

## PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN FASILITAS KESEHATAN MENGGUNAKAN METODE TWO STEP CLUSTER

<sup>1</sup>Putri Indi Rahayu, <sup>2</sup>Pardomuan Robinson Sihombing

<sup>1</sup>Universitas Padjadjaran, <sup>2</sup>Badan Pusat Statistik

e-mail: robinson@bps.go.id

**Abstract:** Health facilities is the most of basic health to help people obtain health services. Based on data from the Ministry of Health in 2017, provincial health facilities in Indonesia, there are 34 provinces which became the object of research. Two Step Cluster (TSC) is a method designed to overcome the problem of a different measurement scale in this type of continuous and categorical. In this study, TSC is used to classify provinces in Indonesia based health facilities. The results of the three groupings are obtained TSC Group 1 has 4 provinces with the frequency of the number of clinics, the number of general practitioners, the number of midwives, their presence and the presence of physical therapy medical technician. Group 2 has 12 provinces with the frequency of power where physical therapy and the presence of a medical technician. While Group 3 has 18 provinces with the frequency of the number of health centers, the number of clinics, the number of general practitioners, the number of pharmacy, physical therapy where power and presence of a medical technician

**Keyword:** Cluster, Health, Indonesia

**Abstrak:** Fasilitas kesehatan merupakan hal yang paling dasar dari kesehatan untuk membantu masyarakat memperoleh layanan kesehatan. Berdasarkan data dari kementerian kesehatan di tahun 2017, fasilitas kesehatan di provinsi indonesia, terdapat 34 provinsi yang menjadi objek penelitian. *Two Step Cluster* (TSC) adalah metode yang dirancang untuk mengatasi masalah skala pengukuran yang berbeda dalam hal ini bertipe kontinu dan kategorik. Dalam penelitian ini TSC digunakan untuk mengelompokkan provinsi di indonesia berdasarkan fasilitas kesehatan. Hasil dari TSC didapatkan 3 pengelompokkan yaitu Kelompok 1 memiliki 4 provinsi dengan frekuensi jumlah klinik, jumlah dokter umum, jumlah bidan, keberadaan tenaga terapi fisik dan keberadaan teknisi medis. Kelompok 2 memiliki 12 provinsi, dimana untuk frekuensi terapi fisik tidak ada provinsi yang tidak mempunyai tenaga terapi fisik, dan 12 provinsi yang mempunyai tenaga terapi fisik sedangkan untuk frekuensi teknisi medis tidak ada provinsi yang tidak mempunyai tenaga teknisi medis dan 12 provinsi yang mempunyai tenaga teknisi medis. Sedangkan Kelompok 3 memiliki 18 provinsi dengan frekuensi jumlah puskesmas, jumlah klinik, jumlah dokter umum, jumlah tenaga kefarmasian, keberadaan tenaga terapi fisik dan keberadaan teknisi medis.

**Kata Kunci:** Kluster, Kesehatan, Indonesia

## PENDAHULUAN

Kesehatan merupakan keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial, dan ekonomis. Kesehatan sebagai salah satu kebutuhan dasar manusia, menjadi salah satu perhatian bagi setiap pemerintah di dunia. Karenanya ditingkat global, kebijakan di bidang kesehatan menjadi perhatian utama dalam pembangunan manusia. Pada tingkatan yang lebih teknis, penerapan kebijakan kesehatan ditunjukkan dalam berbagai hal yaitu dari fasilitas kesehatan, tenaga kesehatan dan anggaran.

Pembangunan dibidang kesehatan merupakan kegiatan integral dari pembangunan nasional yang bertujuan agar semua lapisan masyarakat memperoleh pelayanan kesehatan secara mudah, merata, dan murah. Dengan adanya upaya tersebut pembangunan di bidang kesehatan masyarakat diharapkan akan tercapai, demikian juga derajat kesehatan masyarakat. Salah satu upaya pembangunan kesehatan adalah pendirian Pusat Kesehatan Masyarakat (Puskesmas) untuk terwujudnya perubahan status kesehatan masyarakat, puskesmas menyelenggarakan upaya kesehatan melalui peningkatan, pencegahan, penyembuhan, dan pemulihan disertai dengan upaya penunjang yang diperlukan (Juliana, Yozza, & Rahmi, 2014)

Salah satu yang mendukung pelayanan puskesmas adalah ketersediaan fasilitas yang memadai di tiap provinsi di Indonesia tetapi faktanya fasilitas kesehatan setiap provinsi sangat beragam dan tidak merata. Dari permasalahan tersebut diharapkan pemerintah dapat mengambil kebijakan untuk menanggulangi permasalahan tersebut.

Pada penelitian ini untuk memudahkan pemerintah dalam mengambil kebijakan dilakukan pengelompokan provinsi di Indonesia berdasarkan fasilitas kesehatan menggunakan two step cluster untuk data campuran dengan skala pengukuran variabel yang berbeda yaitu variabel kategorik dan kontinu. Manfaat penelitian ini memudahkan pemerintah dalam mengambil kebijakan berdasarkan hasil pengelompokan yang didapatkan.

## METODE PENELITIAN

**Analisis Cluster:** Analisis cluster merupakan salah satu teknik dalam analisis multivariat metode independen yang mengelompokkan data. Analisis ini mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis cluster mengelompokkan objek-objek sehingga objek yang paling dekat kesamaannya dengan objek lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain (Johnson & Wichern, 2007). Hasil pengelompokan objek harus memperlihatkan tingginya kehomogenan internal atau dalam pengelompokan dan tingginya keheterogenan eksternal atau antar pengelompokan (Hair, Anderson, Tatham, & Black, 1998).

**Metode Berhierarchy:** Metode pengelompokan hierarchy ditujukan untuk ukuran contoh kecil. Pengelompokan hierarchy menghasilkan seluruh kemungkinan terbentuknya kelompok. Metode pengelompokan hierarchy digunakan apabila banyak kelompok yang akan dibentuk belum diketahui sebelumnya. Pada dasarnya, terdapat dua prosedur pada pengelompokan hierarchy, yaitu penggabungan (*agglomerative*) dan prosedur pembagian (*divide*). Dalam metode hierarchy terdapat beberapa ukuran jarak antar

kelompok, antara lain metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, metode *Ward*, dan metode *Centroid*.

**Metode Tak Berhierarchy:** Pada metode pengelompokan tak berhierarchy, peneliti harus menentukan terlebih dahulu jumlah kelompok yang diinginkan. Salah satu contoh dari metode ini adalah metode *K-means*. Analisis kelompok *K-means* menggunakan ukuran jarak euclidean. Penentuan pusat kelompok merupakan langkah awal pada metode ini. Langkah selanjutnya adalah menentukan kelompok dari tiap objek, yaitu berdasarkan atas kedekatan ukuran jarak euclidean terhadap *mean* dari masing-masing kelompok (Sari, 2006).

**Analisis Two Step Cluster:** Metode *two step cluster* adalah metode yang dirancang untuk mengatasi masalah skala pengukuran yang berbeda dalam hal ini bertipe kontinu dan kategorik. Metode ini masih memiliki kelemahan yaitu sensitif terhadap data yang berupa urutan atau tingkatan, sehingga tidak mampu dalam menangani data ordinal. Apabila terdapat peubah yang bertipe ordinal, maka sebelum dianalisis variabel tersebut harus ditransformasi terlebih dahulu. Variabel yang dijadikan sebagai dasar pengelompokan metode berhierarchy dan tak berhierarchy merupakan variabel dengan skala pengukuran yang sama yaitu kontinu atau kategorik saja. Untuk mengatasi permasalahan yang mempunyai tipe variabel campuran kontinu dan kategorik tersebut telah dikembangkan *two step cluster* yang memungkinkan untuk mengolah data tersebut (Shih, Jheng, & Lai, 2010).

Jarak antara dua kelompok didefinisikan sebagai jarak antar pusat dari masing-masing kelompok tersebut. Pusat dari suatu kelompok adalah vektor dari rata-rata masing-masing variabelnya. Jarak yang digunakan dalam metode *two step cluster* adalah jarak log likelihood (Bacher, Wenzig, & Vogler, 2004).

Prosedur pengelompokan mempunyai dua tahapan pengelompokan awal objek ke dalam subcluster-subcluster kecil dan tahap pengelompokan optimal (Mongi, 2015).

**Tahap pertama (Pembentukan Kelompok Awal):** Tahapan ini dilakukan jika data yang digunakan relatif besar. Yang dilakukan pada tahapan ini adalah pembentukan *Cluster Feature (CF) Tree*. Pada *CF Tree*, data diamati satu persatu secara random. Berdasarkan kriteria jarak, data tersebut ditentukan apakah digabungkan dengan gerombol sebelumnya atau membentuk gerombol baru.

**Tahap kedua (Pembentukan Kelompok Optimal):** Suatu kelompok dikatakan optimal apabila memiliki jarak antar kelompoknya paling jauh dan jarak antar objek dalam kelompok paling dekat. Semakin dekat jarak antar objek maka semakin besar kemiripan antar objek dalam satu kelompok. Dalam penentuan jumlah kelompok optimal, ada dua langkah yang harus dilakukan. Langkah pertama yaitu menghitung *Bayesian Information Criterion (BIC)* untuk tiap-tiap kelompok. Kemudian hasil perhitungan tersebut digunakan untuk menduga jumlah kelompok. Langkah kedua yaitu mencari peningkatan jarak terbesar antara dua kelompok terdekat pada masing-masing tahapan kelompok (Sumertajaya & Erfiani, 2007).

Langkah pertama menghitung BIC, rumus BIC untuk kelompok  $j$  adalah:

$$BIC_k = -2 l_k + r_k \log n \quad (1)$$

dimana:  $l_k$  adalah maksimum log-likelihood dan  $r_k$  jumlah parameter bebas

Hasil  $BIC_k$  merupakan dugaan untuk menentukan jumlah kelompok maksimum (Bacher, Wenzig, & Vogler, 2004). Solusi akan memiliki rasio perubahan BIC dan rasio perubahan ukuran yang besar.

Langkah kedua menghitung perubahan rasio jarak untuk k objek kelompok  $R(k)$ , rumusnya adalah:

$$R(k) = d_{k-1}/d_k \quad (2)$$

dimana:  $d_{k-1}$  adalah jarak jika k objek kelompok digabung menjadi k-1 kelompok. Jarak  $d_k$  diperoleh dari perhitungan:

$$d_k = l_{k-1} - l_k \quad (3)$$

dimana: untuk  $v = k, k-1$

$$l_v = (r_v \log n - BIC_v)/2 \quad (4)$$

Jumlah kelompok optimal diperoleh pada saat ditemukannya perbedaan yang nyata pada rasio perubahan kelompok. Rasio perubahan kelompok dihitung sebagai berikut:

$$R(k_1)/R(k_2) \quad (5)$$

Untuk dua nilai terbesar dari  $R(k)$  ( $k=1,2,\dots,k_{max}$ ) didapatkan dari langkah pertama). Jika rasio perubahan lebih besar dari pada nilai batas  $c_2 = 1.15$  jumlah kelompok ditetapkan sama dengan  $k_1$  selainnya jumlah kelompok sama dengan maksimum  $\{k_1, k_2\}$

**Ukuran Jarak:** Jarak log-likelihood digunakan untuk perbedaan skala pengukuran. Rumus jarak antara dua kelompok i dan s adalah:

$$\begin{aligned} d(i, s) &= \xi_i + \xi_s - \xi_{(i,s)} \\ \xi_i &= -n_i \left( \sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{ijl} \log(\hat{\pi}_{ijl}) \right) \\ \xi_s &= -n_s \left( \sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{sj}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{sjl} \log(\hat{\pi}_{sjl}) \right) \\ \xi_{(i,s)} &= -n_{(i,s)} \left( \sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{(i,s)j}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{(i,s)jl} \log(\hat{\pi}_{(i,s)jl}) \right) \end{aligned} \quad (6)$$

Untuk penyingkatan  $\xi_{(i,s)}$  dituliskan sebagai  $\varepsilon_v$  yang artinya sebagai suatu jenis galat penyimpangan didalam kelompok  $v$  ( $v=i,s,(i,s)$ ).  $\varepsilon_v$  terdiri dari dua komponen keragaman. Bagian pertama adalah  $-n_v \sum \frac{1}{2} \log(\hat{\sigma}_{vj}^2 + \hat{\sigma}_j^2)$  yang mengukur total keragaman dari variabel kontinu  $x_m$  di dalam kelompok  $v$  dan bagian kedua yaitu  $entropy - n_v \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{vjl} \log(\hat{\pi}_{vjl})$  mengukur galat pada variabel kategorik (Fitriani, 2009).

**Sumber Data:** Data yang digunakan dalam penelitian ini adalah data sekunder dari Profil Kesehatan di Indonesia Tahun 2017 dari publikasi online pada website Dinas Kesehatan di Indonesia (Departement Kesehatan Indonesia, 2018).

**Variabel Penelitian:** data yang digunakan dalam penelitian ini adalah data fasilitas kesehatan di 33 provinsi di indonesia dengan variabel sebagai berikut:

$X_1$  = Jumlah Puskesmas

$X_2$  = Jumlah klinik

$X_3$  = Jumlah Dokter Umum

$X_4$  = Jumlah Bidan

$X_5$  = Jumlah Tenaga Kefarmasian

$X_6$  = Keberadaan Tenaga Terapi Fisik

$X_7$  = Keberadaan Teknisi Medis

### Langkah Penelitian

Berdasarkan uraian diatas, tahapan penelitian yang dilakukan sebagai berikut:

- Pengkategorian variabel berdasarkan skala pengukuran
- Pembentukan kelompok awal
- Pembentukan kelompok optimal
- Menentukan jumlah kelompok optimal
- Karakteristik masing-masing kelompok

### HASIL DAN PEMBAHASAN

#### Hasil

Tabel 1. Hasil Pengelompokkan Awal

Jumlah Cluster	Bayesian Criterion (BIC)	Perubahan BIC	Rasio Perubahan BIC	Rasio Perubahan Ukuran Jarak
1	219.508			
2	185.946	-33.562	1.000	1.590
3	180.553	-5.393	0.161	3.579
4	209.539	28.986	-0.864	1.558
5	243.300	33.761	-1.006	1.177
6	278.349	35.049	-1.044	1.013
7	313.488	35.139	-1.047	1.754
8	351.713	38.225	-1.139	1.074
9	390.220	38.507	-1.147	1.314
10	429.637	39.418	-1.174	2.946
11	470.970	41.332	-1.232	1.423
12	512.595	41.625	-1.240	1.023
13	554.235	41.640	-1.241	1.240

Tabel 2. Jumlah Kelompok Optimal

Jumlah Cluster	N	Total (%)
1	4	11.8
2	12	35.3
3	18	52.9
<b>Total</b>	<b>34</b>	<b>100</b>

Tabel 3. Karakteristik Cluster

Cluster 1: (11,8%)	Cluster 2 : (35,3%)	Cluster 3 : (52,9%)
Jumlah Klinik	Keberadaan Tenaga Terapi	Jumlah Puskesmas
Jumlah Dokter Umum	Fisik	Jumlah Klinik
Jumlah Bidan	Keberadaan Teknisi Medis	Jumlah Dokter Umum
Keberadaan Tenaga Terapi Fisik		Jumlah Tenaga Kefarmasian
Keberadaan Teknisi Medis		Keberadaan Tenaga Terapi

## Pembahasan

### Pembentukan Kelompok Awal

Hasil pengelompokan awal dapat dilihat pada Tabel 1. Dari Tabel 1 jika jumlah kelompok yang dibentuk adalah 1 maka nilai BIC adalah 219.508. Kemudian jika dibentuk 2 kelompok maka nilai BIC adalah 185.946. Perubahan BIC dari 1 kelompok ke 2 kelompok adalah  $(185.946 - 219.508) = -33.562$ . Sedangkan perubahan BIC dari 2 kelompok ke kelompok 3 adalah  $(180.553 - 185.946) = -5.393$ . Jika dibentuk 2 kelompok, maka rasio perubahan BIC adalah  $(-33.562 / -33.562) = 1$ . Sedangkan jika dibentuk 3 kelompok, maka rasio perubahan BIC adalah  $(-5.393 / -33.562) = 0.161$ . Rasio perubahan BIC digunakan untuk menentukan dugaan jumlah kelompok maksimal yang akan terbentuk pada tahap awal.

### Pembentukan Kelompok Optimal

Berdasarkan Tabel 1 kelompok optimal yang terbentuk adalah 3 kelompok. Hal ini berdasarkan rasio 2 perubahan jarak terbesar, dimana rasio perubahan ukuran jarak terbesar pertama adalah pada kelompok 3 = 3.579 dan kelompok 10 = 2.946 dengan rasio adalah  $3.579 / 2.946 = 1.214$ . Karena nilai rasio yang diperoleh adalah 1.214 lebih besar dari nilai batas  $c_2 = 1.15$  oleh karena itu dalam kasus ini 3 kelompok merupakan solusi optimal.

### Menentukan Jumlah Kelompok Optimal

Berdasarkan Tabel 2. Jumlah kelompok optimal yang terbentuk adalah 3 kelompok. Pada kelompok 1 beranggotakan 4 provinsi (11,8%), kelompok 2 beranggotakan 12 provinsi (35,3%), dan kelompok beranggotakan 18 provinsi (52,8%). Jumlah provinsi yang paling banyak terdapat pada kelompok 3 dan yang paling sedikit terdapat pada kelompok 1.

### Karakteristik masing-masing kelompok

Berdasarkan Tabel 3. Karakteristik masing-masing cluster sebagai berikut:

- Kelompok 1 memiliki 4 provinsi, dimana untuk frekuensi terapi fisik provinsi Sumatera Utara tidak mempunyai tenaga terapi fisik, dan provinsi Jawa Barat, Jawa Tengah dan Jawa Timur yang mempunyai tenaga terapi fisik sedangkan untuk frekuensi teknisi medis semua provinsi dikluster 1 mempunyai tenaga teknisi medis.
- Kelompok 2 memiliki 12 provinsi, dimana untuk frekuensi terapi fisik semua provinsi dikluster 2 mempunyai tenaga terapi fisik, sedangkan untuk frekuensi teknisi medis semua provinsi mempunyai tenaga teknisi medis yaitu provinsi Riau, Lampung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara dan Sulawesi Selatan.
- Kelompok 3 memiliki 18 provinsi, dimana untuk frekuensi terapi fisik provinsi Aceh, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Kepulauan Bangka Belitung, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Utara,

Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, dan Papua Barat tidak mempunyai tenaga terapi fisik sedangkan untuk frekuensi teknisi medis 2 provinsi yaitu Bengkulu dan Sulawesi Barat yang tidak mempunyai tenaga teknisi medis dan 16 provinsi lainnya mempunyai tenaga teknisi medis

## KESIMPULAN

Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan sebagai berikut: Kelompok 1 memiliki 4 provinsi dengan frekuensi jumlah klinik, jumlah dokter umum, jumlah bidan, keberadaan tenaga terapi fisik dan keberadaan teknisi medis. Kelompok 2 memiliki 12 provinsi dengan frekuensi keberadaan tenaga terapi fisik dan keberadaan teknisi medis. Kelompok 3 memiliki 18 provinsi dengan frekuensi jumlah puskesmas, jumlah klinik, jumlah dokter umum, jumlah tenaga kefarmasian, keberadaan tenaga terapi fisik dan keberadaan teknisi medis.

## DAFTAR PUSTAKA

- Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS Two Step Cluster A First Evaluation*. Departement Kesehatan Indonesia. (2018, Desember 1). Retrieved from <http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-Indonesia-tahun-2017.pdf>
- Fitriani, N. (2009). *Aplikasi Metode Two Step Cluster Pada Ukuran Data Berbeda (Studi Kasus: Data Potensi Desa 2006 untuk Wilayah Jawa Barat)*. Bogor: Skripsi, Institut Pertanian Bogor.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate Data Analysis, Five Edition*. New Jersey: Prentice Hall.
- Johnson, R., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis, Sixth Edition*. New Jersey: Pearson Education, Inc.
- Juliana, Yozza, H., & Rahmi, I. (2014). Penggerombolan Kabupaten/Kota Di Indonesia Berdasarkan Fasilitas Kesehatan Dasar Menggunakan Metode Two Step Cluster. *Jurnal Matematika*, 3(2).
- Mongi, C. (2015). Penggunaan Analisis Two Step Clustering untuk Data Campuran. *Jurnal Matematika*, 4(1).
- Sari, N. (2006). *Perbandingan Ukuran Jarak Log-Likelihood dan Euclidean Pada Pembentukan Gerombol dengan Menggunakan Analisis Two Step Cluster*. Bogor: Skripsi. Institut Pertanian Bogor.
- Shih, Y., Jheng, J., & Lai, L. (2010). A Two Step Method For Clustering Mixed Categorical and Numeric Data. *Tamkang Journal Of Science and Engineering*, 13(1).
- Sumertajaya, I., & Erfiani. (2007). Analisis Gerombol Menggunakan Metode Two Step Cluster (Studi Kasus: Data Potensi Desa Sensus Ekonomi 2003 Wilayah Jawa Barat). *Forum Statistika dan Komputasi*, 12(1).