

PENGGUNAAN KORELASI *COPHENETIC* UNTUK PEMILIHAN METODE *CLUSTER* BERTINGKAT PADA MENGELOMPOKKAN KABUPATEN/KOTA BERDASARKAN JENIS PENYAKIT DI PROVINSI SULAWESI TENGGARA TAHUN 2020

Iis*, Irma Yahya, Gusti Ngurah A. Wibawa, Baharuddin, Ruslan, Lilis Laome
Program Studi S1 Statistika FMIPA Universitas Halu Oleo, Kendari

^{*)}Email: iisstatistik@gmail.com

ABSTRACT

Southeast Sulawesi as one of the provinces in Indonesia which is geographically located from one district/city to another, which results in different health level variances. So that in determining policy, the government cannot give equal treatment to all regions, it is necessary to group them. The method used is a hierarchical cluster method (Single Linkage, Average Linkage, Complete Linkage) consisting of 4 clusters, namely districts/cities with low, medium, high, and very high disease cases. The aim is to find out the best cluster method and how the cluster results use the best cluster method. The data used is data on the number of diseases in districts/cities throughout Southeast Sulawesi in 2020 (17 districts/cities) obtained from the Southeast Sulawesi Provincial Health Office. The results showed that the best cluster method was the average linkage method with a cophenetic correlation value of 0.990. cluster results show 14 districts/cities in cluster I, 1 district/city in cluster II, 1 district/city in cluster III, and 1 district/city in cluster IV.

Keywords: *Cluster Analysis, Cophenetic Correlation, Hierarchical Cluster.*

ABSTRAK

Sulawesi Tenggara sebagai salah satu provinsi di Indonesia yang secara geografis letak Kabupaten/Kota yang satu dengan yang lain berbeda yang mengakibatkan adanya variasi tingkat kesehatan yang berbeda. Dalam menentukan kebijakan, pemerintah tidak bisa memberi perlakuan yang sama pada semua wilayah, sehingga perlu dilakukan pengelompokan. Metode yang digunakan adalah metode cluster bertingkat (*Single Linkage, Average Linkage, Complete Linkage*). Tujuannya untuk mengetahui metode cluster terbaik dan bagaimana hasil cluster menggunakan metode cluster terbaik. Terdiri dari 4 cluster yaitu cluster I untuk Kab/Kota dengan rata-rata kasus penyakit Rendah, cluster II untuk Kab/Kota dengan rata-rata kasus penyakit Sedang, cluster III untuk Kab/Kota dengan rata-rata kasus penyakit Tinggi, dan cluster IV untuk Kab/Kota dengan rata-rata kasus penyakit Sangat Tinggi. Data yang digunakan adalah data Jumlah Penyakit di Kab/Kota se-Sulawesi Tenggara tahun 2020 (17 kab/kota) yang diperoleh dari Dinas Kesehatan Provinsi Sulawesi Tenggara. Hasil penelitian menunjukkan metode cluster terbaik adalah metode average linkage dengan nilai korelasi *cophenetic* 0.990. Hasil *cluster* menunjukkan 14 kab/kota pada *cluster* I, 1 kab/kota pada *cluster* II, 1 kab/kota pada *cluster* III, dan 1 kab/kota pada *cluster* IV.

Kata Kunci: Analisis Cluster, Korelasi Cophenetic, Cluster Hierarki.

PENDAHULUAN

Beberapa masalah kesehatan mengenai penyakit, baik itu penyakit menular maupun penyakit tidak menular yang ada di Indonesia, 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 (Ulinuh & Veriani, 2020).

Sulawesi Tenggara sebagai salah satu Provinsi di Indonesia yang secara geografis letak dari Kabupaten/Kota yang satu dengan yang lain berbeda dan mengakibatkan adanya variasi tingkat kesehatan yang berbeda pula antara Kabupaten/Kota satu dengan yang lainnya. Oleh karena itu, dalam hal menentukan kebijakan, pemerintah tentunya tidak bisa semua wilayah mendapatkan perlakuan yang sama. Salah satu upaya yang mampu dilakukan yakni dengan membentuk suatu pengelompokan Kabupaten/Kota ke

dalam suatu kelompok yang memiliki kemiripan atau karakteristik yang hampir sama. Hal ini bermaksud untuk memberikan informasi terkait dengan kesehatan pada Kabupaten/Kota di Provinsi Sulawesi Tenggara. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu menggunakan Analisis *Cluster* (Dinkes Sultra, 2020).

Analisis *Cluster* merupakan salah satu metode statistika yang dapat digunakan untuk melakukan proses pengelompokan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimiliki dengan ciri tertentu relatif homogen, sehingga objek dalam kelompok memiliki kesamaan yang sama sedangkan objek antar kelompok tidak memiliki kesamaan (Fathia et al., 2016).

Tujuan pengelompokan adalah untuk mengelompokkan obyek yang mirip dalam cluster yang sama. Oleh karena itu beberapa ukuran diperlukan untuk mengakses seberapa mirip atau berbeda obyek-obyek tersebut. Pendekatan yang paling biasa ialah mengukur kemiripan dinyatakan dalam jarak (*distance*) antara pasangan obyek. Makin besar nilai ukuran kemiripan atau jarak antar dua buah obyek, makin besar pula perbedaan antara dua obyek tersebut, sehingga cenderung untuk tidak menganggapnya kedalam kelompok yang sama (Rahmayani, 2018).

METODE PENELITIAN

Sumber Data

Data yang digunakan dalam penelitian ini adalah Data Sekunder yang diperoleh dari Dinas Kesehatan Provinsi Sulawesi Tenggara Tahun 2020. Data tersebut adalah data Jumlah Penyakit setiap Kabupaten/Kota se-Sulawesi Tenggara.

Variabel Penelitian

Variabel penelitian yang digunakan pada penelitian ini adalah sebanyak 10 variabel dan merupakan variabel bebas yang merupakan jenis penyakit di provinsi Sulawesi tenggara tahun 2020. Definisi operasional dari variabel-variabel yang akan digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Definisi Operasional Variabel Penelitian

Penyakit	Variabel		Devinisi Operasional	Satuan
Penyakit Menular Langsung	Tuberculosis Paru	X ₁	Tuberculosis Paru Atau TBC Merupakan Penyakit Menular Yang Disebabkan Oleh Infeksi Bakteri <i>Mycobacterium tuberculosis</i> .	Orang
	Pneumonia	X ₂	Pneumonia Adalah Penyakit Yang Disebabkan Oleh Kuman <i>Pneumococcus</i> , <i>Staphylococcus</i> , <i>Streptococcus</i> , Dan Virus.	Orang
	Kusta	X ₃	Penyakit Kusta Atau Lepra Adalah Penyakit Yang Disebabkan Oleh Bakteri <i>Mycobacterium Leprae</i> .	Orang
Penyakit Yang Dapat Dicegah Dengan Imunisasi	Non Folio AFP	X ₄	Non Polio AFP Adalah Kasus Lumpuh Layu Yang Diduga Kasus Polio Sampai Dibuktikan Dengan Pemeriksaan Laboratorium Bahwa Bukan Kasus Polio.	Orang
	Campak	X ₅	Campak Adalah Penyakit Yang Disebabkan Oleh Virus Campak Golongan <i>Paramyxovirus</i> .	Orang
Penyakit Tular Vektor Dan Zoonotik	DBD	X ₆	Penyakit Demam Berdarah Dengue (DBD) Adalah Penyakit Yang Disebebkan Oleh Virus Dengue, Yang Masuk Ke Peredaran Manusia Melalui Gigitan Nyamuk Dari Genus <i>Aedes</i> , Misalnya <i>Aedes Aegypti</i> Atau <i>Aedes Albopictus</i> .	Orang
	Malaria	X ₇	Malaria Adalah Penyakit Infeksi Yang Disebabkan Oleh Parasite <i>Plasmodium</i> Yang Hidup Dan Berkembang Biak Dalam Sel Darah	Orang

Penyakit	Variabel		Devinisi Operasional	Satuan
Penyakit Tidak Menular	Filariasis	X8	Merah Manusia, Ditularkan Oleh Nyamuk Malaria Filariasis Adalah Penyakit Yang Disebabkan Oleh Parasit Berupa Cacing Filaria, Yang Terdiri Dari Tiga Spesies Yaitu Wuchereria Bancrofti, Brugia Malayi, Dan Brugia Tiamori.	Orang
	Hipertensi	X9	Hipertensi Didefinisikan Sebagai Tekanan Sistolik 160 MmHg Dan Tekanan Distolik 90 MmHg. Tekanan Darah Normal Adalah 110/90 MmHg. Hipertensi Merupakan Produk Dari Resistensi Pembuluh Darah Perifer Dan Kardiak Output.	Orang
	Diabete Melitus	X10	Diabetes Mellitus adalah Gangguan Metabolisme Kronis Dengan Multi Etiologi Dan Ditandai Dengan Tingginya Kadar Gula Darah Disertai Dengan Gangguan Metabolisme Karbohidrat, Lipid Dan Protein	Orang

Prosedur Penelitian

Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan pengumpulan data
2. Melakukan analisis deskriptif untuk mengkaji karkteristik kab/kota berdasarkan jenis penyakit di provinsi Sulawesi tenggara tahun 2020
3. Melakukan pemeriksaan multikolinieritas
4. Melakukan analisis komponen utama (PCA)
5. Menghitung Jarak *Euclidean*
6. Menentukan metode *cluster* terbaik berdasarkan nilai koefisien korelasi *cophenetic*
7. Menentukan *cluster* optimum dengan metode *Davies Bouldin Indeks* (DBI)
8. Melakukan pengelompokkan dengan metode *cluster* hierarki (*Single Linkage, Average Linkage, dan Complete Linkage*)
9. Menginterpretasikan hasil *cluster* yang terbentuk pada metode *cluster* terbaik antara *Single Linkage, Average Linkage, dan Complete Linkage*.

Metode Analisis

1. Cluster Hierarki

Metode *Cluster* Hierarki (*hierarchical method*) yaitu metode yang memulai pengelompokkannya dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses dilanjutkan ke obyek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *Cluster* akan membentuk semacam “pohon” dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai dengan yang paling tidak mirip. Biasanya pengelompokkan ini disajikan dalam bentuk dendogram, yang mirip dengan “struktur diagram pohon” (*tree diagram*). Dendogram adalah representasi visual dari langkah-langkah analisis *Cluster* yang menunjukkan bagaimana *Cluster* terbentuk dan nilai koefisien jarak pada setiap langkah (Awaliyah, 2018).

2. Metode Penggabungan (*Agglomerative*)

Metode penggabungan dimulai dengan mengasumsikan bahwa setiap obyek adalah sebuah *Cluster*. Kemudian dua obyek dengan jarak terdekat digabungkan menjadi satu *Cluster*. Selanjutnya obyek ketiga akan bergabung dengan *Cluster* yang ada atau bersama obyek lain dan membentuk *Cluster* baru dengan tetap memperhitungkan jarak kedekatan antar obyek. Perlu diperhatikan bahwa setiap penggabungan dalam metode ini selalu diikuti dengan perbaikan matriks jarak. Hasil analisis gerombol dari metode ini dapat disajikan dalam bentuk dendogram (Utami, 2017). Metode penggabungan

sendiri terbagi menjadi:

➤ **Metode *Single Linkage***

Untuk menentukan jarak antar *Cluster* dengan menggunakan *Single Linkage*, maka dipilih jarak yang paling dekat atau aturan tetangga dekat (*nearest neighbour rule*). Berikut adalah persamaan metode *Single Linkage* untuk menentukan jarak antara kelompok (i,j) dengan k cluster:

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

Keterangan :

d_{ik} : Jarak antara kelompok i dan k

d_{jk} : Jarak antara kelompok j dan k

$d_{(ij)k}$: Jarak antara kelompok ij dan kelompok ke-k

➤ **Metode *Complete Linkage (Pautan Lengkap)***

Metode *Complete Linkage* merupakan metode pembaruan matrik kemiripan yang berdasarkan pada jarak terjauh antar obyeknya (maksimum distance). Berikut adalah persamaan metode complete linkage untuk menentukan jarak antara kelompok (i,j) dengan k cluster:

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

Keterangan :

d_{ik} : Jarak antara kelompok i dan k

d_{jk} : Jarak antara kelompok j dan k

$d_{(ij)k}$: Jarak antara kelompok ij dan kelompok k

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

Metode *average linkage* merupakan metode pembaruan matrik kemiripan yang berdasarkan pada jarak rata-rata antar obyeknya (*average distance*). Metode ini bertujuan meminimumkan rataan jarak semua pasangan pengamatan dari dua kelompok yang digabung. Berikut adalah persamaan metode average linkage untuk menentukan jarak antara kelompok (i,j) dengan k cluster:

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{n_{(ij)} n_k}$$

Keterangan :

d_{ab} : Jarak antara obyek a *cluster* (ij) dan obyek b *cluster* k

$n_{(ij)}$: Jumlah obyek pada *cluster* (ij)

n_k : Jumlah obyek pada *cluster* k

Tahap-tahap pengclusteran data dengan menggunakan metode hierarki (Afandi, 2015) adalah sebagai berikut:

1. Menentukan matriks jarak antar individu $D=(d_{ik})$
2. Mencari nilai jarak terdekat antara dua individu
3. Menghitung jarak antar kelompok dengan beberapa individu diluar kelompok dengan menggunakan metode cluster berhierarki

4. Menyusun kembali matriks jarak baru dengan cara membuat matriks baris dan kolom baru dari nilai jarak antara kelompok dengan beberapa individu diluar kelompok atau kelompok sisa
5. Mengulangi langkah 2-4 sampai semua individu berada dalam satu kelompok dan tidak dapat dikelompokkan lagi
6. Membuat dendogram.

3. Multikolinearitas

Uji *Bartlets* bertujuan untuk mengetahui apakah terdapat hubungan antar variabel dalam kasus multivariate. Jika variabel X_1, X_2, \dots, X_P saling bebas (*independent*), maka matriks korelasi antar variabel sama dengan matriks identitas. Sehingga untuk menguji kebebasan antar variabel ini dilakukan dengan Uji *Bartlets*. Adapun hipotesisnya sebagai berikut:

Hipotesis :

H_0 : $P=I$ (tidak ada korelasi antar variabel)

H_1 : $P \neq I$ (ada korelasi antar variabel)

Statistik Uji :

$$\chi^2 = -\left(n-1-\frac{2p+5}{6}\right) \ln |R|$$

Keterangan:

$\ln|R|$: Nilai determinan dari matriks korelasi

n : Banyaknya observasi

p : Banyaknya variabel

P : Matriks korelasi

I : Matriks identitas

Keputusan :

Tolak H_0 bila $\chi^2_{hitung} > \chi^2_{\alpha; 1/2p(p-1)}$ maka variabel saling berkorelasi, hal ini berarti terdapat hubungan antar variabel.

4. Analisis Komponen Utama (PCA)

PCA merupakan usaha untuk mereduksi p peubah pengamatan menjadi t peubah baru yang saling orthogonal yang masing-masing t peubah baru tersebut merupakan kombinasi linear dari p peubah lama. Analisis komponen utama akan lebih efektif jika antar p peubah asal memiliki korelasi yang cukup tinggi (Astutik & Fernandes, 2012). Analisis komponen utama merupakan suatu teknik analisis statistik untuk mentransformasi peubah-peubah asli yang masih saling berkorelasi satu dengan yang lain menjadi satu set peubah baru yang tidak berkorelasi lagi. Peubah-peubah baru ini disebut komponen utama. Kegunaan analisis komponen utama adalah:

- Identifikasi peubah baru yang mendasari data peubah ganda
- Mereduksi jumlah himpunan peubah yang banyak dan saling berkorelasi menjadi peubah-peubah baru yang tidak berkorelasi dengan mempertahankan sebanyak mungkin keragaman data tersebut.

Peubah-peubah baru ini memanfaatkan informasi dari peubah-peubah asal dan nilai yang diperoleh dari masing-masing objek merupakan koordinat objek-objek tersebut dalam peubah baru yang merupakan suatu sumbu-sumbu koordinat. Ciri-ciri peubah baru adalah sebagai berikut:

- Peubah baru yang terbentuk adalah peubah yang merupakan kombinasi linear dari peubah asal.
- Tidak saling berkorelasi
- Mempunyai ragam terurut dari yang terbesar sampai yang terkecil.

Perhitungan analisis dengan menggunakan PCA adalah masalah memecahkan masalah

eigen. Adapun algoritma PCA secara umum adalah sebagai berikut (Johnson, 2007):

- a. Hitung matriks kovarian dengan persamaan berikut:

$$\text{cov}(x, y) = \frac{\sum xy}{n} - (\bar{x})(\bar{y})$$

- b. Hitung nilai eigen dengan menyelesaikan persamaan berikut:

$$(A - \lambda I) = 0$$

Dimana A merupakan matriks data, λ merupakan nilai eigen, dan I adalah matriks identitas

- c. Hitung vector eigen dengan menyelesaikan persamaan berikut:

$$[A - \lambda I][X] = 0$$

Dimana A merupakan matriks data, λ merupakan nilai eigen, I adalah matriks identitas, dan X adalah vector eigen.

- d. Tentukan variabel baru (PC) dengan mengalikan variabel asli dengan matriks *vector eigen*.

$$Y = XA$$

Dimana Y adalah Matriks Skor Komponen Utama, A adalah Matriks Eigen Vektor, dan X adalah Matriks Variabel Asal. Selanjutnya, hasil dari analisis komponen utama dapat digunakan untuk perhitungan analisis selanjutnya yaitu analisis *cluster*.

5. 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*. Berikut adalah persamaan untuk jarak Euclidean.

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

Keterangan :

- d_{ij} : Jarak *Euclidian* dari objek ke- i dan obyek ke- j
 p : Jumlah variabel *cluster*
 x_{ik} : Nilai objek ke- i pada variable ke- k
 x_{jk} : Nilai objek ke- j pada variable ke- k

6. Korelasi *Cophenetic*

Korelasi *Cophenetic* untuk pohon *cluster* didefinisikan sebagai koefisien korelasi linier antara jarak kofenetik yang diperoleh dari pohon, dan jarak awal (ketidakmiripan) yang digunakan untuk membangun pohon. Ini merupakan ukuran seberapa setia pohon itu mewakili perbedaan diantara pengamatan. Koefisien korelasi *Cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matrix ketidakmiripan (matrix jarak

mahalanobis) dan elemen-elemen yang dihasilkan oleh dendogram (matrix *Cophenetic* berdasarkan ukuran jarak dan metode keterhubungan yang digunakan). Berikut adalah persamaan korelasi *cophenetic*:

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{Cik} - \bar{d}_C)}{\sqrt{\left[\sum_{i < k} (d_{ik} - \bar{d})^2 \right] \left[\sum_{i < k} (d_{Cik} - \bar{d}_C)^2 \right]}}$$

Keterangan:

- r_{coph} : Koefisien korelasi *cophenetic*
- d_{ik} : Jarak *Euclidean* objek ke-*i* dan ke-*k*
- \bar{d} : Rata-rata d_{ik}
- d_{Cik} : Jarak *Cophenetic* objek ke-*i* dan ke-*k*
- \bar{d}_C : Rata-rata d_{Cik}

Nilai korelasi *Cophenetic* yang semakin mendekati 1 berarti solusi yang dihasilkan dari proses gerombol semakin baik. Ukuran jarak *euclidean* dengan metode *Single Linkage*, *Average Linkage*, dan *complete linkage* menghasilkan nilai *Cophenetic* terbesar merupakan solusi terbaik pada metode bergerombol berhierarki (Munthe, 2019).

7. Davies Bouldin Indeks

Davies Bouldin Indeks (DBI) merupakan metode untuk mengecek hasil *clustering*. Pendekatan pengujian nilai DBI berupa nilai separasi dan kohesi. Kohesi berupa jumlah dari kemiripan data terhadap pusat *cluster* dari *cluster* tersebut. Separasi adalah jarak antara pusat *cluster* dari *cluster*. *Cluster* yang optimal adalah *cluster* yang memiliki nilai separasi yang tinggi dan nilai kohesi yang rendah. Nilai *Davies Bouldin Indeks* (DBI) yang semakin mendekati nilai 0 menandakan semakin baik *cluster* yang diperoleh. Semakin rendah nilai DBI menunjukkan hasil *cluster* yang optimal (Wirawan et al., 2016).

Sum of Square Within Cluster (SSW) adalah persamaan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-*i* yang dapat dilihat pada Persamaan:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Keterangan:

- m_i : Jumlah Data dalam *Cluster* ke-*i*
- c_i : *Centroid Cluster* ke-*i*
- $d(x_j, c_i)$: Jarak *Eulidean* Setiap Data ke *Centroid*

Sum of Square Between Cluster (SSB) adalah persamaan untuk mengetahui nilai separasi antara *cluster* yang dapat dilihat pada Persamaan:

$$SSB_{i,j} = d(c_i, c_j)$$

Keterangan :

- $d(c_i, c_j)$: Jarak Antar *Centroid*

Setelah nilai nilai separasi dan kohesi diperoleh, selanjutnya dilakukan

pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *cluster* ke- i dan ke- j , menggunakan Persamaan:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Persamaan untuk menghitung nilai *Davies Bouldin Indeks* (DBI) adalah sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j})$$

Keterangan:

K : Jumlah *Cluster* yang digunakan

Semakin rendah nilai *Davies Bouldin Indeks* (DBI) yang diperoleh, maka semakin baik kualitas *cluster* yang diperoleh dari suatu *clustering* data.

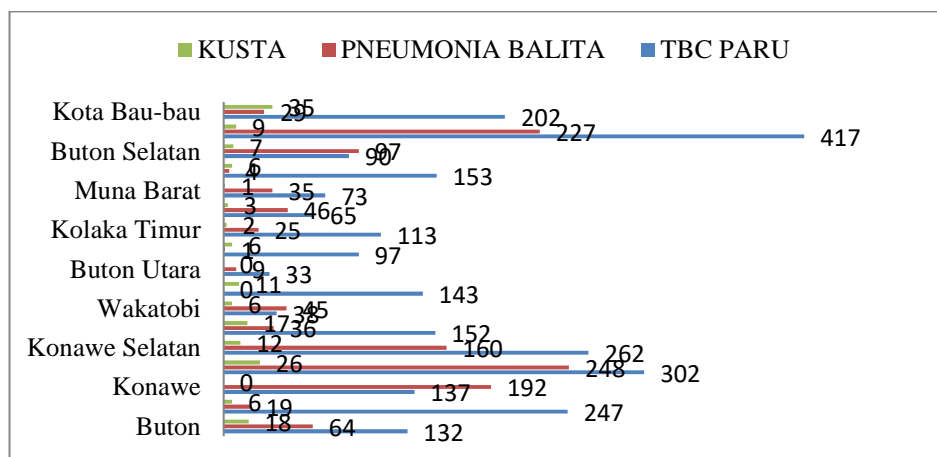
HASIL DAN PEMBAHASAN

Deskripsi variabel penelitian

Data yang digunakan dalam penelitian ini adalah Data Sekunder yang diperoleh dari Dinas Kesehatan Provinsi Sulawesi Tenggara sebanyak 10 variabel bebas yang merupakan jenis penyakit di provinsi Sulawesi Tenggara tahun 2020. Berikut adalah Analisis Deskriptif untuk menggambarkan karakteristik Jumlah Kasus Penyakit per Kabupaten/Kota se-Sulawesi Tenggara Tahun 2020.

1. Penyakit Menular Langsung

Pada penelitian ini, jenis penyakit menular langsung terdiri dari penyakit kusta, penyakit pneumonia pada balita, dan penyakit TBC paru. Berikut adalah Analisis Deskriptif untuk menggambarkan karakteristik Jumlah Kasus Penyakit menular langsung.

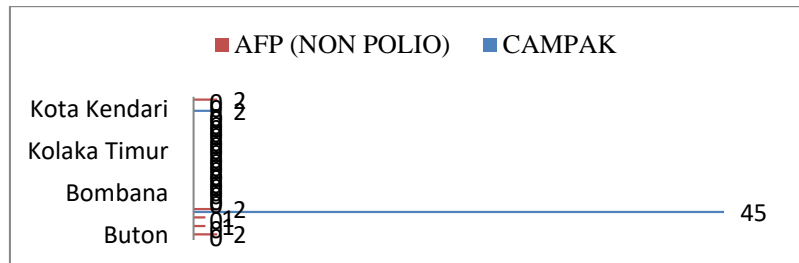


Gambar 1 Jumlah Kasus Penyakit Menular Langsung

Gambar 1 merupakan grafik yang dapat menggambarkan kondisi kesehatan di Provinsi Sulawesi Tenggara tahun 2020 berdasarkan jumlah kasus Penyakit menular langsung. Penyakit Kusta yang paling banyak adalah di Kota Bau-Bau sebanyak 102 kasus. Penyakit Pneumonia Paru yang paling banyak adalah di Kabupaten Kolaka sebanyak 248 kasus. Penyakit TBC Paru yang paling banyak adalah di Kota Kendari sebanyak 417 kasus sedangkan kasus penyakit yang terendah adalah di Kabupaten Buton Utara sebanyak 33 kasus.

2. Penyakit Yang Dapat Dicegah Dengan Imunisasi

Pada penelitian ini, jenis penyakit yang dapat dicegah dengan Imunisasi terdiri dari penyakit AFP (non polio), dan penyakit campak. Berikut adalah Analisis Deskriptif untuk menggambarkan karakteristik Jumlah Kasus Penyakit yang dapat dicegah dengan Imunisasi.

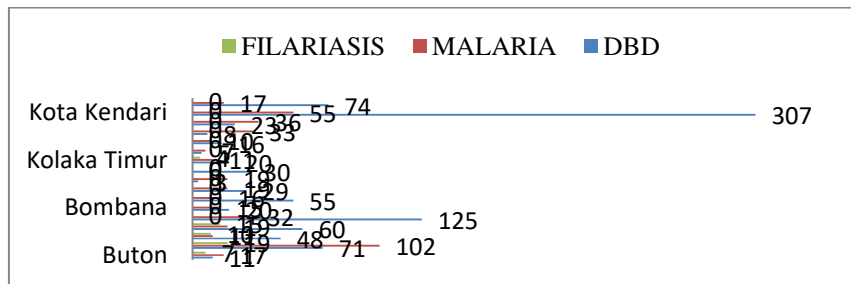


Gambar 2 Jumlah Kasus Penyakit yang dapat dicegah dengan Imunisasi

Gambar 2 merupakan grafik yang dapat menggambarkan kondisi kesehatan di Provinsi Sulawesi Tenggara tahun 2020 berdasarkan Jumlah Kasus Penyakit yang dapat dicegah dengan Imunisasi. Penyakit AFP (non-Polio) ditemukan 9 kasus yaitu 2 di Kota Bau-Bau, 2 di Kabupaten Kolaka, 1 di Kabupaten Konawe, 1 di Kabupaten Muna dan 2 di Kabupaten Buton. Penyakit Campak hanya terdapat kasus di dua kabupaten kota saja, yaitu kabupaten kolaka sebanyak 45 kasus dan kota kendari sebanyak 2 kasus.

3. Penyakit Tular Vektor dan Zoonotik

Pada penelitian ini, jenis penyakit Tular Vektor dan Zoonotik terdiri dari penyakit filariasis, penyakit malaria, dan penyakit DBD. Berikut adalah Analisis Deskriptif untuk menggambarkan karakteristik Jumlah Kasus Penyakit Tular Vektor dan Zoonotik.

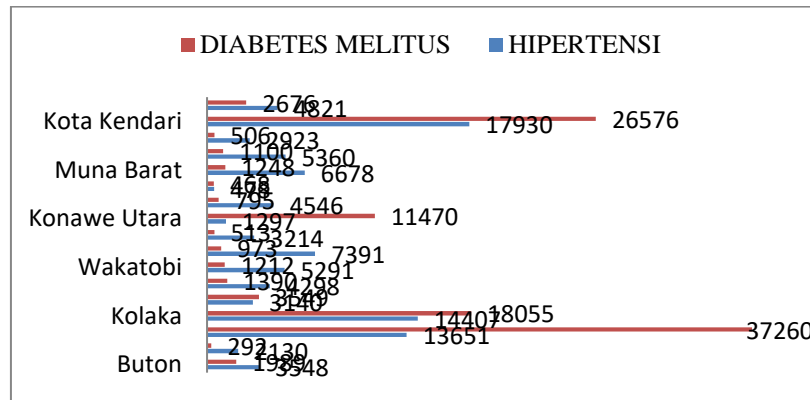


Gambar 3 Jumlah Kasus Penyakit Tular Vektor dan Zoonotik

Gambar 3 merupakan grafik yang dapat menggambarkan kondisi kesehatan di Provinsi Sulawesi Tenggara tahun 2020 berdasarkan Jumlah Kasus Penyakit Tular Vektor dan Zoonotik. Penyakit Filariasis yang paling banyak adalah di Kabupaten Muna sebanyak 19 kasus. Penyakit Malaria yang paling banyak adalah di Kabupaten Muna sebanyak 102 kasus. Penyakit DBD yang paling banyak adalah di Kota Kendari sebanyak 307 kasus sedangkan kasus penyakit yang terendah adalah di Kabupaten Buton Utara 3 kasus dan Konawe Kepulauan sebanyak 5 kasus.

4. Penyakit Tular Vektor dan Zoonotik

Pada penelitian ini, jenis penyakit Tular Vektor dan Zoonotik terdiri dari penyakit filariasis, penyakit malaria, dan penyakit DBD. Berikut adalah Analisis Deskriptif untuk menggambarkan karakteristik Jumlah Kasus Penyakit Tular Vektor dan Zoonotik.



Gambar 4 Jumlah Kasus Penyakit Tidak Menular

Gambar 4 merupakan grafik yang dapat menggambarkan kondisi kesehatan di Provinsi Sulawesi Tenggara tahun 2020 berdasarkan Jumlah Kasus Penyakit tidak menular. Penyakit Hipertensi yang paling banyak adalah di Kota Kendari sebanyak 17.930 kasus sedangkan Penyakit yang terendah adalah di Konawe Kepulauan sebanyak 478 kasus. Penyakit Diabetes Melitus di Provinsi Sulawesi Tenggara tahun 2020 yang paling banyak adalah di Kabupaten Konawe sebanyak 37.260 kasus sedangkan Penyakit yang terendah adalah di Kabupaten Muna sebanyak 292 kasus.

Analisis Data

1. Multikolinearitas

Uji *Bartlets* digunakan untuk mengetahui ada atau tidaknya hubungan antar variabel dalam kasus multivariat. Berikut hipotesisnya:

Hipotesis:

H_0 : Tidak ada korelasi antar variabel

H_1 : Ada korelasi antar variabel

Pengambilan keputusan:

Tolak H_0 bila $\chi^2_{hitung} > \chi^2_{0.05; (1/2(45))(45-1)}$ maka variabel saling berkorelasi, hal ini berarti terdapat hubungan antar variabel.

Pada penelitian ini, nilai χ^2_{hitung} sebesar 136.164 dengan derajat bebas 45, nilai tersebut lebih besar dibandingkan dengan nilai χ^2_{tabel} yaitu sebesar 61.66. sehingga dapat disimpulkan tolak H_0 yang berarti bahwa terdapat hubungan atau korelasi diantara variabel penelitian.

2. Principal Analysis Component (PCA)

Hasil analisis menunjukkan adanya multikolinearitas pada variabel penelitian sehingga untuk melakukan analisis lanjutan yaitu analisis cluster, maka terlebih dahulu dilakukan reduksi data dengan melakukan PCA (*principal analysis component*). PCA dapat mengubah variabel asli yang saling berkorelasi menjadi satu set variabel baru yang saling bebas (tidak berkorelasi signifikan) dan dapat mereduksi dimensi data menjadi lebih kecil untuk memudahkan dalam melakukan interpretasi tanpa adanya informasi data yang hilang.

Tabel 2. Nilai *Eigen*

Komponen	Nilai Eigen	Total Varians (%)	Total Kumulatif (%)
1	4.445	44.45	44.45
2	1.936	19.364	63.814
3	1.466	14.657	78.471
4	1.107	11.066	89.537

Berdasarkan Tabel 2 maka Terdapat 4 komponen yang mempunyai nilai *Eigen* lebih dari 1. Masing-masing komponen dapat menjelaskan keragaman data sebesar 44.450, 19.364, 14.657 dan 11.066 dengan total varians sebesar 89.537%. Artinya 4 faktor yang terbentuk sudah mewakili 10 variabel indikator Penyakit yang menjelaskan tentang kondisi kesehatan di Provinsi Sulawesi Tenggara sebesar 89.537%.

Tabel 3. Komponen yang Terbentuk

Kab/Kota	PC_1	PC_2	PC_3	PC_4
Buton	-0.631	1.109	-0.332	0.671
Muna	-1.286	-0.617	2.475	2.401
Konawe	1.976	-1.195	-1.011	1.159
Kolaka	1.692	1.977	-0.709	1.854
Konawe Selatan	0.133	0.226	0.951	-0.922
Bombana	-0.439	0.459	-0.173	-0.693
Wakatobi	-0.213	-0.426	-0.44	-0.426
Kolaka Utara	-0.322	0.02	-0.119	-0.559
Buton Utara	-0.527	-0.83	-0.605	-0.016
Konawe Utara	-0.176	-0.379	-0.793	-0.445
Kolaka Timur	-0.301	-0.636	-0.427	0.104
Konawe Kepulauan	-0.503	-0.489	-0.731	-0.231
Muna Barat	-0.097	-0.692	-0.689	-0.247
Buton Tengah	-0.511	-0.365	0.131	-0.201
Buton Selatan	-0.411	-0.39	0.055	-0.161
Kota Kendari	2.295	-0.28	2.201	-1.441
Kota Bau-bau	-0.679	2.508	0.217	-0.847

Berdasarkan Tabel 3, maka 4 komponen utama yang terbentuk dan akan dipakai untuk analisis lanjutan yaitu Analisis *Cluster Hierarki*.

3. Jarak *Euclidian*

Misalkan dihitung kemiripan antara Kabupaten Buton dan Kabupaten Muna (objek 1 dan 2) dengan Kabupaten Buton dan Kabupaten Konawe (objek 1 dan 3) sebagai berikut:

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

$$d_{1,2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2}$$

$$= \sqrt{(-0.631 - (-1.286))^2 + (1.109 - (-0.617))^2 + (-0.332 - 2.475)^2 + (0.671 - 2.401)^2}$$

$$= 3.779$$

$$d_{1,3} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2}$$

$$= \sqrt{(-0.631 - 1.976)^2 + (1.109 - (-1.195))^2 + (-0.332 - (-1.011))^2 + (0.671 - 1.159)^2}$$

$$= 3.579$$

Berdasarkan hasil perhitungan jarak antara objek 1 dengan 2 dan objek 1 dengan 3, yang memiliki nilai jarak paling dekat adalah objek 1 dengan 3 yaitu Kabupaten

Buton dan Kabupaten Konawe dengan nilai jarak 3.579. Hal ini menandakan bahwa dari ketiga objek tersebut, objek 1 (Kabupaten Buton) memiliki karakteristik tingkat kesehatan yang lebih mirip dengan objek 3 (Kabupaten Konawe) dibandingkan dengan objek 2 (Kabupaten Muna).

4. Perbaikan Jarak

- Perbaikan matriks jarak pada metode *Single Linkage* dengan menggunakan Persamaan single linkage, Jarak yang melibatkan *cluster* baru mengalami perubahan dengan perhitungan sebagai berikut:

$$d_{(14,15)l} = \min \{d_{(14)l}, d_{(15)l}\} = \min \{1.7781, 1.7707\} = 1.7707$$

- Perbaikan matriks jarak dengan metode *Average Linkage* menggunakan Persamaan average linkage, Jarak yang melibatkan *cluster* baru mengalami perubahan dengan perhitungan sebagai berikut:

$$d_{(14,15)l} = \frac{\{d_{(14)l} + d_{(15)l}\}}{2} = \frac{\{1.778 + 1.770\}}{2} = 1.774$$

- Perbaikan matriks jarak dengan metode *Complete Linkage* menggunakan Persamaan complete linkage, Jarak yang melibatkan cluster baru mengalami perubahan dengan perhitungan sebagai berikut:

$$d_{(14,15)l} = \max \{d_{(14)l}, d_{(15)l}\} = \max \{1.778, 1.771\} = 1.778$$

Seterusnya perhitungan dilakukan sampai perhitungan perbaikan matriks jarak hingga semua objek yang telah digabungkan pada proses *Agglomeration Schedule* telah dilakukan perbaikan.

5. Pemilihan Metode Cluster Terbaik

Setelah dilakukan proses *clustering* pada pengelompokkan kabupaten/kota berdasarkan jenis Penyakit dengan metode *Cluster* Hierarki, maka akan ditentukan metode *cluster* terbaik pada ketiga metode dengan melihat nilai Koefisien Korelasi *Cophenetic*. Berikut hasil perhitungan nilai Korelasi *Cophenetic* dengan menggunakan Persamaan.

Metode	Nilai Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0.985
<i>Average Linkage</i>	0.990
<i>Complete Linkage</i>	0.979

Berdasarkan Tabel 5, nilai korelasi *cophenetic* metode terbaik untuk mengelompokkan kab/kota berdasarkan jenis penyakit di Sulawesi tenggara tahun 2020 adalah metode *Average Linkage*.

6. Penentuan Cluster Optimum

Setelah diketahui metode cluster terbaik. Selanjutnya akan dilakukan penentuan cluster optimum. pada penelitian ini akan dilakukan percobaan pada 2, 3, 4, dan 5 cluster. Sehingga untuk mengetahui jumlah cluster terbaik yang akan digunakan akan dilakukan

penentuan cluster optimum menggunakan metode davies bouldin indeks (DBI) sebagai berikut:

Tabel 6. Nilai DBI

Percobaan	Nilai DBI
2 Cluster	1.71594
3 Cluster	1.38709
4 Cluster	0.40028
5 Cluster	2.5583

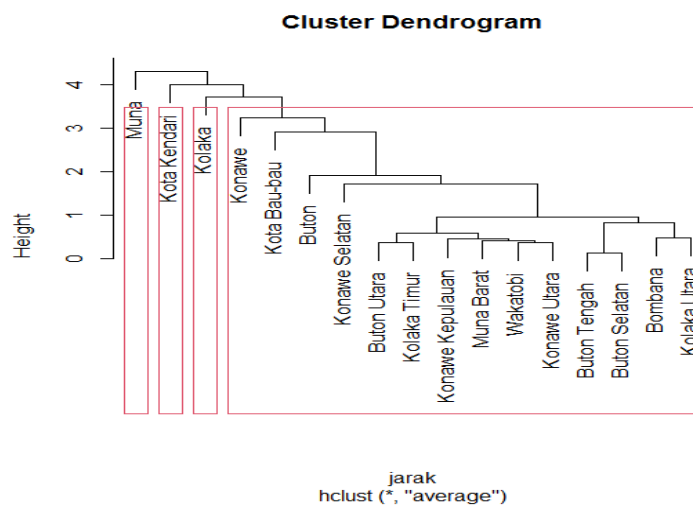
Berdasarkan Tabel 6, Nilai DBI pada masing-masing percobaan *cluster*. Semakin rendah nilai DBI maka akan semakin baik pula kualitas *cluster* yang diperoleh dari suatu *clustering* data. sehingga berdasarkan nilai DBI yang diperoleh, maka jumlah *cluster* optimum adalah 4 *cluster*.

7. Hasil Clustering dengan Metode Cluster Terbaik Average Linkage

Hasil pengelompokkan Kabupaten/Kota berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020 menggunakan metode *Cluster Average Linkage* dengan nilai Korelasi *Cophenetic* yang di hasilkan sebesar 0.9219 dan jumlah *cluster* optimum yaitu 4 *cluster* dapat dilihat pada tabel berikut.

Tabel 7. Hasil Cluster dengan Metode Average Linkage

Kelompok	Jumlah Anggota	Kabupaten/Kota
1	14	Konawe, Kota Bau-Bau, Buton, Konawe Selatan, Buton Utara, Kolaka Utara, Konawe Kepulauan, Muna Barat, Wakatobi, Konawe Utara, Buton Tengah, Buton Selatan, Bombana, Kolaka Utara
2	1	Muna
3	1	Kolaka
4	1	Kota Kendari



Gambar 5. Dendrogram *Cluster Average Linkage*

Setelah didapatkan jumlah *cluster* optimum pada metode *cluster average linkage*, selanjutnya akan dilihat rata-rata kasus penyakit pada setiap *cluster* untuk memudahkan dalam interpretasi hasil cluster.

Tabel 8. Rata-rata Kasus Penyakit Setiap *Cluster*

Jenis Penyakit	Variabel	Cluster			
		1	2	3	4
Penyakit Menular Langsung	TBC Paru	121	247	302	417
	Pneumonia Balita	53	19	248	227
	Kusta	9	6	26	9
Penyakit Yang Dapat Dicegah Dengan Imunisasi	AFP (Non Polio)	0	1	2	0
	Campak	0	0	45	2
Penyakit Tular Vektor Dan Zoonotik	DBD	33	71	60	307
	Malaria	17	102	19	55
	Filariasis	2	19	15	0
Penyakit Tidak Menular	Hipertensi	4760	2130	14407	17930
	Diabetes Melitus	4654	292	18055	26576

Keterangan:

	Kabupaten/Kota dengan Rata-Rata Kasus Rendah
	Kabupaten/Kota dengan Rata-Rata Kasus Sedang
	Kabupaten/Kota dengan Rata-Rata Kasus Tinggi
	Kabupaten/Kota dengan Rata-Rata Kasus Sangat Tinggi

Interpretasi hasil *Cluster*:

➤ *Cluster 1*

Berdasarkan Tabel 4.1 dapat dilihat bahwa pada *Cluster 1* tidak terdapat Penyakit dengan rata-rata kasus Tinggi. Sehingga berdasarkan karakteristik ini dapat dikatakan bahwa Kabupaten/Kota yang masuk kedalam *Cluster 1* merupakan daerah dengan Tingkat Kesehatan yang Baik.

➤ *Cluster 2*

Berdasarkan Tabel 4.1 dapat dilihat bahwa Penyakit dengan rata-rata kasus Sangat Tinggi yaitu Penyakit Malaria, dan Filariasis. dimana kedua Penyakit ini masuk ke dalam jenis Penyakit Tular Vektor dan Zoonotik. Sehingga berdasarkan karakteristik tersebut dapat dikatakan bahwa Kabupaten/Kota yang masuk kedalam *cluster 2* merupakan Daerah dengan Tingkat Kesehatan yang Cukup Baik.

➤ *Cluster 3*

Berdasarkan Tabel 4.1 dapat dilihat bahwa pada bahwa penyakit dengan rata-rata kasus sangat tinggi yaitu penyakit Pneumonia Balita, Campak, dan Kusta. pada cluster ini terdapat 3 jenis Penyakit dengan rata-rata kasus sangat tinggi yaitu penyakit menular langsung (Pneumonia Balita dan Kusta) dan penyakit yang dapat dicegah dengan imunisasi. Sehingga berdasarkan karakteristik tersebut dapat dikatakan bahwa Kabupaten/Kota yang masuk kedalam *cluster 3* merupakan daerah dengan tingkat kesehatan yang kurang baik.

➤ *Cluster 4*

Berdasarkan Tabel 4.1 dapat dilihat bahwa penyakit dengan rata-rata kasus sangat tinggi yaitu Penyakit TBC Paru, DBD, Hipertensi dan Diabetes Melitus. Pada *cluster* ini terdapat 3 jenis Penyakit dengan rata-rata kasus Sangat Tinggi yaitu Penyakit Menular Langsung, Penyakit yang dapat dicegah dengan Imunisasi dan Penyakit Tidak Menular.

Sehingga berdasarkan karakteristik tersebut dapat dikatakan bahwa Kabupaten/Kota yang masuk kedalam *cluster* 4 merupakan daerah dengan Tingkat kesehatan yang Tidak Baik.

KESIMPULAN

Adapun kesimpulan pada penelitian ini adalah:

1. Berdasarkan hasil pembahasan, didapatkan nilai Korelasi *Cophenetic* tertinggi (mendekati 1) pada metode *Cluster Average Linkage*, Sehingga dapat ditarik kesimpulan bahwa metode *Cluster Average Linkage* mampu memberikan solusi *cluster* yang lebih baik bila dibandingkan dengan kedua metode lainnya dalam penerapannya untuk pengelompokan Kabupaten/Kota di Provinsi Sulawesi Tenggara Tahun 2020 berdasarkan Jenis Penyakit.
2. Interpretasi *Cluster* yang terbentuk dengan menggunakan metode *Cluster* terbaik yaitu metode *Cluster Average Linkage* dengan jumlah *cluster* optimum 4 *cluster* pada Data Jumlah Penyakit di Kabupaten/Kota Provinsi Sulawesi Tenggara tahun 2020 adalah sebagai berikut:
 - *Cluster* 1 merupakan daerah dengan tingkat kesehatan yang baik. Karena rata-rata kasus penyakit pada *cluster* ini adalah sedang dan rendah. Daerah yang masuk pada *cluster* 1 adalah Kabupaten Buton, Konawe, Konawe Selatan, Bombana, Wakatobi, Kolaka Utara, Buton Utara, Konawe Utara, Kolaka Timur, Konawe Kepulauan, Muna Barat, Buton Tengah, Buton Selatan, dan Kota Bau-Bau.
 - *Cluster* 2 merupakan daerah dengan tingkat kesehatan yang cukup baik. jenis penyakit dengan rata-rata kasus sangat tinggi yaitu Penyakit Tular Vektor dan Zoonotik . Daerah yang masuk pada *cluster* 2 adalah Kabupaten Muna.
 - *Cluster* 3 merupakan daerah dengan tingkat kesehatan yang kurang baik. Pada *cluster* ini terdapat dua jenis penyakit dengan rata-rata kasus sangat tinggi yaitu penyakit menular langsung dan penyakit yang dapat dicegah dengan imunisasi. Daerah yang masuk pada *cluster* 3 adalah Kabupaten Kolaka.
 - *Cluster* 4 merupakan daerah dengan tingkat kesehatan yang tidak baik. Karena pada *cluster* ini terdapat 3 jenis penyakit dengan rata-rata kasus sangat tinggi yaitu Penyakit Menular Langsung, Penyakit yang dapat dicegah dengan Imunisasi dan Penyakit Tidak Menular. Daerah yang masuk pada *cluster* 4 adalah kota kendari.

DAFTAR PUSTAKA

- Afandi, M., I. 2015. Analisis cluster hierarki dengan metode complete linkage pada kabupaten/kota di provinsi jawa timur berdasarkan indikator kemiskinan [Skripsi]. Malang: Universitas Islam Negeri Maulana Malik Ibrahim, Fakultas Sains Dan Teknologi.
- Awaliyah, R. 2018. Analisis *Clustering* untuk Mengelompokkan Tingkat Kesejahteraan Kabupaten/Kota Berdasarkan Sosial Ekonomi Rumah Tangga di Wilayah Provinsi Sulawesi Selatan [Skripsi]. Makassar: Universitas Islam Negeri (UIN) Alauddin Makassar, Fakultas Sains dan Teknologi.
- Dinas Kesehatan Provinsi Sulawesi Tenggara 2020. *Profil Kesehatan Sulawesi Tenggara Tahun 2020*. Kendari: Dinkes Sultra.
- Fathia, N. A., Rahmawati, R. & Tarno. Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode *Wards* Dan *Single Linkage*. *Jurnal Gaussian*, Volume 5, No.4 Tahun 2016.
- Johnson & Wichern. 2007. *Applied Multivariate Statistical Analysis*. USA: Pearson Prentice Hall.

- Munthe, D. A. 2019. Penerapan *Clustering Time Series* Untuk Menggerombolkan Provinsi di Indonesia Berdasarkan Nilai Produksi Padi. *Jurnal Litbang Sukowati*. Volume 2 No. 2 Tahun 2019.
- Rahmayani, M. T. I. 2018. Analisis *Cluster* Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma *K-Means*. *Jurnal JITI* Volume 1 No. 2.
- Ulinuh, N. & Veriani, R. 2020. Analisis *Cluster* pada Pengelompokkan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular menggunakan Metode *Complete Linkage, Average Linkage* dan *Ward*. *Jurnal Nasional Informatika Dan Teknologi Jaringan*, Volume 5 No.1.
- Utami, N. D. 2017. Perbandingan Hasil Pengelompokkan antara Metode *Average Linkage, Complete Linkage* dan *Single Linkage* [Skripsi]. Yogyakarta: Universitas Islam Indonesia, Fakultas Matematika dan Ilmu Pengetahuan Alam.
- Wirawan, M. F. Fahmi, & Y. K. Suprpto, “*segmentation and distribution of watershed usinf k-modes clusteruing algorithm and davies boulding indeks on geographic information system (GIS)*,” in *proceeding 2016 international seminar on application of technology for information and communication, ISEMANTIC*. 2016, 2017.