

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG MASALAH

Analisis statistik multivariat adalah metode statistik di mana masalah yang diteliti bersifat multidimensional dengan menggunakan tiga atau lebih variabel secara bersamaan. Dengan menggunakan teknik analisis ini, dapat dianalisis pengaruh beberapa variabel terhadap variabel lainnya dalam waktu yang bersamaan. Analisis multivariat digunakan karena pada kenyataannya masalah yang terjadi tidak dapat diselesaikan dengan hanya menghubungkan dua variabel atau melihat pengaruh satu variabel terhadap variabel lainnya. Teknik analisis multivariat diklasifikasikan menjadi dua, yaitu

1. Analisis dependensi (analisis ketergantungan)

Analisis dependensi berfungsi untuk menerangkan atau memprediksi variabel terikat dengan menggunakan dua atau lebih variabel bebas. Metode ini diklasifikasikan berdasarkan pada jumlah variabel terikatnya. Yang termasuk dalam analisis ini adalah Analisis Regresi Linear Berganda, Analisis Diskriminan, Analisis Varians Multivariat (MANOVA) dan lain-lain.

2. Analisis interdependensi (analisis saling ketergantungan)

Analisis interdependensi berfungsi untuk memberikan makna terhadap seperangkat variabel atau membuat kelompok-kelompok secara bersamaan.

Metode ini diklasifikasikan berdasarkan pada jenis masukan variabel dengan skala pengukuran bersifat metrik dan non metrik. Yang termasuk dalam analisis ini adalah Analisis Faktor, Analisis kluster dan *Multidimensional Scaling* (Supranto, 2004:19).

Dari sekian banyak metode statistika, analisis multivariat merupakan analisis yang cocok untuk meringkas data dengan peubah yang banyak. Beberapa analisis dalam analisis multivariat yang dapat digunakan untuk memahami dan mempermudah interpretasi data multivariat di antaranya adalah analisis kluster.

Istilah analisis kluster pertama kali digunakan oleh Tryon pada tahun 1939, analisis kluster meliputi sejumlah algoritma dan metode yang berbeda untuk mengelompokkan objek serupa ke dalam masing-masing kategori. Analisis kluster bertujuan untuk mengalokasikan sekelompok individu pada suatu kelompok-kelompok yang saling bebas sehingga individu-individu di dalam kelompok itu mirip satu dengan yang lain, sementara itu individu-individu di dalam kelompok yang berbeda tidak mirip.

Dalam pengelompokan dibutuhkan suatu ukuran yang dapat menerangkan keserupaan atau kedekatan antar data yaitu ukuran kemiripan atau similaritas. Ukuran similaritas yang sering digunakan adalah ukuran jarak yaitu jarak Euclid.

Dalam teknik-teknik multivariat, yang sangat diperhatikan adalah syarat normalitas, linearitas dan homoskedastisitas tetapi tidak dalam analisis kluster. Dalam analisis kluster, yang lebih diperhatikan adalah data bebas dari *outliers* (pencilan) dan ada tidaknya kolinearitas.

Dalam melakukan pemilihan objek ke dalam kluster-kluster, analisis tidak hanya peka terhadap variabel-variabel yang tidak sesuai dengan kasus yang diteliti, tetapi juga peka terhadap *outliers* yaitu objek-objek yang berbeda dengan objek yang lainnya. *Outliers* dapat mengubah struktur sebenarnya dari populasi sehingga akan memperoleh kluster-kluster yang tidak sesuai dengan struktur sebenarnya dari populasi tersebut. Karena itu, pembuangan *outliers* sangat penting dalam analisis ini.

Kolinearitas terjadi disebabkan karena adanya korelasi antar variabel. Untuk mengatasi variabel-variabel yang berkorelasi dalam data, salah satu variabel harus dibuang. Pembuangan variabel dalam analisis ini tentunya akan mengurangi jumlah variabel. Maka dari itu, diperlukan suatu analisis lain yang dapat menganalisa suatu data, di mana variabel-variabel yang dianalisis dari data tersebut berkorelasi analisis ini disebut analisis komponen utama. Analisis komponen utama merupakan teknik multivariat yang bertujuan untuk mencari kombinasi linear variabel-variabel asal sehingga data dapat diringkas tanpa kehilangan informasi sekecil mungkin. Variabel-variabel tersebut akan ditransformasi menjadi lebih sedikit atau sama dengan jumlah variabel awal.

Proses analisis kluster atau pengelompokan data bisa dilakukan dengan dua metode yaitu:

1. Metode Hirarki.

Dalam metode ini jumlah kelompok yang akan diperoleh belum diketahui. Metode ini dibagi menjadi dua yaitu metode *agglomerative* (pemusatan) dan

metode *divisive* (penyebaran). Beberapa metode kluster hirarki yang sering digunakan dibedakan berdasarkan kemiripan (*similarity*). Ada yang menggunakan *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method*, *Ward's Method*, *Centroid Method*, *Median Method* dan *Equal-Variance Maximum Likelihood Method*.

Hasil dari metode *agglomerative* dan metode *divisive* bisa ditampilkan dalam bentuk diagram yang disebut dendogram (Jhonson, 1982:543). Dendogram menggambarkan proses pembentukan kluster yang dinyatakan dalam bentuk gambar. Garis mendatar di atas dendogram menunjukkan skala yang menggambarkan tingkat kemiripan, semakin kecil nilai skala menunjukkan semakin mirip individu/objek tersebut. Demikian juga sebaliknya.

2. Metode Non-Hirarki.

Metode ini dimulai dengan mengasumsikan ada k kelompok terlebih dahulu. Metode yang digunakan adalah *k-means* dan *fuzzy*. Metode *k-means* merupakan metode pengelompokan yang bertujuan mengelompokkan objek sedemikian sehingga jarak tiap-tiap objek ke pusat kelompok di dalam satu kelompok adalah minimum.

Pembahasan dalam tugas akhir akan dititikberatkan pada metode-metode berikut, yaitu:

1) *Single Linkage Method*

Single Linkage adalah proses pengklasteran yang didasarkan pada jarak terdekat antar objeknya. Jika dua objek terpisah oleh jarak yang pendek, maka kedua objek tersebut akan digabung menjadi satu klaster dan demikian seterusnya.

2) *Complete Linkage Method*

Complete Linkage adalah proses pengklasteran yang didasarkan pada jarak terjauh antar objeknya.

3) *Average Linkage Method*

Average Linkage adalah proses pengklasteran yang didasarkan pada jarak rata-rata antar objeknya.

4) *Ward's Method*

Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis varians untuk menghitung jarak antar klaster. Dengan kata lain, metode ini berusaha untuk memperkecil jumlah kuadrat *Sum Square* dari setiap dua kelompok yang dapat dibentuk. Keuntungan metode ini, lebih efisien dan cenderung menciptakan klaster berukuran kecil.

Ada beberapa kelebihan dan kelemahan dari analisis kluster.

Kelebihan analisis kluster antara lain:

- 1) Dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak, sedemikian sehingga data yang direduksi dengan kelompok akan mudah dianalisis.
- 2) Dapat dipakai dalam skala data ordinal, interval dan rasio.

Kelemahan analisis kluster antara lain:

- 1) Pengelompokan bersifat subjektifitas peneliti karena hanya melihat dari gambar dendogram;
- 2) Untuk data yang terlalu heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang akan dibentuk;
- 3) Metode-metode yang dipakai memberikan perbedaan yang signifikan, sehingga dalam perhitungan biasanya masing-masing metode dibandingkan;
- 4) Semakin besar observasi, biasanya tingkat kesalahan pengelompokan akan semakin besar (hasil penelitian) (Raharto, 2008:3).

Dari beberapa kelebihan dan kelemahan analisis kluster di atas, penulis sangat ingin mengetahui metode-metode yang digunakan untuk membentuk kluster. Meskipun metode-metode yang digunakan berbeda dan akan memberikan perbedaan yang signifikan, namun semua itu tidak menjadi masalah karena tujuan dari metode-metode adalah membentuk kluster dari objek-objek.

1.2. RUMUSAN MASALAH

Berdasarkan latar belakang tersebut, dapat dirumuskan beberapa masalah antara lain:

1. Bagaimanakah proses analisis klaster?
2. Bagaimanakah membentuk klaster dari beberapa objek berdasarkan karakteristiknya dengan menggunakan *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method* dan *Ward's Method*?
3. Bagaimanakah penerapan *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method* dan *Ward's Method* pada data tingkat polusi udara di kota Amerika Serikat?

1.3. BATASAN MASALAH

Dalam penulisan tugas akhir ini akan dibatasi pada masalah ukuran similaritas yang digunakan yaitu jarak Euclid.

1.4. TUJUAN PENULISAN

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

1. Menentukan proses analisis klaster.
2. Membentuk klaster dari beberapa objek berdasarkan karakteristiknya dengan menggunakan *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method* dan *Ward's Method*.

3. Menerapkan *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method* dan *Ward's Method* pada data tingkat polusi udara di kota Amerika Serikat.

1.5. MANFAAT PENULISAN

Adapun manfaat dari penulisan tugas akhir ini adalah sebagai berikut:

1. Untuk menerapkan prinsip-prinsip pengelompokan dengan lebih konsisten dari suatu permasalahan yang cukup besar.
2. Untuk mengembangkan suatu metode generalisasi secara induktif, yaitu pengambilan kesimpulan secara umum berdasarkan fakta-fakta yang lebih spesifik,
3. Mendeskripsikan sifat-sifat atau karakteristik dari klaster.

1.6. SISTEMATIKA PENULISAN

Sistematika penulisan dalam tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penulisan, manfaat penulisan dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini menjelaskan tentang teori pendukung dalam penulisan tugas akhir yang di dalamnya terdapat data multivariat, analisis

komponen utama *z-score*, vektor *mean*, matrik kovariansi, matrik korelasi populasi dan sampel, matriks *sum of square* dan *cross product* (SSCP), ukuran jarak Euclid, *outliers* dan multikolinearitas.

BAB III : METODE SINGLE LINKAGE, COMPLETE LINKAGE, AVERAGE LINKAGE DAN WARD'S METHOD PADA ANALISIS KLASTER

Bab ini menjelaskan inti dari permasalahan dari tugas akhir ini yaitu *Single Linkage Method*, *Complete Linkage Method*, *Average Linkage Method* dan *Ward's Method* Pada Analisis klaster.

BAB IV : STUDI KASUS

Bab ini membahas contoh kasus.

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran.