

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Analisis statistik multivariat merupakan metode-metode statistik yang mempelajari tentang masalah yang menyangkut banyak variabel. Analisis multivariat digunakan karena dalam kehidupan sehari-hari terdapat masalah yang tidak hanya melibatkan satu atau dua variabel, tetapi melibatkan banyak variabel. Analisis multivariat mempelajari hubungan karakteristik yang banyak antar variabel. Aplikasi teknik analisis multivariat saat ini banyak diterapkan pada semua cabang ilmu, baik ilmu pengetahuan alam maupun ilmu pengetahuan sosial.

Teknik analisis statistik multivariat pada dasarnya diklasifikasi menjadi dua, yaitu analisis dependensi (*dependence*) dan analisis interdependensi (*interdependence*). Analisis dependensi bertujuan untuk mengetahui hubungan antara variabel terikat (*dependent*) dan variabel bebas (*independent*). Contoh analisis dependensi yaitu analisis regresi berganda, analisis diskriminan, analisis varians (ANOVA) serta analisis varians berganda (MANOVA). Sedangkan analisis interdependensi bertujuan untuk mengetahui hubungan antar variabel bebas. Contoh analisis interdependensi yaitu *multidimensional scaling*, analisis faktor, dan analisis kluster.

Analisis kluster adalah salah satu metode statistika multivariat yang melakukan sebuah usaha untuk menggabungkan observasi ke dalam beberapa kluster, di mana anggota kluster itu tidak diketahui sebelumnya untuk dianalisis. Dengan kata lain analisis kluster merupakan analisis statistik yang digunakan untuk mengelompokkan n observasi ke dalam k buah kluster, dengan setiap observasi dalam

klaster memiliki kemiripan yang tinggi dibandingkan observasi antar klaster (Wu, 2012:2).

Prinsip dasar dalam analisis klaster adalah mengelompokkan observasi pada suatu klaster yang memiliki kemiripan sangat besar dengan observasi lain dalam klaster yang sama, tetapi tidak mirip dengan observasi lain pada klaster yang berbeda. Hal ini berarti bahwa susunan klaster yang baik akan mempunyai homogenitas yang tinggi antar anggota dalam satu klaster dan heterogenitas yang tinggi antar klaster yang satu dengan yang lainnya (Nuningsih, 2010:2).

Analisis klaster mempunyai beberapa asumsi yang harus dipenuhi, yaitu data bebas dari pencilan (*outliers*) dan bebas dari masalah multikolinieritas. Pencilan merupakan data yang memiliki karakteristik yang berbeda dengan data lainnya. Adanya pencilan dapat mengubah struktur sebenarnya dari populasi sehingga klaster-klaster yang terbentuk menjadi kurang sesuai dengan struktur sebenarnya. Sedangkan multikolinieritas adalah keberadaan hubungan linear yang sempurna atau tepat, di antara sebagian atau seluruh variabel. Adanya multikolinieritas mengakibatkan perkiraan koefisien regresi tidak dapat ditentukan dan varians serta standar *error* menjadi tidak terhingga.

Dalam analisis klaster, metode pengelompokan dibagi menjadi dua, yaitu metode hirarki dan metode non-hirarki. Metode hirarki digunakan apabila belum ada jumlah klaster yang dipilih. Metode hirarki dibedakan menjadi dua pengelompokan, yaitu aglomeratif dan divisif. Pada metode aglomeratif, proses pengelompokan dimulai dengan n klaster sehingga masing-masing objek dipandang sebagai sebuah klaster, kemudian dua klaster terdekat digabungkan yang kemudian membentuk sebuah klaster baru. Proses penggabungan terus dilakukan sampai terbentuk menjadi satu klaster yang memuat semua himpunan data. Beberapa metode aglomeratif antara lain *Single Linkage*, *Average Linkage*, *Complete Linkage* serta *Ward's Method*. Sedangkan pada metode divisif, proses pengelompokan dimulai dengan n objek yang

digabungkan ke dalam satu klaster, kemudian klaster tersebut dipartisi ke dalam dua klaster, seterusnya sampai terbentuk menjadi n klaster dengan tiap klasternya beranggotakan satu objek. Beberapa metode divisif antara lain *monothetic divisive clustering* dan *polythetic divisive clustering*. Metode non-hirarki digunakan untuk mengelompokkan n objek ke dalam k klaster, di mana $k < n$ dan nilai k sudah ditentukan sebelumnya. Beberapa metode non-hirarki antara lain *Fuzzy C-Means*, *K-Medoids*, *K-Means*, dan *K-Medians*.

K-Means merupakan metode pengklasteran secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Metode ini dikembangkan oleh James B Mac-Queen pada tahun 1967. *K-Means* merupakan metode pengelompokan yang paling terkenal karena sederhana dan dapat digunakan dengan mudah di berbagai bidang. Dasar pengelompokan dalam metode ini adalah menempatkan objek berdasarkan rata-rata (mean) klaster terdekat. Sehingga terbentuk suatu kelompok yang antar objeknya memiliki kesamaan karakteristik atau homogenitas yang tinggi (Wu, 2012:2).

Pada dasarnya, mean adalah pengukuran yang sangat rentan terhadap pencilan. Sebuah pencilan yang bernilai ekstrim dapat menggeser rata-rata dari sebagian besar data yang kemudian menjadi tidak seimbang. Menurut Kaufmann & Rosseuw (1990) metode *K-Means* akan lebih sensitif terhadap data yang mengandung pencilan karena menggunakan mean sebagai ukuran nilai tengahnya. Oleh karena itu kajian tentang metode pengelompokan yang tahan terhadap pencilan diperlukan, karena keberadaan pencilan dalam sebuah data terkadang tidak dapat dihindarkan.

Di sisi lain, median adalah statistik deskriptif yang cenderung lebih tahan terhadap pencilan, sehingga berkembanglah metode yang dapat mengelompokkan data yang mengandung pencilan, yaitu metode *K-Medians*, yang merupakan salah satu variasi dari metode *K-Means*. Dalam metode *K-Means*, pengelompokan didasarkan pada nilai mean klaster terdekat, sedangkan dasar pengelompokan dalam

metode *K-Medians* adalah menempatkan objek berdasarkan nilai tengah (median) kluster terdekat. Oleh karena itu, penggunaan metode *K-Medians* akan meminimalkan *error* pada kluster.

K-Medians menggunakan *1-norm* untuk mengukur jarak antar objek, yang juga dikenal dengan jarak Manhattan. Hal ini disebabkan karena titik pusat kluster yang meminimumkan jarak *1-norm* untuk semua objek dalam kluster adalah median dari kluster (Whelan, 2015:1).

Metode *K-Medians* adalah analisis kluster non-hirarki yang bertujuan untuk mengelompokkan objek dengan meminimumkan jarak tiap-tiap objek ke titik pusat kluster (*centroid*), di mana *centroid* tersebut ditentukan berdasarkan nilai tengah (median) dari kluster. Berdasarkan pemaparan di atas, penulis tertarik untuk mengkaji analisis kluster melalui metode *K-Medians* secara mendalam. Oleh karena itu skripsi ini berjudul *K-Medians Clustering* (Studi Kasus pada Data Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Barat Tahun 2015).

1.2 RUMUSAN MASALAH

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

1. Bagaimana kajian teoritis metode *K-Medians* dalam pembentukan kluster?
2. Bagaimana hasil penerapan metode *K-Medians* dalam pembentukan kluster pada data jumlah tenaga kesehatan menurut kabupaten/kota di Provinsi Jawa Barat tahun 2015?

1.3 TUJUAN PENULISAN

Kajian terhadap permasalahan di atas bertujuan untuk:

1. Membahas kajian teoritis metode *K-Medians* dalam pembentukan klaster.
2. Menerapkan metode *K-Medians* dalam pembentukan klaster pada data jumlah tenaga kesehatan menurut kabupaten/kota di Provinsi Jawa Barat tahun 2015.

1.4 MANFAAT PENULISAN

Adapun manfaat penulisan ini sebagai berikut:

1. Teoritis

Secara teoritis manfaat penulisan ini adalah untuk memperdalam dan memperkaya pengetahuan tentang analisis statistik multivariat, khususnya analisis klaster dengan metode *K-Medians*.

2. Praktis

Secara praktis manfaat penulisan ini adalah sebagai bahan pertimbangan dan masukan bagi pihak yang berkepentingan, serta dapat menjadi informasi yang mendukung terlaksananya tujuan dari pihak yang berkepentingan.

1.5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan ini sebagai berikut:

BAB I : PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan penulisan, manfaat penulisan dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini membahas tentang teori-teori pendukung yang akan digunakan dalam pembahasan bab selanjutnya.

BAB III : *K-MEDIANS CLUSTERING*

Bab ini membahas tentang analisis kluster, metode pengelompokan, metode *K-Means*, metode *K-Medians*, interpretasi kluster, dan validasi kluster.

BAB IV : STUDI KASUS

Bab ini membahas tentang penerapan metode *K-Medians* dalam pembentukan kluster pada data jumlah tenaga kesehatan menurut kabupaten/kota di Provinsi Jawa Barat tahun 2015.

BAB V : KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan mengenai keseluruhan isi penulisan dan saran untuk penelitian selanjutnya.