

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis statistika multivariat adalah metode-metode statistik yang mengolah beberapa pengukuran menyangkut individu atau objek sekaligus (*simultaneously*). 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 (Simamora, 2005, hlm. 2).

Tipe-tipe dalam analisis statistika multivariat sendiri terbagi menjadi dua bagian besar, yaitu tipe analisis dependensi dan tipe analisis interdependensi. Hal ini didasari dengan melihat hubungan antarvariabel. Tipe analisis dependensi adalah tipe di mana adanya ketergantungan antarvariabel. Ciri penting dari tipe dependensi adalah adanya dua jenis variabel, yakni variabel dependen dan variabel independen. Tipe interdependensi adalah tipe di mana tidak adanya ketergantungan antara variabel yang satu dengan variabel lainnya. Ciri penting dari tipe interdependensi adalah tidak adanya variabel dependen dan variabel independen. Dengan kata lain, semua variabel bersifat independen (Santoso, 2017, hlm. 11).

Alat dalam analisis dependensi adalah regresi berganda, regresi logistik, analisis diskriminan, SEM (*Structural Equation Modeling*), MANOVA (*Multivariate Analysis of Variance*), dan korelasi kanokinal. Sementara alat dalam analisis interdependensi adalah analisis faktor, analisis kluster, MDS (*Multidimensional Scaling*), dan CA (*Categorical Analysis*), dll (Santoso, 2017, hlm. 12).

Menurut Cornish (Hamad, 2015, hlm. 10), analisis kluster adalah metode multivariat yang bertujuan untuk mengklasifikasikan sampel subjek (atau objek) berdasarkan seperangkat variabel terukur ke dalam sejumlah

klaster (kelompok) yang berbeda sehingga objek serupa ditempatkan pada kelompok

yang sama. Hal ini mengakibatkan objek-objek yang terletak dalam satu kluster memiliki kemiripan 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. Dalam Everitt (Rachmatin, 2014, hlm. 134), alat untuk metode ini adalah *Single Linkage Method (Nearest-Neighbor Method)*, *Complete Linkage Method (Farthest-Neighbor Method)*, *Average Linkage, Centoroid Method, Ward's Method*, dan *Median Clustering Method*. Sementara metode non-hierarki digunakan apabila sudah diketahui banyaknya kluster yang ingin dibentuk. Alat untuk metode non-hierarki adalah *K-Means*, *fuzzymethod*, *K-medoids*, dll.

Menurut Rencher (Nuningsih, 2010, hlm. 12), sebelum melakukan proses analisis kluster ada beberapa asumsi yang harus dipenuhi. Asumsi-asumsi dalam analisis kluster adalah data harus bebas dari pencilan (*outliers*) dan multikolinearitas. *Outliers* atau pencilan adalah sehimunan data yang dianggap memiliki sifat yang berbeda dibandingkan dengan kebanyakan data lainnya baik secara univariat ataupun multivariat. Sedangkan multikolinearitas adalah situasi di mana terdapat dua atau lebih variabel yang saling berkorelasi.

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* (Pramesti, Furqon, dan Dewi, 2017, hlm. 723). *K-Medoids* merupakan pendekatan *clustering* yang berhubungan dengan *k-means clustering* untuk mempartisi kumpulan data ke dalam *k* kelompok atau kluster. Dalam proses pengelompokan *k-medoids*, masing-masing kluster diwakili oleh salah satu titik data dalam setiap kluster. Setiap titik ini disebut medoid dari kluster (Kassambara, 2017, hlm. 48).

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. Pada umumnya dalam analisis data dan statistik, *software* (perangkat lunak) yang sering digunakan adalah SPSS, SAS, Minitab, Matlab, STATA, Splus, dll. Perangkat lunak tersebut merupakan perangkat lunak berbayar yang mana harganya cukup mahal khususnya untuk mahasiswa karena harganya bisa mencapai harga sebuah motor bahkan lebih. Saat ini tersedia *software-software* untuk analisis data dan statistik yang bersumber terbuka (*open source*), salah satunya adalah aplikasi R (Nurmansyah, 2013).

Aplikasi R menggunakan bahasa pemrograman R yang mana bahasa pemrograman ini berbasis pada bahasa pemrograman S. Menurut Gunawan (2012), R memiliki kemampuan menganalisis data dengan sangat efektif dan dilengkapi dengan operator pengolahan *array* dan matriks. Kelebihan lain dari aplikasi R, ukuran *file* yang disimpan jauh lebih kecil dibanding *software* lainnya, lengkap dalam operator perhitungan *array*, lengkap dan terdiri dari koleksi *tools* statistik yang terintegrasi untuk analisis data, di antaranya, mulai statistik deskriptif, fungsi probabilitas, berbagai macam uji statistik, hingga *time series*, tampilan grafik yang menarik dan fleksibel ataupun *costumized*, dapat dikembangkan sesuai keperluan dan 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 (Elfitra, 2017, hlm. 22).

Berdasarkan uraian di atas, penulis tertarik untuk mengkaji analisis kluster melalui metode *k-medoids* secara mendalam dengan menggunakan bahasa R. Oleh karena itu, tulisan ini berjudul **“Program Aplikasi Bahasa R untuk Pengelompokan Objek menggunakan Metode *K-Medoids Clustering*”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penulisan ini adalah sebagai berikut:

1. Bagaimana cara mengkonstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*?
2. Bagaimana hasil konstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*?
3. Bagaimana penerapan program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering* pada data Jumlah Bayi Lahir, Bayi dengan Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015?

1.3 Tujuan Penulisan

Berdasarkan rumusan masalah di atas, tujuan dari penulisan ini adalah sebagai berikut:

1. Mengkonstruksi program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering*.
2. Untuk mengetahui hasil konstruksi program aplikasi bahasa R dalam pengelompokan objek menggunakan metode *k-medoids clustering*.
3. Mengetahui penerapan program aplikasi bahasa R untuk pengelompokan objek menggunakan metode *k-medoids clustering* pada data Jumlah Bayi Lahir, Bayi Berat Badan Lahir Rendah (BBLR), dan Bergizi Buruk Menurut Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2015.

1.4 Manfaat Penulisan

Adapun manfaat dari penulisan ini adalah sebagai berikut:

1.4.1 Teoretis

Secara teoretis manfaat dari penulisan ini adalah untuk memperdalam dan menambah pengetahuan tentang analisis multivariat, terkhusus analisis klaster dengan metode *K-Medoids* menggunakan bahasa pemrograman R.

1.4.2 Praktis

Secara praktis, manfaat penulisan ini bagi mahasiswa adalah sebagai media dalam pembelajaran. Sedangkan bagi individu atau bagi perusahaan/instansi yang menerapkan analisis klaster, diharapkan penelitian ini dapat membantu mempermudah dalam penyelesaian permasalahan pengelompokan objek terkhusus pada individu/perusahaan yang mempunyai keadaan yang sesuai dengan kriteria pada metode *k-medoids*.

1.5 Sistematika Penulisan

Adapun sistematika dalam penulisan ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dibahas tentang latar belakang, rumusan masalah, tujuan penulisan, manfaat penulisan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini dibahas tentang analisis multivariat secara umum, analisis klaster, klasifikasi analisis klaster dan metodenya, jarak yang digunakan dalam analisis klaster, asumsi-asumsi dalam analisis klaster, analisis komponen utama, dan bahasa pemrograman R. Teori atau konsep tersebut berasal dari buku-buku pustaka, laporan penelitian, dan artikel-artikel yang terkait dengan penelitian.

BAB III METODE PENELITIAN

Pada bab ini dibahas tentang metode langkah-langkah penelitian, metode *k-medoids* (*Partitioning Around Medoids*), rancangan program dan pembuatan program serta pemrograman.

BAB IV HASIL DAN PEMBAHASAN

Bab ini dibahas mengenai data yang digunakan serta hasil program aplikasi dengan bantuan bahasa pemrograman R untuk pengelompokan objek menggunakan metode *k-medoids clustering*.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran dari hasil pembahasan materi.