

BAB I

PENDAHULUAN

1.1 Latar Belakang

Masalah dalam kehidupan sehari-hari tidak hanya didasarkan pada hubungan satu variabel atau dua variabel saja, akan tetapi cenderung melibatkan banyak variabel. Analisis tentang banyak variabel dikaji secara mendalam dalam analisis multivariat.

Analisis multivariat adalah metode-metode statistik yang mengolah beberapa pengukuran menyangkut individu atau objek sekaligus (Simamora, 2005:2). Teknik analisis multivariat diklasifikasikan menjadi dua yaitu analisis dependensi/ketergantungan (*dependence methods*) dan analisis interdependensi/saling ketergantungan (*interdependence methods*). Analisis dependensi bertujuan untuk menjelaskan atau memprediksi variabel terikat dengan menggunakan dua atau lebih variabel bebas, sedangkan analisis interdependensi bertujuan untuk mereduksi atau mengelompokkan seperangkat variabel menjadi variabel baru yang lebih sedikit jumlahnya. Analisis dependensi terdiri dari analisis regresi multivariat, analisis diskriminan, analisis varians multivariat (MANOVA) dan analisis korelasi kanonik. Analisis interdependensi terdiri dari analisis faktor, analisis kluster, dan lain-lain.

Dalam perkembangannya analisis kluster telah dipergunakan dalam berbagai disiplin ilmu seperti biologi, ekonomi, psikologi, pemasaran dan lain-lain. Sebagai contoh dalam bidang pemasaran, analisis kluster bertujuan untuk

membuat segmen pasar (*segmenting the market*) dalam mengelompokkan jumlah pembeli berdasarkan keuntungan pembelian barang, memahami perilaku pembeli (*understanding buyer behaviours*) dalam mengelompokkan tempat belanja, mengenali peluang produk baru (*identifying new product opportunities*) dalam mengelompokkan merek suatu produk, memilih uji pasar (*selecting test market*) dalam mengelompokkan jenis kota.

Dillon dan Goldstein (Nurhandayani, 2003:1) mengatakan bahwa Analisis kluster adalah statistik variabel ganda yang digunakan apabila ada n buah individu atau objek yang mempunyai p variabel dan ingin dikelompokkan ke dalam k kluster berdasarkan sifat-sifat yang diamati sehingga individu atau objek yang terletak dalam satu kluster memiliki kemiripan yang lebih besar dibandingkan dengan objek yang terletak dalam kluster lain.

Prinsip dasar dalam analisis kluster adalah mengelompokkan objek (observasi) pada suatu kluster yang memiliki kemiripan sangat besar dengan objek lain dalam kluster yang sama (*similarity*), tetapi sangat tidak mirip dengan objek lain pada kluster yang berbeda (*dissimilarity*). Hal ini berarti bahwa kluster yang baik akan mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu kluster (*within-cluster*) dan heterogenitas (perbedaan) yang tinggi antar kluster yang satu dengan yang lainnya (*between-cluster*).

Sebelum melakukan proses analisis kluster, hendaknya dilakukan pengujian asumsi terlebih dahulu. Asumsi-asumsi dalam analisis kluster yaitu data bebas dari *outliers* (pencilan) dan multikolinieritas. Pencilan merupakan suatu data observasi yang menyimpang dari sekumpulan data yang lain. Adanya

pencilan dapat mengubah struktur sebenarnya dari populasi sehingga kluster-kluster yang terbentuk tidak representatif. Ini berarti bahwa kluster tersebut tidak mencerminkan karakteristik populasi yang sebenarnya. Sedangkan multikolinieritas berarti terdapat hubungan linear di antara beberapa atau semua variabel. Oleh karena itu, variabel-variabel yang bersifat multikolinieritas dalam analisis kluster perlu dipertimbangkan secara seksama.

Analisis kluster terdiri dari metode hirarki (*Hierarchical Clustering method*) dan non-hirarki (*Non-Hierarchical Clustering method*). Metode hirarki digunakan apabila belum ada informasi jumlah kluster yang akan dipilih. Metode ini secara umum dibedakan menjadi dua yaitu metode aglomeratif (penggabungan) dan metode divisif (pemecahan). Metode aglomeratif antara lain *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward's Method*, *Centroid Method*, *Median Method* dan *Equal Variance Maximum Likelihood Method*. Sedangkan metode divisif antara lain *monothic* dan *polithic*. Metode non-hirarki bertujuan untuk mengelompokkan n objek ke dalam k kluster ($k < n$), di mana nilai k telah ditentukan sebelumnya. Metode ini antara lain *K-Means*, dan *Fuzzy C-Means*.

K-Means merupakan metode klustering sederhana dan mudah diimplementasikan karena memiliki kelebihan dalam mengelompokkan data yang berukuran besar serta menghasilkan kluster dengan cepat. Kelemahan metode ini yaitu pada *K-means*, jumlah kluster awal (k) harus ditentukan sebelumnya dan tidak menjamin solusi kluster yang unik karena metode ini sulit mencapai global optimum (Hartigan, 1975:84).

K-Means termasuk dalam teknik *partitioning*. Secara umum teknik ini dimulai oleh penentuan k titik di ruang berdimensi p untuk menentukan estimasi awal pusat klaster (Everitt, 1974:25). Dengan demikian dalam teknik *partitioning* ini, metode *K-Means* memungkinkan setiap data harus termasuk ke klaster tertentu pada suatu tahapan proses, pada tahapan proses berikutnya berpindah ke klaster yang lain.

Metode *K-Means* bertujuan untuk mengelompokkan objek sedemikian hingga jarak tiap-tiap objek ke pusat klaster di dalam suatu klaster minimum. Berdasarkan uraian di atas, penulis tertarik untuk mengkaji analisis klaster melalui metode *K-Means* secara mendalam yang disebabkan oleh algoritma *K-Means* sederhana. Oleh karena itu tugas akhir ini berjudul *K-Means Clustering* (Studi Kasus pada Data Pengujian Kualitas Susu di Koperasi Peternakan Bandung Selatan).

1.2 Rumusan Masalah

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

1. Bagaimana kajian teoritis metode *K-Means* dalam pembentukan klaster?
2. Bagaimana penerapan metode *K-Means* dalam pembentukan klaster pada data pengujian kualitas susu di KPBS?

1.3 Tujuan Penulisan

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

1. Membahas kajian teoritis metode *K-Means* dalam pembentukan kluster
2. Menerapkan metode *K-Means* dalam pembentukan kluster pada data pengujian kualitas susu di KPBS

1.4 Manfaat Penulisan

Adapun manfaat penulisan tugas akhir ini sebagai berikut:

1. Teoritis

Secara teoritis manfaat penulisan tugas akhir ini adalah untuk memperkaya dan memperluas pengetahuan tentang analisis kluster, khususnya metode *K-Means* sebagai salah satu teknik dalam analisis multivariat. Di samping itu tugas akhir ini sebagai bahan evaluasi terhadap kemampuan penulis dalam menerapkan teori-teori tentang analisis multivariat yang telah disampaikan semasa perkuliahan.

2. Praktis

Secara praktis manfaat tugas akhir ini adalah sebagai bahan pertimbangan dan masukan bagi pihak yang berkepentingan serta dapat dijadikan sebagai bahan salah satu informasi yang dapat mendukung tujuan dari pihak yang berkepentingan tersebut.

1.5 Sistematika Penulisan

Adapun sistematika penulisan tugas akhir 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-Means Clustering*

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

BAB IV : STUDI KASUS

Bab ini membahas tentang penerapan metode *K-Means* dalam pembentukan kluster pada data pengujian kualitas susu dari KPBS.

BAB V : KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dan saran-saran. Kesimpulan mengenai keseluruhan isi tugas akhir dan saran-saran untuk kajian selanjutnya.