



Available online at : <http://bit.ly/InfoTekJar>

# InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



## Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward

Nurissaidah Ulinnuh , Rafika Veriani

Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Ahmad Yani No.117, Jemur Wonosari, Kec. Wonocolo, Kota SBY, Jawa Timur 60237

### KEYWORDS

Penyakit, Complete Linkage, Average Linkage, Ward, Euclidean, Squared Euclidean

### CORRESPONDENCE

Phone: +6285707304153

E-mail: rafikaveriani.rv@gmail.com

### A B S T R A K

Penyakit adalah salah satu indikator dalam indeks pembangunan manusia bidang kesehatan. Mengingat bahwa pembangunan bidang kesehatan di Indonesia sedang mengalami beban ganda dimana penyakit menular masih menjadi masalah yang belum dapat diselesaikan, dan masih terdapat penyakit menular yang awalnya masih mampu dikendalikan kini muncul kembali. Hal ini seharusnya mendapatkan perhatian lebih khususnya bagi Dinas Kesehatan maupun Kementerian Kesehatan Republik Indonesia mengenai penyebaran penyakit menular ataupun penyakit tidak menular. Salah satu upaya yang dapat dilakukan adalah dengan membentuk suatu pengelompokan provinsi dalam suatu kelompok yang memiliki karakteristik yang sama dengan maksud memberikan informasi terkait dengan kesehatan pada masing-masing provinsi. Pada penelitian ini bertujuan untuk membentuk suatu *cluster* provinsi di Indonesia berdasarkan variabel jenis penyakit menggunakan metode *Complete Linkage*, *Average Linkage*, dan *Ward*. Ukuran jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean* dan *Squared Euclidean*, dan untuk menentukan cara kerja metode yang terbaik dengan melihat dari nilai simpangan baku dalam kelompok ( $S_w$ ) yang minimum, simpangan baku antar kelompok ( $S_b$ ) yang maksimum, dan rasio  $S_w$  terhadap  $S_b$  yang minimum. Hasil analisis *cluster* yang terbaik adalah dengan metode Ward menggunakan 6 *cluster* dengan nilai  $S_w$  sebesar 0,18405,  $S_b$  sebesar 2,12284 serta rasio  $S_w$  terhadap  $S_b$  sebesar 0,08670.

### PENDAHULUAN

Kesehatan adalah salah satu indikator indeks pembangunan manusia. Mengingat bahwa pembangunan bidang kesehatan di Indonesia sedang dihadapkan pada beban ganda dimana penyakit menular masih menjadi masalah yang belum dapat diselesaikan, dan masih terdapat penyakit menular yang awalnya masih mampu dikendalikan kini muncul kembali dengan penyebaran tanpa mengenal batas daerah maupun batas negara. Di sisi lainnya kasus penyakit tidak menular juga semakin meningkat, akibat gaya hidup serta penyakit-penyakit degeneratif. Pemicu terjadinya kecenderungan ini adalah berubahnya gaya hidup, perubahan tersebut mempengaruhi penentuan makanan yang condong lebih menyukai makanan yang siap saji. Padahal makanan tersebut banyak mengandung lemak dan garam yang tinggi, namun kandungan seratinya rendah [1]. Bukan hanya kecenderungan gaya hidup, hal-hal lain yang dapat menghambat kehidupan khususnya di bidang kesehatan diantaranya adalah jumlah penduduk yang semakin meningkat hal ini dapat menyebabkan keterbatasan akan lahan khususnya pada lahan pemukiman, berkurangnya ketersediaan

air bersih, menurunnya kualitas air karena limbah, dan meningkatnya pencemaran udara akibat dari pemanfaatan transportasi kota.

Sepuluh tahun terakhir ini Indonesia juga menghadapi masalah *triple burden diseases*. Maksudnya Indonesia masih dilanda penyakit infeksi, penyakit tidak menular (PTM) dan penyakit yang seharusnya sudah teratasi selain itu penyakit menular juga masih menjadi masalah. Hal tersebut ditandai dengan sering terjadi KLB (kejadian luar biasa) beberapa penyakit menular tertentu, munculnya kembali beberapa penyakit menular lama (*re-emerging diseases*), serta munculnya penyakit-penyakit menular baru. Diantara penyakit menular yang sering menelan korban jiwa yaitu salah satunya HIV-AIDS. Di Asia Tenggara Indonesia menduduki posisi tertinggi untuk laju penularan HIV-AIDS. Indonesia juga menempati posisi ketiga sebagai negara dengan penularan kasus HIV-AIDS di wilayah Asia Pasifik setelah India dan China [2].

Beberapa masalah kesehatan mengenai penyakit menular maupun tidak menular yang ada di Indonesia ini menunjukkan bahwa masyarakat Indonesia belum menyadari tentang

pentingnya kesehatan. Hal ini seharusnya mendapatkan perhatian lebih, khususnya bagi Dinas Kesehatan maupun Kementerian Kesehatan Republik Indonesia mengenai penyebaran penyakit menular ataupun penyakit tidak menular. Oleh karena itu, diharapkan semua pihak harus lebih menjaga kesehatan diri. Salah satu upaya yang mampu dilakukan yakni dengan membentuk suatu pengelompokan provinsi ke dalam suatu kelompok yang memiliki kemiripan atau karakteristik yang hampir sama. Hal ini bermaksud untuk memberikan informasi terkait dengan kesehatan pada masing-masing provinsi. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu menggunakan analisis *cluster*.

Pada penelitian sebelumnya telah dilakukan penelitian perbandingan kinerja metode *Complete Linkage* dan *Average Linkage* dalam menentukan hasil analisis *cluster* pada produksi palawija Provinsi Kalimantan Timur. Penelitian tersebut menunjukkan hasil bahwa nilai rasio simpangan baku pada metode *Average Linkage* yaitu 0,056 lebih kecil daripada nilai rasio simpangan baku pada metode *Complete Linkage* yaitu 0,083. Berdasarkan hasil tersebut menunjukkan bahwa metode *Average linkage* lebih baik dari metode *Complete Linkage* [3].

Penelitian lainnya dengan judul “Perbandingan Kinerja Metode *Ward* dan *K-Means* dalam Menentukan Cluster Data Mahasiswa Permohonan Beasiswa”. Pada penelitian tersebut diperoleh hasil bahwa metode *Ward* memiliki kinerja yang lebih baik dibandingkan dengan metode *K-Means*. Berdasarkan penelitian tersebut diperoleh nilai rasio simpangan baku ( $S_w/S_b$ ) pada metode *Ward* yaitu 0.5346668% lebih kecil dari metode *K-Means* yang memiliki rasio simpangan baku sebesar 0.831525302 % [4].

Penelitian yang lain dilakukan oleh Lisda, Ika Purnamasari, & Fidia D.T.A (2018) dengan judul “Penerapan Metode *Complete Linkage* dan Metode *Hierarchical Clustering Multiscale Bootstrap* (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016)”. Berdasarkan penelitian tersebut menunjukkan hasil bahwa metode *Complete Linkage* memiliki kinerja yang baik dibandingkan metode *Hierarchical Clustering Multiscale Bootstrap* [5].

Berdasarkan pada uraian tersebut, penulis tertarik dalam melakukan penelitian mengenai analisis *cluster* dengan mengambil kasus penyakit yang sering menyerang kesehatan masyarakat di Indonesia diantaranya adalah Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD (Demam berdarah Dengue), Pneumonia pada balita, dan Malaria. Judul dalam penelitian ini adalah " Analisis *Cluster* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode *Complete Linkage*, *Average Linkage*, dan *Ward*".

## TINJAUAN PUSTAKA

### A. Kesehatan

Kesehatan merupakan keadaan yang sejahtera dari badan, jiwa serta sosial dimana setiap orang hidup dengan produktif baik secara sosial maupun ekonomis. Setiap orang sangat penting dalam menjaga kesehatannya terutama dalam upaya

menanggulangi dan mencegah gangguan kesehatan yang membutuhkan pemeriksaan, perawatan dan pengobatan [6]. Saat seseorang tidak benar-benar menjaga kesehatan seringkali virus penyakit akan datang menghampiri. Diantara beberapa jenis penyakit menular yang sering menyerang kesehatan manusia diantaranya adalah DBD, HIV, Tuberkulosis (TBC), AIDS, Campak, Difteri, Pneumonia, Kusta, dan Malaria.

### B. Analisis Cluster

Analisis *cluster* adalah teknik analisis statistik yang digunakan untuk menempatkan kumpulan obyek ke dalam dua grup bahkan lebih berdasarkan pada karakteristik kesamaan-kesamaan obyek. Analisis *cluster* merupakan teknik analisis yang bertujuan untuk memilih obyek dalam beberapa kelompok dimana antara kelompok satu dengan kelompok lainnya memiliki sifat yang berbeda. Dalam analisis *cluster* tiap kelompok bersifat homogen antar anggota dalam kelompok atau variasi obyek dalam satu kelompok yang terbentuk sekecil mungkin. Secara keseluruhan solusi *cluster* tergantung pada variabel yang digunakan sebagai dasar dalam menilai suatu kesamaan. Pengurangan dan penambahan variabel yang terkait dapat mempengaruhi hasil analisis *cluster*. Umumnya analisis *cluster* dapat dikatakan sebagai suatu proses dalam menganalisa baik atau tidaknya proses pembentuk *cluster*.

### C. Standarisasi Data

Standarisasi data dapat dilakukan jika pada variabel yang diobservasi terdapat perbedaan ukuran satuan yang cukup besar. Perbedaan satuan yang cukup mencolok akan menyebabkan perhitungan pada analisis *cluster* tidak benar-benar valid. Oleh sebab itu hendaknya dilakukan proses standarisasi data melalui transformasi pada data aktual sebelum melakukan analisis yang lebih lanjut. Transformasi dalam bentuk *z-skor* adalah seperti di bawah ini :

$$z = \frac{x_i - \bar{x}}{s} \quad (1)$$

### D. Multikolinearitas

Terdapat asumsi-asumsi yang harus diperhatikan saat melakukan analisis *cluster*, salah satunya adalah uji Multikolinieritas. Istilah Multikolinieritas mula-mula ditemukan oleh Ragnar Frisch. Multikolinieritas adalah suatu masalah dalam teknik multivariat karena akan menyebabkan dampak yang tidak baik. Multikolinieritas merupakan adanya suatu hubungan linear yang pasti atau sempurna diantara beberapa atau semua variabel. Hocking (1996) mengemukakan bahwa ada tiga kriteria yang bisa dipakai untuk mengetahui terjadinya multikolinieritas, dan yang paling umum digunakan adalah *Varians Inflation Factor* (VIF). Apabila nilai VIF lebih besar dari 10 menunjukkan terjadinya multikolinieritas antara variabel-variabel prediktor. Rumus dalam menghitung *Varians Inflation Factor* (VIF) sebagai berikut :

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

Dalam mengetahui adanya multikolinieritas dapat juga dicari menggunakan nilai tolerance dengan rumus :

$$Tolerance = \frac{1}{VIF} \tag{3}$$

Solusi dalam mengatasi adanya multikolinieritas yaitu dengan mengeluarkan variabel yang tidak berhubungan atau saling terkait. Ada atau tidaknya multikolinieritas antar variabel sangat diperhatikan dalam melakukan analisis cluster. Hal ini karena dapat berpengaruh, sehingga variabel-variabel yang bersifat multikolinieritas perlu dipertimbangkan dengan teliti. Jika asumsi multikolinieritas terpenuhi maka ukuran jarak dapat digunakan dalam menentukan ukuran kemiripan [7].

**E. Ukuran Jarak**

Tujuan dari analisis *cluster* adalah mengelompokkan obyek kemiripan suatu obyek dalam *cluster* yang sama. Oleh karena dibutuhkan beberapa ukuran guna mengetahui seberapa kemiripan serta perbedaan obyek-obyek tersebut. Metode ukuran jarak digunakan pada data yang memiliki skala metrik. Ukuran jarak ini sebetulnya adalah ukuran ketidakmiripan, dimana untuk jarak yang besar memiliki kesamaan yang sedikit sedangkan untuk jarak yang kecil menjelaskan bahwasannya suatu obyek akan lebih mirip dengan obyek yang lainnya. Untuk penelitian kali ini ukuran jarak yang dipakai yaitu menggunakan ukuran jarak *Euclidean*, dan *Squared Euclidean* [8].

**Jarak Euclidean**

Jarak *Euclidean* adalah besar suatu jarak pada garis lurus yang menghubungkan antar objek. Sebagai contoh ukuran ketidaksamaan atau jarak antar obyek ke-i dan obyek ke-j dapat disimbolkan dengan  $d_{ij}$  dan untuk variabel ke-k dengan  $k=1, \dots, p$ . nilai  $d_{ij}$  didapatkan dari perhitungan pada jarak *Euclidean* dengan rumus :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{4}$$

**Jarak Squared Euclidean**

Jarak *Squared Euclidean* adalah variasi dari jarak *Euclidean*. Perbedaan dari kedua jarak yaitu jika jarak *Euclidean* diakarkan, sedangkan jarak *Squared Euclidean* tidak terdapat akar. Dengan rumus sebagai berikut :

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \tag{5}$$

**F. Metode Hierarki Clustering**

Metode hierarki merupakan *Agglomerasi* (penggabungan) atau *Divisive* (pemecahan). Pada metode aglomerasi setiap penelitian pada awalnya diasumsikan sebagai *cluster* tersendiri yang mana terdapat *cluster* sesuai jumlah penelitiannya. Selanjutnya untuk dua *cluster* yang memiliki kesamaan terdekat digabungkan menjadi *cluster* baru, dengan begitu jumlah *cluster* akan berkurang satu untuk tiap tahapan. Sedangkan untuk pemecahan akan dimulai dari satu *cluster* besar yang mengandung seluruh observasi. Selanjutnya untuk observasi yang tidak sama akan dipisahkan dan dibentuk kelompok-kelompok yang lebih kecil. Proses ini akan terus dilakukan sampai setiap observasi

terbentuk kelompok sendiri-sendiri [9]. Metode hierarki pengelompokannya dimulai dengan dua objek atau lebih yang memiliki kesamaan terdekat. Selanjutnya perhitungan akan diteruskan ke objek lainnya yang memiliki kedekatan selanjutnya atau kedua. Begitupun seterusnya hingga *cluster* akan membentuk Dendogram atau semacam pohon dengan terdapat hierarki (tingkatan) yang jelas antara objeknya.

**Metode Pautan Lengkap (Complete Linkage)**

Metode pautan lengkap (*Complete Linkage*) berdasarkan pada jarak maksimum. Pengukuran jarak antara cluster satu dengan yang lainnya berdasarkan pada obyek yang memiliki jarak paling jauh. Semua obyek dalam suatu cluster saling dikaitkan satu sama lain pada suatu kesamaan minimum ataupun dengan jarak maksimum. Dalam perhitungan awal, nilai minimum dalam  $D = d_{ij}$  dicari terlebih dulu serta digabungkan pada obyek-obyek yang bersesuaian, contohnya I dan K, guna memperoleh *cluster* (IJ). Jarak antara *cluster* lain K dan (IJ), dihitung menggunakan cara :

$$d_{(IJ)K} = \max\{d_{IK}, d_{JK}\} \tag{6}$$

**Metode Pautan Rata-Rata (Average Linkage)**

Metode *Average Linkage* merupakan proses *clustering Average Distance* atau pada jarak rata-rata antar obyeknya. Untuk metode *Average Linkage*, jarak diantara dua *cluster* dapat diasumsikan sebagai jarak rata-rata dari semua anggota dalam satu *cluster* dengan semua anggota dalam *cluster* lainnya.. Tahapan pertama yang harus dilakukan adalah mencari jarak minimum  $D = d_{ij}$  terlebih dulu serta digabungkan pada obyek-obyek yang bersesuaian, contohnya I dan K, guna memperoleh *cluster* (IJ). Jarak antara *cluster* lain K dan (IJ), dihitung menggunakan cara :

$$d_{(IJ)K} = \frac{\sum_a \sum_b d_{ab}}{N_{(IJ)} N_K} \tag{7}$$

**Metode Ward**

Metode Ward merupakan metode varians yang umum digunakan. Pada metode Ward rata-rata untuk setiap *cluster* dihitung [5]. Kemudian, dihitung jarak *Squared Euclidean* antara setiap obyek dan nilai rata-ratanya, lalu jarak tersebut dihitung semua, setiap tahap, dua *cluster* yang mempunyai kenaikan *sum of squares* (SSE) dalam *cluster* yang terkecil digabungkan. Metode Ward ini dihitung berdasarkan persamaan berikut rumus dalam menghitung SSE [10]:

$$SSE = \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x}) \tag{8}$$

Dalam menggabungkan dua *cluster* misal *cluster* U dan V dengan cara meminimalkan peningkatan SSE yang didefinisikan sebagai jarak antara Cluster U dan Cluster V yaitu seperti di bawah ini :

$$I_{(UV)} = SSE_{UV} - (SSE_U - SSE_V) \tag{9}$$

Apabila U terdiri dari  $X_i$  serta V terdiri dari  $X_j$ , maka  $SSE_U$  dan  $SSE_V$  yaitu bernilai nol. Dalam menentukan jarak dengan metode Ward menggunakan persamaan jarak antara objek dengan rumus seperti di bawah ini :

$$I_{(ij)} = SSE_{ij} = \frac{1}{2} d^2(x_i, x_j) \tag{10}$$

Berdasarkan Persamaan (2.14) jarak antara objek UV dan objek W dengan menggunakan metode Ward yaitu seperti berikut ini:

$$I_{(UV)W} = \frac{n_U+n_W}{n_{UV}+n_W} I_{UW} + \frac{n_V+n_W}{n_{UV}+n_W} I_{VW} + \frac{n_W}{n_{UV}+n_W} I_{UV} \tag{11}$$

**G. Metode Terbaik**

Dalam mengetahui metode mana yang memiliki cara kerja yang paling baik, bisa menggunakan rata-rata simpangan baku dalam cluster ( $S_w$ ) serta simpangan baku antar cluster ( $S_b$ ) [11]. Untuk rumus  $S_w$  adalah sebagai berikut :

$$S_w = \frac{1}{K} \sum_{k=1}^K S_k \tag{12}$$

Apabila diberikan cluster  $c_k$  dimana  $k = 1, \dots, p$ , dan setiap cluster memiliki anggota  $x_i$  dimana  $i = 1, \dots, n$  dan  $n$  adalah jumlah anggota dari setiap cluster, dan  $\bar{x}_k$  adalah rata-rata dari cluster  $k$  maka untuk mencari nilai simpangan baku ke- $k$  ( $S_k$ ) digunakan rumus berikut [12]:

$$S_k = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x}_k)^2} \tag{13}$$

Apabila terdapat rata-rata variabel dalam setiap cluster  $k$  ( $\bar{x}_k$ ) maka komponen dari setiap cluster akan berbeda, dan simpangan baku antar kelompok  $S_b$  dapat dirumuskan sebagai berikut :

$$S_b = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2} \tag{14}$$

Metode pengelompokan yang baik memiliki nilai simpangan baku antar kelompok yang  $S_b$  maksimum dan nilai simpangan baku dalam kelompok  $S_w$  yang minimum. Nilai rasio  $S_w$  terhadap  $S_b$  yang semakin kecil dapat dikatakan metode tersebut mempunyai kinerja yang baik karena memiliki tingkat homogenitas yang tinggi di dalam cluster [4]. Rasio  $S_w$  terhadap  $S_b$  dapat dihitung dengan rumus di bawah ini:

$$S = \frac{S_w}{S_b} \tag{15}$$

**METODE PENELITIAN**

Penelitian ini merupakan penelitian yang bersifat kualitatif. Dilihat dari tujuannya penelitian ini termasuk penelitian komperatif karena membandingkan tiga teori yaitu *Complete Linkage*, *Avarage Linkage* dan *Ward*. Data yang digunakan pada penelitian ini adalah data penyakit yang sering terjadi di Indonesia tahun 2018 di 34 provinsi. Data tersebut diperoleh dari Kementerian Kesehatan Republik Indonesia yang dapat diakses melalui Publikasi Profil Kementerian Kesehatan, 2018. Variabel dalam penelitian adalah jumlah dari beberapa kasus penyakit yang menyerang masyarakat yaitu penyakit Tuberkulosis (X1), HIV (X2), AIDS (X3), Kusta (X4), Campak (X5), Difteri (X6), DBD (Demam berdarah Dengue) (X7), Pnemonia pada balita (X8), dan Malaria (X9).

Langkah-langkah yang harus dilakukan dalam penelitian ini adalah pertama melakukan standarisasi data menggunakan *z-score* karena data yang digunakan di dalam penelitian ini memiliki satuan yang bervariasi, kedua melakukan uji multikolinieritas guna mengetahui suatu data terjadi korelasi antar variabel atau tidak. Ketiga menghitung jarak antar objek untuk metode *Complete Linkage* dan *Avarage Linkage* menggunakan jarak *Euclidean* sedangkan untuk metode *Ward* menggunakan jarak *Squared Euclidean*. Keempat melakukan uji coba 2 sampai 6 cluster dengan tiga metode yaitu *Complete Linkage*, *Avarage Linkage* dan *Ward*. Kelima menghitung nilai simpangan baku dalam kelompok ( $S_w$ ) dan simpangan baku antar kelompok ( $S_b$ ). Keenam menentukan metode terbaik berdasarkan nilai rasio simpangan baku terkecil. Keenam melakukan intepretasi hasil cluster.

**HASIL DAN PEMBAHASAN**

**A. Deskripsi Data**

Sebelum melakukan proses perhitungan data, terlebih dahulu yang harus dilakukan adalah mengumpulkan data. Data yang digunakan dalam penelitian ini secara tidak langsung diperoleh dari publikasi Kementerian Kesehatan Republik Indonesia, yaitu mengenai jumlah kasus penyakit yang sering terjadi di Indonesia tahun 2018. Data 34 provinsi di Indonesia digunakan sebagai objek pengamatan, dengan karakteristiknya adalah jumlah kasus penyakit yang sering terjadi. Kasus penyakit yang dimaksud adalah Tuberkulosis, HIV, AIDS, Kusta, Campak, Difteri, DBD, Pnemonia pada balita, dan Malaria. Jumlah kasus penyakit di Indonesia tahun 2018 secara menyeluruh disajikan pada Gambar 1.



Gambar 1. Jumlah Kasus Penyakit di Indonesia Tahun 2018

**B. Standarisasi Data**

Data yang digunakan dalam penelitian mempunyai jangkauan yang terpaut jauh antara variabel. Oleh sebab itu harus dilakukan standarisasi data. Proses perhitungan standarisasi data yang dimaksud dengan melakukan transformasi pada data asli menggunakan *z-score* berdasarkan Persamaan (1).

Tabel 1. Data Hasil Standarisasi

Provinsi	TBC	HIV	...	Malaria
Aceh	-0,323	-0,581	...	-0,221
Sumatra Utara	0,822	0,299	...	-0,141
Sumatra Barat	-0,201	-0,357	...	-0,212
⋮	⋮	⋮	⋮	⋮
Papua	-0,198	1,0376	...	5,605

**C. Uji Multikolinearitas**

Untuk mengetahui terjadi atau tidaknya multikolinearitas, dapat dilihat dari nilai VIF dan *tolerance* pada masing-masing variabel independen.

Suatu data dikatakan terbebas dari gejala multikolinearitas apabila nilai VIF < 10 dan nilai *tolerance* > 0.10. Berdasarkan hasil perhitungan uji multikolinearitas terhadap data penyakit menular di Indonesia tahun 2018 menggunakan persamaan (2) dan (3) didapatkan hasil yang tertera pada Tabel 2 .

Tabel 2. Hasil Uji Multikolinearitas Semua Variabel

Variabel	VIF	Tolerance
Tuberkulosis	2,6170	0,3821
HIV	2,5974	0,3850
AIDS	1,6387	0,6103
Kusta	1,0177	0,9826
Campak	1,0979	0,9108
Difteri	1,0874	0,9196
DBD	2,8693	0,3485
Pneumonia Balita	1,0114	0,9888
Malaria	1,0025	0,9975

Berdasarkan pada Tabel 1 dapat dilihat bahwa pada perhitungan uji multikolinearitas untuk semua variabel yang digunakan tidak terdapat data yang mengalami multikolinearitas karena nilai VIF < 10 dan nilai *Tolerance* > 0,10 .

**D. Perhitungan Ukuran Jarak Antar Dua Objek**

Ukuran jarak yang digunakan pada metode *Complete Linkage*, *Average Linkage* adalah jarak *Euclidean* sedangkan untuk metode *Ward* adalah *Squared Euclidean*. Jarak antar provinsi yang dihitung menggunakan jarak *Euclidean* sesuai Persamaan (4) diperoleh hasil yang tertera pada Tabel 3.

Tabel 3. Matriks Jarak *Euclidean*

	d <sub>1</sub>	d <sub>2</sub>	...	d <sub>34</sub>
d <sub>1</sub>	0,000	5,690	...	8,365
d <sub>2</sub>	5,690	0,000	...	7,514
d <sub>3</sub>	4,966	2,131	...	7,704
⋮	⋮	⋮	⋮	⋮
d <sub>34</sub>	8,355	7,506	...	0,000

Jarak antar provinsi yang dihitung menggunakan jarak *Squared Euclidean* sesuai Persamaan (5) diperoleh hasil yang tertera pada Tabel 4.

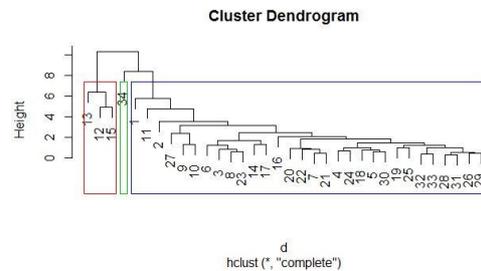
Tabel 4. Matriks Jarak *Squared Euclidean*

	d <sub>1</sub>	d <sub>2</sub>	...	d <sub>34</sub>
d <sub>1</sub>	0,000		...	8,365
d <sub>2</sub>	32,373	0,000	...	7,514
d <sub>3</sub>	24,657	2,131	...	7,704
⋮	⋮	⋮	⋮	⋮
d <sub>34</sub>	69,803	56,335	...	49,954

Berdasarkan perhitungan matriks jarak *Euclidean* dan *Squared Euclidean* dicari nilai terkecil antar objek, karena semakin kecil nilai dari jarak antar dua objek tersebut maka kedua objek tersebut memiliki kemiripan.

**E. Analisis Cluster Metode Complete Linkage**

Pada metode *Complete Linkage* proses pengelompokannya adalah dengan menggabungkan dua obyek atau lebih yang memiliki jarak paling jauh. Berdasarkan pada perhitungan jarak antar provinsi yang paling kecil sebesar 0,27817 yaitu jarak antara *cluster* objek 28 (Provinsi Sulawesi Tenggara) dengan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*. Selanjutnya dilakukan perhitungan kembali jarak terhadap *cluster* baru menggunakan Persamaan (6) hingga mendapatkan matriks baru. Proses pengelompokan 2 sampai 6 *cluster* juga dapat digambarkan dalam bentuk dendrogram. Berikut ini adalah dendrogram pengelompokan provinsi di Indonesia menggunakan metode *Complete Linkage* jika akan membentuk tiga *cluster* yang disajikan pada Gambar 2.



Gambar 2. Dendrogram *Complete Linkage*

Contoh hasil anggota *cluster* jika akan membentuk 3 *cluster* menggunakan metode *Complete Linkage* dapat dilihat pada Tabel 5.

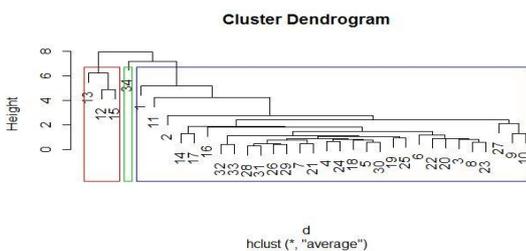
Tabel 5. Anggota Tiga *Cluster Complete Linkage*

Cluster	Jumlah Provinsi	Provinsi
Cluster 1	28	Aceh, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah,

		Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
Cluster 2	3	Jawa Barat, Jawa Tengah, Jawa Timur,
Cluster 3	1	Papua

**F. Analisis Cluster Metode Average Linkage**

Pada metode *Average Linkage* proses pengelompokannya adalah dengan menggabungkan dua objek berdasarkan jarak rata-rata antar obyeknya. Berdasarkan pada perhitungan jarak antar provinsi yang paling kecil sebesar 0,27817 yaitu jarak antara *cluster* objek 28 (Provinsi Sulawesi Tenggara) dengan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*. Selanjutnya dilakukan perhitungan kembali jarak terhadap *cluster* baru menggunakan Persamaan (7) hingga mendapatkan matriks baru. Proses pengelompokan 2 sampai 6 *cluster* juga dapat digambarkan dalam bentuk dendrogram. Dendrogram pengelompokan provinsi di Indonesia menggunakan metode *Average Linkage* jika akan membentuk tiga *cluster* yang disajikan pada Gambar 3.



Gambar 3. Dendrogram Average Linkage

Berdasarkan dendrogram pada Gambar 3. Jika membentuk tiga *cluster* maka akan menghasilkan tiga kelompok anggota provinsi. Hasil anggota provinsi jika akan membentuk 3 *cluster* menggunakan metode *Average Linkage* dapat dilihat pada Tabel 6.

Tabel 6. Anggota Tiga Cluster Average Linkage

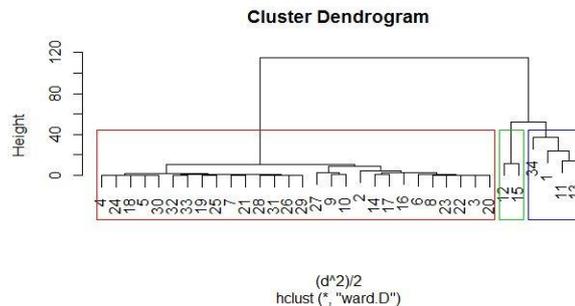
Cluster	Jumlah Provinsi	Provinsi
Cluster 1	28	Aceh, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat

Cluster 2	3	Jawa Barat, Jawa Tengah, Jawa Timur,
Cluster 3	1	Papua

**G. Analisis Cluster Metode Ward**

Berdasarkan pada perhitungan jarak *Squared Euclidean* antar provinsi yang paling kecil sebesar 0,07738 yaitu jarak antar *cluster* objek 28 (Provinsi Sulawesi Tenggara) dan objek 31 (Provinsi Maluku) yang kemudian kedua provinsi tersebut digabung menjadi satu *cluster*. Pada proses Agglomerasi dimulai dari dua objek terdekat, dimana jarak antar dua objek tersebut merupakan jarak yang terdekat dari kombinasi 34 objek yang digunakan. Misalnya Provinsi Sulawesi Tenggara dan Maluku dengan menggunakan Persamaan (10). Proses perhitungan diatas hanya menghitung dua kelompok, dimana untuk setiap kelompok terdiri dari satu objek.

Selanjutnya yaitu melakukan perhitungan pada jarak suatu *cluster* dengan *cluster* baru yang akan terbentuk menggunakan Persamaan (11). Perhitungan objek 28 dan 31 akan dilakukan untuk seluruh provinsi kecuali Provinsi Sulawesi Tenggara dan Maluku itu sendiri. Jarak yang digunakan untuk disatukan menjadi satu cluster adalah jarak yang terkecil. Proses tersebut akan terus berlanjut sampai seluruh provinsi bergabung menjadi satu cluster. Proses pengelompokan 2 sampai 6 *cluster* juga dapat digambarkan dalam bentuk dendrogram. Berikut ini adalah dendrogram pengelompokan provinsi di Indonesia menggunakan metode *Ward* jika akan membentuk tiga *cluster* yang disajikan pada Gambar 4.



Gambar 4. Dendrogram Ward

Contoh hasil anggota *cluster* jika akan membentuk 3 *cluster* menggunakan metode *Complete Linkage* dapat dilihat pada Tabel 7.

Tabel 7. Anggota Tiga Cluster Ward

Cluster	Jumlah Provinsi	Provinsi
Cluster 1	3	Aceh, DKI Jakarta, Jawa Tengah
Cluster 2	29	Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi

Cluster 3	2	Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat Jawa Barat, Jawa Timur,
-----------	---	---

Cluster 3	1	Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
Cluster 4	1	DKI Jakarta
Cluster 5	1	Jawa Tengah
Cluster 6	2	Papua Jawa Barat, Jawa Timur

**H. Metode Terbaik**

Berdasarkan hasil perhitungan diperoleh bahwa uji coba dua sampai enam cluster metode *Complete Linkage* dan *Average Linkage* memiliki hasil cluster yang sama. Maka hasil rasio simpangan baku dalam cluster ( $S_w$ ) rasio simpangan baku antar cluster ( $S_b$ ) dan rasio simpangan baku dalam cluster terhadap simpangan baku antar cluster (S) metode *Complete Linkage* dan *Average Linkage* tertera pada Tabel 8.

Tabel 8. Perbandingan Rasio Simpangan Baku *Complete Linkage* dan *Average Linkage*

Jumlah Cluster	$S_w$	$S_b$	S
2 Cluster	0,92688	1,44625	0,64089
3 Cluster	0,60225	1,48432	0,40573
4 Cluster	0,39295	1,67645	0,23440
5 Cluster	0,29093	1,81692	0,16012
6 Cluster	0,24090	2,03703	0,11826

Hasil rasio simpangan baku dalam cluster ( $S_w$ ) rasio simpangan baku antar cluster ( $S_b$ ) dan rasio simpangan baku dalam cluster terhadap simpangan baku antar cluster (S) metode *Ward* dapat dilihat pada Tabel 9.

Tabel 9. Perbandingan Rasio Simpangan Baku *Ward*

Jumlah Cluster	$S_w$	$S_b$	S
2 Cluster	1,28567	1,10789	1,16046
3 Cluster	0,67600	1,39186	0,48568
4 Cluster	0,39268	1,59351	0,24642
5 Cluster	0,35376	1,69735	0,20847
6 Cluster	0,18405	2,12284	0,08670

Berdasarkan pada Tabel 7 dan 8 hasil dari rasio simpangan baku (S) menunjukkan bahwa metode *Ward* dengan 6 cluster mempunyai kinerja yang paling baik dalam membentuk cluster provinsi diantara dua metode yang lainnya yaitu metode *Complete Linkage* dan *Average Linkage*. Hal tersebut dapat diketahui dari nilai  $S_w$  terhadap  $S_b$  pada metode *Ward* dengan 6 cluster memiliki nilai yang paling minimum diantara ketiga metode tersebut. Dari hasil perbandingan rasio simpangan baku menunjukkan bahwa 6 cluster dari metode *Ward* yang memiliki kinerja terbaik. Maka 6 cluster provinsi yang terbentuk adalah tertera pada Tabel 10.

Tabel 10. Anggota Enam Cluster *Ward*

Cluster	Jumlah Provinsi	Provinsi
Cluster 1	1	Aceh,
Cluster 2	28	Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat,

**KESIMPULAN**

Berdasarkan pada hasil pembahasan mengenai data jumlah kasus penyakit di Indonesia menggunakan ketiga metode yaitu *Complete Linkage*, *Average Linkage*, dan *Ward* dapat ditarik kesimpulan sebagai berikut :

1. Apabila ditinjau berdasarkan nilai simpangan baku dalam kelompok ( $S_w$ ) terhadap simpangan baku antar kelompok ( $S_b$ ). Hasil rasio simpangan baku paling kecil terdapat pada metode *Ward* dengan 6 cluster yaitu sebesar 0,08670. Oleh karena itu metode *Ward* merupakan metode yang paling baik jika dibandingkan dengan metode *Complete Linkage* dan *Average Linkage* karena pada metode *Ward* memiliki nilai rasio simpangan baku terkecil.
2. Hasil dari uji coba 2 hingga 6 cluster yang terbentuk berdasarkan metode *Complete Linkage*, *Average Linkage*, dan *Ward*, menunjukkan bahwa yang terbaik adalah dengan membentuk 6 cluster yakni untuk hasil cluster dengan metode *Ward* didapatkan hasil cluster pertama terdiri dari 1 provinsi, cluster kedua terdiri dari 28 provinsi, cluster ketiga, cluster keempat, dan cluster kelima, masing-masing terdiri dari 1 provinsi, serta cluster keenam terdiri dari 2 provinsi. Urutan cluster provinsi berdasarkan kelompok yang telah terbentuk diperoleh bahwa untuk tingkat kesehatan sangat baik, baik, cukup baik, kurang baik, tidak baik, dan sangat tidak baik berturut-turut yaitu cluster 1, cluster 2, cluster 3, cluster 4, cluster 5, dan cluster 6.

**REFERENCES**

[1] A. A. Irfa, "Gambaran Prevalensi dan Risiko Penyakit Kardiovaskuler pada Penduduk Dewasa di Provinsi DKI Jakarta," 2011.

[2] W. Blair Bell, "An Address ON THE UNSOLVED PROBLEMS IN GYNÆCOLOGY AND OBSTETRICS. Read by invitation before the American Gynœcological Society on June 2nd, 1921," *The Lancet*, vol. 197, no. 5104. pp. 1339–1341, 1921, doi: 10.1016/S0140-6736(01)12146-X.

[3] S. Ningsih, S. Wahyuningsih, and Y. N. Nasution, "Perbandingan kinerja metode complete linkage dan average linkage dalam menentukan hasil analisis cluster ( Studi Kasus : Produksi Palawija Provinsi Kalimantan Timur 2014 / 2015 )," *Pros. Semin. Sains dan Teknol. FMIPA Unmu*, vol. 1, no. 1, 2016.

[4] P. Kinerja et al., "PEMOHON BEASISWA ( STUDI KASUS : STMIK PRINGSEWU )," no. August, 2017.

[5] L. Ramadhani, I. Purnamasari, and F. D. T. Amijaya, "Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016)," *Eksponensial*, vol. 9, no. 2016, pp. 1–10, 2018, [Online]. Available: [https://fmipa.unmul.ac.id/files/docs/\[1\]](https://fmipa.unmul.ac.id/files/docs/[1]) LISDA

- RAMADHANI 1307015041\_Edit.pdf.
- [6] Y. F. Wardhani and A. Paramita, "Pelayanan Kesehatan Mental Dalam Hubungannya Dengan Disabilitas dan Gaya Hidup Masyarakat Indonesia (Analisis Lanjut Riskesdas 2007 Dan 2013)," *Bul. Penelit. Sist. Kesehat.*, vol. 19, no. 1, pp. 99–107, 2016.
- [7] A. Sholiha, "Perbandingan Analisis Klaster Menggunakan Metode Single Linkage , Complete Linkage , Average Linkage Dan K-Means Untuk Pengelompokan," 2015.
- [8] J. Matematika, F. Matematika, D. A. N. Ilmu, P. Alam, and U. N. Semarang, *Perbandingan Kinerja Metode Single Linkage , Metode Complete Linkage Dan Metode K-Means Dalam Analisis Cluster*. 2011.
- [9] V. Machine, "Perbandingan kinerja," vol. 9, no. 2, pp. 62–68, 2013.
- [10] A. C. Rencher and F. C. William, *Methods of multivariate analysis: Third edition*. 2012.
- [11] A. Fadliana and F. Rozi, "Penerapan Metode Angglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana," *Cauchy*, vol. 4, no. 1, p. 25, 2015, doi: 10.18860/ca.v4i1.3172.
- [12] A. R. Barakbah and K. Arai, "Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering," *Ind. Electron. Semin*. 2004, no. October, pp. 409–413, 2004.