

# Penggerombolan Daerah 3T di Indonesia Berdasarkan Rasio Tenaga Kesehatan dengan Metode Penggerombolan Berhierarki dan Cluster Ensemble\*

Kesuma Millati<sup>1</sup>, Cici Suhaeni<sup>2‡</sup>, Budi Susetyo<sup>3</sup>

<sup>123</sup>Department of Statistics, IPB University, Indonesia  
<sup>‡</sup>corresponding author: [cici\\_suhaeni@apps.ipb.ac.id](mailto:cici_suhaeni@apps.ipb.ac.id)

Copyright © 2021 Kesuma Millati, Cici Suhaeni, and Budi Susetyo. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Health is a major factor in community development. Inequality on health is most felt by people living in disadvantaged, outermost, and leading areas (3T) because of the difficulty of access to transportation and communication. Effective efforts are needed to achieve the optimal distribution of health services, one of which is by clustering 3T areas based on the ratio of health workers to see which areas are experiencing shortage of health workers and know the adequacy of the number of health workers spread in 3T areas. The object used in this research is 27 provinces 3T region in Indonesia and the applied statistical method is various hierarchical methods and Cluster Ensemble. Based on the results of this study, 3T area is divided into four clusters. The first cluster consists of 22 provinces and has good characteristics because all categories of the variables are in the medium category. The second and the third cluster consists of two provinces. The characteristics of the second cluster are good enough. The characteristics of the third cluster are not been good enough because there is one variable in the low category. The fourth cluster consists of one province and has characteristics that are not been good enough because there are several categories of the variables are in the low category.

**Keywords:** Cluster Analysis, Cluster Ensemble, Health Workers, Link-Based Cluster Ensemble, Underdeveloped Outer and Frontier Areas (3T).

## 1. Pendahuluan

### 1.1 Latar Belakang

Kesehatan merupakan salah satu faktor penting untuk meningkatkan kesejahteraan masyarakat di suatu negara, sehingga kesehatan menjadi indikator utama dalam upaya pemerataan pelayanan kesehatan. Reputasi pelayanan kesehatan di Indonesia kerap kali dinilai buruk oleh masyarakat ekonomi kelas bawah, terlebih kemajuan dan keberhasilan pelayanan kesehatan di Indonesia hanya dapat dirasakan oleh masyarakat ekonomi kelas atas yang tinggal di daerah perkotaan. Saat ini, ketidakmerataan pelayanan kesehatan paling banyak dirasakan oleh masyarakat yang tinggal di daerah tertinggal, terluar, dan terdepan (3T) di Indonesia.

---

\* Received: Sep 2018; Reviewed: May 2021; Published: May 2021

Daerah terdepan dan terluar merupakan daerah-daerah yang berada di garis terdepan dalam hubungan dengan negara lain maupun daerah-daerah terluar, sedangkan daerah tertinggal merupakan daerah kota/kabupaten yang wilayah serta masyarakatnya kurang berkembang dibandingkan dengan daerah lain dalam skala nasional serta daerah yang jauh dari akses, baik transportasi maupun komunikasi. Sulitnya akses transportasi serta komunikasi di daerah 3T menjadi salah satu penyebab ketidakmerataan pembangunan kesehatan. Dalam pembangunan kesehatan, tenaga kesehatan masyarakat merupakan bagian dari sumber daya manusia yang sangat penting perannya guna meningkatkan kesadaran yang lebih tinggi pada pelayanan kesehatan yang bersifat promotif dan preventif (Sukowati and Shinta 2003).

Rasio tenaga kesehatan terhadap jumlah penduduk merupakan indikator guna mengukur ketersediaan tenaga kesehatan untuk mencapai target pembangunan kesehatan tertentu. Berdasarkan Keputusan Menteri Koordinator Bidang Kesejahteraan Rakyat No.54 Tahun 2013 tentang Rencana Pengembangan Tenaga Kesehatan Tahun 2011-2025, target rasio tenaga kesehatan terhadap jumlah penduduk pada tahun 2019 di antaranya: rasio dokter umum 45 per 100.000 penduduk, rasio dokter gigi 13 per 100.000 penduduk, rasio perawat 180 per 100.000 penduduk, dan rasio bidan 120 per 100.000 penduduk (Kemenkes (2016)). Target rasio tenaga kesehatan tersebut harus terpenuhi agar pemerataan tenaga kesehatan di Indonesia dapat tercapai.

Salah satu upaya yang dapat dilakukan agar pemerataan tenaga kesehatan di seluruh daerah dapat tercapai adalah dengan melakukan analisis dan memetakan kondisi daerah 3T yang mengalami kekurangan tenaga kesehatan. Hal ini dilakukan untuk melihat daerah mana saja yang mengalami kekurangan jumlah tenaga kesehatan serta mengetahui tingkat kecukupan jumlah tenaga kesehatan berdasarkan jumlah penduduk yang ada di daerah tersebut. Salah satu metode statistika yang dapat digunakan adalah penggerombolan sehingga untuk menggerombolan daerah 3T berdasarkan karakteristik yang ditentukan digunakan analisis gerombol (*clustering*).

Analisis gerombol adalah alat statistik populer untuk memperoleh kelompok responden, objek, atau kasus yang memiliki kemiripan satu sama lain dan berbeda dari kelompok yang lain (Orme and Johson 2008). Gerombol yang baik memiliki karakteristik yang sama antar objek di dalam gerombol dan karakteristik objek yang berbeda antar gerombol (Hu and Yoo 2004). Penggerombolan yang dilakukan dengan menggunakan beberapa metode akan memberikan hasil yang beragam. Strehl dan Ghosh (2002) memperkenalkan sebuah metode yang digunakan untuk menggabungkan sekumpulan hasil gerombol yang disebut *Cluster Ensemble*. Penelitian yang dilakukan oleh Strehl dan Ghosh (2002) menunjukkan bahwa metode *Cluster Ensemble* mampu meningkatkan kualitas dan kekekaran hasil gerombol. Selain itu, metode ini juga mampu meningkatkan kestabilan hasil gerombol.

Berdasarkan uraian tersebut maka penggerombolan daerah 3T di Indonesia menggunakan metode *Cluster Ensemble* merupakan kajian yang menarik untuk dilakukan. Kemudian, penting juga dilakukan evaluasi hasil penggerombolan metode *Cluster Ensemble* dengan metode berhierarki, yang dalam hal ini merupakan basis input dari *Cluster Ensemble*. Dari evaluasi tersebut dapat diperoleh gambaran pemetaan kondisi kecukupan tenaga kesehatan di setiap gerombol.

## 1.2 Tujuan Penelitian

Tujuan penelitian ini adalah menggerombolkan daerah tertinggal, terluar, dan terdepan (3T) di Indonesia berdasarkan rasio tenaga kesehatan dengan menggunakan metode penggerombolan berhierarki dan *Cluster Ensemble*.

## 2. Tinjauan Pustaka

### 2.1 Tujuan Penelitian

Analisis gerombol merupakan salah satu analisis peubah ganda untuk mengelompokkan objek dengan teknik yang lebih primitif karena tidak ada asumsi yang dibuat mengenai jumlah kelompok atau struktur kelompok. Pengelompokan dilakukan berdasarkan pada sifat kemiripan atau sifat ketidakmiripan antar objek. Objek yang berada dalam kelompok yang sama akan lebih mirip dibandingkan dengan objek antar kelompok. Jarak *Euclid* adalah jarak yang sering digunakan sebagai ukuran kemiripan atau ketidakmiripan antar objek. Jarak *Euclid* didefinisikan sebagai berikut:

$$d_{ab} = \{(\vec{x}_a - \vec{x}_b)^T(\vec{x}_a - \vec{x}_b)\}^{\frac{1}{2}} \quad (1)$$

dengan  $d_{ab}$  = jarak antara objek ke- $a$  dan objek ke- $b$ ;  $\vec{x}_a$  = vektor peubah untuk objek ke- $a$ ;  $\vec{x}_b$  = vektor peubah untuk objek ke- $b$ ,  $a \neq b$ ;

Menurut Johnson dan Winchern (2002), metode penggerombolan terbagi menjadi dua jenis yaitu:

#### Metode penggerombolan berhierarki

Terdapat dua prosedur pada metode berhierarki, yaitu prosedur penggabungan (*aglomeratif*) dan prosedur pemisahan gerombol (*devisive*). Hasil penggerombolan dengan metode berhierarki dapat digambarkan dalam sebuah diagram pohon yang biasa disebut dendogram. Penentuan banyaknya kelompok dapat dilihat dari pemotongan dendogram pada selisih jarak penggabungan terbesar atau dengan melihat tingkat kemiripan dalam kelompok. Penggerombolan berhierarki terdiri dari beberapa metode yaitu (Johnson and Winchern 2002); (Mullner 2013):

##### 1. Pautan Tunggal

Pautan tunggal mendefinisikan jarak dua gerombol sebagai jarak terdekat antara sebuah objek dalam gerombol yang satu dengan sebuah objek dalam gerombol yang lain.

$$d_{(ij)k} = \min(d_{ik}, d_{jk}) \quad (2)$$

dengan  $d_{(ij)k}$  adalah jarak antara gerombol ( $ij$ ) dan gerombol  $k$ ,  $d_{ik}$  dan  $d_{jk}$  adalah jarak antara tetangga terdekat gerombol  $i$  dan  $k$ , serta gerombol  $j$  dan  $k$ .

##### 2. Pautan Lengkap

Pautan lengkap mendefinisikan jarak dua gerombol sebagai jarak terjauh antara sebuah objek dalam gerombol yang satu dengan sebuah objek dalam gerombol yang lain.

$$d_{(ij)k} = \max(d_{ik}, d_{jk}) \quad (3)$$

### 3. Pautan *Centroid*

Pautan *centroid* mendefinisikan jarak dua buah gerombol sebagai jarak Euclidean antara kedua rata-rata (*centroid*) gerombol.

$$d_{(ij)k} = \frac{n_i}{n_i + n_j} d_{ik} + \frac{n_j}{n_i + n_j} d_{jk} - \frac{n_i n_j}{n_i + n_j} d_{ij} \quad (4)$$

dengan  $n_i$  adalah jumlah objek gerombol ke- $i$  dan  $n_j$  adalah jumlah objek gerombol ke- $j$ .

### 4. Pautan Median

Pautan median mendefinisikan jarak dua gerombol sebagai jarak antar median, dan gerombol-gerombol dengan jarak terkecil akan digabungkan.

$$d_{(ij)k} = \frac{d_{ik} + d_{jk}}{2} - \frac{d_{ij}}{4} \quad (5)$$

### 5. Pautan Rataan

Pautan rata-rata mendefinisikan jarak dua gerombol sebagai jarak rata-rata antara sebuah objek dalam gerombol yang satu dengan sebuah objek dalam gerombol yang lain.

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{N_{ij} N_k} \quad (6)$$

dengan  $d_{ab}$  adalah jarak antara objek ke- $a$  dalam gerombol ( $ij$ ) dan objek ke- $b$  dalam gerombol ke  $k$ , dan  $N_{ij}$  dan  $N_k$  adalah jumlah objek dalam gerombol ( $ij$ ) dan  $k$ .

### 6. Metode Ward

Metode ward mendefinisikan jarak dua gerombol sebagai total jumlah kuadrat dua gerombol pada masing masing variabel. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis varians untuk menghitung jarak antar gerombol atau metode ini meminimumkan jumlah kuadrat (ESS).

$$ESS = \sum_{a=1}^n \{(\vec{x}_a - \vec{\bar{x}})^T (\vec{x}_a - \vec{\bar{x}})\}^{\frac{1}{2}} \quad (7)$$

dengan ;  $\vec{x}_a$  = vektor peubah untuk objek ke- $a$ ;  $\vec{\bar{x}}$  = vektor rata-rata objek dalam sebuah gerombol.

## Metode penggerombolan tak-berhierarki

Metode ini biasanya digunakan jika banyaknya gerombol yang akan dibentuk telah ditentukan jumlahnya ataupun dapat ditentukan terlebih dahulu sebagai prosedur penggerombolan dan data yang digunakan berjumlah besar. Salah satu metode penggerombolan tak-berhierarki yang paling sering digunakan adalah metode K-rataan.

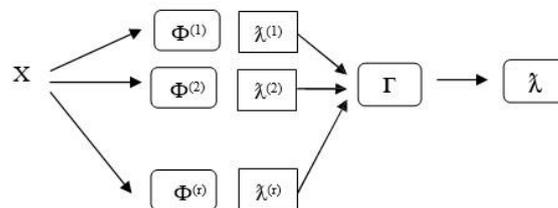
MacQueen dalam Johnson dan Winchern (2002) menyarankan penggunaan K-rataan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam gerombol tertentu berdasarkan rata-rata terdekat. Dalam bentuk yang paling sederhana, proses ini terdiri dari tiga tahap:

1. Bagi objek-objek ke dalam K gerombol awal.
2. Masukkan tiap objek ke suatu gerombol berdasarkan rata-rata terdekat. Jarak biasanya ditentukan dengan menggunakan *Euclidean*. Hitung kembali rata-rata untuk gerombol yang mendapat objek dan yang kehilangan objek.
3. Ulangi langkah 2 sampai tidak ada lagi pemindahan objek antar gerombol.

Sebelum membagi objek ke dalam K gerombol awal (pada langkah a), dapat ditentukan K *centroid* awal dan kemudian dilanjutkan ke langkah b.

## 2.2 Cluster Ensemble

*Cluster Ensemble* adalah salah satu metode yang digunakan untuk menggabungkan berbagai solusi penggerombolan tanpa melihat karakteristik data awal. Fungsi dari menggabungkan berbagai metode penggerombolan ini adalah untuk meningkatkan kualitas dan kekekaran solusi gerombol (Strehl and Gosh 2002).



Gambar 1: Kerangka Cluster Ensemble. X (pengamatan),  $\varphi^q$  = metode penggerombolan ke-q,  $\lambda^{(q)}$  = solusi gerombol metode ke-q,  $\Gamma$  = fungsi *Consensus*, dan  $\lambda$  = solusi akhir.

### Fungsi *Consensus*

Fungsi *Consensus*  $\Gamma$  didefinisikan sebagai fungsi  $N^{n \times r} \rightarrow N^n$  yang memetakan sekumpulan solusi gerombol menjadi satu solusi gabungan (Strehl and Gosh 2002).

$$\Gamma : \{\lambda(q) | q \in \{1, 2, \dots, r\}\} \rightarrow \lambda \tag{8}$$

Tujuan dari fungsi *consensus* adalah menggabungkan solusi gerombol dan mencari solusi gabungan yang memberikan informasi yang mampu mewakili keseluruhan solusi (Orme and Johnson 2008). Fungsi *consensus* memiliki beragam algoritma, salah satunya algoritma *pairwise similarity*. Ide dasar dari algoritma ini adalah menggunakan hubungan kesamaan berpasangan antara semua titik data. Untuk suatu data  $X = x_1, x_2, \dots, x_n$ , akan dibangun *Cluster Ensemble*  $\pi = \pi_1, \pi_2, \dots, \pi_M$  dengan melakukan penggerombolan dasar M ke data X. Lalu, matriks kesamaan berukuran  $N \times N$  dibangun untuk setiap anggota *ensemble*. Input dalam setiap matriks ini merupakan hubungan antara dua titik data. Lebih tepatnya, kesamaan antara dua titik data. Matriks kesamaan ini dibangun menggunakan algoritma *Link-Based Cluster Ensemble* (Iam-on and Garret 2010).

### *Link-Based Cluster Ensemble*

Algoritma *Link-Based Cluster Ensemble* sebagai berikut (Iam-on and Garret 2010):

1. Melakukan pelabelan ulang anggota *ensemble* yaitu jika terdapat kesamaan hasil yang terbentuk, hanya satu hasil penggerombolan yang digunakan sebagai anggota *ensemble*. Anggota *ensemble* merupakan keseluruhan hasil

- yang diperoleh dari berbagai metode penggerombolan.
2. Membentuk *Binary Association Matrix* (BM) yang didapatkan dari hasil pelabelan ulang anggota *ensemble*. *Binary Association Matrix* adalah matriks yang kolom-kolomnya menggambarkan gerombol dari setiap solusi. Matriks ini berukuran  $n \times q$ . Dengan  $n$  merupakan baris yang menggambarkan jumlah objek pengamatan dan  $q$  merupakan kolom yang menggambarkan jumlah label. objek bernilai 1 jika merupakan anggota yang bersesuaian dengan label tersebut dan 0 jika sebaliknya.
  3. Membentuk matriks pembobot (WCT) dengan langkah sebagai berikut:

- (a) Membentuk matriks ( $w$ ) dengan persamaan

$$w_{ij} = \frac{X_{c_i} \cap X_{c_j}}{X_{c_i} \cup X_{c_j}} \quad (9)$$

dengan  $X_{c_i}$  = himpunan objek pada gerombol ke- $i$ ,  $i = 1, 2, \dots, k$ ;  $X_{c_j}$  = himpunan objek pada gerombol ke- $j$ ,  $j = 1, 2, \dots, k$ ,  $i \neq j$ ;

- (b) Membentuk matriks WCT (*Weighted Connected-Triple*) dengan persamaan

$$WCT_{ij}^k = \min(w_{ik}, w_{jk}) \quad (10)$$

$$WCT_{ij} = \sum_{k=1}^q WCT_{ij}^k \quad (11)$$

dengan  $k$  adalah indeks label,  $k = 1, 2, \dots, q$

$$SIM^{WCT}(C_i, C_j) = \frac{WCT_{ij}}{WCT_{max}} \quad (12)$$

dengan  $WCT_{max}$  adalah nilai tertinggi yang berada pada matriks WCT.

4. Membuat matriks kesamaan (CTS).

$$S_m(x_a, x_b) = \begin{cases} 1 & \text{for } C(x_a) = C(x_b) \\ SIM^{WCT}(C(x_a), C(x_b)) \times DC & \text{for lainnya} \end{cases} \quad (13)$$

dengan  $C(x_a)$ ,  $C(x_b)$  adalah label objek ke- $a$  dan label objek ke- $b$  pada matriks BM.  $S_m(x_a, x_b)$  adalah nilai kesamaan antara objek ke- $a$  dengan objek ke- $b$ .  $SIM^{WCT}(C(x_a), C(x_b))$  adalah nilai kesamaan antara solusi gerombol label objek ke- $a$  dengan label objek ke- $b$ . DC adalah tingkat kepercayaan menerima dua objek non-identik sebagai dua objek yang serupa,  $DC \in (0, 1]$ . Berdasarkan rangkaian persamaan tersebut matriks kesamaan dapat diperoleh dengan cara:

$$CTS(x_a, x_b) = \frac{1}{M} \sum_{m=1}^M S_m(x_a, x_b) \quad (14)$$

dengan  $M$  adalah banyaknya gerombol yang terbentuk,  $m = 1, 2, \dots, M$ .

5. Mentransformasi matriks kesamaan menjadi matriks jarak.

$$d_{ab} = 1 - CTS(x_a, x_b) \quad (15)$$

6. Melakukan penggerombolan dengan menggunakan metode pautan tunggal, pautan lengkap, dan pautan rata-rata untuk mendapatkan solusi gerombol akhir.

### 2.3 Evaluasi Hasil Gerombol

Adapun tiga kategori pengukuran validitas yang digunakan yaitu :

### 1. Compactness (CP)

CP mengukur jarak rata-rata antara setiap pasang titik data yang termasuk dalam kelompok yang sama.

$$C = \frac{1}{N} \sum_{k=1}^K \left( \frac{\sum_{x_a, x_b \in C_k} d(x_a, x_b)}{n_k(n_k - 1)/2} \right) \quad (16)$$

dengan  $K$  adalah banyaknya gerombol yang terbentuk,  $n_k$  adalah jumlah data yang termasuk ke dalam gerombol ke- $K$ ,  $d(x_a, x_b)$  adalah jarak antara objek ke- $a$  dan objek ke- $b$ , dan  $N$  adalah jumlah seluruh objek. Semakin kecil nilai indeks CP, maka struktur gerombol yang dihasilkan semakin baik (Iam-on and Garret 2010).

### 2. Davies-Bouldin (DB)

Indeks DB merupakan salah satu metode evaluasi internal yang mengukur evaluasi gerombol pada suatu metode penggerombolan yang didasarkan pada nilai kohesi dan separasi. Indeks DB dirumuskan sebagai berikut:

$$DB = \frac{1}{k} \sum_{i=1}^K R_i \quad (17)$$

dengan  $R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}$

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (18)$$

dengan  $d_{ij} = d(v_i, v_j)$ ,  $v_i$  = nilai rata-rata pada gerombol  $v_i$ ,  $v_j$  = nilai rata-rata pada gerombol  $v_j$ .

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} d(x, v_i) \quad (19)$$

dengan  $|C_i|$  adalah jumlah data yang termasuk ke dalam gerombol ke- $i$ . Semakin kecil nilai pada indeks, maka hasil penggerombolan semakin baik (Iam-on and Garret 2010).

### 3. Rasio $S_w/S_b$

Untuk menilai validitas hasil analisis gerombol secara objektif, dapat digunakan rata-rata simpangan baku dalam gerombol ( $S_w$ ) dan simpangan baku antar gerombol ( $S_b$ ). Rata-rata simpangan baku dalam gerombol ( $S_w$ ) dirumuskan sebagai berikut:

$$S_w = \frac{1}{k} \sum_{k=1}^K |S_k| \quad (20)$$

dengan  $K$  adalah banyaknya gerombol yang terbentuk dan  $S_k$  adalah simpangan baku gerombol ke- $k$

Sedangkan simpangan baku antar gerombol ( $S_b$ ) dirumuskan sebagai berikut:

$$S_b = \left[ (K - 1)^{-1} \sum_{k=1}^K (\bar{X}_k - \bar{X}) \right]^{1/2} \quad (21)$$

dengan  $\bar{X}_k$  adalah rata-rata gerombol ke- $k$  dan  $\bar{X}$  adalah rata-rata keseluruhan gerombol. Metode yang mempunyai rasio  $S_w/S_b$  merupakan metode terbaik (Bunkers and Miller 1996).

## 3. Metodologi

### 3.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder tahun 2017 yang diperoleh dari website Departemen Kesehatan serta BPS masing-masing provinsi. Objek yang digunakan dalam penelitian ini adalah 27 provinsi daerah tertinggal, terluar, dan terdepan (3T) di Indonesia. Daftar objek penelitian diperoleh dari laman [lpdp.qualbase.co.id/pib/wp-content/uploads/2017/05/Daftar-Daerah-3T-2015.pdf](http://lpdp.qualbase.co.id/pib/wp-content/uploads/2017/05/Daftar-Daerah-3T-2015.pdf).

Peubah yang digunakan sebagai dasar penggerombolan adalah 15 peubah rasio tenaga kesehatan di daerah tertinggal, terluar, dan terdepan (3T)/100000 penduduk yang ditampilkan pada Tabel 1.

Tabel 1: Daftar peubah karakteristik

Rasio Tenaga Kesehatan/100.000 penduduk	Kode
Rasio Dokter Spesialis/100.000 penduduk	A
Rasio Dokter Umum/100.000 penduduk	B
Rasio Dokter Gigi Spesialis/100.000 penduduk	C
Rasio Dokter Gigi/100.000 penduduk	D
Rasio Tenaga Psikologi Klinis/100.000 penduduk	E
Rasio Tenaga Keperawatan/100.000 penduduk	F
Rasio Tenaga Kebidanan/100.000 penduduk	G
Rasio Tenaga Kefarmasian/100.000 penduduk	H
Rasio Tenaga Kesmas/100.000 penduduk	I
Rasio Tenaga Kesling/100.000 penduduk	J
Rasio Tenaga Gizi/100.000 penduduk	K
Rasio Tenaga Keterampilan Fisik/100.000 penduduk	L
Rasio Tenaga Keteknisian Medis/100.000 penduduk	M
Rasio Tenaga Teknik Biomedika/100.000 penduduk	N
Rasio Tenaga Penunjang Kesehatan/100.000 penduduk	P

### 3.2 Metode Analisis Data

Tahapan analisis data pada penelitian ini adalah:

1. Melakukan eksplorasi data dengan statistika deskriptif untuk mengetahui gambaran umum rasio tenaga kesehatan pada tahun 2017.
2. Melakukan analisis gerombol dengan menggunakan metode hierarki pautan tunggal, lengkap, rata-rata, *centroid*, median dan *ward*.
3. Melakukan analisis *Cluster Ensemble* dengan menggunakan algoritma *Link-Based Cluster Ensemble*
4. Melakukan evaluasi hasil penggerombolan dari Tahap 2 dan Tahap 3 menggunakan indeks CP, DB, dan rasio  $S_w/S_b$ .
5. Mendeskripsikan hasil gerombol akhir yang diperoleh dari Tahap 4.
6. Melakukan pengkategorian peubah pada gerombol tertentu, pengkategorian tersebut adalah sebagai berikut:

(a) Tinggi (T). Peubah ke-i pada gerombol ke-j dikategorikan tinggi jika:

$$\bar{x}_{ij} > \bar{x}_i + \frac{1}{2} S_i \quad (22)$$

(b) Sedang (S). Peubah ke-i pada gerombol ke-j dikategorikan sedang jika:

$$\bar{x}_i - \frac{1}{2} S_i < \bar{x}_{ij} \leq \bar{x}_i + \frac{1}{2} S_i \quad (23)$$

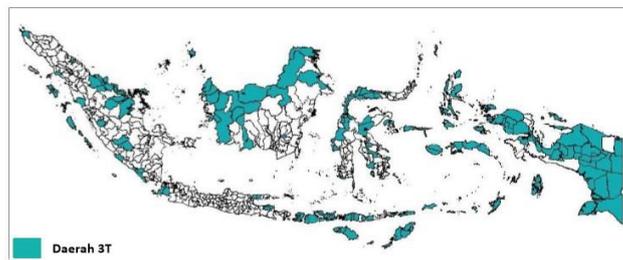
(c) Rendah (R). Peubah ke-i pada gerombol ke-j dikategorikan rendah jika:

$$\bar{x}_{ij} \leq \bar{x}_i + \frac{1}{2} S_i \quad (23)$$

## 4. Hasil dan Pembahasan

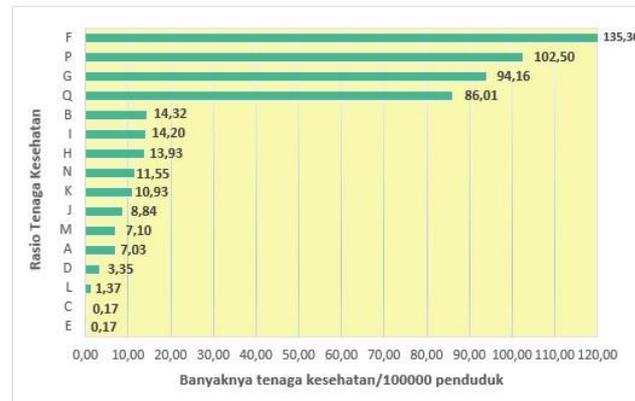
### 4.1 Analisis Deskriptif

Daerah 3T meliputi 143 kabupaten/kota yang berada di 27 provinsi. Jumlah kabupaten/kota daerah 3T yaitu 27,8% dari total kabupaten/kota di Indonesia. Kabupaten/kota yang termasuk dalam daftar daerah tertinggal sebanyak 122 daerah sedangkan yang termasuk dalam daftar daerah terdepan dan terluar sebanyak 43 daerah. Adapun kabupaten/kota yang termasuk ke dalam daftar keduanya sebanyak 22 daerah. Gambaran yang lebih jelas mengenai daftar daerah 3T yang ada di Indonesia dapat dilihat pada Gambar 2.



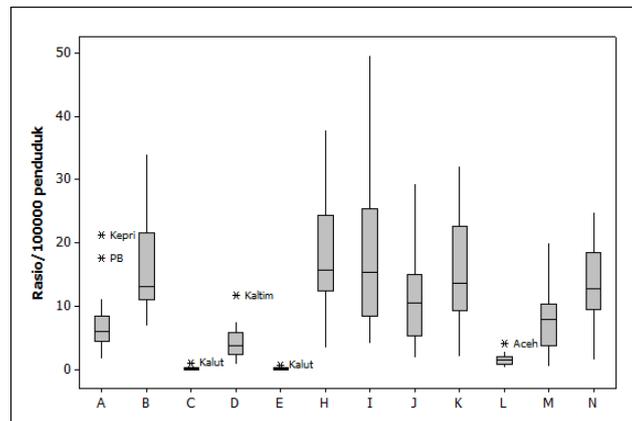
Gambar 2: Peta sebaran daerah 3T di Indonesia

Pada Gambar 3, diketahui bahwa rasio dokter umum di daerah 3T terhadap 100.000 penduduk masih jauh dari target rasio dokter umum di Indonesia pada tahun 2019 yaitu 14,32 per 100.000 penduduk. Begitu pula dengan rasio dokter gigi, rasio perawat, dan rasio bidan di daerah 3T yang berturut-turut sebesar 3,35 per 100.000 penduduk, 135,36 per 100.000 penduduk, dan 94,16 per 100.000 penduduk.



Gambar 3: Rasio tenaga kesehatan daerah 3T terhadap 100.000 penduduk di Indonesia tahun 2017

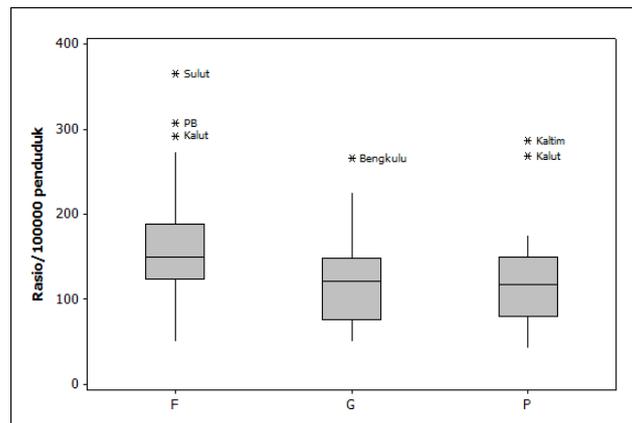
Tenaga kesehatan dengan rasio tertinggi yaitu tenaga keperawatan (135,36 per 100.000 penduduk) dan tenaga kesehatan dengan rasio terendah yaitu tenaga psikologi klinis (0,17 per 100.000 penduduk).



Gambar 4: Diagram *box-plot* peubah A, B, C, D, E, H, I, J, K, L, M, N

Deskripsi masing-masing peubah rasio tenaga kesehatan di daerah 3T secara rinci disajikan menggunakan diagram *box-plot* seperti terlihat pada Gambar 4 dan Gambar 5. Lebar diagram *box-plot* menunjukkan bahwa keragaman peubah tinggi, artinya karakteristik peubah tiap daerah 3T heterogen. Pada Gambar 4 terlihat bahwa peubah dokter umum (B), tenaga kefarmasian (H), tenaga kesehatan masyarakat (I), dan tenaga gizi (K) memiliki ragam yang besar dibandingkan dengan peubah lainnya. Peubah dengan ragam terbesar yaitu tenaga kesehatan masyarakat (I) dan peubah dengan ragam terkecil yaitu peubah dokter gigi spesialis (C), tenaga psikologi klinis (E), dan tenaga keterampilan fisik (L). Dari gambar tersebut juga terlihat bahwa pada beberapa peubah terdapat daerah 3T yang memiliki nilai di luar batas atas *box-plot*. Pada peubah dokter spesialis (A), daerah Kepulauan Riau dan Papua Barat memiliki nilai di luar batas atas jika dibandingkan dengan daerah lainnya. Kalimantan Utara memiliki nilai di luar batas atas pada peubah dokter gigi spesialis (C) dan tenaga psikologi klinis (E), Kalimantan Timur memiliki nilai di luar batas atas pada peubah

dokter gigi (D) dan Aceh memiliki nilai di luar batas atas pada peubah tenaga keterampilan fisik (L).



Gambar 5: Diagram *box-plot* peubah F, G, P

Pada Gambar 5 terlihat bahwa peubah tenaga keperawatan (F), tenaga kebidanan (G), dan tenaga penunjang kesehatan (P) memiliki ragam yang besar dan hampir sama. Daerah Sulawesi Utara, Papua Barat, dan Kalimantan Utara memiliki nilai di luar batas atas pada peubah tenaga keperawatan (F), Bengkulu memiliki nilai di luar batas atas pada peubah tenaga kebidanan (G), dan Kalimantan Timur serta Kalimantan Utara memiliki nilai di luar batas atas pada peubah tenaga penunjang kesehatan (P).

#### 4.2 Pembentukan Anggota *Ensemble*

Tahapan awal dalam melakukan penggerombolan menggunakan *Cluster Ensemble* adalah membentuk sekumpulan hasil gerombol yang diperoleh dari berbagai metode berbeda sebagai anggota *ensemble*. Pada penelitian ini, anggota *ensemble* dibentuk dari hasil penggerombolan menggunakan enam metode hierarki, yaitu metode pautan tunggal, pautan lengkap, pautan rata-rata, pautan median, pautan *centroid*, dan metode *ward*. Hasil dari keenam metode ini digambarkan dalam bentuk dendrogram dimana jumlah gerombol beserta anggota gerombolnya diperoleh dari pemotongan dendrogram pada selisih jarak penggabungan terbesar (Rafikasari 2016). Pemotongan dendrogram dapat juga dilakukan berdasarkan subjektivitas peneliti. Pada penelitian ini, jarak *Euclid* digunakan sebagai ukuran kemiripan atau ketidakmiripan antar objek. Jarak *Euclid* dapat digunakan apabila setiap peubah memiliki satuan pengukuran yang sama. Hal ini disebabkan agar nilai setiap peubah bisa dijumlahkan pada perhitungan nilai jarak. Peubah pada penelitian ini sudah memiliki satuan yang sama sehingga tidak perlu dilakukan transformasi ke dalam bentuk baku untuk menghilangkan satuan pengukuran. Hasil penggerombolan pada tahap awal, terdapat beberapa metode yang memberikan hasil yang sama yaitu:

1. Metode hierarki pautan tunggal dan pautan median.
2. Metode hierarki pautan rata-rata dan pautan *centroid*.

Tidak semua hasil penggerombolan yang diperoleh digunakan sebagai anggota *ensemble* karena dari beberapa hasil yang sama, hanya akan diambil salah satu hasil.

Dengan demikian, hasil gerombol yang digunakan sebagai anggota *ensemble* hanya berjumlah 4 hasil. Tahapan selanjutnya yaitu mentransformasi anggota *ensemble* menjadi *Binary-Association Matrix* (BM). Matriks BM yang diperoleh selanjutnya digunakan untuk mendapatkan gerombol akhir. Ukuran kesamaan antar provinsi pada tahap ini menggunakan algoritma *Link-Based Cluster Ensemble*. Matriks kesamaan yang telah terbentuk kemudian ditransformasi menjadi matriks jarak. Setelah matriks jarak diperoleh kemudian dilakukan penggerombolan dengan metode pautan tunggal, pautan lengkap, dan pautan rataan untuk memperoleh hasil penggerombolan akhir. Dendrogram yang diperoleh pada penggerombolan akhir memiliki hasil yang hampir sama. Pemotongan dendrogram pada tahap ini didasarkan pada unsur subjektivitas. Hasil akhir dari pemotongan dendrogram membentuk empat gerombol, karena dari empat gerombol yang terbentuk terlihat jelas persamaan karakteristik dari setiap gerombolnya. Hasil penggerombolan menggunakan *Cluster Ensemble* dapat dilihat pada Tabel 2. Gerombol pertama terdiri dari satu provinsi, gerombol kedua terdiri 20 provinsi, gerombol ketiga terdiri dari dua provinsi, dan gerombol keempat terdiri dari empat provinsi.

Tabel 2: Hasil gerombol *cluster ensemble*

Gerombol	Jumlah	Anggota Gerombol
Gerombol 1	1	Maluku.
Gerombol 2	20	Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Lampung, Kepulauan Riau, Jawa Timur, Banten, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua.
Gerombol 3	2	Aceh, Bengkulu.
Gerombol 4	4	Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Papua Barat.

### 4.3 Evaluasi Hasil Penggerombolan

Evaluasi hasil penggerombolan dilakukan untuk melihat metode mana yang mampu memberikan hasil penggerombolan yang terbaik. Metode yang dibandingkan pada penelitian ini adalah *Cluster Ensemble* dengan berbagai jumlah gerombol, pautan rataan, pautan tunggal, pautan lengkap, dan metode ward. Hasil dari penggerombolan dievaluasi menggunakan indeks CP, DB dan rasio  $S_w/S_b$ . Nilai indeks validitas tersebut dapat dilihat pada Tabel 3. Nilai indeks CP, DB, dan rasio  $S_w/S_b$  yang diperoleh dari masing-masing metode menunjukkan hasil yang berbeda. Metode *Cluster Ensemble* memiliki indeks CP, DB, dan rasio  $S_w/S_b$  sebesar 91.8934, 0.4829, dan 0.8766. Metode pautan lengkap memiliki nilai indeks CP terkecil yaitu sebesar 81.7762, sedangkan nilai indeks DB dan rasio  $S_w/S_b$  terkecil dimiliki oleh

metode pautan tunggal masing-masing sebesar 0.3791 dan 0.7763. Berdasarkan hasil tersebut, *Cluster Ensemble* bukan merupakan hasil terbaik karena nilai indeks dari CP, DB, dan rasio  $S_w/S_b$  yang dihasilkan bukan merupakan yang terkecil. Hasil evaluasi menunjukkan bahwa metode pautan tunggal memberikan hasil yang lebih tepat dalam penggerombolan daerah 3T berdasarkan rasio tenaga kesehatan karena nilai indeks DB dan rasio  $S_w/S_b$  yang dihasilkan merupakan yang terkecil.

Tabel 3: Nilai evaluasi hasil gerombol

Metode	Jumlah	CP	DB	$S_w/S_b$
CE dgn PT, PL, PR	4	91.8934	0.4829	0.8766
Pautan Rataan	4	90.4519	0.5579	0.9960
Pautan Tunggal	4	96.2182	<b>0.3791</b>	<b>0.7763</b>
Pautan Lengkap	4	<b>81.7762</b>	0.7879	1.0479
Metode Ward	4	82.3862	0.8959	1.1355

Berdasarkan hasil evaluasi tersebut pula penggerombolan akhir pada penelitian ini menggunakan metode pautan tunggal dengan empat gerombol. Hasil penggerombolan menggunakan metode pautan tunggal dapat dilihat pada Tabel 4.

Tabel 4: Hasil Gerombol metode pautan tunggal

Gerombol	Jumlah	Anggota Gerombol
Gerombol 1	22	Aceh, Bengkulu, Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Lampung, Kepulauan Riau, Jawa Timur, Banten, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua.
Gerombol 2	2	Kalimantan Timur, Kalimantan Utara.
Gerombol 3	2	Sulawesi Utara, Papua Barat.
Gerombol 4	2	Maluku.

#### 4.4 Karakteristik Gerombol Akhir

Hasil akhir dari Cluster Ensemble membagi daerah 3T ke dalam empat gerombol. Gerombol akhir yang telah terbentuk kemudian dideskripsikan berdasarkan karakteristik masing-masing gerombol. Nilai rata-rata tiap gerombol pada tiap peubah dihitung untuk mengkategorikan gerombol berdasarkan karakteristiknya. Nilai rata-rata dan kategori tiap gerombol pada tiap peubah dapat dilihat pada Tabel 4. Rataan tersebut dikategorikan berdasarkan tiga kategori yaitu rendah (R), sedang (S), dan tinggi (T) dengan membandingkan nilai rata-rata dan simpangan baku tiap gerombol pada tiap peubah dengan nilai rata-rata dan simpangan baku keseluruhan daerah 3T tiap peubah.

Tabel 5: Rataan dan kategori peubah tiap gerombol

Rataan Gerombol Peubah				
	1	2	3	4
A	6.627 (S)	9.483 (T)	12.890 (T)	<b>4.163 (R)</b>
B	14.813 (S)	31.441 (T)	26.849 (T)	<b>12.490 (R)</b>
C	0.153 (S)	0.541 (T)	0.226 (S)	0.088 (S)
D	3.852 (S)	9.660 (T)	3.665 (S)	2.569 (R)
E	0.153 (S)	0.360 (T)	<b>0.000(R)</b>	0.354 (T)
F	137.718 (S)	283.009 (T)	336.192 (T)	238.829 (T)
G	119.002 (S)	149.614 (T)	93.577 (S)	95.673 (S)
H	15.494 (S)	33.116 (T)	25.208 (T)	17.097 (S)
I	17.955 (S)	23.949 (S)	21.733 (S)	24.006 (S)
J	8.694 (S)	17.819 (T)	24.444 (T)	23.475 (T)
K	13.013 (S)	20.569 (T)	25.540 (T)	26.310 (T)
L	1.468 (S)	2.658 (T)	1.749 (S)	1.240 (S)
M	7.973 (S)	5.743 (S)	8.938 (S)	<b>3.809 (R)</b>
N	12.611 (S)	24.192 (T)	17.568 (T)	<b>10.010 (R)</b>
P	104.175 (S)	277.562 (T)	162.962 (T)	<b>68.831 (R)</b>

##### Karakteristik Gerombol Satu

Provinsi yang merupakan anggota dari gerombol satu adalah Aceh, Bengkulu, Sumatera Barat, Riau, Sumatera Selatan, Kepulauan Riau, Jawa Timur, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sumatera Utara, Lampung, Banten, Sulawesi Selatan, Sulawesi Barat, Maluku Utara dan Papua. Berdasarkan hasil yang diperoleh dari Tabel 5, gerombol satu merupakan gerombol yang memiliki karakteristik dimana kategori semua peubah rasio tenaga kesehatannya sedang.

##### Karakteristik Gerombol Dua

Provinsi yang merupakan anggota dari gerombol dua adalah Provinsi Kalimantan Timur, Kalimantan Utara. Berdasarkan hasil yang diperoleh dari Tabel 5, gerombol dua merupakan gerombol yang memiliki karakteristik dimana kategori peubah rasio dokter spesialis (A), rasio dokter umum (B), rasio dokter gigi spesialis (C), rasio dokter gigi (D), rasio tenaga psikologi klinis (E), rasio tenaga keperawatan (F), rasio tenaga

kebidanan (G), rasio tenaga kefarmasian (H), rasio tenaga kesehatan lingkungan (J), rasio tenaga gizi (K), rasio tenaga keterampilan fisik (L), rasio tenaga teknik (N), rasio tenaga penunjang kesehatan (P) yang tinggi, sedangkan kategori peubah rasio tenaga kesehatan masyarakat (I), dan rasio tenaga keteknisian medis (M) yang sedang, serta tidak ada kategori peubah rasio tenaga kesehatannya yang rendah.

### Karakteristik Gerombol Tiga

Provinsi yang merupakan anggota dari gerombol tiga adalah Provinsi Sulawesi Utara dan Papua Barat. Berdasarkan hasil yang diperoleh dari Tabel 5, gerombol tiga merupakan gerombol yang memiliki karakteristik dimana kategori peubah rasio dokter spesialis (A), rasio dokter umum (B), rasio tenaga keperawatan (F), rasio tenaga kefarmasian (H), rasio tenaga kesehatan lingkungan (J), rasio tenaga gizi (K), rasio tenaga teknik biomedika (N), rasio tenaga penunjang kesehatan (P) yang tinggi, sedangkan kategori peubah rasio dokter gigi spesialis (C), rasio dokter gigi (D), rasio tenaga kebidanan (G), rasio tenaga kesehatan masyarakat (I), rasio tenaga keterampilan fisik (L), rasio tenaga keteknisian medis (M) yang sedang, serta kategori peubah rasio tenaga psikologi klinis (E) yang rendah.

### Karakteristik Gerombol Empat

Provinsi yang merupakan anggota dari gerombol empat adalah Provinsi Maluku. Berdasarkan hasil yang diperoleh dari Tabel 5, gerombol empat merupakan gerombol yang memiliki karakteristik dimana kategori peubah rasio psikologi klinis (E), rasio tenaga keperawatan (F), rasio tenaga kesehatan lingkungan (J), rasio tenaga gizi (K) yang tinggi, sedangkan kategori peubah rasio dokter gigi spesialis (C), rasio tenaga kebidanan (G), rasio tenaga kefarmasian (H), rasio tenaga kesehatan masyarakat (I), dan rasio tenaga keterampilan fisik (L) yang sedang, serta kategori peubah rasio dokter spesialis (A), rasio dokter umum (B), rasio dokter gigi (D), rasio tenaga keteknisian medis (M), rasio tenaga teknik biomedika (N), dan rasio tenaga penunjang kesehatan (P) yang rendah.

Setiap peubah karakteristik telah diwakili oleh tiap gerombol. Deskripsi penggerombolan tersebut akan memudahkan para peneliti untuk melihat daerah mana saja yang mengalami kekurangan tenaga kesehatan serta mengetahui tingkat kecukupan tenaga kesehatan berdasarkan jumlah penduduk yang ada di daerah tersebut. Hal ini merupakan salah satu upaya agar pemerataan tenaga kesehatan di seluruh daerah dapat tercapai. Gambaran yang lebih jelas mengenai hasil gerombol akhir yang terbentuk berdasarkan rasio tenaga kesehatan di daerah 3T pada provinsi di Indonesia dapat dilihat pada Gambar 6.



Gambar 6: Peta hasil penggerombolan daerah 3T berdasarkan rasio tenaga kesehatan

## 5. Simpulan

### 5.1 Simpulan

Berdasarkan evaluasi hasil penggerombolan antara metode penggerombolan berhierarki dengan metode *Cluster Ensemble* diperoleh hasil bahwa metode pautan tunggal menghasilkan nilai indeks DB dan rasio  $S_w/S_b$  yang paling baik. Hasil penggerombolan menggunakan metode pautan tunggal membagi daerah 3T menjadi empat gerombol. Gerombol pertama merupakan gerombol dengan provinsi terbanyak yang terdiri dari 22 provinsi sebagai anggota gerombolnya. Gerombol pertama memiliki karakteristik yang sudah baik karena semua kategori peubah rasio tenaga kesehatannya tidak ada yang rendah. Gerombol kedua terdiri dari dua provinsi yaitu Kalimantan Timur dan Kalimantan Utara. Karakteristik yang dimiliki gerombol kedua sudah baik karena tidak ada peubah yang berada pada kategori rendah. Gerombol ketiga terdiri dari dua provinsi yaitu Sulawesi Utara dan Papua Barat. Gerombol ketiga memiliki karakteristik yang sudah baik namun ada satu kategori peubah rasio tenaga kesehatannya yang masih rendah. Gerombol keempat terdiri dari satu provinsi yaitu Provinsi Maluku. Berdasarkan hasil pengkategorian, gerombol keempat memiliki karakteristik yang belum baik karena ada beberapa kategori peubah rasio tenaga kesehatannya yang masih rendah.

### Daftar Pustaka

- Bunkers, M. and J. Miller (1996). Definition of climate regions in the northern plains using an objective cluster modification technique. *Journal of Climate* **9**: 130–146.
- Hu, X. and I. Yoo (2004). *Cluster Ensemble and Its Applications in Gene Expression Analysis*. Philadelphia: Drexel University.
- Iam-on, N. and S. Garret (2010). Linkclue: a matlab package for link-based cluster ensemble. *Journal of Statistical Software* **36**(9): 1–36.
- Johnson, R. and D. Winchurn (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Kemenkes (2016). *Profil Kesehatan Indonesia Tahun 2016*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Mullner, D. (2013). fastcluster: fast hierarchical, agglomerative clustering routines for R and python. *Journal of Statistical Software* **53**(9): 1–18.
- Orme, B. and R. Johnson (2008). *Improving K-Means Cluster Analysis: Ensemble Analysis Instead Of Highest Reproducibility Replicates*. Sawtooth Software.
- Rafikasari, E. (2016). Pengelompokan kabupaten/kota di Jawa Timur berdasarkan indikator pendidikan tahun 2013 menggunakan analisis hierarchial cluster. *Dinamika Penelitian* **16**(2): 247–262.
- Strehl, A. and J. Gosh (2002). *A Knowledge Reuse Framework for Combining Partitionings*. Austin: The University of Texas.

Sukowati, S. and Shinta (2003). Peran tenaga kesehatan masyarakat dalam mengubah perilaku masyarakat menuju hidup bersih dan sehat. *Media Litbang Kesehatan* **13**(2): 31–37.