

Perbandingan Hasil Pengelompokan menggunakan Analisis *Cluster* Berhirarki, *K-Means Cluster*, dan *Cluster Ensemble* (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)

Cici Suhaeni¹, Anang Kurnia², Ristiyanti³

^{1,2}Dosen, Institut Pertanian Bogor
Jl. Meranti Wing 22 Level 4 Kampus IPB Dramaga Bogor 16680 (Telp/fax (0251) 8624535;
e-mail: cici_suhaeni@apps.ipb.ac.id

³Alumni Institut Pertanian Bogor
Jl. Meranti Wing 22 Level 4 Kampus IPB Dramaga Bogor 16680 (Telp/fax (0251) 8624535;
e-mail: ristiyanti83@gmail.com

Abstract—Clustering is an activity in the field of research that is widely used nowadays. Especially in the era of big data. Many methods developed for this purpose. This study compared the results of clustering using hierarchical clustering method, k-means cluster, and cluster ensemble on clustering of provinces in Indonesia based on maternal health service indicator. The results of the research indicated that cluster ensemble is the most appropriate method in clustering the provinces. There was 3 (three) clusters obtained by the method.

Keyword: cluster analysis, cluster ensemble, hierarchical cluster, k-means cluster .

Intisari—Pengelompokan merupakan kegiatan di bidang riset yang banyak digunakan hingga saat ini. Terlebih di era big data seperti sekarang. Banyak metode yang berkembang untuk keperluan tersebut. Penelitian ini membandingkan hasil pengelompokan menggunakan metode cluster hierarki, *k-means cluster*, dan *cluster ensemble* pada pengelompokan provinsi di Indonesia berdasarkan indikator pelayanan kesehatan ibu hamil. Hasil analisis menunjukkan bahwa cluster ensemble merupakan metode yang paling tepat dalam mengelompokkan provinsi-provinsi tersebut. Cluster yang dihasilkan adalah 3 (tiga) cluster.

Kata Kunci: analisis cluster, cluster ensemble, cluster hierarki, k-means cluster.

I. PENDAHULUAN

Analisis cluster saat ini semakin berkembang pesat seiring dengan kemajuan teknologi dan informasi yang melahirkan data besar (*big data*). Pemanfaatan analisis cluster pada berbagai bidang adalah untuk pengelompokan individu yang memiliki karakteristik serupa di dalam satu kelompok, dan berbeda

karateristik dengan kelompok lain. Dengan terbentuknya kelompok-kelompok demikian, identifikasi karakteristik kelompok mudah dilakukan sehingga diperoleh gambaran karakteristik setiap kelompok. Hal ini akan sangat berarti pada saat kita berhadapan dengan data yang banyak, terlebih lagi di era *big data* ini.

Banyak metode *clustering* yang telah dikembangkan oleh para pakar dan telah banyak pula diterapkan pada berbagai bidang. Selain metode konvensional, yaitu analisis cluster berhirarki dan non hirarki, salah satu metode yang hingga saat ini banyak dikembangkan adalah metode *cluster ensemble*. Metode ini diperkenalkan oleh Strehl dan Gosh (2002). Ide dasar dari *cluster ensemble* adalah menggabungkan sekumpulan hasil gerombol dari metode-metode lain. Menurut penelitian Strehl dan Gosh tersebut, *cluster ensemble* dapat memberikan hasil penggerombolan yang lebih berkualitas.

Berkembangnya metode *cluster* hingga saat ini disebabkan oleh banyaknya bidang kehidupan yang membutuhkan analisis *cluster* untuk pengelompokan objek. Salah satu bidang yang juga sering membutuhkannya adalah bidang kesehatan. Misalnya pengelompokan daerah berdasarkan indikator pelayanan kesehatan ibu hamil, penting dilakukan

guna melihat daerah-daerah mana saja yang membentuk kelompok dengan tingkat pelayanan yang sudah memadai dan yang belum memadai. Informasi ini dapat menjadi dasar bagi pemerintah dalam menentukan prioritas kebijakan.

Menariknya, berbeda metode analisis cluster yang digunakan akan berakibat pula pada kemungkinan terjadinya perbedaan hasil pengelompokan. Oleh karena itu, membandingkan hasil dari beberapa metode pengelompokan menjadi suatu yang penting dan lebih bermanfaat.

Salah satu penerapan dari analisis cluster adalah pada bidang kesehatan. Pengelompokan daerah berdasarkan indikator pelayanan kesehatan ibu hamil merupakan isu yang menarik untuk dikaji. Hal ini disebabkan karena pentingnya memberikan pelayanan kesehatan terhadap ibu hamil dengan maksimal. Hasil Survei Penduduk Antar Sensus (SUPAS) tahun 2015, menyebutkan bahwa angka kematian ibu sebesar 305 kematian per 100000 Kelahiran Hidup (KH) dari seluruh wilayah Indonesia (Kemenkes RI 2015). Ini merupakan kondisi yang sangat memprihatinkan, sehingga diperlukan upaya yang handal dalam pemberian layanan kesehatan ibu hamil di setiap daerah dengan tepat. Salah satu upaya yang dapat dilakukan adalah dengan mengelompokkan daerah-daerah di Indonesia agar prioritas pelayanan kesehatan dapat diutamakan pada kelompok daerah yang paling buruk.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk membandingkan hasil pengelompokan provinsi di Indonesia berdasarkan indikator pelayanan kesehatan ibu hamil menggunakan metode berhierarki, *k-means*, dan *cluster ensemble*.

II. TINJAUAN PUSTAKA

A. Analisis Cluster

Analisis cluster merupakan metode dengan analisis peubah ganda untuk mengelompokkan n objek ke

dalam m cluster ($m \leq n$) berdasarkan karakteristiknya. 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:

$$d_{ij} = [(x_i - x_j)' (x_i - x_j)]^{1/2}$$

dengan d_{ij} adalah jarak antara objek ke- i dan ke- j , \mathbf{x}_i adalah vektor peubah untuk objek ke- i , dan \mathbf{x}_j adalah vektor peubah untuk objek ke- j (Johnson & Winchern 2002). Metode pengklasteran terbagi menjadi dua jenis, yaitu:

1. Metode berhierarki

Metode berhierarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan cluster yang diinginkan belum diketahui banyaknya (Mattjik & Sumertajaya (2011)). Terdapat dua prosedur pada metode berhierarki, yaitu prosedur aglomeratif dan prosedur divisive (Johnson & Winchern 2002). Sedangkan beberapa ukuran kemiripan atau ketidakmiripan antar cluster dapat dilihat dengan menggunakan Pautan Tunggal, Pautan Lengkap, Pautan Centroid, Pautan Rataan, Pautan Median dan Ward. Hasil pengklasteran dengan metode berhierarki dapat digambarkan dalam sebuah diagram pohon yang biasa disebut dendrogram. Menurut Mattjik & Sumertajaya (2011), banyaknya cluster yang terbentuk ditentukan dari dendrogram yang terjadi dan tergantung subyektivitas peneliti. Namun demikian pemisahan cluster biasanya ditentukan berdasarkan jarak penggabungan terbesar.

Berikut penjelasan mengenai beberapa pautan dalam metode pengklasteran berhierarki (Johnson & Winchern 2002; Rachmatin 2014):

a. Pautan Tunggal

Jarak dua cluster diukur dengan jarak terdekat antara sebuah objek dalam cluster yang satu dengan sebuah objek dalam cluster yang lain.

$$d(uv)w = \min(d_{uw}, d_{vw})$$

dimana $d(uv)w$ adalah jarak antara cluster (UV) dan cluster W, d_{uw} dan d_{vw} adalah jarak antara tetangga terdekat cluster U dan W, serta cluster V dan W.

b. Pautan Lengkap

Jarak dua cluster diukur dengan jarak terjauh antara sebuah objek dalam cluster yang satu dengan sebuah objek dalam cluster yang lain.

$$d(uv)w = \max(d_{uw}, d_{vw})$$

c. Pautan Centroid

Jarak antara dua buah cluster diukur sebagai jarak *Euclidian* antara kedua rata-rata (*centroid*) cluster.

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

d. Pautan Median

Jarak antar cluster didefinisikan sebagai jarak antar median, dan cluster-cluster dengan jarak terkecil akan digabungkan.

e. Pautan Rataan

Jarak antara dua cluster diukur dengan jarak rata-rata antara sebuah objek dalam cluster yang satu dengan sebuah objek dalam cluster yang lain.

$$d(uv)w = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w}$$

dimana d_{ik} adalah jarak antara objek ke-i dalam cluster (UV) dan objek ke-k dalam cluster ke W, dan $N_{(uv)}$ dan N_w adalah jumlah objek dalam cluster (UV) dan W.

f. Metode Ward

Jarak antar dua cluster adalah total jumlah kuadrat dua cluster pada masing-masing variabel. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis variansi untuk menghitung jarak antar cluster atau metode ini meminimumkan jumlah kuadrat (ESS).

$$ESS = \sum_{j=1}^n x_j^2 - \frac{1}{n} (\sum_{j=1}^n x_j)^2$$

2. Metode Non Hierarki

Menurut Johnson & Wicern (2002), metode non hierarki digunakan untuk pengelompokan objek dimana

banyaknya cluster yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur penggerombolan. Kemudian, metode ini dapat diterapkan pada data yang lebih besar dibandingkan metode hirarki.

Secara ringkas, langkah-langkah pengklasteran menggunakan metode non hierarki adalah sebagai berikut (Johnson & Wicern 2002):

1. Bagi objek-objek tersebut 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.

B. Cluster Ensemble

Cluster Ensemble diperkenalkan oleh Strehl dan Gosh (2002), yaitu sebuah metode yang digunakan untuk menggabungkan sekumpulan solusi cluster. Metode ini memiliki keunggulan dibanding metode pengklasteran lainnya, yakni mampu meningkatkan kualitas dan kekekaran solusi cluster. Tantangan untuk mendapatkan solusi cluster dengan kualitas yang baik dan adanya keragaman solusi cluster yang dihasilkan dari metode yang berbeda merupakan motivasi dikembangkannya *Cluster Ensemble*.

Pengklasteran pada *Cluster Ensemble* dilakukan dengan menggabungkan berbagai solusi dari berbagai metode pengklasteran hingga diperoleh satu pengklasteran akhir yang lebih baik. Input yang dibutuhkan adalah solusi pengklasteran yang telah diperoleh dengan menggunakan berbagai hasil pengklasteran tanpa melihat karakteristik data awal. Secara umum, pengklasteran objek dengan metode *Cluster Ensemble* dilakukan dalam dua tahap menurut Lam-on dan Garret (2010), yaitu:

1. Membentuk anggota *ensemble* yang anggotanya adalah solusi dari berbagai metode pengklasteran yang berbeda.
2. Mengombinasikan seluruh anggota *ensemble* untuk memperoleh satu solusi akhir yang dinamakan fungsi *Consensus*.

Fungsi *Consensus* didefinisikan sebagai fungsi yang memetakan sekumpulan solusi cluster menjadi solusi gabungan. Fungsi ini 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 pengklasteran 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 $x_i, x_j \in X$. Matriks kesamaan ini dibangun menggunakan algoritma *Link-Based Cluster Ensemble* (Iam-on dan Garret 2010).

Hasil pengklasteran akhir akan dibentuk menggunakan matriks kesamaan. Matriks kesamaan dibuat dengan algoritma *Link-Based Cluster Ensemble* (Iam-on dan Garret 2010). Algoritma *Link-Based Cluster Ensemble* diantaranya:

1. Melakukan pelabelan ulang anggota *ensemble* yaitu jika terdapat kesamaan solusi yang terbentuk hanya satu solusi yang digunakan sebagai anggota *ensemble*.
2. Membentuk *Binary Association Matrix* (BM) yang didapatkan dari hasil pelabelan ulang anggota *ensemble*. *Binary Association Matrix* adalah matriks yang kolom-kolomnya menggambarkan cluster 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 dengan langkah sebagai berikut:

- a. Membentuk matriks (w) dengan persamaan

$$w_{ij} = \frac{|X_{C_i} \cap X_{C_j}|}{|X_{C_i} \cup X_{C_j}|}$$

Keterangan:

X_{C_i} = himpunan objek pada cluster ke- i , $i = 1, 2, \dots, k$.

X_{C_j} = himpunan objek pada cluster ke- j , $j = 1, 2, \dots, k, i \neq j$.

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

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

$$WCT_{ij} = \sum_{k=1}^q WCT_{ij}^k$$

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

$$Sim^{WCT}(i, j) = \frac{WCT_{ij}}{WCT_{max}}$$

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

- c. Membuat matriks kesamaan.

$$S_m(x_i, x_j) = \begin{cases} 1 & \text{jika } C(x_i) = C(x_j) \\ Sim^{WCT}(C(x_i), C(x_j)) & \text{DC lainnya} \end{cases}$$

dengan $C(x_i), C(x_j)$ adalah label objek ke- i dan label objek ke- j pada matriks BM. $S_m(x_i, x_j)$ adalah nilai kesamaan antara objek ke- i dengan objek ke- j . $Sim^{WCT}(C(x_i), C(x_j))$ adalah nilai kesamaan antara solusi cluster label objek ke- i dengan label objek ke- j . 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:

$$CTS(x_i, x_j) = \frac{1}{M} \sum_{m=1}^M S_m(x_i, x_j)$$

dengan M adalah banyaknya cluster yang terbentuk.

- d. Mentransformasi matriks kesamaan menjadi matriks jarak

$$d_{ij} = 1 - CTS(x_i, x_j)$$

- e. Melakukan pengklasteran dengan menggunakan metode pautan tunggal, pautan lengkap, dan pautan rata-rata untuk mendapatkan solusi cluster akhir.

C. Evaluasi Hasil Clustering

Setelah diperoleh hasil dari pengklasteran, dapat dinilai kualitas struktur data dari cluster yang terbentuk menggunakan berbagai jenis ukuran validitas (Iam-on & Garret 2010). Adapun 2 (dua) kategori pengukuran validitas yang digunakan yaitu:

1. Compactness (CP)

CP merupakan salah satu kategori pengukuran yang paling umum digunakan. CP mengukur jarak rata-rata antara setiap pasang titik data yang termasuk dalam kelompok yang sama.

$$CP = \frac{1}{N} \sum_{k=1}^K n_k \left(\frac{\sum_{x_i, x_j \in C_k} d(x_i, x_j)}{n_k(n_k-1)/2} \right)$$

dengan K adalah banyaknya cluster yang terbentuk, n_k adalah jumlah data yang termasuk ke dalam cluster ke- K , $d(x_i, x_j)$ adalah jarak antara objek ke- i dan objek ke- j , dan N adalah jumlah seluruh objek. Semakin kecil nilai indeks *Compactness*, maka struktur cluster yang dihasilkan semakin baik.

2. Davies-Bouldin (DB)

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

$$DB = \frac{1}{k} \sum_{i=1}^k R_i$$

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

$$R_{ij} = \frac{S_i + S_j}{d_{ij}}$$

dengan $d_{ij} = d(v_i, v_j)$

v_i adalah nilai rata-rata pada cluster C_i

v_j adalah nilai rata-rata pada cluster C_j

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

dengan $|C_i|$ adalah jumlah data yang termasuk ke dalam cluster ke- i . Skema pengklasteran Davies Bouldin yang optimal adalah yang memiliki nilai paling kecil. Semakin kecil nilai pada indeks, maka hasil pengklasteran semakin baik.

III. METODOLOGI

A. Data

Data yang digunakan pada penelitian ini adalah data indikator pelayanan kesehatan ibu hamil dari seluruh provinsi se-Indonesia. Data tersebut diperoleh dari Kementerian Kesehatan Republik Indonesia tahun 2015, yang meliputi 34 provinsi dan 7 variabel berskala numerik. Daftar variabel ditampilkan pada Tabel 1.

B. Analisis Data

Tahap-tahap analisis data pada penelitian ini adalah sebagai berikut:

- 1) Melakukan eksplorasi data untuk melihat gambaran umum pelayanan kesehatan ibu hamil di Indonesia.
- 2) Melakukan analisis cluster berhirarki (pautan tunggal, pautan lengkap, pautan median, pautan rata-rata, pautan *centroid*, dan *ward*) dan *k-means cluster*. Hasil dari pengklasteran tersebut merupakan anggota ensemble bagi analisis cluster ensemble.
- 3) Membuat tabulasi anggota *ensemble*. Tabulasi ini merupakan solusi cluster yang dihasilkan dari setiap metode.
- 4) Membentuk matriks kesamaan dengan algoritma *Link-Based Cluster Ensemble*.
- 5) Mentransformasi matriks kesamaan yang

terbentuk menjadi matriks jarak.

$$d_{ij} = 1 - CTS(x_i, x_j)$$

- 6) Melakukan pengklasteran berdasarkan matriks jarak pada langkah 5 (lima) menggunakan metode pautan lengkap mendapatkan solusi cluster akhir.
- 7) Melakukan evaluasi hasil pengclusteringan *Cluster Ensemble*.
- 8) Mendeskripsikan masing-masing cluster berdasarkan karakteristik peubahnya.

Tabel 1. Daftar Variabel Penelitian

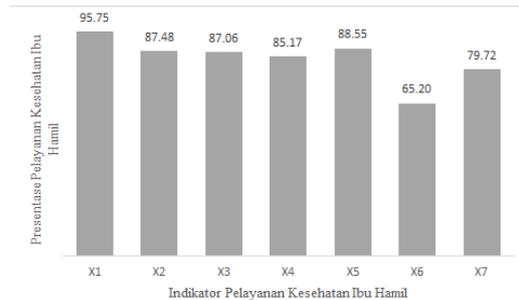
Peubah	Keterangan
X1	Persentase ibu hamil memperoleh pelayanan pertama kali oleh tenaga kesehatan (K1)
X2	Persentase ibu hamil memperoleh pelayanan minimal empat kali oleh tenaga kesehatan (K4)
X3	Persentase ibu nifas melakukan kunjungan nifas 3 kali
X4	Persentase ibu hamil mendapatkan 90 tablet tambah darah
X5	Persentase ibu bersalin mendapatkan pertolongan persalinan oleh tenaga kesehatan
X6	Persentase ibu hamil mendapatkan imunisasi toksoid (TT2+) pada ibu hamil
X7	Persentase ibu bersalin mendapatkan pertolongan persalinan oleh tenaga kesehatan di Fasyankes

IV. HASIL DAN PEMBAHASAN

A. Hasil Eksplorasi Data

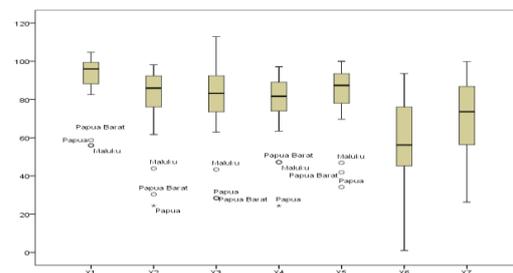
Gambaran pelayanan kesehatan ibu hamil untuk setiap variabel ditampilkan pada Gambar 1. Pelayanan kesehatan yang paling banyak diperoleh ibu hamil di Indonesia adalah K1 (X1), yaitu sebanyak 95.75%. Hal tersebut menunjukkan bahwa ibu hamil yang melakukan pelayanan kesehatan ibu hamil pertama kali oleh tenaga kesehatan sudah baik. Sebanyak 95.75% ibu hamil telah mengerti mengenai pentingnya pelayanan kesehatan yang diberikan oleh tenaga kesehatan untuk perkembangan kehamilan mereka. Selanjutnya secara berturut-turut peubah dengan presentase tertinggi pada indikator pelayanan kesehatan ibu hamil yaitu jumlah ibu bersalin mendapatkan pertolongan persalinan oleh tenaga kesehatan (X5) sebesar 88.55% dan jumlah ibu hamil mendapatkan pelayanan K4 (X2) sebesar 87.48%.

Sedangkan peubah terendah yaitu jumlah ibu hamil mendapatkan imunisasi tetanus toksoid pada ibu hamil (X6) sebesar 65.20%. Hal tersebut menunjukkan bahwa jumlah ibu hamil yang mendapatkan imunisasi toksoid masih sangat rendah sehingga perlu adanya informasi yang lebih jelas mengenai pentingnya imunisasi tetanus toksoid untuk ibu hamil.



Gambar 1. Gambaran Pelayanan Kesehatan Ibu Hamil.

Kemudian, melalui diagram kotak garis (Gambar 2) ditampilkan informasi tentang gambaran masing-masing variabel untuk semua provinsi. Variabel X2, X3, X4, X5, X6, dan X7 memiliki ragam yang besar. Sedangkan variabel X1 memiliki ragam yang kecil. Variabel X6 memiliki ragam terbesar dibandingkan variabel lainnya dan variabel dengan ragam terkecil yaitu X1. Terlihat bahwa dari beberapa variabel terdapat pencilan. Provinsi Papua Barat, Papua, dan Maluku merupakan pencilan untuk beberapa variabel yaitu X1, X2, X3, X4, dan X5 dengan nilai yang jauh lebih kecil dibandingkan dengan provinsi lainnya.

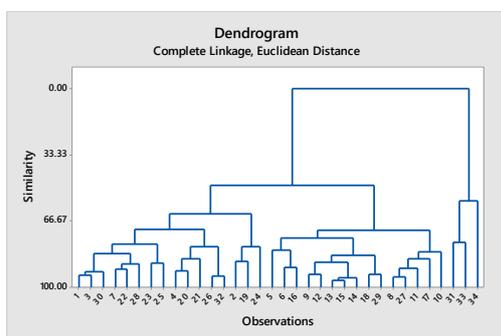


Gambar 2. Sebaran provinsi di Indonesia untuk setiap variabel

B. Perbandingan Hasil Pengelompokan

Pengelompokan provinsi-provinsi di Indonesia berdasarkan indikator pelayanan kesehatan ibu hamil ini dilakukan menggunakan analisis cluster berhirarki terlebih dahulu. Berdasarkan dendrogram dari semua

pautan yang digunakan, diperoleh kemiripan banyaknya cluster yang terbentuk, yaitu antara 2 (dua) atau 3(cluster). Untuk memberikan hasil yang lebih memuaskan, dicobakan lagi pengklasteran hierarki dengan 4 (empat) cluster. Banyaknya cluster yang terbentuk melalui cluster berhierarki ini, selanjutnya juga dijadikan dasar penentuan inisial banyaknya cluster pada metode *k-means*. Dendrogram salah satu pautan, yaitu hasil pautan lengkap ditampilkan pada Gambar 3 sebagai gambaran mengenai banyaknya cluster yang terbentuk.



Gambar 3. Dendrogram Hasil Pengklasteran Menggunakan Complete Linkage

Selanjutnya, analisis *cluster ensemble* dilakukan menggunakan solusi dari analisis cluster hierarki dan non hierarki. Pengelompokan dengan metode cluster hierarki dilakukan menggunakan pautan tunggal, pautan lengkap, pautan rata-rata, pautan median, pautan *centroid*, dan ward. Solusi gerombol dari metode-metode tersebut selanjutnya disebut sebagai anggota *ensemble*.

Dari 7 (tujuh) metode pengklasteran, diperoleh hasil yang sama untuk beberapa metode. Berikut rinciannya:

1. Solusi 2 cluster : pautan tunggal, pautan lengkap, pautan rata-rata, pautan median, dan pautan *centroid*.
2. Solusi 3 cluster : pautan tunggal, pautan rata-rata, pautan median, dan pautan *centroid*.
3. Solusi 4 cluster : pautan tunggal, pautan median, dan pautan *centroid*.

Dari solusi gerombol yang sama tersebut hanya diambil salah satu wakilnya untuk membentuk anggota

ensemble. Oleh karena itu, solusi gerombol yang digunakan sebagai anggota *ensemble* berjumlah 12 solusi.

Hasil pengklasteran menggunakan *cluster ensemble* diperoleh hasil 3 cluster. Selanjutnya, hasil cluster metode hierarki dan *k-means* dipilih hasil 3 cluster untuk dibandingkan dengan metode *cluster ensemble*. Hasil pengelompokan dari metode hierarki, *k-means*, dan *cluster ensemble* dibandingkan menggunakan indeks CP dan indeks DB. Hasil perbandingannya ditampilkan pada Tabel 1.

Tabel 1. Hasil Evaluasi Ketiga Metode Cluster

Metode	CP	DB
<i>Cluster Ensemble</i>	35.95	0.75
Berhierarki	50.88	0.92
K-Rataan	42.81	1.49

Kriteria evaluasi kebaikan cluster yaitu metode yang mendapatkan nilai CP dan DB terkecil merupakan metode dengan hasil cluster yang lebih tepat. Pada Tabel 1 terlihat bahwa *Cluster Ensemble* memiliki nilai CP dan DB terkecil. Oleh karena itu, dapat disimpulkan bahwa dalam pengelompokan provinsi-provinsi di Indonesia berdasarkan indikator pelayanan kesehatan ibu hamil, metode yang memberikan hasil lebih tepat adalah *Cluster Ensemble*.

C. Deskripsi Hasil *Cluster Ensemble*

Deskripsi hasil pengklasteran dilakukan dengan melihat rata-rata setiap variabel pada setiap cluster seperti yang ditampilkan pada Tabel 2. Cluster yang memiliki rata-rata variabel tertinggi dan di atas rata-rata variabel untuk Indonesia (rata-rata dari 34 provinsi) diidentifikasi sebagai cluster terbaik, yaitu provinsi-provinsi yang memiliki pelayanan kesehatan ibu hamil terbaik bila dibandingkan dengan cluster lainnya. Sedangkan, cluster yang memiliki rata-rata variabel yang terendah dan di bawah rata-rata Indonesia, diidentifikasi sebagai cluster yang belum baik dalam hal pelayanan kesehatan ibu hamil.

Cluster pertama terdiri dari 16 provinsi yang didominasi oleh provinsi-provinsi di Pulau Sumatera, Jawa dan sebagian di Pulau Sulawesi. Anggota cluster pertama yaitu Provinsi Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Sulawesi Utara, Sulawesi Selatan, dan Gorontalo. Nilai rata-rata untuk semua indikator pelayanan kesehatan ibu hamil pada cluster pertama lebih tinggi dibanding dengan nilai rata-rata provinsi di Indonesia dan juga paling tinggi bila dibandingkan dengan rata-rata cluster lainnya. Oleh karena itu, cluster pertama dapat dikatakan sebagai cluster yang memiliki karakteristik pelayanan kesehatan ibu hamil yang paling baik dibanding cluster kedua dan ketiga.

Tabel 2. Rata-rata variabel pada setiap cluster

Variabel	Cluster			Indonesia
	1	2	3	
X1	99.27	91.04	56.88	95.75
X2	92.16	79.12	32.91	87.48
X3	91.74	76.13	33.41	87.06
X4	87.95	76.79	39.6	85.17
X5	92.9	81.63	40.98	88.55
X6	72.04	42.88	36.66	65.2
X7	84.59	63.11	29.43	79.72

Anggota cluster kedua terdiri dari provinsi yang cukup menyebar di berbagai pulau di Indonesia kecuali provinsi yang terletak di wilayah timur. Cluster kedua terdiri dari 15 provinsi yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat, dan Maluku Utara. Cluster kedua dapat dikatakan cluster yang memiliki karakteristik pelayanan kesehatan ibu hamil yang cukup baik, karena nilai rata-rata semua indikator pelayanan kesehatan ibu hamil di cluster kedua masih relatif tidak terlalu jauh dari nilai rata-rata Indonesia dan nilai-nilainya lebih tinggi dari rata-rata cluster ketiga.

Cluster ketiga terdiri dari provinsi yang secara geografis terletak di wilayah timur Indonesia, yaitu Provinsi Maluku, Papua Barat, dan Papua. Ketiga provinsi tersebut merupakan pencilan berdasarkan Gambar 2. Ketiga provinsi yang merupakan pencilan ini membentuk cluster sendiri dikarenakan nilai ketiga provinsi tersebut jauh dari provinsi lainnya. Dari nilai rata-rata setiap variabel dapat diidentifikasi bahwa cluster ketiga ini memiliki tingkat pelayanan kesehatan ibu hamil yang belum baik.

V. KESIMPULAN

Metode yang lebih tepat dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator pelayanan kesehatan ibu hamil adalah *Cluster Ensemble*. Hasil pengelompokan yang diperoleh adalah 3 (tiga) cluster. Cluster pertama terdiri dari 16 provinsi yang memiliki karakteristik pelayanan kesehatan ibu hamil terbaik. Cluster kedua terdiri dari 15 provinsi yang memiliki karakteristik pelayanan kesehatan ibu hamil cukup baik. Cluster ketiga terdiri dari 3 provinsi yang secara geografis terletak di wilayah timur Indonesia, yang memiliki karakteristik pelayanan kesehatan ibu hamil yang belum.

DAFTAR PUSTAKA

- [Kemenkes] Kementerian Kesehatan Republik Indonesia. 2015. *Profil Kesehatan nesia Tahun 2015*. Jakarta : Kementerian Kesehatan Republik Indonesia.
- Strehl A, Gosh J. 2012. A Knowledge Reuse Framework for Combining Partitionings. *The Journal of Machine learning Research*. 3(1):583-6
- Mattjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Wibawa GNA, Hadi AF, editor. Bogor (ID): IPB Press.
- Johson RA, Winchern DW. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Rachmatin D. 2014. Aplikasi metode-metode *agglomerative* dalam analisis klaster pada data tingkat polusi udara. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*. 3(2):133-149
- Iam-on N, Garret S. 2010. LinkCluE: A MATLAB Package for Link-Based Cluster Ensemble. *Journal of Statistical Software*. 36(9):1-3