

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Tinjauan pustaka merupakan acuan utama pada penelitian ini berupa beberapa studi yang pernah dilakukan yang berkaitan dengan penelitian ini yaitu bisa dilihat pada tabel 2.1.

Sebelumnya telah dilakukan penelitian oleh Emha Taufiq Luthfi (2007) dengan judul “Fuzzy C-means untuk Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)”. Data yang diperoleh melalui mahasiswa dalam memberikan penilaian terhadap *performance* dosen yang terdiri dari lima kriteria. Dari nilai kelima kriteria tersebut, didapat nilai indeks prestasi kumulatif dosen dalam mengampu suatu mata kuliah selama satu semester. Setelah dilakukan clustering maka diperoleh *cluster* untuk nilai matriks input.

Kemudian ada yang melakukan penelitian oleh Ratna Ekawati dan Nurul Yulis (2013) dengan judul “Klasifikasi Usaha Kecil dan Menengah (UKM) Sektor Industri Dengan Metode Fuzzy C-Means Clustering Wilayah Kota Cilegon”. Data yang diperoleh merupakan data sekunder dari usaha kecil dan menengah (UKM) sektor industri yang berada di Kota X. Data-data yang telah didapatkan kemudian diolah berdasarkan metode Fuzzy C-Means. Hasil pengelompokkan UKM dengan 8 pusat cluster.

Setelah itu ada yang melakukan penelitian oleh Monica Dias Pangestika, Drs.Prihanto Ngesti Basuki, M.Kom. , dan Charitas Fibriani, S.Kom., M.Eng.

penelitian tersebut memiliki judul “Implementasi Fuzzy C-Means untuk Clustering Data dalam Sistem Informasi Promosi”. Pada penelitian ini menggunakan data promosi di UKSW. Data yang diperlukan adalah dari data jumlah masing-masing sekolah menengah atas calon mahasiswa dan jumlah program studi yang diminati per tahunnya selama 3 tahun. Hasil dari penelitian ini akan dikelompokkan menjadi 3 cluster, yaitu rendah, sedang, dan tinggi dan keterangan informasi bagi biro promosi.

Kemudian ada yang membuat penelitian dengan judul “Penerapan Fuzzy C-Means dalam Sistem Pendukung Keputusan untuk Penentuan Penerima Bantuan Langsung Masyarakat (BLM) PNPMMPd (Studi Kasus PNPM-MPd Kec. Ngadirojo Kab. Pacitan)”. Data yang dibutuhkan adalah data usulan desa dan kriteria tiap desa. Untuk nilai pangkat 2, 3, 4, 5, 6, 10 dan 51 menghasilkan 3 cluster dengan jumlah anggota tiap clusternya yang berbeda.

Tabel 2.1 Daftar Tinjauan Pustaka

Peneliti	Tahun	Data	Metode	Hasil	Teknologi
Emha Taufiq Luthfi	2007	Nilai indeks prestasi mengajar seorang dosen	FCM	nilai matriks input (IPK)	Desktop
Ratna Ekawati dan Nurul Yulis	2013	Usaha kecil dan menengah (UKM)	FCM	pengelompokan UKM	Desktop

Peneliti	Tahun	Data	Metode	Hasil	Teknologi
Monica Dias Pangestika, Drs. Prihanto Ngesti Basuki, M.Kom. , Charitas Fibriani, S.Kom., M.Eng.	2013	Informasi Promosi	FCM	Mengetahu i status promosi	Desktop
Aziz Ahmadi dan Sri Hartati	2013	Penerima Bantuan Langsung Masyarak-at	FCM	Mengetahu i kategori kelayakan	Desktop

2.2. Dasar Teori

2.2.1. *Clustering*

Data clustering merupakan salah satu metode Data mining yang bersifat tanpa arah (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam poses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan non-*hierarchical* (non hirarki) data *clustering*. *Clustering* merupakan cara memasukkan suatu pola yang diamati ke suatu kelas pola yang belum diketahui dan disebut sebagai kluster pola. Tujuan dari *clustering (unsupervised classification)* adalah berusaha untuk mengelompokan data dalam ruang ciri (*feature space*) secara natural ke dalam sebuah kluster.

Kluster merupakan suatu kelompok yang homogen, dimana tiap unit di dalamnya memiliki kemiripan satu sama lain. Untuk membentuk *clustering* dari sekumpulan data, maka kriteria dari kluster harus mempunyai kumpulan data yang homogen dan tidak serupa dengan kumpulan data lainnya, sedangkan *cluster* yang berbeda secara umum akan mengarah kepada kluster yang berbeda pula. Aplikasi dari *clustering* antara lain : *engineering, bioinformatics, social sciences (sociology, archeology), medicine sciences (psychiatry, pathology)* data dan *web mining*.

2.2.2. Logika Fuzzy

Istilah Fuzzy yang didasarkan pada logika Boolean yang umum digunakan dalam komputasi. Secara ringkas, teorema Fuzzy memungkinkan komputer “berpikir” tidak hanya dalam skala hitam-putih(0 dan 1, mati atau hidup) tetapi juga dalam skala abu-abu. Dalam Logika Fuzzy suatu preposisi dapat dipresentasikan dalam derajat kebenaran (*truthfulness*) atau kesalahan (*falsehood*) tertentu. Ada beberapa metode untuk mempresentasikan hasil logika fuzzy yaitu Tsukamoto, Sugeno, dan Mamdani. (Sri Kusumadewi, Hari Purnomo, 2004)

2.2.3. Fuzzy C-means

Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA, merupakan salah satu metode *clustering* yang merupakan bagian dari metode Hard K-Means. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. (Emha Taufiq Luthfi, 2007)

Konsep dasar FCM, pertama kali adalah menentukan pusat *cluster*, yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat cluster

ini masih belum akurat. Tiap-tiap titik data mewakili derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. (Sri Kusumadewi, Hari Purnomo, 2004).

2.2.4. Algoritma Fuzzy C-Means

Algoritma FCM adalah sebagai berikut :

1. Inisialisasi : tentukan jumlah cluster ($k \geq 2$), tentukan jumlah maksimal iterasi, tentukan ambang batas perubahan nilai fungsi obyektif (jika perlu juga perubahan nilai centroid).
2. Berikan nilai awal matriks fuzzy pseudo-partition, dengan syarat seperti pada persamaan :

$$\sum_{j=1}^k u_{ij} = 1 \dots\dots\dots (\text{Persamaan 2.1})$$

Keterangan :

k = jumlah *cluster*

j = menyatakan *centroid cluster* ke-j

u_{ij} = derajat keanggotaan

3. Lakukan langkah 4 sampai 5 selama syarat masih terpenuhi :
 - a. Apabila perubahan pada nilai fungsi obyektif masih di atas nilai ambang batas yang ditentukan; atau

- b. Perubahan pada nilai centroid masih di atas nilai ambang batas yang ditentukan; atau
 - c. Iterasi maksimal belum tercapai.
4. Hitung nilai centroid dari masing-masing cluster menggunakan persamaan :

$$C_{lj} = \frac{\sum_{i=1}^N (u_{il})^w x_{ij}}{\sum_{i=1}^N (u_{il})^w} \dots\dots\dots \text{(Persamaan 2.2)}$$

Keterangan :

N = jumlah data

w = bobot pangkat

u_{ij} = nilai derajat keanggotaan data x_i ke *cluster* c_i

x_i = data ke-i

5. Hitung kembali matriks fuzzy pseudo-partition (derajat keanggotaan setiap data pada setiap cluster) menggunakan persamaan :

$$u_{ij} = \frac{D(X_I, C_J)^{\frac{-2}{w-1}}}{\sum_{l=1}^k D(X_I, C_J)^{\frac{-2}{w-1}}} \dots\dots\dots \text{(Persamaan 2.3)}$$

Keterangan :

c_j = centroid *cluster* ke-j

D = jarak antara data dengan centroid

w = bobot pangkat