

Pengelompokan Puskesmas Berdasarkan Kasus Balita *Stunting* di Kabupaten Paser Menggunakan Metode *K-medoids*

Grouping Community Health Centers Based on Stunting Toddler Cases in Paser Regency Using the K-medoids Method

Ika Puspita^{a)}, Memi Nor Hayati, dan Darnah Andi Nohe

Laboratorium Statistika Terapan, FMIPA, Universitas Mulawarman

^{a)}Corresponding author: ikapuspita311999@gmail.com

ABSTRACT

The number of cases of stunting toddler in Paser Regency increased by 6.66% from 2018 to 2019%. The increased in the number of stunting toddler in Paser Regency shows that the efforts made by the Paser Regency Government have not been effective in reducing the prevalence of stunting toddler because the stunting toddler rate in Paser Regency is still above the threshold set by the World Health Organization (WHO), which is a maximum of 20%. Therefore, an appropriate strategy is needed to find out which areas receive special attention and treatment, one of method to be used is cluster analysis. Cluster analysis is divided into two methods, namely the hierarchical method and the non-hierarchical method. The non-hierarchical method begins by establishing the number of groups. One of the methods included in the non-hierarchical method is *K-medoids*. In this study, clustering will be carried out in cases of stunting toddlers in Paser Regency using the *K-medoids* method. This study aims to determine the optimal cluster formed by selecting the smallest Davies Bouldin Index (DBI) value from the 2019 Community Health Center grouping in Paser Regency. The clusters formed for the *K-medoids* method in this study were 2 clusters, 3 clusters, and 4 clusters. Based on the results of the analysis, the *K-medoids* method for 2 clusters, 3 clusters and 4 clusters was based on the DBI values of 0.977, 1.470, and 1.670, respectively. The optimal group for classifying stunting toddler cases in Paser Regency in 2019 is 2 cluster using *K-medoids* method.

Keywords: *Stunting* Toddlers, *K-medoids*, Davies Bouldin Index.

1. Pendahuluan

Analisis statistika yang digunakan untuk membagi rangkaian objek menjadi beberapa kelompok berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya disebut analisis kluster (Metisen & Sari, 2015). Analisis kluster terbagi atas dua metode, yaitu metode hierarki dan metode non-hierarki. Menurut Anderberg (1973), metode non-hierarki dimulai dengan mengasumsikan untuk memutuskan pembentukan jumlah kelompok terlebih dahulu dan umumnya digunakan jika jumlah satuan pada objek pengamatan besar. Salah satu metode yang termasuk ke dalam metode non-hierarki adalah *K-medoids*.

K-medoids merupakan salah satu metode berbasis partisi yang menggunakan objek perwakilan (*representatif*) yang disebut *medoids* sebagai titik pusat atau *centroid*. *K-medoids* melakukan partisi dengan cara meminimalkan ketidakmiripan atau *dissimilarity* antar setiap objek *i* dan objek representatif terdekat (Suyanto, 2017). Metode *K-medoids* didasarkan pada penggunaan *medoids*, dengan tujuan dapat mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang terdapat dalam *dataset* (Triyanto, 2015).

Metode *K-medoids* dapat diaplikasikan dalam berbagai bidang, salah satunya pada bidang kesehatan. Berdasarkan data Dinas Kesehatan Provinsi Kalimantan Timur (2021), jumlah kasus *stunting* di Kabupaten Paser mengalami peningkatan sebesar 6,66% dari tahun 2018 ke tahun 2019, masing-masing sebesar 27,54 % dan 34,2%. Peningkatan jumlah kasus *stunting* di Kabupaten Paser menunjukkan bahwa upaya-upaya yang dilakukan pemerintah Kabupaten Paser belum efektif dalam menurunkan prevalensi balita *stunting*, karena angka balita *stunting* di Kabupaten Paser masih berada di atas ambang yang ditetapkan oleh *World Health Organization* (WHO) yaitu maksimal 20% (Riskesdas, 2019). Melalui klusterisasi pada kasus balita *stunting* di Kabupaten Paser menggunakan metode metode *K-medoids*, dapat diperoleh hasil pengelompokan wilayah di Kabupaten Paser berdasarkan penyebaran Puskesmas.

Berdasarkan latar belakang tersebut, maka penulis tertarik untuk melakukan penelitian ilmiah dengan judul “Pengelompokan Puskesmas Berdasarkan Kasus Balita *Stunting* di Kabupaten Paser Menggunakan Metode *K-medoids*”. Penelitian ini bertujuan untuk Memperoleh hasil kluster optimal yang dihasilkan *K-medoids* berdasarkan nilai DBI. Kelompok optimal yang terbentuk dapat dilihat berdasarkan nilai DBI terkecil pada kasus balita *stunting* di Kabupaten Paser tahun 2019.

2. Tinjauan Pustaka

2.1 Analisis Multivariat

Analisis multivariat berasal dari kata *multi* dan *variate* yang berarti alat analisis dengan menggunakan lebih dari dua variabel. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks. Analisis multivariat merupakan analisis yang berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran objek (Widarjono, 2015).

2.2 Data Mining

Data *mining* adalah langkah analisis terhadap proses penemuan pengetahuan di dalam basis data atau *Knowledge Discovery in Databases* (KDD). *Data mining* dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari sejumlah data besar yang membantu dalam pengambilan keputusan. *Data mining* bertujuan untuk mengekstrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan manajemen data, pra pemrosesan data, pertimbangan model, dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan *online updating* (Suyanto, 2017).

2.3 Jarak Antar Objek

Terdapat sejumlah formula untuk mengukur jarak antara dua objek data yang memiliki atribut numerik. Ukuran jarak yang paling banyak digunakan adalah jarak *euclidean*. Jarak *euclidean* digunakan ketika ingin memberikan jarak terpendek antara dua titik (jarak lurus). Adapun persamaan untuk menghitung jarak *euclidean* adalah sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^q (x_{ip} - x_{jp})^2} \tag{1}$$

dimana:

- $d(x_i, x_j)$: Jarak *euclidean* data pengamatan ke- i dengan data pengamatan ke- j dimana $i, j=1,2,3,\dots,n$
- x_{ip} : Data pengamatan ke- i pada variabel ke- p
- x_{jp} : Data pengamatan ke- j pada variabel ke- p dimana $p=1,2,3,\dots, q$

2.4 Analisis Kluster

Menurut Prasetyo (2012), analisis kluster (*cluster analysis*) adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya. Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (berhubungan) satu sama lain dan berbeda (tidak berhubungan) dengan objek dalam kelompok lain. Lebih besar kemiripan (homogenitas) dalam kelompok dan lebih besar perbedaannya di antara kelompok lainnya.

Asumsi dalam analisis kluster yaitu sampel yang diambil harus mewakili populasi (representatif) dan tidak adanya variabel penelitian yang memiliki hubungan linier yang besar dengan variabel lainnya (non multikolinieritas). Menurut Gujarati (2003), multikolinieritas adalah terjadinya hubungan linier yang kuat (hampir sempurna) antara satu variabel dengan variabel yang lainnya. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinieritas adalah dengan melihat nilai korelasi antar variabel penelitian. Menurut Gujarati dan Porter (2010), jika pada variabel penelitian tersebut terdapat korelasi yang cukup tinggi yaitu di atas 0,8 maka dapat dikatakan adanya gejala multikolinieritas. Perhitungan Koefisien korelasi (korelasi *pearson product moment*) adalah sebagai berikut:

$$r_{x_p x_l} = \frac{n(\sum_{i=1}^n x_{ip} x_{il}) - (\sum_{i=1}^n x_{ip}) \cdot (\sum_{i=1}^n x_{il})}{\sqrt{n(\sum_{i=1}^n x_{ip}^2) - (\sum_{i=1}^n x_{ip})^2} \cdot \sqrt{n(\sum_{i=1}^n x_{il}^2) - (\sum_{i=1}^n x_{il})^2}} \tag{2}$$

dimana:

- $r_{x_p x_l}$: Nilai koefisien korelasi antara variabel X ke- p dan ke- l
- n : Banyaknya data
- x_{il} : Data Pengamatan ke- i pada variabel ke- l , dengan $l=1,2,3,\dots,q$
- x_{ip} : Data pengamatan ke- i pada variabel ke- p , dengan $p=1,2,3,\dots,q$

2.5 Metode K-Medoids

Metode *K-medoids* merupakan salah satu metode pengelompokan dengan mengelompokkan sekumpulan n objek menjadi K kluster. Setelah pemilihan nilai awal *K-medoids* tersebut, dilakukan proses berulang untuk membuat pilihan yang lebih baik dari *medoids* sebelumnya dengan menganalisis semua kemungkinan pasangan objek, sedemikian sehingga satu objek adalah *medoids* dan yang lainnya tidak. Ukuran kualitas pengelompokan terbaik dihitung untuk setiap kombinasi tersebut, pilihan terbaik dari titik dalam satu iterasi dipilih sebagai *medoids* untuk iterasi berikutnya.

Algoritma *K-medoids* adalah sebagai berikut:

1. Memilih jumlah kluster.

2. Memilih secara acak objek pengamatan sebagai objek representatif O_m (*medoids*).
3. Menghitung jarak *euclidean* untuk setiap objek terhadap masing-masing *medoids* seperti dinyatakan oleh persamaan (2) berikut:

$$d(x_{ip}, O_{mp}) = \sqrt{(x_{i1} - O_{m1})^2 + (x_{i2} - O_{m2})^2 + \dots + (x_{iq} - O_{mq})^2} \quad (3)$$

dengan $d(x_{ip}, O_{mp})$ adalah jarak dari data ke- i pada variabel ke- p terhadap *medoids* ke- m pada variabel ke- p dimana $m=1,2,\dots,K$ serta $p=1,2,\dots,q$.

4. Menetapkan setiap objek ke gugus yang sesuai dengan *medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap *medoids*.
5. Memilih secara acak objek pengamatan yang tidak representatif O_h (*non-medoids*).
6. Menghitung jarak *euclidean* untuk setiap objek terhadap masing-masing *non-medoids* seperti dinyatakan oleh persamaan (3) sebagai berikut:

$$d(x_{ip}, O_{hp}) = \sqrt{(x_{i1} - O_{h1})^2 + (x_{i2} - O_{h2})^2 + \dots + (x_{iq} - O_{hq})^2} \quad (4)$$

dengan $d(x_{ip}, O_{hp})$ adalah jarak dari data ke- i pada variabel ke- p terhadap *non-medoids* ke- h pada variabel ke- p dimana $h=1,2,\dots,n-K$ serta $p=1,2,\dots,q$.

7. Menetapkan setiap objek ke gugus yang sesuai dengan *non-medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap *non-medoids*.
8. Menghitung selisih dari fungsi objektif dengan cara mengurangi fungsi objektif *non-medoids* dengan fungsi objektif *medoids*.
9. Mengganti *medoids* O_m dengan *non-medoids* O_h apabila didapatkan nilai fungsi objektif lebih kecil dari nol.
10. Mengulangi langkah (5-9) sampai didapatkan nilai fungsi objektif lebih besar dari nol.
11. Pengelompokan selesai jika didapatkan fungsi objektif lebih dari nol atau sudah tidak terdapat perubahan objek representatif.

2.6 Davies-Bouldin Index (DBI)

DBI merupakan salah satu metode yang digunakan untuk mengukur validitas kluster pada suatu metode pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* terhadap klasternya. Pengukuran dengan DBI dapat memaksimalkan jarak antar kluster dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah kluster (Wani & Riyaz, 2017). Adapun tahapantahapan untuk perhitungan menggunakan DBI:

1. *Sum of Square Within-cluster* (SSW)

SSW merupakan persamaan yang digunakan untuk mengetahui kohesi (homogenitas) dalam sebuah kluster ke- k yang dirumuskan sebagai berikut:

$$SSW_k = \sqrt{\frac{1}{n} \sum_{i=k}^{n_k} d(x_i, c_k)} \quad (5)$$

dimana:

SSW_k : Nilai Kohesi pada kluster ke- k

n_k : Jumlah data dalam kluster ke- k

$d(x_i, c_k)$: Jarak *euclidean* data pengamatan ke- i dengan *centroid* ke- k .

2. *Sum of Square Between-cluster* (SSB)

SSB merupakan persamaan yang digunakan untuk mengetahui separasi (heterogenitas) antar kluster.

$$SSB_{k,t} = d(c_k, c_t) \quad (6)$$

dimana:

$SSB_{k,t}$: Nilai separasi dari kluster ke- k dan kluster ke- t

$d(c_k, c_t)$: Jarak antara *centroid* pada kluster ke- k dengan *centroid* pada kluster ke- t dengan $t=1,2,3,\dots,K$.

3. Rasio (*Ratio*)

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio ($R_{k,t}$) untuk mengetahui nilai perbandingan antar kluster ke- k dan kluster ke- t .

$$R_{k,t} = \frac{SSW_k + SSW_t}{SSB_{k,t}} \quad (7)$$

dimana:

$R_{k,t}$: Rasio antar kluster ke- k dengan kluster ke- t

SSW_k : Nilai kohesi pada kluster ke- k

SSW_t : Nilai kohesi pada kluster ke- t

$SSB_{k,t}$: Nilai separasi dari kluster ke- k dan kluster ke- t

4. Nilai DBI

Nilai rasio yang diperoleh dari persamaan (6) digunakan untuk mencari nilai DBI dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{k \neq t} (R_{k,t}) \tag{8}$$

dimana:

K : Banyaknya kluster

\max : Maksimum nilai rasio antar kluster

Semakin kecil nilai DBI yang diperoleh (non-negatif), maka semakin baik kluster yang diperoleh dari pengelompokan menggunakan algoritma *clustering* (Bates & Kalita, 2016).

2.7 Definisi Stunting

Stunting (pendek) atau kurang gizi kronis adalah suatu bentuk lain dari kegagalan pertumbuhan. Kurang gizi kronik adalah keadaan yang sudah terjadi sejak lama, bukan seperti kurang gizi akut. Anak yang mengalami *stunting* sering terlihat memiliki badan normal yang proporsional, namun sebenarnya tinggi badannya lebih pendek dari tinggi badan normal yang dimiliki anak seusianya (Wirasmoyo, 2019).

2.8 Penyebab Stunting

Stunting disebabkan oleh beberapa faktor multidimensi dan tidak hanya disebabkan oleh faktor gizi buruk yang dialami oleh ibu hamil maupun anak balita. Andriana dan Wirjatmadi (2014), membagi penyebab terjadinya *stunting* pada anak menjadi 4 kategori besar yaitu faktor keluarga dan rumah tangga, makanan tambahan yang tidak memenuhi syarat, menyusui dan infeksi.

3. Bahan dan Metode

3.1 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

X_1 : Persentase balita *underweight*

X_2 : Persentase balita *stunting*

X_3 : Persentase ibu hamil anemia

X_4 : Persentase bayi dengan Berat Badan Lahir Rendah (BBLR) (berat badan < 2500 gram)

X_5 : Persentase bayi usia 6 bulan mendapat ASI

X_6 : Persentase ibu hamil yang mendapatkan Tablet Tambah Darah (TTD) minimal 90 tablet selama masa kehamilan

X_7 : Persentase balita kurus yang mendapat makanan tambahan

X_8 : Persentase remaja putri yang mendapat Tablet Tambah Darah (TTD)

X_9 : Persentase bayi yang baru lahir mendapat Inisiasi Menyusui Dini (IMD)

X_{10} : Persentase balita mempunyai buku Kesehatan Ibu dan Anak (buku KIA)

X_{11} : Persentase balita yang ditimbang berat badannya

X_{12} : Persentase balita yang ditimbang naik berat badannya

X_{13} : Persentase balita ditimbang yang tidak naik berat badannya dua kali berturut-turut

X_{14} : Persentase balita 6-59 bulan mendapat kapsul vitamin A

X_{15} : Persentase ibu nifas mendapat kapsul vitamin A

X_{16} : Persentase ibu nifas mendapat kapsul vitamin A

X_{17} : Persentase balita di Bawah Garis Merah (BGM)

3.2 Tahap Analisis Data

Langkah-langkah analisis data dalam penelitian ini adalah sebagai berikut:

1. Mendeteksi multikolinieritas.
2. Melakukan pengelompokan pada objek pengamatan dengan menerapkan algoritma *K-medoids* sebagai berikut:
 - a. Menentukan banyaknya kluster (K) dengan $K=2, 3$, dan 4 yang akan dibentuk.
 - b. Memilih objek representatif O_m (*medoids*) secara acak dari objek pengamatan.
 - c. Menghitung jarak *euclidean* untuk setiap objek pengamatan terhadap masing-masing *medoids* dengan menggunakan persamaan (1).
 - d. Menetapkan setiap objek ke gugus yang sesuai dengan *medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap *medoids*.
 - e. Memilih secara acak objek yang tidak representatif O_h (*non-medoids*).
 - f. Menghitung jarak *euclidean* untuk setiap objek pengamatan terhadap masing-masing *non-medoids* dengan menggunakan persamaan (4).

- g. Menetapkan setiap objek ke gugus yang sesuai dengan *non-medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap *medoids*.
 - h. Menghitung selisih dari fungsi objektif dengan cara mengurangi fungsi objektif *non-medoids* dengan fungsi objektif *medoids*.
 - i. Mengganti *medoids* dengan *non-medoids* apabila didapatkan nilai fungsi objektif < 0.
 - j. Mengulangi langkah e, f, g, h dan i sampai didapatkan nilai fungsi objektif > 0.
 - k. Pengelompokan selesai jika didapatkan fungsi objektif lebih dari nol atau sudah tidak terdapat perubahan objek representatif.
 - l. Melakukan interpretasi hasil kluster terbaik.
3. Menghitung nilai BDI untuk mengetahui kluster optimum dari hasil pengelompokan pada pengelompokan *K-means* dan *K-medoids* dengan tahapan sebagai berikut:
- a. Menghitung nilai SSW setiap kluster ke-*k* dengan menggunakan persamaan (5).
 - b. Menghitung nilai SSB dengan menggunakan persamaan (6).
 - c. Melakukan pengukuran rasio untuk mengetahui nilai perbandingan antar kluster ke-*k* dan kluster ke-*t* dengan menggunakan persamaan (7).
 - d. Menghitung nilai DBI dengan menggunakan persamaan (8).

4. Hasil dan Pembahasan

4.1 Pendeteksian Multikolinieritas

Salah satu cara pendeteksian adanya multikolinieritas adalah dengan melihat nilai koefisien korelasi. Sebagai contoh perhitungan korelasi antara variabel X_1 dan variabel X_2 adalah sebagai berikut:

$$r_{X_1, X_2} = \frac{19(\sum_{i=1}^{19} X_{11}X_{12}) - (\sum_{i=1}^{19} X_{11}) \cdot (\sum_{i=1}^{19} X_{12})}{\sqrt{19(\sum_{i=1}^{19} X_{11}^2) - (\sum_{i=1}^{19} X_{11})^2} \cdot \sqrt{19(\sum_{i=1}^{19} X_{12}^2) - (\sum_{i=1}^{19} X_{12})^2}}$$

$$= \frac{19(6932,511) + (266,53) + (419,41)}{\sqrt{19(5052,583) - (71038,24)} \sqrt{19(11128,37 - (175904,7))}} = 0,669$$

Perhitungan yang sama dilakukan sampai variabel ke-16 dengan variabel ke-17. Nilai koefisien korelasi antar variabel diperoleh dengan menggunakan bantuan *software R* dan disajikan dalam Tabel 2 sebagai berikut:

Tabel 1. Koefisien Korelasi Antar Variabel

Variabel	X_1	X_2	...	X_{17}
X_1	1	0,669	...	0,077
X_2	0,669	1	...	0,160
X_3	0,137	0,187	...	0,100
X_4	-0,296	-0,375	...	-0,334
X_5	-0,250	-0,444	...	-0,199
⋮	⋮	⋮	⋮	⋮
X_{17}	0,077	0,160	...	1

Berdasarkan hasil Tabel 1, terlihat bahwa nilai koefisien korelasi antar variabel penelitian di bawah 0,8 yang artinya tidak terdapat multikolinieritas antar variabel dalam penelitian dan dapat dilanjutkan ke proses pengelompokan dengan metode *K-medoids*.

4.2 Analisis Pengelompokan dengan Menggunakan Metode *K-medoids*

Analisis kluster dengan menggunakan metode *K-medoids* perlu menentukan banyak kluster yang akan terbentuk. Pada penelitian ini, banyak kluster yang akan digunakan adalah $K=2, 3$, dan 4.

– Analisis Pengelompokan dengan Metode *K-medoids* 2 Kluster

Proses pengelompokan menggunakan metode *K-medoids* 2 kluster adalah sebagai berikut:

1. Memilih dua objek representatif (O_m) secara acak dari data pengamatan.
 Pada percobaan pertama ini, digunakan *software R* untuk menentukan *medoids* secara acak yang akan terpilih memilih *medoids* 1 dan *medoids* 2. Masing-masing *medoids* (O_m) yang telah terpilih secara acak yaitu untuk *medoids* 1 menggunakan data ke-1 dan *medoids* 2 menggunakan data ke-13.
2. Menghitung jarak *euclidean* untuk setiap data pengamatan terhadap *medoids* 1 dan *medoids* 2 menggunakan persamaan (3). Sebagai contoh perhitungan jarak *euclidean* data pengamatan ke-1 terhadap *medoids* 1 adalah sebagai berikut:

$$d(x_{ip}, O_{1p}) = \sqrt{(x_{11} - O_{11})^2 + (x_{12} - O_{12})^2 + \dots + (x_{117} - O_{117})^2}$$

$$= \sqrt{(6,39 - 6,39)^2 + (5,11 - 5,11)^2 + \dots + (5,4 - 5,4)^2}$$

$$= 0$$

Contoh perhitungan jarak *euclidean* data pengamatan ke-1 terhadap *medoids* 2 adalah sebagai berikut:

$$\begin{aligned}
 d(x_{ip}, O_{2p}) &= \sqrt{(x_{11} - O_{21})^2 + (x_{12} - O_{22})^2 + \dots + (x_{17} - O_{217})^2} \\
 &= \sqrt{(6,39 - 5,4)^2 + (5,11 - 3,24)^2 + \dots + (5,4 - 6,9)^2} \\
 &= 52,815
 \end{aligned}$$

Perhitungan yang sama dilakukan pada data ke-2 sampai dengan data ke-19 untuk semua variabel. Hasil perhitungan dapat dilihat pada Tabel 2 dengan keseluruhan hasil perhitungan jarak *euclidean* terhadap masing-masing *medoids* 1 dan *medoids* 2.

Tabel 2. Jarak *Euclidean* Data Pengamatan ke *Medoids* 1 dan *Medoids* 2 untuk 2 Klaster

Data Pengamatan	<i>Medoids</i> 1	<i>Medoids</i> 2
1	0,000	52,814
2	119,597	115,268
3	59,623	64,855
4	121,890	120,518
5	148,125	149,690
6	113,122	118,592
7	96,796	103,351
8	76,614	69,272
9	106,170	114,993
10	62,874	50,714
11	88,554	92,804
12	134,295	143,252
13	52,814	0,000
14	76,358	56,569
15	112,372	107,468
16	66,535	85,911
17	78,305	86,358
18	120,458	127,695
19	80,623	96,720

- Menetapkan setiap data pengamatan yang sesuai dengan *medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua data pengamatan untuk setiap *medoids*.

Tabel 3. Penetapan Data Pengamatan ke *Medoids* Terdekat untuk 2 Klaster

Data Pengamatan	Jarak <i>Euclidean</i> Data Pengamatan		Jarak Minimum
	<i>Medoids</i> 1	<i>Medoids</i> 2	
	1	0,000	
2	119,597	115,268	115,268
3	59,623	64,855	59,623
4	121,890	120,518	120,518
6	113,122	118,592	113,122
7	96,796	103,351	96,796
8	76,614	69,272	69,272
9	106,170	114,993	106,170
10	62,874	50,714	50,714
11	88,554	92,804	88,554
12	134,295	143,252	134,295
13	52,814	0,000	0,000
14	76,358	56,569	56,569
15	112,372	107,468	107,468
16	66,535	85,911	66,535
17	78,305	86,358	78,305
18	120,458	127,695	120,458
19	80,623	96,720	80,623
Total Jarak Minimum			1.612,422

- Memilih secara acak dua data pengamatan yang tidak representatif *non-medoids* (O_h).

Jumlah data pengamatan yang digunakan sebanyak 19 data. Data yang sudah digunakan sebagai *medoids* sebanyak 2 data sehingga tersisa 17 data pengamatan yang akan menjadi *non-medoids*. Dari 17 data, digunakan *software R* untuk menentukan *non-medoids* secara acak. Data pengamatan yang terpilih menjadi *non-medoids 1* yaitu data ke-6 dan yang menjadi *non-medoids 2* yaitu data ke-15.

- Menghitung jarak *euclidean* untuk setiap data pengamatan terhadap *non-medoids 1* dan *non-medoids 2* dengan menggunakan persamaan (4). Sebagai contoh perhitungan jarak *euclidean* data pengamatan 1 ke *non-medoids 1* adalah sebagai berikut:

$$\begin{aligned} d(x_{ip}, O_{1p}) &= \sqrt{(x_{11} - O_{11})^2 + (x_{12} - O_{12})^2 + \dots + (x_{117} - O_{117})^2} \\ &= \sqrt{(6,39 - 8,75)^2 + (5,11 - 11,41)^2 + \dots + (5,4 - 12,6)^2} \\ &= 113,122 \end{aligned}$$

Contoh perhitungan jarak data pengamatan 1 ke *non-medoids 2* adalah sebagai berikut:

$$\begin{aligned} d(x_{ip}, O_{2p}) &= \sqrt{(x_{11} - O_{21})^2 + (x_{12} - O_{22})^2 + \dots + (x_{117} - O_{217})^2} \\ &= \sqrt{(6,39 - 23,26)^2 + (5,11 - 17,38)^2 + \dots + (5,4 - 14,7)^2} \\ &= 112,373 \end{aligned}$$

Perhitungan yang sama dilakukan pada data ke-2 sampai dengan data ke-19 untuk semua variabel. Hasil perhitungan dapat dilihat pada Tabel 4 dengan keseluruhan hasil perhitungan jarak terhadap *non-medoids 1* dan *non-medoids 2*.

Tabel 4. Jarak *Euclidean* Data Pengamatan ke *Non-medoids 1* dan *Non-medoids 2* untuk 2 Klaster

Data Pengamatan	<i>Non-medoids 1</i>	<i>Non-medoids 2</i>
1	113,122	112,372
2	48,500	48,713
3	103,476	110,758
4	31,271	57,123
5	78,338	91,728
6	0,000	50,874
7	66,197	87,681
8	119,838	117,437
9	128,037	133,709
10	108,398	109,195
11	111,400	122,839
12	51,063	81,549
13	118,592	107,468
14	122,203	117,117
15	50,874	0,000
16	103,837	123,560
17	91,547	107,520
18	39,122	61,734
19	115,219	131,569

- Menetapkan setiap data pengamatan yang sesuai dengan *non-medoids* terdekat dan menghitung fungsi objektif yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua data pengamatan untuk setiap *non-medoids*.

Tabel 5. Penetapan Data Pengamatan ke *Non-medoids* Terdekat untuk 2 Klaster

Data Pengamatan	Jarak <i>Euclidean</i> Data Pengamatan		Jarak Minimum
	<i>Non-medoids 1</i>	<i>Non-medoids 2</i>	
1	113,122	112,372	112,372
2	48,500	48,713	48,500
3	103,476	110,758	103,476
4	31,271	57,123	31,271
5	78,338	91,728	78,338
6	0,000	50,874	0,000
8	119,838	117,437	117,437
7	66,197	87,681	66,197
8	119,838	117,437	117,437
9	128,037	133,709	128,037
10	108,398	109,195	108,398

Tabel 5. Penetapan Data Pengamatan ke *Non-medoids* Terdekat untuk 2 Klaster (Lanjutan)

Data Pengamatan	Jarak Euclidean Data Pengamatan		Jarak Minimum
	Non-medoids 1	Non-medoids 2	
11	111,400	122,839	111,400
12	51,063	81,549	51,063
13	118,592	107,468	107,468
14	122,203	117,117	117,117
15	50,874	0,000	0,000
16	103,837	123,560	103,837
17	91,547	107,520	91,547
18	39,122	61,734	39,122
19	115,219	131,569	115,219
Total Jarak Minimum			1.530,807

- Menghitung selisih dari fungsi objektif dengan cara mengurangi fungsi objektif *non-medoids* dengan fungsi objektif *medoids*.
Berdasarkan Tabel 4, didapatkan fungsi objektif (total jarak minimum) *medoids* sebesar 1.612,422 dan berdasarkan Tabel 6 didapatkan fungsi objektif (total jarak minimum) *non-medoids* sebesar 1.530,807. Sehingga selisih fungsi objektif pada iterasi pertama sebagai berikut:
Selisih = fungsi objektif *non-medoids* – fungsi objektif *medoids*
= 1.530,807 - 1.612,422
= - 81,615
- Mengganti *medoids* dengan *non-medoids* apabila didapatkan nilai fungsi objektif < 0.
Selisih fungsi objektif yang didapatkan sebesar -81,615 yaitu lebih kecil dari 0, maka perlu ditukar nilai *non-medoids* menjadi *medoids* baru karena objek pengamatan *non-medoids* memiliki jarak antar data pengamatan lebih dekat dengan pusat klasternya dibandingkan dengan jarak antar data pengamatan pada pusat klaster *medoids*.
- Mengulangi langkah 4 sampai 8 hingga didapatkan nilai fungsi objektif lebih besar dari 0.
Setelah dilakukan 5 kali iterasi didapatkan nilai fungsi objektif lebih besar dari 0 yang artinya proses pengelompokan dihentikan dan didapatkan objek representatif terbaik pada data pengamatan ke-6 dan data pengamatan ke-10 dengan jumlah *dissimilaritas* sebesar 601,667. Data hasil pengelompokan *K-medoids* untuk 2 klaster dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengelompokan *K-medoids* 2 Klaster

Klaster	Anggota Klaster	
	Data Pengamatan	Puskesmas
1	2, 4, 5, 6, 7, 12, 15, 18	Senaken, Lolo, Kuaro, Long Ikis, Krayan, Muara Komam, Pasir Belengkong, dan Kerang. Tanah Grogot, Pandang Pangrapat, Long Long Ikis, Krayan, Kayungo, Mendik, Sebakung taka, Batu Kajang, Muser, Muara Komam, Pasir Belengkong, Suatang Baru, Selilir Baru, dan Tanjung Aru.
2	1, 3, 8, 9, 10, 11, 13, 14, 16, 17, 19	

Pada Tabel 6 dapat dilihat bahwa pengelompokan Puskesmas di Kabupaten Paser pada kasus balita *stunting* tahun 2019 menggunakan metode *K-medoids* didapatkan klaster 1 beranggotakan 11 Puskesmas, sedangkan klaster 2 beranggotakan 8 Puskesmas.

4.3 Menghitung Nilai DBI dengan Metode *K-medoids* 2 Klaster

- Menghitung nilai SSW setiap klaster ke-*k* dengan menggunakan persamaan (5). Adapun langkah-langkah menghitung nilai SSW yaitu data dikelompokkan berdasarkan klaster yang diikuti kemudian dihitung jarak data ke pusat klaster masing-masing dengan menggunakan jarak *euclidean* sebagai berikut:
 - Jarak *euclidean* data pengamatan terhadap pusat klaster ke-1

$$d(x_1, c_2) = ((13,95 - 8,75)^2 + (25,52 - 11,41)^2 + \dots + (9,7 - 12,6)^2) = 2.352,298$$

$$d(x_2, c_2) = ((13,89 - 8,75)^2 + (14,35 - 11,41)^2 + \dots + (14,9 - 12,6)^2) = 977,898$$

$$\vdots$$

$$d(x_8, c_2) = ((11,17 - 8,75)^2 + (25,53 - 11,41)^2 + \dots + (26,8 - 12,6)^2) = 1.530,549$$

- b. Jarak *euclidean* data pengamatan terhadap pusat kluster ke-2

$$d(x_1, c_1) = ((6,39 - 8,33)^2 + (5,11 - 29,17)^2 + \dots + (5,4 - 13,6)^2) = 3.953,194$$

$$d(x_2, c_1) = ((17,2 - 8,33)^2 + (21,87 - 29,17)^2 + \dots + (9,3 - 13,6)^2) = 1.441,690$$

⋮

$$d(x_{11}, c_1) = ((11,27 - 8,33)^2 + (25,49 - 29,17)^2 + \dots + (14,4 - 13,6)^2) = 5.380,458$$

Setelah didapatkan perhitungan jarak *euclidean*, menghitung nilai rata-rata untuk mendapatkan nilai SSW dengan menggunakan persamaan (5) sebagai berikut:

$$SSW_1 = \sqrt{\frac{1}{8}(2.352,298 + 977,898 + \dots + 1.530,549)} = 50,714$$

$$SSW_2 = \sqrt{\frac{1}{11}(3.953,194 + 1.441,690 + \dots + 5.380,458)} = 55,254$$

2. Menghitung nilai SSB dengan menggunakan persamaan (6).

SSB didapatkan dengan menghitung jarak *euclidean* antar pusat kluster, berikut perhitungan SSB yaitu menggunakan persamaan (6) sebagai berikut:

$$SSB_{1,2} = \sqrt{(8,75 - 8,33)^2 + (11,41 - 29,17)^2 \dots + (12,6 - 13,6)^2} = 108,398$$

3. Melakukan pengukuran rasio untuk mengetahui nilai perbandingan antar kluster ke-*k* dan kluster ke-*t* dengan menggunakan persamaan (7).

$$\begin{aligned} R_{1,2} &= \frac{SSW_1 + SSW_2}{SSB_{1,2}} \\ &= \frac{50,714 + 55,254}{108,398} = 0,977 \end{aligned}$$

Tabel 7. Hasil Perhitungan Nilai Rasio *K-medoids* 2 Kluster

Rasio	1	2	R_{max}
1	0,000	0,977	0,977
2	0,977	0,000	0,977

4. Menghitung nilai DBI dengan menggunakan persamaan (8).

$$\begin{aligned} DBI &= \frac{1}{2} \sum_{k=1}^2 \max_{k \neq t} (R_{k,t}) \\ &= \frac{1}{2} (0,977 + 0,977) \\ &= 0,977 \end{aligned}$$

Langkah-langkah yang sama dilakukan pada pengelompokan 3 kluster dan 4 kluster dengan menggunakan metode *K-medoids* dan metode DBI.

4.4 Perbandingan Hasil Pengelompokan

Pada bagian ini dilakukan perbandingan dari 3 kluster dengan metode *K-medoids* berdasarkan nilai DBI. Perbandingan hasil perhitungan dengan menggunakan validitas DBI dapat dilihat pada Tabel 8 berikut ini.

Table 8. Hasil Nilai DBI dengan Menggunakan Metode *K-medoids*

Metode	Jumlah Kluster	Nilai DBI
<i>K-medoids</i>	2	0,977
	3	1,470
	4	1,672

Berdasarkan Tabel 9, dapat diketahui bahwa dengan menggunakan metode *K-medoids* didapatkan hasil bahwa metode *K-medoids* 2 kluster memiliki nilai DBI sebesar 0,977, *K-medoids* 3 kluster memiliki nilai DBI sebesar 0,1,472, dan *K-medoids* 4 kluster memiliki nilai DBI sebesar 1,670. Nilai DBI terkecil adalah pengelompokan 2 kluster yaitu sebesar 0,977. Oleh karena itu, pengelompokan yang paling optimal pada pengelompokan Puskesmas di Kabupaten Paser berdasarkan kasus balita *stunting* pada tahun 2019 adalah metode *K-medoids* 2 kluster.

5. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang diperoleh yaitu kluster optimal yang terbentuk pada pengelompokan Puskesmas di Kabupaten Paser berdasarkan kasus balita *stunting* dengan

menggunakan metode *K-medoids* adalah sebanyak 2 kluster dengan nilai DBI sebesar 0,977. Kluster 1 beranggotakan 8 Puskesmas dan kluster 2 beranggotakan 11 Puskesmas.

Daftar Pustaka

- Andriana, M., & Wirjatmadi, B. (2014). *Gizi dan Kesehatan Balita Peranan Micro Zinc pada Pertumbuhan Balita*. Jakarta: Kencana Prenadamedia Group.
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Academic Press, New York.
- Bates, A., & Kalita, J. (2016). *Counting Clusters in Twitter Posts*. Proceedings of the 2 International Conference on Information Technology for Competitive Strategies, pp. 85
- Dinas Kesehatan Provinsi Kalimantan Timur. (2021). *Data Prevalensi Stunting Provinsi Kaltim Tahun 2018-2020*. Samarinda: Dinas Kesehatan Prov. Kaltim.
- Metisen, B. M., & Sari, H. L. (2015). *Analisis Clustering Menggunakan Metode K-means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila*. Jurnal Media Infortama. Vol.11 No. 2.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset.
- Riskesdas, K. (2019). Hasil Utama Riset Kesehatan Dasar (RISKESDAS). Hasil Utama Riset Kesehatan Dasar (RISKESDAS), 44.
- Supranto, J. (2004). *Analisis Multivariat Arti dan Interpretasi*. Jakarta: PT Rineka Cipta.
- Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- Triyanto, W. A. (2015). *Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk*. Jurnal SIMETRIS. Vol. 6 No. 1 Hal. 183 – 188.
- Wani, M. A. & Riyaz, R. (2017). *A novel point density based validity index for clustering gene expression datasets*. International Journal of Data Mining and Bioinformatics 17 (1): 66–84.
- Widarjono, A. (2015). *Analisis Multivariat Terapan dengan Program SPSS, AMOS, dan SMARTPLS Edisi Kedua*. Yogyakarta: UPM STIM YKPN.
- Wirasmoyo, B. (2019). *ANALISIS FAKTOR PENYEBAB KASUS STUNTING (Studi Kasus di Desa Tamanharjo Kecamatan Singosari Kabupaten Malang Provinsi Jawa Timur)*. Undergraduate (S1) thesis, University of Muhammadiyah Malang.