

BAB III

METODOLOGI PENELITIAN

Pada bab ini dibahas metode penelitian dan metode *k-medoids* atau yang biasa dikenal sebagai *Partitioning Around Medoids* (PAM) yang akan digunakan untuk melakukan proses pengelompokan.

3.1 Metodologi Penelitian

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 (Diresdikot, 2018, hlm. 39).

Objek penelitian pada skripsi ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) Sulawesi Selatan dalam publikasi Provinsi Sulawesi Selatan dalam Angka 2017. Data yang digunakan adalah data Jumlah Bayi Lahir, Bayi Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk menurut Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2015. Banyaknya objek pada data adalah 24 buah.

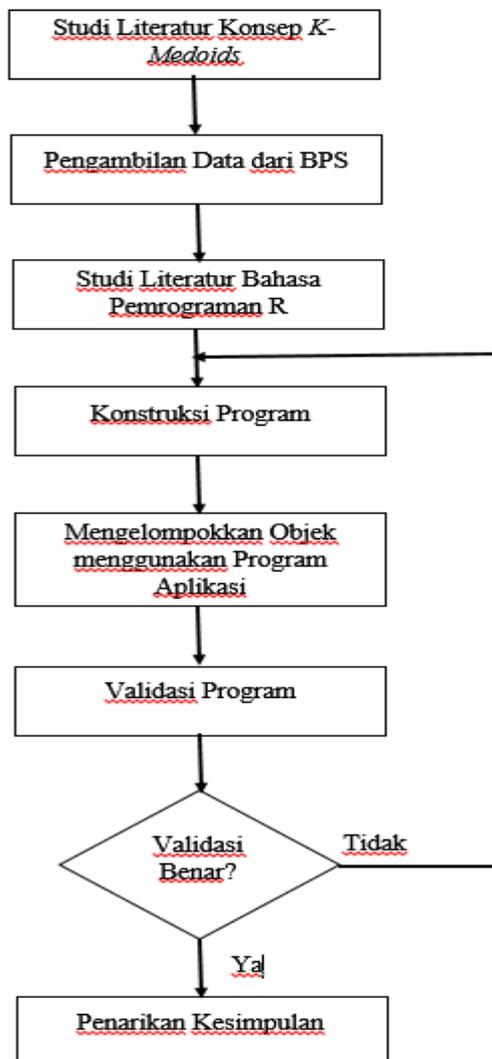
Sebagaimana telah disebutkan sebelumnya bahwa proses pengelompokan menggunakan metode *k-medoids* (PAM). Metode ini dipilih karena sifatnya yang tahan terhadap pencilan dibandingkan dengan metode *k-means*. Namun sebelum melakukan proses pengelompokan, dilakukan pengujian asumsi, yaitu ada atau tidak pencilan dan multikolinearitas.

Langkah-langkah penelitian dalam skripsi ini adalah sebagai berikut:

1. Studi literatur mengenai konsep dasar analisis klaster non hierarki dengan metode *k-medoids*.
2. Pengambilan data dari BPS.
3. Studi literatur mengenai bahasa pemrograman R.
4. Mengkonstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*.

5. Melakukan proses pengelompokkan menggunakan program aplikasi yang sudah dikonstruksi.
6. Setelah program aplikasi selesai dikonstruksi, maka selanjutnya dilakukan proses validasi, yaitu dengan cara membandingkan hasil pengelompokkan menggunakan program aplikasi dengan perhitungan manual. Tujuan dari validasi ini adalah untuk mengetahui apakah pengelompokkan menggunakan program telah sesuai dengan
7. Penarikan kesimpulan, yaitu pengambilan beberapa kesimpulan yang bersesuaian dengan tujuan penelitian.

Langkah-langkah dari metodologi penelitian di atas disajikan dalam bentuk diagram alur berikut:



Gambar 3.1 Diagram Alur Metodologi Penelitian

3.1 *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, hlm. 17). 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, 2015, hlm. 509). 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, hlm. 11).

Strategi dasar dari algoritma *k-medoids clustering* adalah untuk menemukan *k cluster* 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/kemiripan. Ukuran kesamaan/kemiripan yang digunakan adalah jarak Manhattan. Jarak Manhattan antara objek *i* dan objek *j* adalah sebagai berikut:

$$d_{ij} = \sum_{a=1}^p |X_{ia} - X_{ja}|, i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (3.1)$$

Menurut Kaufman & Rousseeuw (Flowrensia, 2010, hlm. 3) 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:

$$\text{minimize } \sum_{i=1}^n \sum_{j=1}^n d_{ij} C_{ij} \quad (3.2)$$

di mana:

$$\sum_{i=1}^n C_{ij} = 1, \quad j = 1, 2, \dots, n \quad (3.3)$$

$$C_{ij} \leq y_i, \quad i, j = 1, 2, \dots, n \quad (3.4)$$

$$\sum_{i=1}^n y_i = k, \quad k = \text{banyak kluster} \quad (3.5)$$

$$y_i C_{ij} \in \{0,1\}, \quad i, j = 1, 2, \dots, n \quad (3.6)$$

Persamaan (3.3) menyatakan bahwa setiap objek j harus di tempatkan ke hanya satu medoid awal. Persamaan (3.3) dan (3.6) berimplikasi bahwa untuk suatu j , C_{ij} akan bernilai 1 atau 0. Persamaan (3.5) menyatakan bahwa hanya ada k objek yang akan dipilih sebagai medoid.

Kluster akan terbentuk dengan menempatkan setiap objek ke medoid awal yang terdekat. Persamaan (3.3) 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.

Setiap objek yang bukan *medoid* dikelompokkan ke dalam kelompok dengan *medoid* terdekat dengan dirinya. Setelah inisialisasi pemilihan *medoid* awal secara sembarang, algoritma yang ada secara berulang berusaha untuk menemukan pilihan *medoid* lain yang lebih baik (Velmurugan dan Santhanam, 2010, hlm. 364).

3.2 Algoritma K-Medoids

Misalkan terdapat n objek dan p variabel dan 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 (3.1).

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:

$$\text{Total cost} = \sum d_{ij} \quad (3.7)$$

di mana:

d_{ij} = merujuk pada rumus (3.1).

Nilai S dinyatakan dengan persamaan:

$$S = \text{Total cost baru} - \text{Total cost lama} \quad (3.8)$$

di mana:

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

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

3.3 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.

3.3.1 Data Masukan

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

Tabel 3.1 Data Masukan

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

3.3.2 Data Keluaran

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

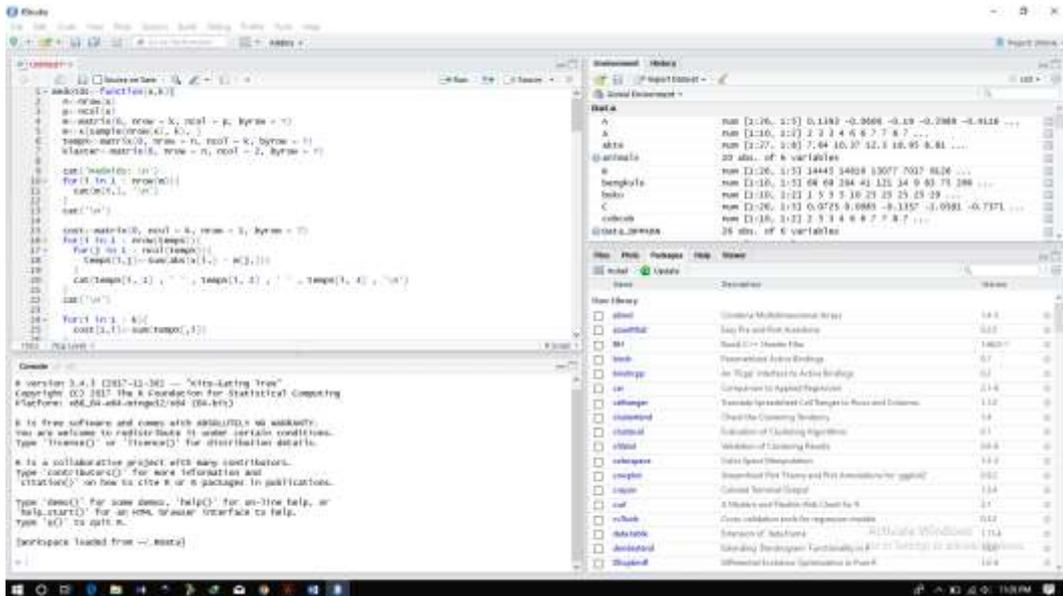
Tabel 3.2 Data Keluaran

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

3.3.3 Langkah-Langkah Pembuatan Program

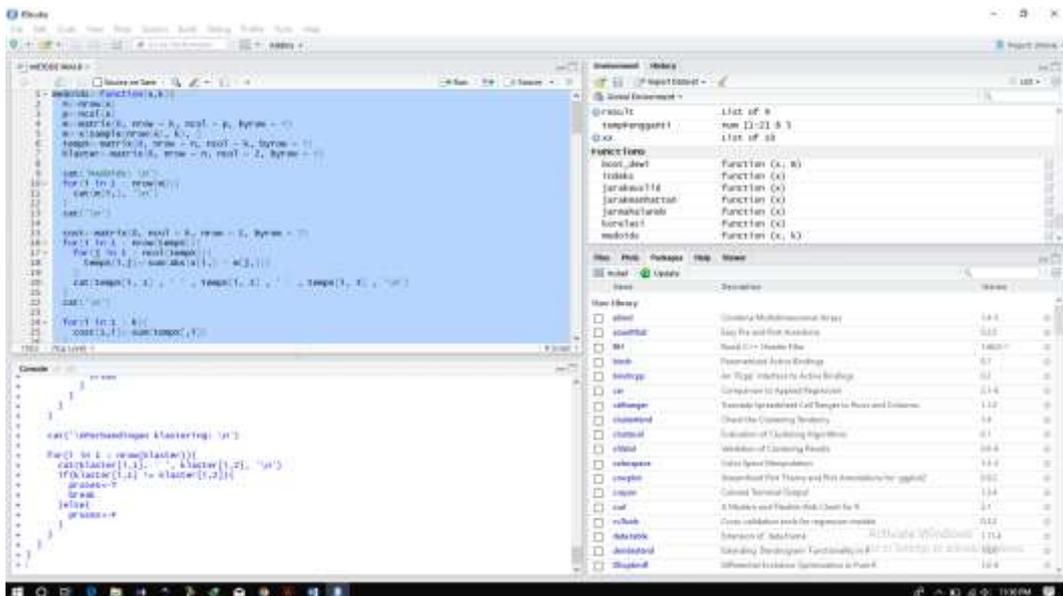
Pada bagian ini dibahas langkah-langkah yang harus dilakukan untuk mengkonstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*. Langkah-langkah tersebut, yaitu:

1. Buka program aplikasi RStudio pada layar *desktop* sehingga muncul tampilan utama program aplikasi RStudio seperti yang ditunjukkan pada gambar 3.2



Gambar 3.4 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 3.5 Menjalankan Program

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

Rika Elizabet Sihombing, 2018

PROGRAM APLIKASI BAHASA R UNTUK PENGELOMPOKAN OBJEK MENGGUNAKAN METODE K-MEDOIDS CLUSTERING

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu