

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Sekarang ini penelitian sering kali melibatkan beberapa variabel pengamatan. Data yang diperoleh dengan mengukur lebih dari satu variabel pengamatan pada setiap individu anggota sampel disebut data multivariat. Pada data multivariat, seringkali terdapat suatu hubungan atau korelasi antar variabel-variabel yang diamati. Teknik yang memperlakukan sekelompok variabel pengamatan yang saling berhubungan sebagai satu sistem, dengan memperhitungkan hubungan antar variabel disebut Analisis Multivariat. Analisis multivariat memungkinkan peneliti untuk mencari pemecahan atas masalah-masalah yang lebih kompleks sehingga mencerminkan keadaan yang sebenarnya.

Dalam analisis multivariat, terdapat berbagai macam metode yang digunakan untuk menganalisis data, tergantung pada jenis masalah yang ingin dipecahkan dan pada skala pengukuran yang terlibat. Jika di antara variabel yang diukur tidak dibedakan antara variabel bebas dan tak bebas, sehingga persoalan pokoknya adalah tentang saling ketergantungan, maka metode analisisnya disebut metode saling ketergantungan. Salah satu metode saling ketergantungan yang sering digunakan yaitu analisis kluster (Suryanto, 1988).

Analisis kluster banyak digunakan di berbagai bidang. Tujuannya adalah mengklasifikasikan objek-objek ke dalam jumlah kelompok yang lebih kecil yang berbeda dengan tujuan untuk meyakinkan bahwa akan terdapat kesamaan yang

besar dalam kelompok-kelompok tersebut dan perbedaan antar suatu kelompok dengan kelompok yang lain sehingga sangat membantu dalam pembuatan suatu keputusan (Sarwono, 2002).

Seperti halnya melakukan analisis lain, analisis kluster juga menetapkan adanya suatu asumsi. Asumsi yang diperhatikan dalam melakukan analisis kluster adalah data yang akan digunakan tidak terdapat pencilan (*outlier*). Pencilan merupakan suatu pengamatan yang menyimpang cukup jauh dari pengamatan lainnya. Adanya pencilan dapat mengubah struktur sebenarnya dari populasi sehingga kluster-kluster yang terbentuk tidak sesuai dengan struktur yang sebenarnya. Selain itu, multikolinearitas pada data yang akan digunakan sebaiknya tidak terlalu tinggi karena multikolinearitas menunjukkan adanya hubungan dalam dua variabel atau lebih, sehingga variabel-variabel yang bersifat multikolinearitas harus dipertimbangkan secara seksama.

Terdapat dua jenis data klustering yang digunakan dalam analisis kluster, yaitu hirarki dan non-hirarki. Pada hirarki klustering, pembentukan kluster diilustrasikan dengan bentuk pohon atau biasa disebut dendogram, yaitu dengan menempatkan objek-objek ke dalam kluster sekaligus kemudian terbentuk sejumlah kluster tertentu sehingga banyaknya kluster yang terbentuk pada awalnya tidak dapat ditentukan. Sedangkan non-hirarki klustering digunakan jika banyaknya kelompok sudah diketahui. Dalam non-hirarki klustering terdapat beberapa algoritma yang dapat digunakan, contohnya yaitu *K-Means*, *Fuzzy C-Means*, *Gaussian Mixture Models*, dan lain-lain. Pemilihan penggunaan dari salah satu jenis data klustering ini tidak dapat ditentukan karena dua alasan. Pertama,

penelitian yang sedang diamati dimungkinkan dapat diselesaikan dengan kedua metode (hirarki atau non-hirarki). Kedua, pengamatan yang diamati mungkin sebaiknya diselesaikan dengan suatu metode yang selanjutnya dengan metode lainnya. Penentuan metode mana yang akan dipakai tergantung kepada peneliti dan konteks penelitian dengan tidak mengabaikan substansi, teori dan konsep yang berlaku. Hasil dari klastering, yaitu pengklasifikasian objek-objek, dan interpretasinya tersebut sangat berpengaruh dari jenis data klastering yang digunakan dalam pengelompokan objek-objek ke dalam klaster-klaster.

Pada metode *K-Means* klastering dalam jenis non-hirarki klastering, suatu data secara tegas atau pasti dinyatakan sebagai anggota satu klaster tertentu dan tidak menjadi anggota klaster yang lain. Sedangkan *Fuzzy C-Means* diartikan masing-masing data mempunyai nilai kemungkinan untuk dapat bergabung ke setiap klaster yang ada. Pada metode *K-Means* klastering, fungsi ketaksamaan (*dissimilarity*) diminimumkan sehingga mampu mengelompokkan data yang terpisah jauh. Hasil dari algoritma *K-Means* klastering, yang berupa pengklasifikasian objek-objek, bergantung pada inisial dari posisi centroid. Sehingga algoritma tersebut tidak memberikan jaminan solusi yang jelas untuk batas-batas antar kelompok.

Pada kenyataannya terkadang ditemukan suatu objek yang bukan hanya tergolong dalam satu kelompok, objek tersebut bisa saja memiliki sifat-sifat dari kelompok lain. Sehingga harus ada batasan yang jelas agar tidak terjadi tumpang tindih (*overlap*) terhadap data tersebut.

Perkembangan terakhir dari analisis kluster dengan mempertimbangkan tingkat keanggotaan yang mencakup himpunan *fuzzy* sebagai dasar pembobotan bagi pengelompokan disebut *fuzzy* klustering (Bezdek, 1981). Keunggulan utama *fuzzy* klustering adalah dapat memberikan hasil pengelompokan bagi objek-objek yang tersebar tidak teratur. Himpunan *fuzzy* dapat mengubah suatu data yang tumpang tindih (*overlap*) ke dalam model matematis sehingga dapat diproses lebih lanjut untuk dapat diterapkan dalam proses penalaran. Dengan menggunakan teori himpunan *fuzzy*, logika bahasa dapat diwakili oleh sebuah daerah yang mempunyai jangkauan tertentu yang menunjukkan derajat keanggotaannya. Sehingga lebih fleksibel dan memiliki toleransi terhadap suatu data.

Fuzzy C-Means merupakan salah satu metode *fuzzy* klustering yang mengadopsi himpunan *fuzzy* sebagai logika pengambilan keputusan. Secara umum metode *Fuzzy C-Means* meminimumkan fungsi objektif

$$P_t = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ik})^m \|x_k - v_i\|^2, \text{ yaitu fungsi yang menggambarkan jarak dari suatu}$$

titik data ke pusat kluster yang terboboti oleh derajat keanggotaan titik data tersebut sehingga dapat mengatasi masalah *overlap* yang terjadi dalam analisis kluster. Berdasarkan uraian di atas penulis tertarik untuk mengkaji suatu metode klustering yang dapat mengatasi masalah *overlap* dengan menjelaskan keanggotaan tiap-tiap data, selanjutnya tugas akhir ini penulis beri judul “METODE FUZZY C-MEANS PADA ANALISIS KLASSTER”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan dalam penulisan tugas akhir ini dirumuskan sebagai berikut:

- 1.2.1 Bagaimanakah pembentukan kluster dengan algoritma *Fuzzy C-Means*?
- 1.2.2 Bagaimanakah penerapan pembentukan kluster dengan algoritma *Fuzzy C-Means* pada data industri susu dan makanan dari susu di kota Bandung berdasarkan rata-rata biaya yang harus dikeluarkan perbulan?

1.3 Batasan Masalah

Pada penulisan tugas akhir ini jarak yang digunakan adalah jarak kuadrat *Euclid*.

1.4 Tujuan Penulisan

Tujuan penulisan tugas akhir ini adalah:

- 1.4.1 Untuk membentuk kluster dengan algoritma *Fuzzy C-Means*.
- 1.4.2 Untuk menerapkan pembentukan kluster dengan algoritma *Fuzzy C-Means* pada data industri susu dan makanan dari susu di kota Bandung berdasarkan rata-rata biaya yang harus dikeluarkan perbulan.

1.5 Manfaat Penulisan

Manfaat penulisan yang penulis harapkan pada tugas akhir ini adalah :

- 1.5.1 Bagi keilmuan : Memperkaya wacana dalam bidang statistika, khususnya pada metode analisis kluster.
- 1.5.2 Bagi peneliti : Mengembangkan konsep statistika dalam analisis kluster yaitu dengan algoritma *Fuzzy C-Means*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan yang digunakan pada penulisan tugas akhir ini adalah sebagai berikut:

BAB I Pendahuluan

Mengemukakan latar belakang, rumusan masalah, batasan masalah, tujuan penulisan, manfaat penulisan dan sistematika penulisan.

BAB II Landasan Teori

Mengemukakan landasan teori yang mendukung Bab III di antaranya menjelaskan analisis kluster dan himpunan *fuzzy*.

BAB III *Fuzzy C-Means*

Membahas tentang metode pengklusteran menggunakan algoritma *Fuzzy C-Means*.

BAB IV Studi Kasus

Pembahasan studi kasus penggunaan dari bahasan dalam Bab III, studi kasus yang diambil oleh penulis adalah data yang berasal dari industri susu dan makanan dari susu di kota Bandung.

BAB V Kesimpulan dan Saran

Mencoba merangkum keseluruhan hasil pembahasan dalam bentuk kesimpulan dan saran.

