

**SKRIPSI**

**PENGELOMPOKAN PUSKESMAS DI KABUPATEN  
POLEWALI MANDAR TAHUN 2023 BERDASARKAN  
INDIKATOR GIZI BURUK BALITA DENGAN MODEL  
*AGGLOMERATIVE HIERARCHICAL CLUSTERING***



**NURUL ZALFA LATIF  
E0220307**

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS SULAWESI BARAT  
TAHUN 2024**

## HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Nurul Zalfa Latif

NIM : E0220307

Judul : Pengelompokan Puskesmas di Kabupaten Polewali Mandar Tahun 2023 Berdasarkan Indikator Gizi Buruk Balita dengan Model *Agglomerative Hierarchical Clustering*

Telah berhasil dipertanggungjawabkan di hadapan Tim Penguji (SK Nomor 48/UN55.7/HK.04/2024, tanggal 10 Juli 2024) dan diterima sebagai bagian persyaratan memperoleh gelar Sarjana S1 Statistika pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat.

Disahkan oleh:

Dekan FMIPA

Universitas Sulawesi Barat



\*Musafira S.Si., M.Sc.

NIP. 197709112006042002

Tim Penguji:

Ketua Penguji : Musafira, S.Si., M.Sc.

(.....)

Sekretaris : Muh. Hijrah, S.Pd., M.Si.

(.....)

Pembimbing 1: Hikmah, S.Sd., M.Sc.

(.....)

Pembimbing 2: Nur Hilal A.Syahrir, S.Si., M.Si.

(.....)

Penguji 1 : Darma Ekawati, S.Pd., M.Sc.

(.....)

Penguji 2 : Muh. Hijrah, S.Pd., M.Si.

(.....)

Penguji 3 : Retno Mayapada, S.Si., M.Si.

(.....)

## ABSTRAK

Keadaan gizi yang baik merupakan salah satu faktor penting dalam upaya mencapai derajat kesehatan yang optimal khususnya pada balita yang mengalami *stunting* dan gizi buruk akan lebih rentan terhadap penurunan daya tahan tubuh dan perkembangan yang tidak optimal. Berdasarkan wilayah, Kabupaten Polewali Mandar menempati peringkat kedua wilayah dengan prevalensi balita *stunting* terbesar di Sulawesi Barat sebesar 39,3%. Tujuan dari penelitian ini adalah untuk mengelompokkan puskesmas berdasarkan indikator gizi buruk balita di Kabupaten Polewali Mandar. Pada penelitian ini diterapkan metode *agglomerative hierarchical clustering* untuk pengelompokan puskesmas. Dari 20 puskesmas di Kabupaten Polewali Mandar diperoleh hasil nilai korelasi *cophenetic* metode *average linkage* sebesar 0,746 yang paling mendekati 1 dan diperoleh nilai indeks *connectivity* dan indeks *silhouette* dengan jumlah *cluster* 2 dipilih sebagai metode pengelompokan terbaik diantaranya *cluster* 1 dengan anggota kelompok puskesmas yang tingkat gizi buruk tinggi. Sebaliknya, *cluster* 2 dengan anggota kelompok puskesmas yang tingkat gizi buruk rendah.

**Kata Kunci** : Gizi Buruk Balita, *Agglomerative Hierarchical Clustering*, *Average Linkage*.

## **ABSTRACT**

*Good nutrition is one of the important factors in achieving optimal health status, especially in toddlers who are stunted and malnourished will be more vulnerable to decreased endurance and suboptimal development. Based on region, Polewali Mandar Regency ranks second in the region with the highest prevalence of stunted toddlers in West Sulawesi at 39.3%. The purpose of this study is to group health centers based on indicators of malnutrition among children under five in Polewali Mandar Regency. In this study, the agglomerative hierarchical clustering method was applied to group health centers. From the 20 health centers in Polewali Mandar District, the results of the average linkage method cophenetic correlation value of 0.746 are closest to 1 and the obtained connectivity index value and silhouette index with the number of clusters 2 are selected as the best clustering method including cluster 1 with members of the health center group with high malnutrition rates. Conversely, cluster 2 with members of the puskesmas group with low malnutrition rates.*

**Keywords** : *Malnutrition of Toddlers, Agglomerative Hierarchical Clustering, Average Linkage.*

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Keadaan gizi yang baik merupakan salah satu faktor penting dalam upaya mencapai derajat kesehatan yang optimal khususnya pada balita. Namun, masalah gizi buruk yang dialami balita dapat mengganggu proses tersebut. Balita yang mengalami gizi buruk akan lebih rentan terhadap penurunan daya tahan tubuh dan perkembangan yang tidak optimal, bahkan dapat menyebabkan kematian. Menurut Riset Kesehatan Dasar (2018), balita adalah anak berumur 0 sampai dengan 59 bulan yang sangat membutuhkan suplai makanan dan gizi sesuai dengan kebutuhan untuk membantu proses pertumbuhan dan perkembangan. Selain itu, gizi buruk pada balita juga dapat mengganggu pertumbuhan dan perkembangan mental pada anak.

Faktor penyebab gizi buruk dapat dikelompokkan menjadi 2 yaitu penyebab langsung dan penyebab tidak langsung. Penyebab langsung gizi buruk meliputi kurangnya jumlah dan kualitas makanan yang dikonsumsi dan menderita penyakit infeksi, sedangkan penyebab tidak langsung gizi buruk yaitu ketersediaan pangan rumah tangga dan kemiskinan. Riwayat Berat Badan Lahir Rendah (BBLR) juga merupakan faktor yang dapat berpengaruh terhadap kejadian gizi buruk. Hal ini dikarenakan bayi yang mengalami BBLR akan mengalami komplikasi penyakit karena kurang matangnya organ, menyebabkan gangguan pertumbuhan fisik dan gangguan gizi saat balita (Oktavia dkk., 2017). Gizi buruk merupakan kelainan gizi yang dapat berakibat fatal pada kesehatan balita. Kejadian gizi buruk ini apabila tidak diatasi akan menyebabkan dampak yang buruk bagi balita. Gizi buruk akan menimbulkan dampak hambatan bagi pertumbuhan anak (Wahyudi dkk., 2015).

Sulawesi Barat merupakan provinsi dengan prevalensi balita *stunting* tertinggi kelima di Indonesia pada 2023. Menurut hasil Survei Status Gizi Indonesia (SSGI) Kementerian Kesehatan, prevalensi balita *stunting* di provinsi ini sebesar 30,3% pada tahun 2023. Selain peringkat kelima nasional, angka *stunting* di Sulawesi Barat berada di bawah ambang batas yang ditetapkan standar Organisasi Kesehatan Dunia (WHO) sebesar 20%. Ini mengindikasikan bahwa penanganan

*stunting* di Sulawesi Barat masih buruk. Berdasarkan wilayahnya, Kabupaten Polewali Mandar menempati peringkat kedua wilayah dengan prevalensi balita *stunting* terbesar di Sulawesi Barat sebesar 39,3%.

Upaya yang dilakukan untuk mengatasi gizi buruk salah satunya dengan merekomendasikan perencanaan program dari puskesmas untuk mengatasi gizi buruk. Oleh karena itu, perlu dilakukan analisis data *cluster* untuk mengelompokkan tiap-tiap puskesmas berdasarkan indikator gizi buruknya. Diharapkan hasil tersebut dapat digunakan oleh pemerintah untuk membuat program yang tepat sasaran sesuai dengan karakteristik atau kondisi puskesmas.

Analisis *cluster* atau analisis kelompok merupakan teknik analisis data yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relatif (Talaku dkk 2017). Dalam *clustering* pengumpulan objek yang memiliki kemiripan satu sama lain menempati tempat atau posisi berdekatan. Analisis *cluster* juga dilakukan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Secara umum, analisis *cluster* dibagi menjadi dua, yaitu *hierarchical clustering methods* (metode *clustering* hierarki) dan *nonhierarchical clustering methods* (metode *clustering* nonhierarki). Metode *clustering* hierarki terdiri atas dua bagian, yaitu metode *agglomerative* (penyatuan) dan *divisive* (penyebaran). Metode *agglomerative* dikenal beberapa metode untuk membentuk *cluster*, di antaranya yaitu metode *single linkage*, *complete linkage*, *average linkage*, dan *ward*.

Beberapa penelitian sebelumnya menggunakan metode *agglomerative hierarchical clustering* diantaranya yaitu penelitian yang dilakukan oleh Widodo dkk., (2020) yang merupakan metode *average linkage*, *complete linkage*, dan *ward* pada pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan Indikator Indeks Pembangunan Manusia, hasil penelitian menunjukkan bahwa metode *clustering* terbaik adalah *average linkage* dengan nilai korelasi sebesar 0,865 metode hierarki *average linkage* menghasilkan tiga *cluster* dimana pada *cluster* 1 sebanyak 20 kabupaten/kota, *cluster* 2 sebanyak 12 kabupaten/kota dan *cluster* 3 sebanyak 3 kabupaten/kota. Mustafidah dan Purnama., (2024), tentang

pengelompokan kabupaten/kota di Jawa Timur berdasarkan Indikator kasus DBD menggunakan *complete linkage* dan *average linkage*, hasil penelitian menunjukkan adanya empat *cluster*, yakni *cluster* 1 dengan 28 kabupaten/kota, *cluster* 2 dengan 5 kabupaten/kota, *cluster* 3 dengan 4 kabupaten/kota dan *cluster* 4 dengan 1 kabupaten/kota.

Melanjutkan penelitian sebelumnya, dengan adanya kasus gizi buruk di Kabupaten Polewali Mandar maka perlu pengelompokan puskesmas agar masyarakat khususnya yang sudah mempunyai balita agar lebih memperhatikan pola gizi balita. Oleh karena itu maka dibuatkan pengolahan data yang dapat menyelesaikan permasalahan tersebut yaitu pengelompokan puskesmas status gizi buruk menggunakan analisis *cluster*, yang mana metode digunakan adalah *agglomerative hierarchical clustering* yang telah disebutkan sebelumnya, dengan tujuan untuk mengetahui solusi *cluster* terbaik yang dihasilkan atau mengetahui puskesmas yang sangat berdampak pada gizi buruk. Selanjutnya, penulis menuliskan dalam bentuk Tugas Akhir dengan judul “Pengelompokan Puskesmas di Kabupaten Polewali Mandar Tahun 2023 berdasarkan Indikator Gizi Buruk Balita dengan Metode *Agglomerative Hierarchical Clustering*”.

## **1.2 Rumusan Masalah**

Masalah yang di kaji dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana pemilihan model terbaik dan penerapannya dalam metode *agglomerative hierarchical clustering* untuk pengelompokan puskesmas di Kabupaten Polewali Mandar berdasarkan indikator gizi buruk balita Tahun 2023?
2. Bagaimana pembagian jumlah kelompok-kelompok puskesmas yang diperoleh menggunakan metode terbaik *agglomerative hierarchical clustering*?
3. Bagaimana karakteristik pengelompokan puskesmas yang diperoleh menggunakan metode terbaik *agglomerative hierarchical clustering*?

### 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Mengetahui model terbaik dan penerapan dalam metode *agglomerative hierarchical clustering* untuk pengelompokan puskesmas di Kabupaten Polewali Mandar berdasarkan indikator gizi buruk balita tahun 2023.
2. Mengetahui pembagian jumlah kelompok-kelompok puskesmas yang diperoleh menggunakan metode terbaik *agglomerative hierarchical clustering*.
3. Mendeskripsikan karakteristik pengelompokan puskesmas yang diperoleh menggunakan metode terbaik *agglomerative hierarchical clustering*.

### 1.4 Manfaat Penelitian

Adapun manfaat tugas akhir ini adalah sebagai berikut:

1. Bagi instansi, dapat dijadikan sebagai bahan pertimbangan dalam upaya penurunan jumlah gizi buruk dengan harapan dapat meningkatkan kualitas kesehatan khususnya kesehatan gizi pada balita.
2. Bagi peneliti, dapat memberikan kontribusi akademis untuk pengembangan ilmu pengetahuan dalam bidang matematika dan statistika, khususnya yang berkaitan dengan analisis *cluster* metode *agglomerative hierarchical clustering* dan aplikasinya.

### 1.5 Batasan Masalah

Agar mendekati tujuan yang diharapkan, maka perlu dilakukan pembatasan permasalahan, antara lain:

1. Studi kasus penelitian ini dibatasi yaitu 20 puskesmas di Kabupaten Polewali Mandar.
2. Data dalam penelitian ini adalah data sekunder dari Dinas Kesehatan Kabupaten Polewali Mandar yang diolah dengan menggunakan software *Microsoft Excel* dan *R STUDIO*.

## BAB II KAJIAN PUSTAKA

### 2.1 Analisis *Cluster*

Menurut Talakua dkk., (2017), *clustering* adalah proses pengelompokan objek-objek dalam dataset ke dalam beberapa kelompok berdasarkan kesamaan. Analisis *cluster* atau analisis kelompok merupakan teknik analisa data yang bertujuan untuk mengelompokan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relatif homogen. Tujuan analisis *cluster* adalah mengelompokan objek-objek yang memiliki karakteristik yang sama ke dalam *cluster* yang sama.

Menurut Talakua dkk., (2017) analisis *cluster* memiliki beberapa kelebihan dan juga kekurangan sebagai berikut:

- 1) Kelebihan
  - a. Dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak. Data yang direduksi dengan kelompok akan mudah dianalisis.
  - b. Dapat dipakai dalam skala data ordinal, interval dan rasio.
- 2) Kekurangan
  - a. Pengelompokan bersifat subjektifitas peneliti karena hanya melihat dari gambar *dendogram*.
  - b. Untuk data heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang di bentuk.
  - c. Metode-metode yang dipakai memberikan perbedaan yang signifikan, sehingga dalam perhitungan biasanya masing-masing metode dibandingkan.
  - d. Semakin besar observasi, biasanya tingkat kesalahan akan semakin besar.

Adapun ciri-ciri *cluster* adalah *homogenitas* (kesamaan) yang tinggi antara anggota dalam satu *cluster* (*within cluster*) dan *heterogenitas* (perbedaan) yang tinggi antara *cluster* yang satu dengan *cluster* lain (*between-cluster*). Analisis *cluster* memiliki beberapa istilah penting, antara lain:

- 1) *Distances between cluster centers*, yaitu jarak yang menunjukkan bagaimana terpisahnya pasangan individu *cluster*.
- 2) Keanggotaan *cluster* (*cluster membership*), ialah keanggotaan yang menunjukkan *cluster* untuk setiap objek yang menjadi anggotanya.
- 3) Pusat *cluster* (*cluster centers*), ialah titik awal di mulai pengelompokan dalam *cluster* nonhierarki.
- 4) Rata-rata lama *cluster* (*cluster centroid*), ialah nilai rata-rata variabel dari semua objek atau observasi dalam *cluster* tertentu.
- 5) Skedul aglomerasi (*agglomerative schedule*), ialah jadwal yang memberikan informasi tentang objek atau kasus yang dikelompokkan pada setiap tahap pada suatu proses analisis *cluster* yang hierarki.

### **2.1.1 Proses Analisis Cluster**

Menurut Talakua dkk., (2017), analisis *cluster* terdiri dari beberapa proses dasar, yaitu:

- 1) Menetapkan ukuran jarak atau similaritas antar data

Oleh karena tujuan *cluster* ialah untuk mengelompokkan objek yang mirip dalam *cluster* yang sama, maka beberapa ukuran diperlukan untuk mengakses seberapa mirip atau berbeda objek-objek tersebut. Pendekatan yang paling biasa ialah mengukur kemiripan dinyatakan dalam jarak (*distance*) antara pasangan objek. Objek dengan jarak yang lebih pendek antara mereka akan lebih mirip satu sama lain dibandingkan dengan pasangan dengan jarak yang lebih panjang.

- 2) Proses standarisasi data

Proses standarisasi dilakukan apabila di antara variabel yang diteliti terdapat perbedaan ukuran satuan yang besar. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis *cluster* menjadi tidak valid. Untuk itu, perlu dilakukan proses standarisasi dengan melakukan transformasi (standarisasi pada data asli sebelum dianalisis lebih lanjut). Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk Z-skor.

### 3) Melakukan pengelompokan

Setelah data yang dianggap mempunyai satuan yang sangat berbeda diseragamkan, dan metode *cluster* ditentukan (misal dipilih *Euclidian*), langkah selanjutnya adalah pengelompokan data.

### 4) Melakukan interpretasi terhadap *cluster* yang telah terbentuk

Setelah jumlah *cluster* terbentuk dengan metode hierarki atau nonhierarki, langkah selanjutnya adalah melakukan interpretasi terhadap *cluster* yang telah terbentuk. Yang pada intinya memberi nama spesifik untuk menggambarkan isi *cluster* tersebut. Tahap interpretasi meliputi pengujian tiap *cluster* terminologi untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan kealamian *cluster*. Membuat profil dan interpretasi *cluster* tidak hanya untuk memperoleh suatu gambaran saja, melainkan untuk menyediakan rata-rata untuk menilai korespondensi pada *cluster* yang terbentuk serta profil *cluster* memberikan arahan bagi signifikansi praktis.

### 5) Melakukan validasi dan profiling data

#### a. Proses validasi solusi *cluster*

Proses validasi bertujuan menjamin bahwa solusi yang dihasilkan dari *cluster* analisis dapat mewakili populasi dan dapat digeneralisasi untuk objek lain. Pendekatan ini membandingkan solusi *cluster* dan melalui korespondensi hasil.

#### b. Pembuatan profil (*profiling cluster*)

*Profiling cluster* adalah proses untuk menganalisis dan menggambarkan karakteristik masing-masing *cluster* yang telah terbentuk dalam proses *clustering*. Tujuan dari *profiling cluster* adalah memahami perbedaan utama antara *cluster* dan mengidentifikasi fitur atau atribut unik yang mendefinisikan setiap kelompok. Ini sangat penting dalam penelitian, di mana pemahaman mendalam tentang karakteristik masing-masing *cluster* membantu dalam pengambilan keputusan.

## 2.1.2 Metode Analisis *Cluster*

Analisis *cluster*, terdapat banyak metode untuk mengelompokkan observasi ke dalam *cluster*. Secara umum metode pengelompokkan dalam metode hierarki

(*Hierarchical Clustering Method*) dan analisis non hierarki (*Nonhierarchical Clustering Method*). Metode hierarki digunakan apabila belum ada informasi jumlah *cluster* yang dipilih. Sedangkan metode non hierarki bertujuan untuk mengelompokkan  $n$  objek ke dalam  $k$  *cluster* ( $k < n$ ), dimana nilai  $k$  telah ditentukan sebelumnya.

## 2.2 Metode Hierarki

Metode hierarki (*hierarchical method*) yaitu metode yang memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses dilanjutkan ke objek 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 pengelompokan ini disajikan dalam bentuk dendogram, yang mirip dengan “struktur diagram pohon”.

Praktik metode *hierarchical* yaitu penggabungan (*agglomerative*) lebih banyak digunakan karena pembentukan *cluster* yang bersifat ilmiah. *Agglomerative* dimulai dengan menganggap setiap data observasi dengan memasukkan sebagai *cluster* tersendiri. Kemudian mencari pasangan *cluster* yang berbeda disatukan sehingga menjadi satu *cluster* yang baru dengan tetap memperhitungkan jarak kedekatan antar objek. Berikut ini adalah algoritma *cluster* secara *hierarchical clustering* (Mulyaningrum dkk., 2018).

- 1) Menentukan  $K$  *cluster* yang akan dibentuk, dimana  $N$  adalah banyaknya objek yang diamati dan di dalamnya terdapat jarak  $D = \{d_{ij}\}$ .
- 2) Mencari matriks jarak untuk pasangan *cluster* terdekat kemudian tentukan jarak matriksnya.
- 3) Menggabungkan masing-masing *cluster* yang terbukti memiliki jarak kedekatan.
- 4) Mengulangi langkah 2 dan 3, sebanyak  $(n - 1)$  kali hingga semua objek membentuk satu *cluster*.

### 2.2.1 *Agglomerative Hierarchical Clustering*

Metode *agglomerative* dimulai dengan mengasumsikan bahwa setiap objek adalah sebuah *cluster*. Kemudian dua objek dengan jarak terdekat digabungkan menjadi satu *cluster*. Selanjutnya objek ketiga akan bergabung dengan *cluster* yang ada atau bersama objek lain dan membentuk *cluster* baru dengan tetap memperhitungkan jarak kedekatan antar objek. 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*.

#### 2.2.1.1 *Single Linkage*

*Single Linkage* merupakan pengelompokannya berdasarkan jarak minimum. Dimulai dengan dua objek yang memiliki jarak paling dekat, keduanya akan ditempatkan pada *cluster* pertama, langkah ini dilakukan terus sehingga seluruh objek membentuk satu *cluster*. Prosedur pengelompokan *single linkage* pada awalnya dipilih jarak terkecil dalam  $D = dij$  dan menggabungkan objek-objek yang bersesuaian misalnya  $U$  dan  $V$  untuk mendapatkan *cluster* ( $UV$ ). Langkah selanjutnya, jarak di antara ( $UV$ ) dan *cluster* lainnya (Johnson dan Wichern., 2007) dapat dirumuskan sebagai berikut:

$$d_{(UV)W} = \min (d_{uW}, d_{vW}) \quad (2.2)$$

dengan :

$d_{uW}$  = jarak antara tetangga terdekat dari *cluster*  $U$  dan  $W$ .

$d_{vW}$  = jarak antara tetangga terdekat dari *cluster*  $V$  dan  $W$ .

#### 2.2.1.2 *Complete Linkage*

Algoritma metode *complete linkage* hampir sama dengan metode *single linkage*, hanya saja pada metode *complete linkage*, proses pengelompokan didasarkan pada jarak maksimum atau jarak terjauh antar objek. Metode *complete linkage* akan mengelompokkan dua objek yang mempunyai jarak terjauh lebih awal. Metode ini memberikan kepastian bahwa semua objek-objek dalam satu *cluster* berada dalam jarak paling jauh (similaritas terkecil) satu sama lain. Algoritma metode *complete linkage* dimulai dengan menemukan elemen minimum dalam  $D = dij$ , selanjutnya menggabungkan objek-objek yang bersesuaian

misalnya  $U$  dan  $V$  untuk mendapatkan *cluster* ( $UV$ ). Tahap selanjutnya, jarak di antara ( $UV$ ) dan *cluster* lainnya (Johnson dan Wichern., 2007) dapat dirumuskan sebagai berikut:

$$d_{(UV)W} = \max (d_{uw}, d_{vw}) \quad (2.3)$$

dengan :

$d_{uw}$  = jarak antara tetangga terdekat dari *cluster*  $U$  dan  $W$ .

$d_{vw}$  = jarak antara tetangga terdekat dari *cluster*  $V$  dan  $W$ .

### 2.2.1.3 Average Linkage

*Clustering* dengan metode *average linkage* memperlakukan jarak antar dua *cluster* sebagai jarak rata-rata antara semua pasangan objek data dalam satu *cluster* dengan seluruh objek pada *cluster* lain. Prosedur *average linkage* dimulai dengan mendefinisikan matrik  $D = \{d(ij)\}$  untuk memperoleh objek-objek paling dekat, sebagai contoh  $U$  dan  $V$ . Kemudian objek ini digabung ke dalam bentuk *cluster* ( $UV$ ). Selanjutnya, jarak antara ( $UV$ ) dan *cluster* lainnya (Johnson dan Wichern., 2007) dapat dirumuskan sebagai berikut:

$$d_{(UV)W} = \frac{\sum_i \sum_j d_{ij}}{n_{(UV)}n_w} \quad (2.4)$$

dengan :

$d_{ij}$  = jarak antar objek  $i$  pada *cluster* ( $UV$ ) dan objek  $j$  pada *cluster*  $W$ .

$n_{(UV)}$  = banyaknya anggota dalam *cluster*  $UV$ .

$n_w$  = banyaknya anggota dalam *cluster*  $W$

### 2.2.1.4 Ward

Menurut Rachmatin (2014) metode *cluster* bersifat *agglomerative* untuk memperoleh kelompok yang memiliki varian internal sekecil mungkin. Untuk pengelompokan metode *ward*, ukuran yang digunakan adalah *Sum Of Square* (SEE), dapat dirumuskan sebagai berikut:

$$SEE = \sum_{i=1}^n (X_i - \bar{x})'(X_i - \bar{x}) \quad (2.5)$$

dengan:

$X_i$  = vektor obyek ke- $i$ .

$\bar{x}$  = vektor rata-rata semua obyek dalam *cluster*.

Adapun jarak yang digunakan dalam analisis *cluster* adalah jarak *euclidian* (Charles dan Mongi., 2015). Jarak *euclidian* adalah jarak yang paling umum dan paling sering digunakan dalam analisis *cluster*. Jarak *euclidian* antara dua titik dapat terdefiniskan dengan jelas. Jarak yang digunakan adalah peubah kontinu. Dalam metode ini, pengukuran jarak dilakukan dengan menghitung akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing peubah. Jarak *euclidian* dapat dirumuskan sebagai berikut:

$$d_{i,j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

Dengan:

$d_{i,j}$  = jarak antar objek  $i$  ke objek  $k$ .

$x_{ik}$  = nilai objek  $i$  pada peubah ke- $k$ .

$x_{jk}$  = nilai objek  $j$  pada peubah ke- $k$ .

$P$  = banyaknya peubah yang diamati.

Jarak *euclidian* mempunyai kelebihan dibanding dengan metode perhitungan lain berikut adalah kelebihan jarak *euclidian* (Fajriah dkk., 2019):

1. Perhitungan jarak menggunakan *euclidian* lebih umum digunakan dalam metode *clustering*.
2. Menghitung jarak *euclidean* cukup sederhana dari sudut pandang komputasi, karena hanya memerlukan operasi dasar (penjumlahan, pengurangan, dan akar kuadrat). Hal ini memungkinkan penghitungan jarak ini menjadi lebih cepat dibandingkan dengan beberapa ukuran jarak lain yang lebih kompleks.
3. Jarak *euclidean* sangat cocok digunakan untuk data numerik yang berada dalam ruang *euclidean*, di mana perbedaan antara nilai-nilai data merepresentasikan jarak fisik atau logis yang relevan.

### 2.3 Koefisien Korelasi *Cophenetic*

Setelah mendapatkan hasil *cluster* dari proses *clustering* data, hal yang penting untuk dilakukan adalah menguji validitas *cluster*. Uji validitas *cluster* diperlukan untuk melihat kebaikan (*goodness*) atau kualitas (*quality*) hasil analisis *cluster*. Salah satu ukuran yang dapat digunakan untuk menguji validitas hasil *clustering* metode hierarki adalah koefisien korelasi *cophenetic*. Koefisien korelasi *cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matriks ketidakmiripan (matriks jarak *Euclidean*) dan elemen-elemen yang dihasilkan oleh *dendogram* (matriks *cophenetic*) (Da Silva dan Dias., 2013).

Formula untuk menghitung koefisien korelasi *cophenetic* sebagai berikut (Harnanto dkk., 2017).

$$r_{Coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{C_{ik}} - \bar{d}_C)}{\sqrt{\left[ \sum_{i < k} (d_{ik} - \bar{d})^2 \right] \left[ \sum_{i < k} (d_{C_{ik}} - \bar{d}_C)^2 \right]}} \quad (2.6)$$

dengan:

$r_{Coph}$  : koefisien korelasi *cophenetic*.

$d_{ik}$  : jarak asli (jarak *euclidian*) antara objek  $i$  dan  $k$ .

$\bar{d}$  : rata-rata  $d_{ik}$ .

$d_{C_{ik}}$  : jarak *cophenetic* objek  $i$  dan  $k$ .

$\bar{d}_C$  : rata-rata  $d_{C_{ik}}$ .

Nilai  $r_{Coph}$  berkisaran antar -1 dan 1. Nilai 1 : Menunjukkan bahwa *dendogram* adalah representasi sempurna dari jarak asli. Nilai 0 : Menunjukkan bahwa tidak ada hubungan antara *dendogram* dan jarak asli. Nilai -1 : Menunjukkan bahwa *dendogram* berlawanan arah dengan jarak.

Interpretasi koefisien korelasi *cophenetic*:

- Tinggi (dekat 1): Semakin tinggi nilai korelasi *cophenetic*, semakin baik *dendogram* dalam merepresentasikan struktur data asli.
- Rendah (dekat 0 atau negatif): Semakin rendah nilai korelasi *cophenetic*, semakin kurang tepat *dendogram* dalam merepresentasikan struktur data asli.

## 2.4 Validasi Cluster

Validasi *cluster* adalah prosedur yang mengevaluasi hasil analisis *cluster* secara kuantitatif dan objektif. Validasi digunakan untuk mengetahui apakah hasil kelompok *cluster* yang terbentuk mampu menjelaskan dan mewakili populasi secara umum. Validasi *cluster* digunakan untuk memecahkan permasalahan utama yaitu menentukan jumlah kelompok optimum. Kelompok optimum merupakan kelompok yang mempunyai jarak yang padat atau jarak terpendek antar individu atau objek dalam *cluster* dan terisolasi atau memiliki jarak yang jauh dari *cluster* lainnya. Terdapat beberapa pendekatan untuk menentukan jumlah *cluster* optimum yaitu diantaranya adalah indeks *connectivity*, indeks *dunn*, dan *silhouette*. (Irwansyah dan Faisal., 2015), formula dari masing-masing pendekatan indeks tersebut sebagai berikut:

### 1). Indeks *Connectivity*

Indeks *Connectivity* adalah melihat seberapa terhubung suatu titik data dengan titik data tetangga dalam *cluster* yang sama.

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)} \quad (2.7)$$

dengan:

- $Conn(C)$  : Indeks *Connectivity*.
- $X_i$  : jumlah titik data  $i$ .
- $nni(j)$  : pengamatan terdekat  $i$  ke  $j$  dan  $l$ .
- $N$  : banyak pengamatan.
- $L$  : banyak *cluster*.

### 2). Indeks *Silhouette*

Indeks *Silhouette* adalah dengan mengukur derajat kepercayaan dalam proses *clustering* pada pengamatan baru, jika dikatakan baik dengan nilai indeks mendekati 1 dan kondisi sebaiknya jika nilai indeks mendekati -1.

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (2.8)$$

dengan:

- $S(i)$  : indeks *Silhouette*.  
 $a(i)$  : rata-rata kemiripan antara objek ke  $i$  dengan objek lain di dalam *cluster* nya.  
 $b(i)$  : nilai minimum dari rata-rata kemiripan antar objek ke  $i$  dengan objek lain diluar *cluster*.

### 3). Indeks *Dunn*

Indeks *dunn* adalah rasio jarak terkecil antara observasi pada *cluster* yang berbeda dengan jarak terbesar pada masing-masing *cluster*.

$$C = \frac{d_{min}}{d_{max}} \quad (2.9)$$

dengan:

- $d_{min}$  : jarak terkecil antara observasi pada *cluster* yang berbeda.  
 $d_{max}$  : jarak terbesar pada masing-masing *cluster*.

### DAFTAR PUSTAKAI

- Charles, E., dan Mongi., 2015, Penggunaan Analisis *Two Step Clustering* untuk Data Campuran, *JdC*, No.1, Vol.4, 10-18, : <https://doi.org/10.35799/dc.41.2015.7251>.
- Da Silva, A. R., dan Dias, C. T., 2013, *A cophenetic correlation coefficient for tocher's method*, *Pesquisa Agropecuaria Brasileira*, No.6, Vol.148, 589–596, : <https://doi.org/10.1590/S0100-204X2013000600003>.
- Fajriah, R. I., Sutisna, H., dan Simpony, B. K., 2019, Perbandingan *Distance Space Manhattan* Dengan *Euclidean* Pada *K-Means Clustering* Dalam Menentukan Promosi, *IJCIT (Indonesian Journal on Computer and Information Technology)*, No.1, Vol.4, 36–49, : <https://doi.org/10.31294/ijcit.v4i1.4630>.
- Ghozali, I., 2016, Aplikasi Analisis *Multivariate* dengan Program IBM SPSS 23 Edisi 8, Badan Penerbit Universitas Diponegoro, Semarang.
- Harnanto, Y. I., Rusgiyono, A., dan Wuryandari, T., 2017, Penerapan Analisis Kluster Metode Ward Terhadap Kabupaten/Kota Di Jawa Tengah Berdasarkan Pengguna Alat Kontrasepsi, *Jurnal Gaussian*, No.4, Vol.6, 528–537, : <http://ejournal-s1.undip.ac.id/index.php/gaussian>.
- Irwansyah, E., dan Faisal, M., 2015, *Advanced Clustering Teori dan Aplikasi Edisi 1*, Penerbit Deepublish, Yogyakarta.
- Johnson, R. A., dan Wichern, D. W., 2007, *Applied Multivariate Statistical Analysis*, Pearson Prentice Hall, Amerika.
- Mahmudan, A., 2020, *Clustering of District or City in Central Java Based COVID19 Case Using K-Means Clustering*, *Jurnal Matematika, Statistika, dan Komputasi*, No.1, Vol.17, 1-13, : <https://doi.org/10.20956/jmsk.10727>.
- Mulyaningrum, D., Nusrang, M., dan Sudarmin., 2018, Analisis *Cluster* Pendekatan Metode *Hierarchical Clustering* Terhadap Pertumbuhan Ekonomi Di Provinsi Sulawesi Selatan, *Universitas Negeri Makassar*, No.7, Vol.27, 1–

9, : <http://eprints.unm.ac.id/id/eprint/19073>.

Mustafidah, M., dan Purnama, M., 2024, Pengelompokan Kabupaten /Kota di Jawa Timur Berdasarkan Indikator Kasus DBD menggunakan *Complete Linkage* dan *Average Linkage*, *Jurnal Ilmiah Matematika*, No.02, Vol.12, 337–343, : <https://doi.org/10.20956/ejsa.v5i2.27091>.

Oktavia, S., Widajanti, L., dan Aruben, R., 2017, Faktor yang Berhubungan dengan Status Gizi Buruk Pada Balita di Kota Semarang, *Jurnal Kesehatan Masyarakat*, No.3, Vol.5, 2356–3346, : <https://doi.org/10.30598/ba119-128>.

Rachmatin, D., 2014, Aplikasi Metode-Metode *Agglomerative* Dalam Analisis Kluster Pada Data Tingkat Polusi Udara, *Infinity Journal*, No.2, Vol.3, 133-149, : <https://doi.org/10.22460/infinity.v3i2.59>.

Talakua, M. W., Leleury, Z. A., dan Taluta, A. W., 2017, Analisis *Cluster* Dengan Menggunakan Metode *K-Means* Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014, *Jurnal Ilmu Matematika Dan Terapan*, No.2, Vol.11, 119–128, : <https://doi.org/10.30598/barekengvol11iss2pp119-128>.

Wahyudi, Budi,F., Sriyono, dan Indarwati, R., 2015, Analisis Faktor yang Berkaitan dengan Kasus Gizi buruk Pada Balita, *Jurnal Pediaternal*, No. 1, Vol.3, 83–91, : <https://doi.org/10.20473/pmnj.v3i1.11773>.

Widodo, E., Mashita, S. N., dan Prasetyowati, Y. G., 2020, Perbandingan Metode *Average Linkage*, *Complete Linkage*, dan *Ward'S* pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia, *Faktor Exacta*, No.2, Vol.13, 81-87, : <https://doi.org/10.30998/faktorexacta.v13i2.6581>.