

# PROGRAM APLIKASI BAHASA R UNTUK PENGELOMPOKAN OBJEK MENGUNAKAN METODE *K-MEDOIDS* *CLUSTERING*

Rika Elizabet Sihombing<sup>1</sup>, Dewi Rachmatin<sup>2</sup>, Jarnawi Afgani Dahlan<sup>3</sup>  
Departemen Pendidikan Matematika FPMIPA UPI  
Email: [rikaelizabet@student.upi.edu](mailto:rikaelizabet@student.upi.edu)

**ABSTRAK.** *Cluster* atau dalam bahasa Indonesia dapat ditulis sebagai klaster adalah nama lain dari kelompok. Pada hakekatnya, analisis klaster adalah salah satu alat pengelompokan (*clustering*) dalam analisis statistika multivariat yang pada akhirnya menghasilkan sejumlah klaster (kelompok). Pengelompokan dilakukan pada objek/pengamatan (baris) dalam data yang memiliki kemiripan sangat besar dengan objek/pengamatan lainnya dalam satu klaster. Kemiripan tersebut diukur menggunakan ukuran kemiripan. Analisis klaster terbagi menjadi metode hierarki dan metode non-hierarki. Metode hierarki menggunakan struktur seperti pohon pada keputusan pengelompokan objeknya. Metode non-hierarki menempatkan objek-objek ke pusat klaster terdekat ketika banyak klaster sudah ditentukan terlebih dahulu. Ada dua buah asumsi yang harus dipenuhi agar analisis klaster dapat digunakan, yaitu data bebas pencilan (*outlier*) dan bebas dari masalah multikolinearitas. Namun, pada kenyataannya asumsi pencilan sering kali tidak terpenuhi. Salah satu metode analisis klaster yang tahan terhadap pencilan adalah metode *k-medoids* atau biasa dikenal dengan nama *Partitioning Around Medoids* (PAM). *K-medoids* merupakan salah satu bagian dari metode non-hierarki. Untuk menyelesaikan pengelompokan objek menggunakan *k-medoids*, diperlukan iterasi yang tidak sedikit. Jika menggunakan cara manual, tentu memakan banyak waktu dan kemungkinan mengalami kesalahan yang cukup tinggi. Untuk mengatasi hal tersebut, penelitian menggunakan bahasa pemrograman R yang dirancang untuk membentuk sebuah program aplikasi pengelompokan objek berdasarkan metode *k-medoids* pada suatu data. Hasil dari penelitian ini adalah sebuah program aplikasi dalam bahasa R. Program ini dapat digunakan sebagai alat bantu pengelompokan objek dengan memperlihatkan proses iterasi dari pengelompokan sampai akhirnya terbentuk sejumlah klaster yang sebelumnya telah ditentukan.

**Kata Kunci:** Analisis Klaster, Pengelompokan, *K-Medoids*, PAM, bahasa R.

**ABSTRACT.** *Cluster* or in Indonesian can be written as a 'klaster' is another name of the group. In essence, cluster analysis is one of the clustering tools in multivariate statistical analysis that ultimately produces

*a number of clusters (groups). Clustering is done on objects / observations (rows) in data that have very similar resemblance to other objects / observations in a cluster. The similarities are measured using similarity measurement. Cluster analysis is divided into hierarchical methods and non-hierarchical methods. The hierarchical method uses a tree-like structure in its object clustering decisions. The non-hierarchical method places objects to the nearest cluster center when the number of clusters are predefined. There are two assumptions that must be fulfilled for cluster analysis to be used, those are the data free from outliers and free from multicollinearity problems. In reality, however, the assumption of outliers is often unfulfilled. One method of cluster analysis that is robust to outliers is k-medoids method or commonly known as Partitioning Around Medoids (PAM). K-medoids are one part of a non-hierarchical method. To accomplish the object grouping using k-medoids, it takes a lot of iteration. If using a manual way, would take a lot of time and the possibility of experiencing a fairly high error. To overcome this, research using R programming language designed to form an object clustering application program based on k-medoids method on a data. The result of this research is an application program in R language. This program can be used as a tool clustering object by showing the iteration process from clustering until finally formed a number of clusters that have been previously determined.*

**Keywords:** Cluster Analysis, Clustering, K-Medoids, PAM, R.

## 1. PENDAHULUAN

Analisis statistika multivariat merupakan metode dalam melakukan penelitian terhadap lebih dari dua variabel secara bersamaan. Analisis statistika multivariat digunakan karena dalam kehidupan sehari-hari, permasalahan-permasalahan yang terjadi melibatkan banyak variabel. Implementasi dari ilmu ini tidak hanya pada ilmu pengetahuan alam, melainkan mencakup ilmu pengetahuan sosial juga.

Salah satu metode dalam analisis statistika multivariat adalah analisis kluster. Analisis kluster adalah metode multivariat yang bertujuan untuk mengklasifikasikan sampel subjek (atau objek) berdasarkan seperangkat variabel terukur ke dalam sejumlah kluster (kelompok) yang berbeda sehingga objek serupa ditempatkan pada kelompok yang sama. Hal ini mengakibatkan objek-objek yang terletak dalam satu kluster memiliki kemiripan (homogenitas) yang lebih besar. Analisis kluster terdiri dua metode, yaitu metode hierarki dan metode non-hierarki. Metode hierarki digunakan apabila belum diketahui banyaknya kluster yang akan dibentuk. Sementara metode non-hierarki digunakan apabila sudah diketahui banyaknya kluster yang ingin dibentuk.

Sebelum melakukan proses analisis kluster ada beberapa asumsi yang harus dipenuhi. Asumsi-asumsi dalam analisis kluster adalah data harus bebas dari penculan (*outliers*) dan multikolinieritas. Pada kenyataannya, tidak semua

data memenuhi kedua asumsi di atas, khususnya untuk asumsi pencilan. Untuk mengatasi masalah pencilan tersebut, terdapat salah satu metode analisis kluster non-hierarki yang tahan terhadap pencilan, yaitu *K-Medoids*.

Untuk menyelesaikan pengelompokan objek menggunakan *k-medoids*, diperlukan iterasi yang tidak sedikit. Jika menggunakan cara manual, tentu akan memakan banyak waktu dengan kemungkinan mengalami kesalahan yang cukup tinggi. Guna mengatasi masalah efisiensi dan efektivitas dalam penyelesaian masalah pengelompokan objek ini, diperlukan sebuah program aplikasi yang dapat membantu.

Aplikasi R menggunakan bahasa pemrograman R yang mana bahasa pemrograman ini berbasis pada bahasa pemrograman S. R memiliki kemampuan menganalisis data dengan sangat efektif dan dilengkapi dengan operator pengolahan *array* serta matriks. Kelebihan lain dari aplikasi R, ukuran *file* yang disimpan jauh lebih kecil dibanding *software* lainnya, lengkap dalam operator perhitungan *array*, dan terdiri dari koleksi *tools* statistik yang terintegrasi untuk analisis data, dapat dikembangkan sesuai kebutuhan dan sifatnya yang terbuka, setiap orang dapat menambahkan fitur-fitur tambahan dalam bentuk paket ke dalam *software* R dan dapat dipasang pada sistem operasi Windows, Mac OS, Mac OS X, Linux, Free BSD, NetBSD, Irix, Solaris, AIX, dan HP-UX. Artikel ini membahas cara dan hasil konstruksi program serta contoh penerapannya dalam data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015

## 2. METODOLOGI

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian dasar. Penelitian dasar dikategorikan pada penelitian yang menghasilkan prinsip dasar dari teknologi, formulasi konsep dan/atau aplikasi teknologi, hingga pembuktian konsep (*proof-of-concept*) fungsi dan/atau karakteristik penting secara analitis dan eksperimental (Direktorat Riset dan Pengabdian Masyarakat, 2018).

### 2.1 Asumsi pada Analisis Kluster

Sebelum melakukan proses pengelompokan, terdapat dua asumsi yang harus dipenuhi, yaitu data bebas pencilan dan bebas multikolinearitas (Wicaksono, 2017). Uraian untuk masing-masing asumsi adalah sebagai berikut:

#### 2.1.1 Pencilan

Pencilan (*outliers*) merupakan pengamatan atau tindakan yang mencurigakan karena jauh lebih kecil atau jauh lebih besar daripada sebagian besar pengamatan (Cousineau & Chartier, 2010). Pencilan terbagi menjadi dua,

yaitu pencilan univariat atau pencilan multivariat. Dalam analisis kluster, pendeteksian pencilan ditekankan pada pencilan multivariat. Pendeteksian pencilan tersebut dilakukan dengan cara menghitung kuadrat jarak Mahalanobis ( $d_{ij}^2$ ), dari masing-masing pengamatan. Jarak Mahalanobis didefinisikan sebagai:

$$d_{ij}^2 = (\mathbf{x}_{ij} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_{ij} - \bar{\mathbf{x}}) \quad (2.1)$$

di mana  $\mathbf{S}$  merupakan matriks variansi-kovariansi sampel.

Suatu data bebas dari pencilan multivariat apabila  $d_{ij}^2 < \chi_{\alpha;p}^2$  dengan  $p$  banyaknya variabel dan  $\alpha$  merupakan taraf signifikansi (Bowo, Hoyyi, & Mukid, 2013).

### 2.1.2 Multikolinearitas

Multikolinearitas adalah adanya hubungan linear yang hampir sempurna di antara beberapa atau semua variabel independen (Gujarati & Porter, 2010). Dalam analisis kluster variabel-variabel yang terlibat harus bebas dari masalah multikolinearitas. Salah satu cara untuk mendeteksi masalah multikolinearitas adalah dengan menggunakan matriks korelasi. Apabila nilai koefisien korelasi antar variabel dalam matriks korelasi bernilai dari 0.8 hingga 1.0 ( $0.8 \leq r \leq 1.0$ ), maka multikolinearitas terjadi (Prihastuti, 2014).

Apabila data yang diuji mengalami masalah multikolinearitas, maka salah satu cara untuk mengatasi masalah tersebut adalah dengan melakukan transformasi variabel menggunakan Analisis Komponen Utama (AKU). AKU merupakan salah satu metode yang mampu mereduksi dimensi data yang besar dan saling berkorelasi menjadi dimensi data yang lebih kecil dan tidak saling berkorelasi, tanpa kehilangan banyak informasi.

Prinsip dari AKU adalah mentransformasikan  $p$  variabel  $X_1, X_2, \dots, X_p$  yang berkorelasi menjadi  $p$  variabel baru  $Z_1, Z_2, \dots, Z_p$  yang tidak berkorelasi. Rumus yang digunakan untuk transformasi ke dalam AKU adalah:

$$\mathbf{z} = \mathbf{U}' [\mathbf{x} - \bar{\mathbf{x}}] \quad (2.2)$$

dengan  $\mathbf{x}$  dan  $\bar{\mathbf{x}}$  berturut-turut merupakan vektor berukuran  $p \times 1$  dari objek pada variabel asal dan rata-ratanya serta  $\mathbf{U}$  merupakan matriks yang berisi vektor eigen ortonormal.

## 2.2 *K-Medoids*

*K-Medoids* atau *Partitioning Around Medoids* (PAM) adalah algoritma *clustering* yang mirip dengan *K-Means*. Algoritma *K-Medoids* ini diusulkan pada tahun 1987 dan dikembangkan oleh Leonard Kaufmann dan Peter J. Rousseeuw (Setiyawati, 2017). Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* atau PAM menggunakan objek sebagai perwakilan (*medoid*) sebagai

pusat kluster untuk setiap kluster, sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat kluster. Algoritma *K-Medoids* memiliki kelebihan untuk mengatasi kelemahan pada algoritma *K-Means* yang sensitif terhadap *outlier* (Arora, Deepali, & Varshney, 2016). Selain itu, hasil pengelompokan pada metode ini tetap meskipun dilakukan pengacakan terhadap data. Namun, kekurangan dari metode *k-medoid* ini adalah pengelompokannya bekerja dengan baik pada data berskala kecil (Firda, 2017).

Strategi dasar dari algoritma *k-medoids clustering* adalah untuk menemukan *k* kluster dalam *n* objek dengan pertama kali secara sembarang menemukan objek yang representatif untuk tiap-tiap kluster. Objek yang mewakili setiap kluster disebut dengan *medoid*. Kluster dibangun dengan menghitung kemiripan yang dimiliki antara *medoid* dengan objek non-medoid menggunakan ukuran kesamaan. Ukuran kesamaan yang digunakan adalah jarak Manhattan. Jarak Manhattan antara objek *i* dan objek *j* adalah sebagai berikut:

$$d_{ij} = \sum_{a=1}^p |X_{ia} - X_{ja}|, i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (2.3)$$

Menurut Kaufman & Rousseuw (Flowrensia, 2010) menyatakan bahwa, misalkan  $n \times p$  adalah gugus data yang mempunyai *n* objek dan *p* variabel. Jarak antara objek ke-*i*,  $x_i$ , dan objek ke-*j*,  $x_j$ , dinotasikan dengan  $d_{ij}$ . Dalam pemilihan suatu objek sebagai objek yang representatif dalam suatu kluster (*medoid* awal),  $y_i$  didefinisikan sebagai variabel biner 0 dan 1, dimana  $y = 1$  jika dan hanya jika objek *i* ( $i = 1, 2, \dots, n$ ) dipilih sebagai *medoid* awal. Penempatan setiap objek *j* ke salah satu *medoid* awal dapat dituliskan sebagai  $C_{ij}$ , dimana  $C_{ij}$  didefinisikan sebagai variabel biner 0 dan 1, bernilai 1 jika dan hanya jika objek *j* ditempatkan ke kluster dimana objek *i* sebagai *medoid* awal. Model optimasi dalam *k-medoid* yang dapat dituliskan sebagai berikut:

$$\text{minimize } \sum_{i=1}^n \sum_{j=1}^n d_{ij} C_{ij} \quad (2.4)$$

di mana:

$$\sum_{i=1}^n C_{ij} = 1, j = 1, 2, \dots, n \quad (2.5)$$

$$C_{ij} \leq y_i, i, j = 1, 2, \dots, n \quad (2.6)$$

$$\sum_{i=1}^n y_i = k, k = \text{banyak kluster} \quad (2.7)$$

$$y_i C_{ij} \in \{0,1\}, i, j = 1, 2, \dots, n \quad (2.8)$$

Persamaan (2.5) menyatakan bahwa setiap objek *j* harus di tempatkan ke hanya satu *medoid* awal. Persamaan (2.5) dan (2.8) berimplikasi bahwa untuk suatu *j*,  $C_{ij}$  akan bernilai 1 atau 0. Persamaan (2.7) menyatakan bahwa hanya ada *k* objek yang akan dipilih sebagai *medoid*.

Kluster akan terbentuk dengan menempatkan setiap objek ke *medoid* awal yang terdekat. Persamaan (2.5) berimplikasi bahwa jarak antara objek *j* dan *medoid* awal didefinisikan sebagai:

$$\sum_{i=1}^n d_{ij} C_{ij}$$

Karena semua objek harus ditempatkan ke medoid terdekat, total jarak didefinisikan sebagai:

$$\sum_{i=1}^n \sum_{j=1}^n d_{ij} C_{ij}$$

dimana fungsi tersebut merupakan fungsi objektif yang harus diminimalkan dalam metode ini.

### 2.3 Algoritma K-Medoids

Misalkan terdapat  $n$  objek dan  $p$  variabel, lalu akan dikelompokkan ke dalam  $k$  ( $k < n$ ) kluster (kelompok), di mana  $k$  diberikan. Kita definisikan variabel ke- $j$  dari objek  $i$  sebagai  $X_{ij}$  ( $i = 1, 2, \dots, n; j = 1, 2, \dots, p$ ). Jarak Manhattan akan digunakan sebagai ukuran kesamaan. Rumus untuk jarak Manhattan dapat dilihat pada persamaan (2.3).

#### Input:

$k$ : banyak kluster yang ingin dibentuk

D: data sebanyak  $n$  objek

$t$  : maksimum iterasi

#### Output:

Satu set  $k$  kluster dengan sejumlah data yang paling dekat dengan medoidnya.

#### Algoritma:

1. Pilih sembarang  $k$  objek dari  $n$  objek yang ada sebagai inisialisasi *medoid*. Misalkan setiap medoids sebagai  $O_i$ .
2. Hitung jarak setiap objek terhadap setiap *medoid* dan tempatkan setiap objek ke kluster yang terdekat dengan *medoid*-nya.
3. Pilih sembarang objek non-*medoid* sebagai  $O_{random}$ .
4. Hitung total biaya (*cost*) dan  $S$  dari pertukaran *medoid*  $O_i$  dan  $O_{random}$ .
5. Jika  $S < 0$ , maka tukar  $O_i$  dengan  $O_{random}$  untuk membentuk diri sebagai *medoid* baru.
6. Ulangi langkah 2-5 sampai tidak ada perubahan.

Nilai total biaya (*cost*) dinyatakan dengan persamaan:

$$\text{Total cost} = \sum d_{ij} \quad (2.9)$$

di mana:

$d_{ij}$  = merujuk pada rumus (2.3).

Nilai  $S$  dinyatakan dengan persamaan:

$$S = \text{Total cost baru} - \text{Total cost lama} \quad (2.10)$$

di mana:

Total *cost* baru = jumlah biaya (*cost*) non-*medoid*

Total *cost* lama = jumlah biaya (*cost*) *medoid*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Perancangan Program

Pada bagian ini akan dibahas mengenai data masukan, data keluaran dan langkah-langkah pembuatan program aplikasi pengelompokan metode *k-medoids clustering* dengan menggunakan bahasa pemrograman R.

##### 3.1.1 Data Masukan

Data masukan yang akan dimuat dalam program aplikasi pengelompokan ini adalah sebagai berikut:

**Table 1 Data Masukan**

Data	Notasi	Tipe Data
Data yang akan diolah	x	Matriks
Banyak klaster	k	Numerik
Maksimum Iterasi	t	Numerik

##### 3.1.2 Data Keluaran

Data keluaran yang akan ditampilkan dalam program aplikasi pengelompokan ini adalah sebagai berikut:

**Table 2 Data Keluaran**

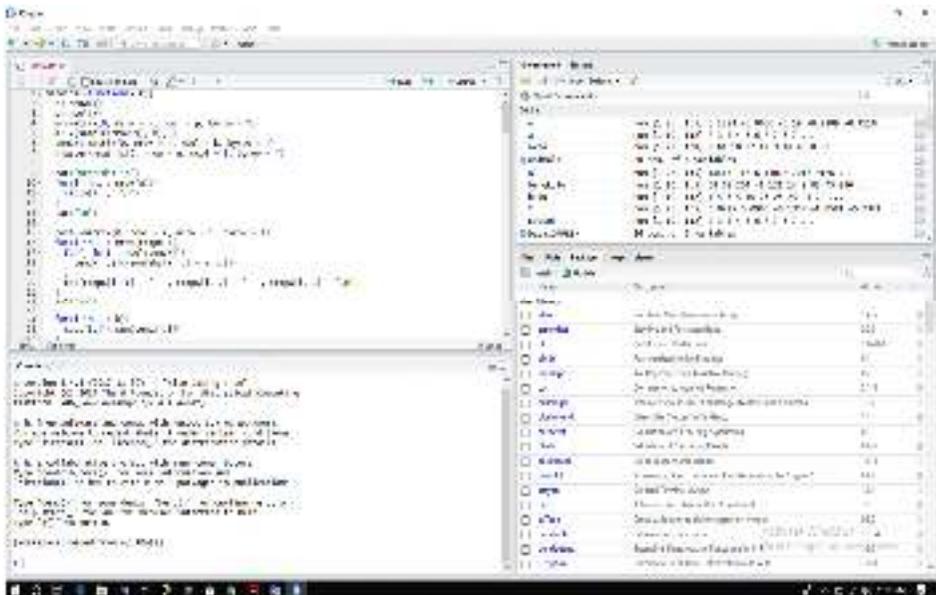
Data	Notasi	Tipe Data
Hasil Pengelompokan	klaster	<i>Character</i>

##### 3.1.3 Langkah-Langkah Pembuatan Program

Pada bagian ini dibahas langkah-langkah yang harus dilakukan untuk mengkonstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*. Langkah-langkah tersebut, yaitu:

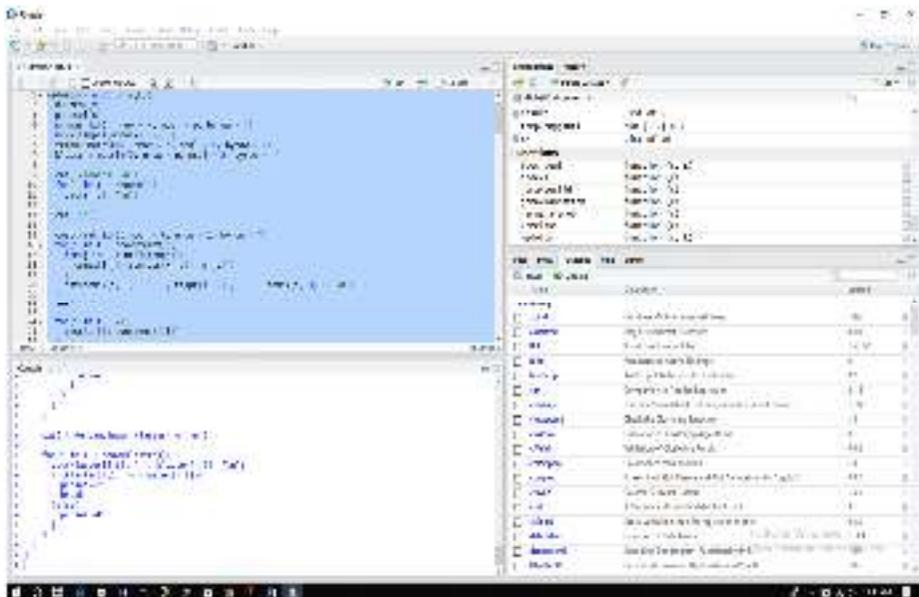
1. Buka program aplikasi RStudio pada layar *desktop* sehingga muncul tampilan utama program aplikasi RStudio seperti yang ditunjukkan pada gambar 3.1





**Gambar 3 Hasil Terjemahan Algoritma ke dalam Bahasa R**

3. Simpan lembar kerja dalam *R Script* tersebut dengan menekan tombol *save*. Misalkan file dengan nama 'Metode PAM'.
4. Jalankan program yang telah dibuat dengan cara memblok seluruh hasil terjemahan di *R Script*, kemudian menekan tombol *Run* atau menekan *Ctrl+Enter*.



**Gambar 4 Menjalankan Program**

- Selanjutnya, apabila ingin menjalankan program dapat dilakukan dengan mengetikkan nama fungsi dari program yang telah dibuat pada layar *Console* diikuti data masukan yang diperlukan.

### 3.2 Data

Data yang digunakan dalam skripsi ini adalah data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015. Data tersebut merupakan data sekunder yang berasal dari Badan Pusat Statistika (BPS).

### 3.3 Uji Pencilan

Setelah dilakukan perhitungan jarak Mahalanobis diperoleh bahwa jarak minimum sebesar 0.00576311 dan jarak maksimum sebesar 17.9898434. Dengan menggunakan  $\alpha = 0.01$  dan  $p = 3$ , didapat nilai dari  $\chi^2_{(3;0.01)} = 11.34487$ . Karena terdapat jarak Mahalanobis yang mempunyai nilai lebih 11.34487, maka dapat disimpulkan bahwa data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015 mengandung pencilan.

### 3.4 Uji Multikolinearitas

Dengan menggunakan matriks korelasi, diperoleh hasil sebagai berikut:

**Table 3 Matriks Korelasi**

Variabel	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
X <sub>1</sub>	1	0.9034573	0.2784903
X <sub>2</sub>	0.9034573	1	0.3013984
X <sub>3</sub>	0.2784903	0.3013984	1

Berdasarkan tabel di atas, diketahui bahwa nilai dari koefisien korelasi X<sub>2</sub> dengan X<sub>1</sub> adalah 0.9034573. Nilai tersebut lebih dari 0.8, sehingga pada data ini terjadi masalah multikolinearitas. Oleh sebab itu, perlu dilakukan transformasi menggunakan Analisis Komponen Utama (AKU). Selanjutnya, data yang digunakan untuk pengelompokan adalah data hasil transformasi menggunakan AKU. Hasil transformasi tersebut dapat dilihat pada di bawah ini.

Table 4 Transformasi AKU

No	Kabupaten/Kota	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>
1	Kepulauan Selayar	3746.4152	3.50203	-4.2421666
2	Bulukumba	426.3672	26.6547	4.690858
3	Bantaeng	2636.5831	-87.666919	3.9408454
4	Jeneponto	197.586	-18.912925	1.1256429
5	Takalar	198.2775	-42.885516	-0.2296038
6	Gowa	-7052.1866	-61.996023	2.9589587
7	Sinjai	1737.7068	68.471962	5.6820349
8	Bantaeng	2636.5831	-87.666919	3.9408454
9	Kepulauan Selayar	3746.4152	3.50203	-4.2421666
10	Bulukumba	426.3672	26.6547	4.690858
11	Bantaeng	2636.5831	-87.666919	3.9408454
12	Jeneponto	197.586	-18.912925	1.1256429
13	Takalar	198.2775	-42.885516	-0.2296038
14	Gowa	-7052.1866	-61.996023	2.9589587
15	Sinjai	1737.7068	68.471962	5.6820349
16	Bantaeng	2636.5831	-87.666919	3.9408454
17	Kepulauan Selayar	3746.4152	3.50203	-4.2421666
18	Bulukumba	426.3672	26.6547	4.690858
19	Bantaeng	2636.5831	-87.666919	3.9408454
20	Jeneponto	197.586	-18.912925	1.1256429
21	Takalar	198.2775	-42.885516	-0.2296038
22	Gowa	-7052.1866	-61.996023	2.9589587
23	Sinjai	1737.7068	68.471962	5.6820349
24	Bantaeng	2636.5831	-87.666919	3.9408454

### 3.5 Implementasi Program Aplikasi

Sebelum melakukan pengelompokan menggunakan program yang telah dibuat, terlebih dahulu mempersiapkan data masukan yang dibutuhkan. Data masukan tersebut adalah data yang akan diolah ( $x$ ), banyaknya klaster yang ingin dibentuk ( $k$ ) dan maksimum iterasi ( $t$ ). Untuk menunjukkan proses pengelompokan yang terjadi menggunakan program aplikasi, maka diambil 10 data pertama dari data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015. Berikut 10 data masukan yang digunakan untuk menunjukkan proses pengelompokan pada program aplikasi:

**Table 5 Sepuluh Buah Data Masukan**

No	Kabupaten/Kota	Bayi Lahir ( $X_1$ )	BBLR ( $X_2$ )	Bayi Gizi Buruk ( $X_3$ )
1	Kepulauan Selayar	1877	97	9
2	Bulukumba	5195	216	2
3	Bantaeng	2989	38	0
4	Jeneponto	5425	177	5
5	Takalar	5425	153	6
6	Gowa	12673	343	6
7	Sinjai	3883	220	1
8	Maros	5685	202	7
9	Pangkajene Kepulauan	4795	134	6
10	Barru	2524	28	7

Data yang akan diolah berupa matriks sehingga untuk menyimpannya menjadi sebuah matriks dalam bahasa R dapat dilakukan dengan mengetikkan *syntax* berikut ini pada *Console*:

```
> bayisulsel<-matrix(scan(), nrow = 10, ncol = 3, byrow =  
T)  
1: 1877      97      9  
4: 5195     216     2  
7: 2989      38     0  
10: 5425     177     5  
13: 5425     153     6  
16: 12673    343     6  
19: 3883     220     1  
22: 5685     202     7  
25: 4795     134     6  
28: 2524      28     7  
31:  
Read 30 items
```

di mana *nrow* menyatakan banyak baris, *ncol* menyatakan banyaknya kolom dan *byrow* menunjukkan bahwa data yang akan dijadikan matriks dibaca dari atas ke bawah. Setelah *syntax* diketik, tekan tombol *Enter* lalu isikan nilai-nilai dari matriks tersebut. Banyak klaster (*k*) yang akan dibentuk adalah sebanyak dua buah dengan maksimum iterasi (*t*) sebanyak lima kali. Namun, sebelum itu tetap dilakukan pengujian asumsi. Data yang diolah merupakan data hasil transformasi menggunakan AKU sebagai berikut:

```
> zskor(bayisulse1)
[[1]]
      [,1]      [,2]      [,3]
[1,] 3170.6746 20.57032 -4.5477052
[2,] -149.3169 51.27125 2.3614841
[3,] 2060.6401 -67.91037 5.4803800
[4,] -378.1973 6.13257 -0.1386333
[5,] -377.5585 -17.86826 -0.8815658
[6,] -7628.0477 -20.88806 -0.3299887
[7,] 1162.1116 90.20572 2.8504422
[8,] -638.7709 24.17965 -2.3134639
[9,] 252.7240 -20.08985 -0.9028969
[10,] 2525.7410 -65.60296 -1.5780526
```

selanjutnya, data ini akan disimpan dengan nama `bayisulse1z`.

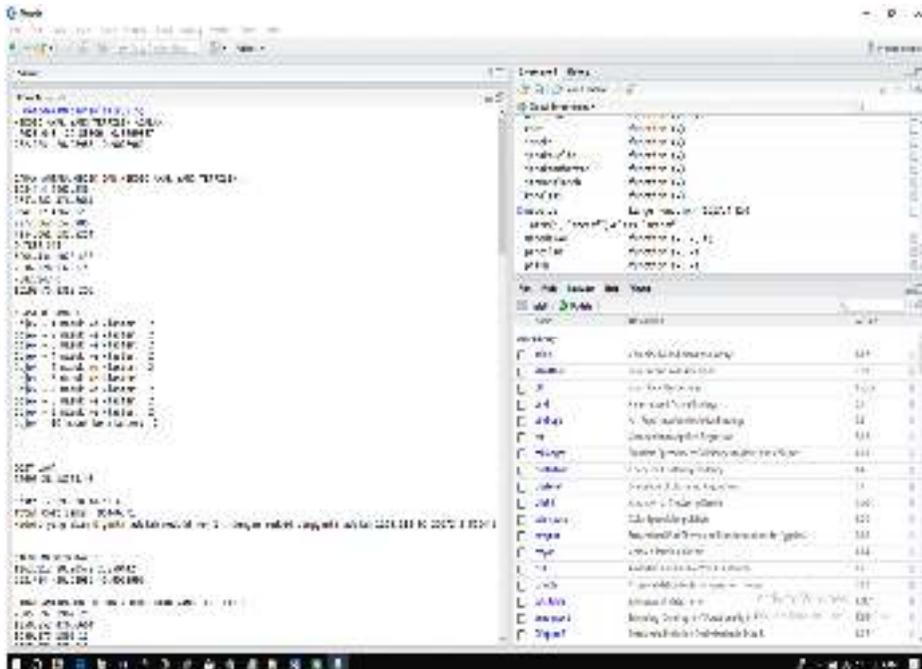
Bentuk umum dari penulisan sintaks untuk program aplikasi yang telah dikonstruksi adalah:

```
> metodePAM(x, k, t)
```

sehingga, cara menjalankan program aplikasi PAM yang telah dibuat adalah dengan mengetikkan sintaks sebagai berikut pada layar *Console* dan dilanjutkan dengan menekan tombol *Enter*:

```
> metodePAM(bayisulse1z, 2, 5)
```

setelah menekan tombol *Enter* pada *keyboard*, muncullah tampilan seperti di bawah ini:



**Gambar 5 Tampilan Keluaran Program Aplikasi**

Dengan menggunakan program aplikasi yang dikonstruksi, dapat dilihat perjalanan dari setiap iterasi. Pada iterasi terakhir diperoleh medoid akhir untuk data 10 objek pengamatan, yaitu objek ke delapan dan sepuluh dengan *cost* masing-masing medoid berturut-turut sebesar 21098,38 dan 18753,43. Hasil pengelompokan 10 data pengamatan menggunakan program aplikasi adalah sebagai berikut :

- objek - 1 masuk ke klaster: 1
- objek - 2 masuk ke klaster: 2
- objek - 3 masuk ke klaster: 1
- objek - 4 masuk ke klaster: 2
- objek - 5 masuk ke klaster: 2
- objek - 6 masuk ke klaster: 2
- objek - 7 masuk ke klaster: 1
- objek - 8 masuk ke klaster: 2
- objek - 9 masuk ke klaster: 2
- objek - 10 masuk ke klaster: 1

**Table 6 Hasil Program Aplikasi 10 Data**

<b>Klaster 1</b>	Kepulauan Selayar, Bantaeng, Sinjai, Barru
<b>Klaster 2</b>	Bulukumba, Jenepono, Takalar, Gowa, Maros, Pangkajene Kepulauan

### 3.6 Contoh Perhitungan Manual Metode K-Medoids

Berikut ini adalah contoh penerapan algoritma PAM atau K-Medoids pada 10 data pengamatan Jumlah Bayi Lahir, Bayi Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015.

**Table 7 Data Contoh**

No	Kabupaten/Kota	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>
1	Kepulauan Selayar	3746.4152	3.50203	-4.2421666
2	Bulukumba	426.3672	26.6547	4.690858
3	Bantaeng	2636.5831	-87.666919	3.9408454
4	Jeneponto	197.586	-18.912925	1.1256429
5	Takalar	198.2775	-42.885516	-0.2296038
6	Gowa	-7052.1866	-61.996023	2.9589587
7	Sinjai	1737.7068	68.471962	5.6820349
8	Maros	-63.0277	-1.394086	-0.4905113
9	Pangkajene Kepulauan	828.5633	-43.708214	-0.543551
10	Barru	3101.6747	-84.148861	-3.2305316

Adapun langkah-langkah yang harus dilakukan adalah:

- Menginisialisasi pusat kluster sebanyak k buah dari data contoh. Pada kasus ini akan dibentuk dua buah kluster. Dipilih objek ke-6 (Kabupaten Gowa) dan objek ke-9 (Kabupaten Pangkajene Kepulauan) sebagai medoid awal.
- Menghitung jarak menggunakan jarak Manhattan.
  - Untuk C1

**Table 8 Medoid C1**

C1 = Gowa	-7628.0477	-20.88806	-0.3299887
-----------	------------	-----------	------------

**Table 9 Tabel Cost C1**

C1	Jarak			Cost
Kepulauan Selayar	10798.7223	41.45838	4.2177165	10844.3984
Bulukumba	7478.7308	72.15931	2.6914728	7553.581583
Bantaeng	9688.6878	47.02231	5.8103687	9741.520479
Jeneponto	7249.8504	27.02063	0.1913554	7277.062385
Takalar	7250.4892	3.0198	0.5515771	7254.060577
Gowa	0	0	0	0

C1	Jarak			Cost
Sinjai	8790.1593	111.09378	3.1804309	8904.433511
Maros	6989.2768	45.06771	1.9834752	7036.327985
Pangkajene Kepulauan	7880.7717	0.79821	0.5729082	7882.142818
Barru	10153.7887	44.7149	1.2480639	10199.75166
<b>Total C1</b>				<b>0</b>

b. Untuk C2

**Table 10 Medoid C2**

C2 = Pangkajene Kepulauan	252.724	-20.08985	-0.9028969
---------------------------	---------	-----------	------------

**Table 11 Tabel Cost C2**

C2	Jarak			Cost
Kepulauan Selayar	2917.9506	40.66017	3.6448083	2962.255578
Bulukumba	402.0409	71.3611	3.264381	476.666381
Bantaeng	1807.9161	47.82052	6.3832769	1862.119897
Jeneponto	630.9213	26.22242	0.7642636	657.9079836
Takalar	630.2825	2.22159	0.0213311	632.5254211
Gowa	7880.7717	0.79821	0.5729082	7882.142818
Sinjai	909.3876	110.29557	3.7533391	1023.436509
Maros	891.4949	44.2695	1.410567	937.174967
Pangkajene Kepulauan	0	0	0	0
Barru	2273.017	45.51311	0.6751557	2319.205266
<b>Total C2</b>				<b>14572.10966</b>

Setelah menghitung *cost*, kemudian menentukan *cost* yang terdekat dengan medoids. *Cost* yang terdekat dengan medoids terlihat dengan data yang diberi warna kuning dalam tabel.

- Pilih sembarang objek non-medoids pada data pengamatan. Objek non-medoids yang dipilih dapat sebanyak banyak kluster yang ingin dibentuk atau satu objek saja. Pada kasus ini, akan dipilih satu objek non-medoid, yaitu objek ke-10. Kemudian, ulangi langkah 2 hingga diperoleh hasil seperti di bawah ini.

a. Untuk Non-Med 1

**Table 12 Non-Med 1**

Non-Med 1 = Barru	2525.741	-65.60296	-1.5780526
-------------------	----------	-----------	------------

**Table 13 Tabel Cost Non-Med 1**

<b>Non-Med 1</b>	<b>Jarak</b>			<b>Cost</b>
Kepulauan Selayar	644.9336	86.17328	2.9696526	734.0765326
Bulukumba	2675.0579	116.87421	3.9395367	2795.871647
Bantaeng	465.1009	2.30741	7.0584326	474.4667426
Jeneponto	2903.9383	71.73553	1.4394193	2977.113249
Takalar	2903.2995	47.7347	0.6964868	2951.730687
Gowa	10153.7887	44.7149	1.2480639	10199.75166
Sinjai	1363.6294	155.80868	4.4284948	1523.866575
Maros	3164.5119	89.78261	0.7354113	3255.029921
Pangkajene Kepulauan	2273.017	45.51311	0.6751557	2319.205266
Barru	0	0	0	0
			<b>Total Non-Med 1</b>	<b>474.4667426</b>

b. Untuk C2

**Table 14 Medoid C2**

C2 = Pangkajene Kepulauan	252.724	-20.08985	-0.9028969
---------------------------	---------	-----------	------------

**Table 15 Tabel Cost C2**

<b>C2</b>	<b>Jarak</b>			<b>Cost</b>
Kepulauan Selayar	2917.9506	40.66017	3.6448083	2962.255578
Bulukumba	402.0409	71.3611	3.264381	476.666381
Bantaeng	1807.9161	47.82052	6.3832769	1862.119897
Jeneponto	630.9213	26.22242	0.7642636	657.9079836
Takalar	630.2825	2.22159	0.0213311	632.5254211
Gowa	7880.7717	0.79821	0.5729082	7882.142818
Sinjai	909.3876	110.29557	3.7533391	1023.436509
Maros	891.4949	44.2695	1.410567	937.174967
Pangkajene Kepulauan	0	0	0	0
Barru	2273.017	45.51311	0.6751557	2319.205266
			<b>Total C2</b>	<b>14572.10966</b>

- Menghitung nilai selisih ( $S$ ) antara total *cost* baru dengan total *cost* lama.

**Table 16 Tabel Selisih (S) Total Cost**

Total Cost C1, C2	10871.292
Total Cost Non-Med 1, C2	15046.5764
<b>Selisih</b>	<b>4175.284398</b>

- Jika selisih yang didapat kurang dari 0, maka tukar C1 dengan Non-Med 1. Lalu, ulangi langkah 3 dengan menggunakan non-medoid baru sehingga pada langkah 4 didapatkan nilai selisih lagi. Apabila nilai selisih lebih dari atau sama dengan 0, maka pilihan objek non-medoid yang akan ditukarkan dengan objek medoid merupakan pilihan yang kurang tepat sehingga perlu dicari objek non-medoid lainnya. Namun, apa bila nilai baik selisih kurang dari 0 maupun lebih dari 0 dan anggota klaster tidak lagi berubah, maka iterasi berhenti serta klaster yang terbentuk pada iterasi terakhir merupakan hasil dari pengelompokan.
- Apabila iterasi dilanjutkan, akan diperoleh klaster dan anggota klaster, sebagai berikut:

**Table 17 Hasil Klaster dengan Manual**

<b>Klaster 1</b>	Kepulauan Selayar, Bantaeng, Sinjai, Barru
<b>Klaster 2</b>	Bulukumba, Jeneponto, Takalar, Gowa, Maros, Pangkajene Kepulauan

Berdasarkan perhitungan menggunakan program aplikasi dan perhitungan manual, diperoleh hasil pengelompokan pada 10 data pengamatan Jumlah Bayi Lahir, Bayi Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015 dengan iterasi maksimum 5, sebagai berikut:

**Table 18 Hasil Klaster dengan Program Aplikasi dan Manual**

<b>Klaster</b>	<b>Program Aplikasi</b>	<b>Manual</b>
1	4	4
2	6	6

Dengan demikian, diketahui bahwa jarak dan hasil akhir pengelompokan dengan menggunakan program aplikasi menghasilkan jarak dan hasil akhir yang sama dengan perhitungan manual.

### 3.7 Perbandingan Hasil Klaster

Pada bagian ini, perbandingan hasil klaster dilakukan dengan membandingkan hasil klaster pada data lengkap (24 objek) menggunakan program aplikasi yang sudah dikonstruksi sebelumnya (pada kasus ini, menggunakan maksimum iterasi sebanyak 20) dengan hasil klaster

menggunakan *package* yang telah tersedia pada aplikasi R. Adapun hasil pengelompokan data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015 adalah:

**Table 19 Hasil Kluster dengan Aplikasi**

<b>APLIKASI</b>		
<b>Kluster</b>	<b>Banyak Anggota Kluster</b>	<b>Anggota Kluster (Objek ke-)</b>
1	3	6, 11, 22
2	21	1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24

**Table 20 Hasil Kluster dengan Package**

<b>PACKAGE</b>		
<b>Kluster</b>	<b>Banyak Anggota Kluster</b>	<b>Anggota Kluster (Objek ke-)</b>
1	21	1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24
2	3	6, 11, 22

Berdasarkan hasil pengelompokan menggunakan program aplikasi yang dikonstruksi dan *package* yang sudah ada di dalam R, diperoleh hasil pengelompokan yang tidak berbeda. Perbedaan hanya terdapat pada letak kluster, di mana dengan menggunakan program aplikasi, anggota kluster 1 sebanyak 3 buah dan anggota kluster 2 sebanyak 21 buah sedangkan dengan menggunakan *package* pada *software* R diperoleh anggota kluster 1 sebanyak 21 buah dan anggota kluster 2 sebanyak 3 buah. Anggota masing-masing kluster tidaklah berbeda. Dengan menggunakan program aplikasi yang sudah dikonstruksi sendiri, anggota kluster 1 terdiri dari Kabupaten Gowa, Kabupaten Bone, Kota Makassar dan anggota kluster 2 terdiri dari Kepulauan Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Sinjai, Maros, Pangkajene Kepulauan, Barru, Soppeng, Wajo, Sindereng Rappang, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Toraja Utara, Parepare, dan Palopo.

Pada program aplikasi yang dikonstruksi, diperoleh beberapa temuan, yaitu:

1. Setiap kali melakukan pengelompokan dengan menggunakan program aplikasi yang dikonstruksi dengan data, banyak kluster, dan iterasi maksimum yang sama, dapat menghasilkan hasil akhir yang sedikit berbeda dari sebelumnya. Hal ini dikarenakan medoid awal atau

pertukaran antara medoid dengan objek non-medoid yang digunakan berbeda atau bahkan dikarenakan keduanya berbeda. Dengan menggunakan data sebelumnya, banyak kluster 2 buah, dan iterasi maksimum 5, diperoleh:

**Table 21 Temuan 1**

<b>Percobaan I</b>				
<b>Medoid Awal</b>			<b>Kluster 1</b>	<b>Kluster 2</b>
198.2775	-42.885516	-0.2296038	17	7
3101.6747	-84.148861	-3.2305316		
<b>Percobaan II</b>				
<b>Medoid Awal</b>			<b>Kluster 1</b>	<b>Kluster 2</b>
2994.9721	42.723828	4.6989912	6	18
2962.4654	59.821183	2.9673848		

2. Dengan data, banyak kluster, dan iterasi maksimum yang sama, namun dengan medoid awal yang berbeda, dapat juga memberikan hasil akhir dengan jumlah keanggotaan yang sama, namun isi dari keanggotaan tiap kluster sedikit berbeda. Berikut contoh dengan data yang sebelumnya digunakan, dibentuk dua buah kluster dengan iterasi maksimum 5, diperoleh:

**Table 22 Temuan 2**

<b>Percobaan I</b>				
<b>Medoid Awal</b>			<b>Kluster 1</b>	<b>Kluster 2</b>
198.2775	-42.885516	-0.2296038	17 anggota	7 anggota
3101.6747	-84.148861	-3.2305316	(1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 14, 16, 18, 19, 21, 23, 24)	(6, 11, 13, 15, 17, 20, 22)
<b>Percobaan II</b>				
<b>Medoid Awal</b>			<b>Kluster 1</b>	<b>Kluster 2</b>
-1181.2771	129.795405	-25.0164031	17 anggota	7 anggota
-15748.0413	50.307629	4.7841064	(2, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 17, 18, 19, 20, 21, 22)	(1, 3, 10, 12, 16, 23, 24)

3. Penggunaan iterasi maksimum juga dapat memberikan hasil akhir yang berbeda apabila dibandingkan dengan hasil akhir menggunakan

*package* yang telah ada. Berikut contoh dengan data yang sebelumnya digunakan, dibentuk dua buah klaster dengan iterasi maksimum 5, diperoleh:

**Table 23 Temuan 3**

Klaster	Program Aplikasi	Package
1	8	21
2	16	3

- Objek awal yang menjadi medoid dan pertukaran antara medoid dan objek non-medoid berpengaruh juga terhadap lamanya iterasi. Medoid awal dan pertukaran antara medoid dan non-medoid dapat mempercepat proses iterasi. Hal ini berlaku pula sebaliknya.

Kekurangan dari program aplikasi ini adalah adanya pengulangan dari pasangan medoid yang telah digunakan sebelumnya.

#### 4. KESIMPULAN

Berdasarkan uraian dan pembahasan sebelumnya, maka dapat disimpulkan bahwa, konstruksi program dilakukan dengan terlebih dahulu melakukan studi literatur terhadap metode *k-medoids* (PAM) dan bahasa pemrograman R. Selanjutnya merancang data masukan, data keluaran, serta algoritma yang akan digunakan, lalu terjemahkan algoritma tersebut ke dalam bahasa R. Setelah program aplikasi selesai dibuat, lakukan validasi dengan membandingkan hasil perhitungan program aplikasi dengan perhitungan manual atau dapat pula membandingkannya dengan *package* yang telah tersedia di dalam *software* R.

Hasil dari konstruksi program aplikasi ini berupa program komputer yang berbasis bahasa R dan sistem operasi *windows*. Program aplikasi ini berguna untuk melakukan pengelompokan (*clustering*) menggunakan metode *k-medoids* (PAM). Keluaran dari program aplikasi ini adalah klaster-klaster beserta dengan anggota klaster serta jarak antara objek terhadap medoid.

Implementasi dari program aplikasi ini menggunakan data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015. Adapun hasil pengelompokan sebanyak dua kelompok dengan maksimum iterasi sebanyak 20 pada program aplikasi ini adalah sebagai berikut:

- Wilayah klaster 1, yaitu Kabupaten Gowa, Kabupaten Bone, Kota Makassar. Wilayah-wilayah tersebut merupakan wilayah dengan jumlah bayi lahir, bayi dengan BBLR, dan bayi dengan gizi buruk yang cenderung lebih tinggi dari klaster 2.

2. Wilayah klaster 2, yaitu Kepulauan Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Sinjai, Maros, Pangkajene Kepulauan, Barru, Soppeng, Wajo, Sindereng Rappang, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Toraja Utara, Pare-Pare, dan Palopo. Wilayah-wilayah tersebut merupakan wilayah dengan jumlah bayi lahir, bayi dengan BBLR, dan bayi dengan gizi buruk yang cenderung lebih rendah dari klaster 1.

## DAFTAR PUSTAKA

- Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Procedia Computer Science* 78, 507-512.
- Bowo, K., Hoyyi, A., & Mukid, M. (2013). Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian dan Kepuasan Konsumen pada Notebook Merek Acer (Studi Kasus Mahasiswa Universitas Diponegoro). *Jurnal Gaussian*, 29-38.
- Cousineau, D., & Chartier, S. (2010). Outliers Detection and Treatment: a Review. *International Journal of Psychological Research* 3 (1), 58-67.
- Direktorat Riset dan Pengabdian Masyarakat. (2018). *Panduan Penelitian dan Pengabdian kepada Masyarakat Edisi XII*. Jakarta: Diresdikat.
- Firda, R. A. (2017). *Penggerombolan Daerah di Indonesia dengan K-Medoid dan CLARA (Studi Kasus Peubah Indeks Pembangunan Manusia Tahun 2015)*. Bogor: (Skripsi) Departemen Statistika Institut Pertanian Bogor.
- Flowrensia, Y. (2010). *Perbandingan Penggerombolan K-Means dan K-Medoid pada Data yang Mengandung Pencilan*. Bogor: (Skripsi) Departemen Statistika Institut Pertanian Bogor.
- Gujarati, D., & Porter, D. (2010). *Dasar-Dasar Ekonometrika, Edisi 5, Buku I*. Jakarta: Salemba Empat.
- Prihastuti, D. (2014). *Analisis Generalized Two Stages Ridge Regression (GTSRR) untuk Mengatasi Multikolinearitas dan Autokorelasi berserta Aplikasinya*. Yogyakarta: (Skripsi) Program Studi Matematika Universitas Negeri Yogyakarta.
- Setiyawati, W. (2017). *Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY berdasarkan Nilai Daya Serap Ujian Nasional*. Yogyakarta: (Skripsi) Jurusan Teknik Informatika Universitas Sanata Dharma.
- Wicaksono, A. (2017). *K-Medians Clustering (Studi Kasus pada Data Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Barat Tahun 2015)*. Bandung: Universitas Pendidikan Indonesia.