menggunakan rata-rata jarak antara semua pasangan obyek sebagai jarak antara dua kluster. Metode ini jarang digunakan jika dibandingkan *single linkage* dan *complete linkage* karena metode ini membutuhkan informasi pada semua pasangan jarak. Namun ke tiga metode *linkage* ini seringkali memberikan hasil yang hampir sama (Anonim, 2003c).

Jarak antara dua kluster  $B_r$  dan  $B_s$  didefinisikan sebagai berikut:

$$h(B_r, B_s) = \frac{1}{n_1 n_2} \sum_{x_i \in B_r} \sum_{x_j \in B_s} d(x_i, x_j)$$

#### 2. Ward method (Metode Ward)

Ward (1963) mengusulkan penggunaan metode yang didasarkan pada hasil informasi yang minimum dari kenaikan pada jumlah kuadrat deviasi rata-rata kluster. Proses berhenti pada kenaikan yang menyebabkan *error sum of squares* (ESS) dari gabungan tiap kluster yang mungkin. Nilai ESS digunakan sebagai fungsi obyektif dan didefinisikan sebagai berikut (Dillon & Goldstein, 1984):

$$ESS = \sum_{j=1}^k \left( \sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j} \left( \sum_{i=1}^{n_j} x_{ij} \right)^2 \right)$$

$x_{ij}$  : Nilai obyek ke  $i$  pada kluster ke  $j$

$k$  : Jumlah kluster tiap *stage*

$n_j$  : Jumlah obyek ke  $i$  pada kluster ke  $j$

Metode ini juga dikenal dengan metode varian minimum dan harus menggunakan jarak kuadrat *euclid* namun sulit untuk menggunakannya tanpa bantuan komputer (Anonim, 2005c).

#### 3. Centroid Method (Metode Centroid)

Sokal dan Michener (1958), King (1996, 1997) dalam Seber (1984) menyatakan bahwa jarak antara dua kluster didefinisikan sebagai jarak *euclid* antar kedua rata-rata (*centroid*) kluster. Jika  $\bar{x}_r$  dan  $\bar{x}_s$

adalah vektor rata-rata (*centroid*) kluster  $B_r$  dan  $B_s$ , maka jarak antar dua kluster didefinisikan sebagai  $h(B_r, B_s) = d(\bar{x}_r, \bar{x}_s)$ . *Centroid* kluster baru yang terbentuk didapat dengan rumus

$$\frac{n_r \bar{x}_r + n_s \bar{x}_s}{n_r + n_s}$$

$n_r$  dan  $n_s$  adalah banyaknya anggota kluster  $B_r$  dan  $B_s$ .

#### 4. Median Method (Metode Median)

Terkadang terdapat ukuran kluster  $B_s$  jauh lebih kecil dari pada kluster  $B_r$  yaitu  $n_s \ll n_r$  bila kedua kluster digabungkan maka *centroid* dari kluster baru tidak akan jauh berbeda dengan  $\bar{x}_r$ . Untuk menghindari kontribusi  $B_s$  terhadap pembentukan jarak yang baru tidak terlalu besar, Gower menyarankan penggunaan *median* antara kluster yang digabungkan sebagai titik untuk menghitung jarak yang baru.

Jika kluster  $B_r$  dan  $B_s$  digabung maka akan diperoleh median baru yang didefinisikan sebagai berikut (Anonim, 2003b):

$$m_{baru} = \frac{m_r + m_s}{2}$$

Median ini dihitung sebagai titik tengah pada garis yang menghubungkan median lama dan median baru. Dengan demikian jarak antar kluster didefinisikan sebagai jarak antar median.

Metode berhirarki devisif merupakan kebalikan metode aglomeratif. Metode pengklasteran ini dimulai dengan menganggap bahwa hanya ada satu kluster yang memuat semua obyek.

William dan Lance dalam Seber (1984) menyatakan metode devisif lebih menguntungkan dibandingkan aglomeratif. Namun metode ini jarang digunakan (Anonim, 2005c) dan tidak semua *software* menyediakan fasilitas metode ini. Oleh karena itulah, metode berhirarki devisif tidak penulis bahas pada skripsi ini.

#### Metode Non Hirarki (Pengklasteran *K-means*)

Berbeda dengan metode hirarki, metode ini justru dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan dan *centroid* di tiap kluster. Pada beberapa *software*, *centroid* yang digunakan adalah  $k$  pengamatan pertama namun ada juga *software* yang menentukan *centroid* secara acak. Pengklasteran *K-means* pertama kali dipopulerkan oleh Hartigan pada tahun 1975 (Anonim, 2004b).

Venable & Ripley (1995), Anonim (2003b) dan Wichern (2002) menyatakan bahwa Mac Queen menyarankan penggunaan pengklasteran *K-means* untuk menjelaskan algoritma dalam penentuan suatu obyek ke dalam kluster tertentu berdasarkan rata-rata terdekat.

Asumsikan  $n$  adalah obyek dan  $p$  adalah variabel yang dinotasikan dengan  $x(i, j)$   $i=1, 2, \dots, n$  dan  $j=1, 2, \dots, p$  dan dengan menggunakan jarak *euclid* antar obyek. Jika  $p(n, k)$  adalah partisi yang merupakan hasil pada tiap obyek dialokasikan untuk salah satu dari kluster ke  $1, 2, \dots, k$ . Rata-rata variabel ke  $j$  pada kluster ke  $l$  dinotasikan dengan  $\bar{x}(i, j)$ , dan jumlah obyek pada kluster ke  $l$  dinotasikan dengan  $n(l)$ . Maka jarak antara obyek ke  $i$  dan kluster ke  $l$  didefinisikan sebagai berikut:

$$d(i, l) = \left( \sum_{j=1}^p [x(i, j) - \bar{x}(i, j)]^2 \right)^{1/2}$$

dengan

$$E[p(n, k)] = \sum_{i=1}^n D[i, l(i)]^2$$

adalah *error* partisi.  $l(i)$  adalah kluster yang memuat obyek ke  $i$ ,  $D[i,l(i)]$  adalah jarak *euclid* antara obyek ke  $i$  dan rata-rata kluster yang memuat obyek.

*Output* metode hirarki adalah semacam ringkasan yang digambarkan oleh dendogram. Dendogram merupakan diagram seperti pohon dua dimensi yang mengilustrasikan pemisahan atau penggabungan dengan tingkat yang berjenjang (Dillon & Goldstein, 1984).

Pada metode hirarki, dendogram dapat membantu peneliti untuk menentukan jumlah kluster yang ideal. Posisi garis pada skala menunjukkan jarak. Jarak pada tahapan awal mempunyai nilai yang hampir sama sehingga sukar untuk menentukan urutan beberapa kluster awal dibentuk. Namun, jelas sekali bahwa dua tahap terakhir mempunyai jarak yang besar untuk digabung. Sebenarnya tidak ada aturan baku dalam menentukan banyak kluster tergantung subjektivitas peneliti. Peneliti juga dapat menggunakan pertimbangan teoritis, konseptual dan praktis.

### **Interpretasi Profil dan Akses Validitas Kluster**

Interpretasi profil kluster meliputi pengkajian mengenai *centroid* yaitu rata-rata nilai obyek yang terdapat dalam kluster pada tiap variabel. Nilai *centroid* memungkinkan untuk menguraikan setiap kluster dengan cara memberi suatu label atau nama. Label suatu kluster juga dapat didasarkan pada manfaat yang akan di cari.

Pengecekan mutu hasil pengklasteran dapat dilakukan dengan analisis kluster yang menggunakan ukuran jarak yang berbeda, menggunakan metode pengklasteran yang berbeda dan membandingkan hasilnya.

## **METODE PENELITIAN**

Penelitian ini merupakan penelitian deskriptif yang bersifat eksploratif dengan variabel adalah macam-macam keluhan kesehatan yang dialami penduduk selama sebulan sebelum survei dilakukan pada tahun 2003. Adapun variabelnya yaitu

1. Panas
2. Sakit Kepala
3. Batuk
4. Pilek
5. Diare
6. Sesak Napas
7. Asma
8. Sakit Gigi
9. Kejang-Kejang
10. Lumpuh
11. Telinga Berair
12. Campak
13. Sakit Kuning
14. Pikun
15. Kecelakaan
16. Lainnya

Data diperoleh dari buku Statistik Potensi Desa 2003 terbitan Badan Pusat Statistik (BPS) tahun 2003. Data tersebut diperoleh BPS dari hasil pengolahan data di tingkat provinsi.

Data dianalisis melalui tahap sebagai berikut

1. Menilai perlunya melakukan transformasi data
2. Analisis kluster
  - a. Metode hirarki
  - b. Metode pengklasteran *K-means*
3. Interpretasi profil kluster
4. Akses validitas kluster

## HASIL DAN PEMBAHASAN

### Pendahuluan

Pada data tersebut tidak terlihat adanya variasi dalam satuan atau tidak terdapat perbedaan yang mencolok. Perbedaan data yang besar akan membuat perhitungan menjadi tidak valid (Santoso, 2004). Oleh karena itu, data tidak perlu ditransformasi.

Pengklasteran dilakukan dengan dua metode yaitu metode hirarki dan metode pengklasteran *K-means*. Metode – metode yang terdapat dalam metode hirarki digunakan sebagai bahan eksplorasi dan perbandingan bagi peneliti. Walaupun metode – metode hirarki tersebut mempunyai fungsi dan peran yang sama tetapi terkadang berbeda dalam hasil pengklasterannya.

### Hasil Uji Asumsi

Penelitian ini menggunakan total sampling yaitu populasi. Dengan demikian asumsi bahwa sampel yang digunakan harus benar benar bisa mewakili populasi telah dipenuhi. Multikolinieritas terjadi bila terdapat dua atau lebih variabel yang saling berkaitan. Untuk mengetahui ada tidaknya multikolinieritas antar variabel dapat dilakukan dengan matrik korelasi (Lampiran 2).

Pada tabel matrik korelasi dapat dilihat bahwa masih terdapat nilai korelasi yang lebih besar dari 0,5. Variabel tersebut adalah batuk dan pilek (0,636), sakit kepala dan lumpuh (0,6155), sakit kepala dan diare (0,59766), panas dan kejang (0,55661).

Dengan menggunakan  $\alpha=0,001$  pada ke empat pasang variabel tersebut diperoleh hasil bahwa  $H_0$  diterima. Artinya, variabel-variabel tersebut *independent* atau tidak terdapat hubungan antar variabel. Dengan Demikian tidak terdapat multikolinieritas.

### Metode Hirarki

Karena data berbentuk matrik dan jumlah obyek tidak sama dengan jumlah variabel maka metode yang digunakan adalah mengukur korelasi antar sepasang obyek pada beberapa variabel.

Nilai matrik *proximity* tergantung dari jarak yang digunakan. Metode ward harus menggunakan jarak kuadrat *euclid* begitu juga dengan metode median dan metode *centroid* (Anonim (2005c) dan output SPSS versi 11.0). Sedangkan metode *linkage* tidak ada sumber yang mengharuskan menggunakan jarak tertentu namun jarak yang paling sering digunakan adalah jarak *euclid*. Nilai matrik *proximity* terlampir untuk jarak *euclid* dan jarak kuadrat *euclid* (Lampiran 3 dan lampiran 4). Semakin kecil nilai pada tabel matrik *proximity* berarti semakin mirip kedua obyek dalam karakteristiknya (keluhan kesehatan masyarakat dalam sebulan pada tahun 2003). Sebaliknya, semakin besar nilai jarak pada tabel matrik *proximity* maka semakin berbeda kedua obyek tersebut.

Tabel 3 Aglomerasi Metode *Single linkage*

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	4,126	0	0	4
2	14	22	5,139	0	0	4
3	6	13	5,485	0	0	6
4	3	14	5,532	1	2	9
5	25	26	5,6	0	0	21
6	6	11	5,68	3	0	8
7	15	19	5,935	0	0	10
8	6	12	6,125	6	0	9
9	3	6	6,711	4	8	10
10	3	15	6,799	9	7	11
11	3	7	6,877	10	0	13
12	4	20	6,961	0	0	14
13	1	3	7,231	0	11	14
14	1	4	7,698	13	12	15
15	1	2	7,715	14	0	16
16	1	10	7,758	15	0	17
17	1	30	7,982	16	0	18
18	1	9	9,18	17	0	20
19	28	29	9,594	0	0	26
20	1	21	9,847	18	0	21
21	1	25	9,86	20	5	22
22	1	23	9,993	21	0	23
23	1	17	10,415	22	0	24
24	1	16	10,419	23	0	25
25	1	24	12,594	24	0	26
26	1	28	14,229	25	19	27
27	1	18	14,634	26	0	28
28	1	27	14,67	27	0	29
29	1	8	51,198	28	0	0

Tabel aglomerasi diatas menunjukkan cara penggabungan kluster tiap tahap pada metode *single linkage* dengan menggunakan jarak *euclid*. Tabel aglomerasi metode lain dapat dilihat pada lampiran 5.

Pada tahap pertama terbentuk satu kluster dengan anggota obyek 3 (Sumatra Barat (Sumbar)) dan obyek 5 (Jambi). Kolom koefisien menyatakan jarak antara Sumatra Barat dan Jambi seperti yang terlihat pada tabel matrik *proximity* pada lampiran 3. Nilai koefisien tersebut didapat dari jarak yang digunakan. Karena proses aglomerasi dimulai dengan dua obyek terdekat maka kedua obyek tersebut adalah yang terdekat dari 30 kombinasi jarak 30 provinsi. Se makin kecil nilai koefisien pada tabel aglomerasi berarti semakin mirip satu obyek dengan obyek yang lain, sebaliknya semakin besar nilai koefisien pada tabel aglomerasi berarti semakin tak mirip satu obyek dengan obyek yang lain.

Kolom *next stage* menyatakan tahapan suatu obyek digabung dengan kluster yang baru saja dibentuk. Pada tahap pertama terlihat angka 4. Hal ini berarti langkah pengklasteran berikutnya dengan melihat langkah 4. Demikian seterusnya sampai langkah terakhir yakni langkah ke 29.

Output metode kluster digambarkan seperti diagram pohon yang dikenal dengan nama dendogram. Dendogram merupakan visualisasi proses pengklasteran yang terjadi. Dendogram masing-masing metode hirarki terlampir (Lampiran 6).

Dendogram dibaca dari kiri ke kanan. Posisi garis pada skala menunjukkan jarak suatu kluster digabung. Skala yang digunakan bukanlah koefisien yang ada pada tabel aglomerasi namun telah dilakukan penskalaan ulang dengan batasan 0 sampai 25. Garis vertikal mewakili kluster yang digabung bersama. Dendogram berguna untuk menunjukkan anggota kluster yang ada jika akan ditentukan berapa kluster yang seharusnya dibentuk. Dari dendogram dapat dilihat bahwa dua tahap terakhir, jarak suatu kluster digabung angkanya besar. Informasi ini berguna untuk memutuskan banyaknya kluster.

Dari dendogram metode *single linkage* dengan jarak *euclid* dapat dilihat bahwa ada empat kluster awal yang terbentuk dengan jarak yang paling dekat yaitu (Sumbar, Jambi, Jawa Timur (Jatim) dan Kalimantan Timur (Kaltim)), (Sumatra Selatan (Sumsel), Yogya dan Jawa Barat (Jabar)), (Banten dan Kalimantan Barat (Kalbar)) dan (Sulawesi Selatan (Sulsel) dan Sulawesi Tenggara (Sultra)). Kluster terakhir yang terbentuk yaitu kluster yang paling jauh jaraknya atau provinsi yang paling tidak mirip dibandingkan provinsi lain. Kluster ini adalah kluster provinsi Lampung kemudian diikuti dengan kluster Nusa Tenggara Timur (NTT) dan Gorontalo dengan jarak sama yang masing-masing membentuk kluster sendiri dengan anggota tunggal yaitu dirinya.

Dendogram metode *complete linkage* dengan jarak *euclid*, awalnya membentuk 5 kluster yaitu (Sumbar dan Jambi), (Sumsel dan Yogya), (Jatim dan Kaltim), (Banten dan Kalbar) dan (Sulsel dan Sultra). Kluster terakhir yang terbentuk yaitu Lampung.

Pada metode *average linkage* juga terbentuk lima kluster pada awal tahapannya, yaitu pada skala 0. Anggota kluster tersebut sama dengan metode *complete linkage* sedangkan kluster yang terakhir terbentuk sama dengan metode *single linkage*.

Dendogram metode ward juga menyatakan bahwa Lampung adalah kluster terakhir yang terbentuk dengan anggota tunggal yaitu dirinya. Pada metode median, kluster yang terakhir terbentuk sama dengan metode lainnya yaitu Lampung pada skala 25 diikuti Gorontalo pada skala 5-10 dan NTT pada skala 0-5. Pada awal tahapannya, metode ini membentuk empat kluster dengan salah satu kluster memuat anggota yang banyak. Pada akhir tahapan metode *centroid*, kluster yang terbentuk sama dengan metode *average linkage*. Kluster awal yang terbentuk ada 5 yaitu (Sumbar, Jambi, Bengkulu, NAD, Sumut, Banten, Kalbar, Jatim, Kaltim, Sumsel, Yogya, Jabar, Jateng, Jakarta dan Babel), (Riau dan Kalteng), (NTB dan Sulut), (Maluku dan Maluku Utara) dan (Sulsel, Sultra dan Kalsel).

Dari dendogram semua metode hirarki dapat dilihat bahwa metode-metode tersebut memberikan hasil yang mirip. Adapun provinsi yang selalu tergabung dalam satu kluster pada awal tahapannya atau provinsi-provinsi yang paling mirip yaitu (Sumbar dan Jambi), (Jatim dan Kaltim), (Sulsel dan Yogya), (Banten dan Kalbar) dan (Sumsel dan Sultra). Sedangkan provinsi yang paling tidak mirip dengan provinsi lain atau kluster yang terakhir terbentuk karena jaraknya paling jauh yaitu provinsi Lampung diikuti NTT dan Gorontalo.

Pada metode *single linkage*, kluster terkecil yang dapat dibentuk adalah 5 kluster. Anggota kluster tersebut adalah (Lampung), (Gorontalo), (NTT), (Maluku dan Maluku Utara) dan (provinsi-provinsi selain 4 kluster sebelumnya).

Metode *complete linkage* dapat membentuk 3, 4 atau 5 kluster. Jika diambil 4 maka anggotanya yaitu adalah (Lampung), (Sulsel, Sultra dan Kalsel), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng, dan Gorontalo) dan (provinsi-provinsi selain 3 kluster sebelumnya). Jika diambil 5 adalah pemecahannya maka anggota kluster tersebut adalah (Lampung), (Sulsel, Sultra dan Kalsel), (NTB, Sulut dan NTT), (Maluku, Maluku Utara, Bali, Sulteng, dan Gorontalo) dan (provinsi-provinsi selain 4 kluster sebelumnya).

Metode *average linkage* juga dapat membentuk 3, 4 atau 5 klaster. jika yang diambil 3 klaster adalah pemecahannya maka anggota klaster tersebut adalah (Lampung), (NTT) dan (provinsi-provinsi selain 2 klaster sebelumnya). Jika diambil 4 klaster maka anggotanya yaitu adalah (Lampung), (NTT), (Gorontalo) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster adalah pemecahannya maka anggota klaster tersebut adalah (Lampung), (NTT), (Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 4 klaster sebelumnya).

Metode median dapat memberikan 4, 5 atau 6 hasil pengklasteran. Jika diambil 4 klaster maka anggota adalah (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster maka anggotanya sama dengan metode *average linkage* yaitu (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel), (NTT) dan (provinsi-provinsi selain 4 klaster sebelumnya). Jika diambil 6 klaster maka anggotanya (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel), (NTT), (Nusa Tenggara Barat (NTB), Sulut, Maluku, Maluku Utara, Bali dan Sulteng) dan (provinsi-provinsi selain 5 klaster sebelumnya).

Dendogram metode ward mirip dengan metode *complete linkage* namun anggota klaster yang terbentuk tidak sama. Jarak yang terbentuk antar pemecahan tiap klaster lebih besar dibandingkan metode lainnya. Metode ini juga dapat membentuk 3, 4 atau 5 klaster. Jika diambil 3 klaster maka anggotanya yaitu (Lampung), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng dan Gorontalo) dan (provinsi-provinsi selain 2 klaster sebelumnya). Jika diambil 4 klaster maka anggotanya yaitu (Lampung), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng dan Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster maka anggotanya sama dengan metode *complete linkage*.

Pada metode *centroid*, jika diambil 3 atau 4 klaster adalah pemecahan yang ideal maka anggotanya sama dengan metode *average linkage*. Jika diambil 5 maka anggotanya sama dengan metode median.

Dari hasil dendogram dapat dilihat bahwa perubahan dari 3 klaster menjadi 4 klaster atau dari 4 klaster menjadi 5 klaster adalah penggabungan atau pemecahan provinsi yang sudah ada, bukan mengacak provinsi-provinsi dari awal.

### **Metode Pengklasteran *K-means***

Bila pengguna analisis klaster tidak mampu secara tegas menentukan berapa nilai k, banyaknya klaster maka disarankan untuk melakukan analisis komponen utama (Anonim, 2003c). Dengan menggunakan Minitab 13 *for window* diperoleh ada 5 komponen utama yang terbentuk.

Seperti telah disebutkan pada bab sebelumnya, bahwa metode ini dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Penulis menginginkan 5 klaster sesuai dengan analisis komponen utama yang telah dilakukan. Nilai *centroid* ditentukan oleh *software* secara acak. Dengan SPSS versi 11.0 *for window* diperoleh :

Tabel 4 *Cluster Membership*

<i>Case Number</i>	<i>Provinsi</i>	<i>Cluster</i>	<i>Distance</i>
1	NAD	4	7,660
2	Sumut	4	8,143
3	Sumbar	4	4,905
4	Riau	4	11,696
5	Jambi	4	6,115
6	Sumsel	4	7,612
7	Bengkulu	4	8,319
8	Lampung	1	,000
9	Babel	4	10,365
10	Jakarta	4	6,921
11	Jabar	4	10,098
12	Jateng	4	8,412
13	Yogya	4	8,677
14	Jatim	4	6,283
15	Banten	4	5,477
16	Bali	3	11,677
17	NTB	3	7,580
18	NTT	3	17,732
19	Kalbar	4	5,990
20	Kalteng	4	12,111
21	Kalsel	5	7,014
22	Kaltim	4	3,651
23	Sulut	3	7,409
24	Sulteng	2	7,335
25	Sulsel	5	3,610
26	Sultra	5	5,220
27	Gorontalo	2	7,335
28	Maluku	3	11,525
29	M/Utara	3	11,554
30	Papua	4	13,204

Hasil output diatas ternyata tidak sama dengan hasil metode-metode hirarki. Jika 5 dipilih sebagai pemecahan hasil klaster yang ideal maka hasil output diatas, 3 klaster diantaranya sama dengan 3 diantara 5 pemecahan klaster metode *complete linkage* atau metode ward. Jika dibandingkan dengan 5 klaster pada metode *average linkage*, metode median atau metode *centroid* yang sama hanya 2 klaster. Sedangkan bila dibandingkan dengan metode *single linkage* hanya 1 klaster yang sama yaitu klaster provinsi Lampung.

#### **Interpretasi Profil dan Akses Validitas Klaster**

Dari hasil pengklasteran metode hirarki dan metode pengklasteran *K-means*, dengan alasan *reliability*, dapat diambil kesimpulan bahwa 5 adalah pemecahan hasil pengklasteran. Hal ini sesuai dengan analisis komponen utama dan hasil metode *complete linkage* dan metode ward.

Oleh karena itu, penulis mengambil kesimpulan bahwa 5 adalah penyelesaian pengklasteran dengan anggota sebagai berikut:

Tabel 5 Anggota kluster

1	2	3	4	5
Lampung	Sulawesi Selatan	Maluku	NTB	NAD
	Sulawesi Tenggara	Maluku Utara	Sulut	Sumut
	Kalimantan Selatan	Bali	NTT	Sumbar
		Sulawesi Tengah		Jambi
		Gorontalo		Sumsel
				Bengkulu
				Babel
				Jakarta
				Jabar
				Jateng
				Yogya
				Jatim
				Banten
				Kalimantan Barat
				Papua
				Kalimantan Tengah
				Riau
				Kalimantan Timur

Nama atau label dapat dicari dengan menguraikan *centroid* masing - masing kluster pada tiap variabel. Penamaan kluster dan pengambilan kesimpulan pengklasteran bersifat subyektif dan tergantung tujuan penelitian. Berikut nilai *centroid* masing-masing kluster dan *centroid* rataan populasi.

Tabel 6 *Centroid*

No	Variabel	<i>Centroid</i> Populasi	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Klaster5
1	Panas	40.2960	31.38	35.1233	50.9140	52.3667	36.6922
2	Sakit Kepala	14.9630	13.45	15.5733	18.7980	19.8000	13.1739
3	Batuk	46.3287	44.59	35.6233	47.5340	53.0967	46.7467
4	Pilek	44.2607	45.84	31.3967	36.1600	51.9667	47.2828
5	Diare	4.9267	3.54	4.9833	5.3860	6.8933	4.5389
6	Sesak Napas	3.5937	2.11	2.6633	4.3940	3.9700	3.5461
7	Asma	3.2433	2.00	2.8467	3.4320	3.2200	3.3300
8	Sakit Gigi	6.8187	5.90	6.0467	8.8360	6.3800	6.5111
9	Kejang	0.9370	0.31	0.9967	1.7100	2.0433	0.5628
10	Lumpuh	0.7173	0.45	0.9433	0.9460	0.8867	0.6028
11	Telinga Berair	0.3210	0.24	0.2633	0.2900	0.5100	0.3122
12	Campak	2.3543	52.00	0.5133	0.6040	0.4867	0.7006
13	Sakit Kuning	0.5777	0.83	0.4900	0.7220	1.1100	0.4494
14	Pikun	1.1633	1.50	1.0100	1.2280	1.1800	1.1494
15	Kecelakaan	0.8470	0.67	1.4500	0.9760	0.4167	0.7922
16	Lainnya	25.3477	29.32	27.8300	22.8260	25.7667	25.3439

Pada klaster pertama (provinsi Lampung) terdapat nilai variabel keluhan kesehatan masyarakat yang mencolok yaitu campak dengan nilai 52,00 dari persentase penduduk yaitu 22,01 yang mempunyai keluhan kesehatan. Nilai ini sangat besar bila dibandingkan dengan *centroid* populasi dan klaster lain. Dibandingkan klaster lain, keluhan kesehatan lainnya, selain yang merupakan penyakit dan kecelakaan juga tinggi dibandingkan *centroid* populasi. Selain 2 variabel tersebut, sakit kuning, pikun dan pilek juga diatas *centroid* populasi. Nilai variabel lain, masih di bawah nilai *centroid* populasi.

Dari ciri-ciri diatas dapat dikatakan bahwa angka kesakitan (persentase penduduk yang mengalami keluhan kesehatan atau kesakitan) masih rendah. Dengan demikian klaster ini dapat diberi nama atau profil, klaster dengan angka kesakitan yang rendah.

Pada klaster 2, nilai *centroid* variabel yang berada diatas *centroid* populasi yaitu sakit kepala, diare, kejang, lumpuh, kecelakaan dan keluhan lainnya. Klaster ini dapat diberi nama klaster dengan angka kesakitan sedang.

Pada klaster 3, banyak terdapat nilai *centroid* variabel diatas *centroid* populasi yaitu ada 12 variabel. Variabel tersebut adalah panas, sakit kepala, batuk, diare, sesak napas, asma, sakit gigi, kejang, lumpuh, sakit kuning, pikun dan kecelakaan. Namun, variabel klaster ini tidak ada yang mencolok seperti nilai *centroid* campak pada klaster 1. Dari sini, dapat dilihat bahwa angka kesakitan pada klaster ini tinggi. Dengan demikian klaster ini dapat diberi nama, klaster dengan angka kesakitan tinggi.

Klaster 4 sama dengan klaster 3. Klaster ini juga mempunyai 12 nilai *centroid* variabel diatas rata-rata yaitu panas, sakit kepala, batuk, pilek, diare, sesak napas, kejang, lumpuh, telinga berair, sakit kuning, pikun dan keluhan lainnya. Namun, nilai-nilai pada variabel ini umumnya lebih tinggi dibandingkan klaster 3 bahkan jika dijumlahkan maka nilainya paling tinggi dibandingkan *centroid* populasi dan klaster lain. Dengan demikian klaster ini dapat diberi nama klaster dengan angka kesakitan yang lebih tinggi.

Angka kesakitan klaster ini lebih rendah dibandingkan klaster lain. Klaster ini hanya mempunyai 3 keluhan kesehatan yang nilainya diatas *centroid* populasi yaitu batuk, pilek dan asma. Selisih nilai batuk dan asma pun tidak jauh berbeda dengan nilai *centroid* populasi. Provinsi-provinsi pada klaster ini mempunyai derajat kesehatan yang secara umum lebih baik dibandingkan provinsi lain. Klaster ini mempunyai anggota yang lebih banyak bila dibandingkan klaster lain yaitu 18 provinsi. Dengan demikian, angka kesakitan bangsa Indonesia pada tahun 2003 masih dapat tergolong rendah. Kegagalan-kegagalan dibidang kesehatan dalam rangka meningkatkan derajat kesehatan masyarakat mungkin disebabkan faktor lain yaitu status gizi, umur harapan hidup dan angka kematian. Dari ciri-ciri diatas, maka klaster 5 dapat diberi nama klaster dengan angka kesakitan rendah sekali.

Akses validitas klaster secara tidak langsung telah dilakukan yaitu dengan melakukan anali sis klaster dengan metode yang berbeda dan membandingkan hasilnya.

## KESIMPULAN DAN SARAN

### Kesimpulan

1. Data persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003 berdasarkan provinsi dan jenis keluhan dapat dibagi menjadi 5 klaster dengan rincian sebagai berikut:
  - i. Klaster 1 dengan profil klaster dengan angka kesakitan rendah. Klaster ini hanya mempunyai satu anggota yaitu provinsi Lampung
  - ii. Klaster 2 dengan profil klaster dengan angka kesakitan sedang. Anggota klaster ini yaitu provinsi Sulawesi Selatan, Sulawesi Tenggara dan Kalimantan Selatan
  - iii. Klaster 3 dengan profil klaster dengan angka kesakitan tinggi. Anggota klaster ini yaitu provinsi Maluku, Maluku Utara, Bali, Sulawesi Tengah dan Gorontalo.

- iv. Klaster 4 dengan profil klaster dengan angka kesakitan lebih tinggi. Anggota klaster ini yaitu provinsi NTB, NTT dan Sulawesi Utara.
  - v. Klaster 5 dengan profil klaster dengan angka kesakitan rendah sekali. Klaster ini mempunyai 18 anggota yaitu provinsi-provinsi selain provinsi yang ada pada 4 klaster sebelumnya. Provinsi tersebut adalah Nangro Aceh Darussalam, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Bangka Belitung, Jakarta, Jawa Barat, Jawa Tengah, Yogya, Jawa Timur, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur dan Papua.
2. Secara umum, angka kesakitan masyarakat Indonesia pada tahun 2003 masih tergolong rendah. Kegagalan - kegagalan dibidang kesehatan, dalam rangka meningkatkan derajat kesehatan guna mensukseskan program pemerintah di bidang kesehatan Indonesia Sehat 2010 mungkin disebabkan faktor status gizi, umur harapan hidup dan angka kematian.
  3. Semua metode pengklasteran memberikan hasil yang mirip. Provinsi Lampung diikuti NTT dan Gorontalo selalu saja menjadi klaster terakhir yang terbentuk dan berbeda bila dibandingkan provinsi lain.
  4. Metode *complete linkage* lebih mirip dengan metode ward. Hasil metode *average linkage* sesuai dengan definisinya yang menggunakan seluruh informasi pada pasangan jarak. Metode metode median mirip dengan metode *centroid*.
  5. Hasil pengklasteran metode pengklasteran *K-means* untuk 5 klaster yang ideal lebih mirip dengan metode *complete linkage* atau metode ward yaitu 3 diantara klaster tersebut memberikan hasil anggota klaster yang sama. Bila dibandingkan dengan metode *average linkage*, metode median atau metode *centroid* yang hanya memberikan 2 klaster dengan anggota yang sama.
  6. Hasil pengklasteran yang lebih baik adalah dengan metode Ward atau metode *complete linkage* karena jarak yang terbentuk antar klaster lebih besar dibandingkan metode lainnya. Hal ini dapat dilihat pada dendogram (lampiran 6). Jarak yang besar menunjukkan klaster yang terbentuk memiliki nilai perbedaan yang besar pula.

## Saran

1. Pemerintah dapat memberikan informasi (promotif) berupa penyuluhan kepada masyarakat dalam upaya mencegah penyakit yang datang terutama penyakit yang paling sering muncul dengan meningkatkan ketahanan diri dan menjaga kebersihan lingkungan (preventif).
2. Hasil masing-masing metode pengklasteran tidak sama persis, oleh karena itu akses validitas pengklasteran memang harus dilakukan.
3. Penamaan atau profil klaster sangat subyektif dan tergantung tujuan penelitian. Untuk dapat memberikan pengertian yang mendalam, diperlukan informasi atau fakta-fakta terbaru dalam bidang yang akan diteliti.

## DARTAR PUSTAKA

- [1]. Anonim, 1991. *Informasi Ringkas Kesehatan*. Depkes RI. Jakarta.
- [2]. Anonim, 2002b. *Profil Kesehatan Provinsi Bengkulu 2001*, Dinas Kesehatan. Bengkulu.
- [3]. Anonim, 2003a. *Statistik Potensi Desa 2003*. CV Nasional. Jakarta.
- [4]. Anonim, 2004b. *Cluster Analysis*. <http://www.statsoft.com/textbook/stcluan.html>.
- [5]. Anonim, 2005b. *Cluster Analysis*. <http://149.170.199.144/multivar/ca.htm>.
- [6]. Anonim, 2005c. *Cluster Analysis*. <http://paleo.cortland.edu/class/stats/LectureNotes/11-cluster.pdf>.
- [7]. Anonim. 2002a. *Bengkulu Sehat 2010*. Dinkes Provinsi Bengkulu, Bengkulu.
- [8]. Anonim. 2003b. *Analisis Peubah Ganda*. Jurusan Statistika IPB, Bogor.
- [9]. Anonim. 2003c. *Modul Praktikum Pelatihan Analisis Multivariat*. Jurusan Statistika, Bogor.

- [10]. Anonim. 2004a. *Indonesia Sehat 2010*. <http://www.depkes.go.id>.
- [11]. Argadiredja, D. 2003. *Program Penanggulangan Kemiskinan Bidang Kesehatan*. <http://www.bppt.go.id/rakorbangnas03/depkes4.pdf>.
- [12]. Dillon, W.R. & M. Golstein. 1984. *Multivariate Analysis Method & Applications*. John Wiley & Sons, inc, Canada.
- [13]. Milne, G.R. 2005. *ClusterAnalysis*. <http://intra.som.umass.edu/georgemilne/powerpoint/cluster%anal%ysis.ppt>
- [14]. Moloe, 1999. ----. <http://www.sehat2010.com>.
- [15]. Notoatmodjo, S. 2002. *Metodologi Penelitian Kesehatan*. Rineka Cipta. Jakarta.
- [16]. Portier, K.M. 2003. *Some Old and New Approaches to Cluster Analysis*. Florida. <http://www.ifasstat.ufl.edu/sta4702/pdf/lecture10.pdf>.
- [17]. Sambamoorthi, N. 2005a. *Hierarchical Cluster Analysis: Some Basic and Algorithms*. <http://www.crmportals.com>.
- [18]. Santoso, S. 2004. *Buku Latihan SPSS Statistik Multivariat*. Elex Media Komputindo, Jakarta.
- [19]. Seber, G.A.F. 1984. *Multivariate Observations*. John Wiley & Sons. New York. USA.
- [20]. Sudjana. 2002. *Metoda Statistika*. Tarsito. Bandung
- [21]. Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Rineka Cipta. Jakarta.
- [22]. Venables, W.N. & B.D. Ripley. 1995. *Modern Applied Statistics with S-Plus*, Springer, Verlag. New York.
- [23]. Wichern, D.W. 2002. *Multivariate Statistical Analysis*. Prentice Hall, New Jersey.