

## **BAB III**

### **LANDASAN TEORI**

Teori-teori yang dijelaskan dalam landasan teori ini merupakan teori-teori yang terkait dengan penelitian yang dilakukan dan akan digunakan sebagai dasar atau landasan untuk memecahkan permasalahan dalam penelitian. Permasalahan clustering penduduk miskin, yang menjadi topik penelitian yang dilakukan akan diselesaikan dengan menerapkan teori data mining dengan fuzzy c-means..

#### **1.1.Data Clustering**

Hermawati (2013) dalam bukunya mengungkapkan “Clustering merupakan salah satu teknik data mining. Clustering akan mempartisi data-set menjadi beberapa sub-set atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set properti yang di share bersama, dengan tingkat similaritas yang tinggi dalam satu kelompok dan tingkat similaritas antar kelompok yang rendah”. Tujuan utama dari metode cluster adalah pengelompokkan sejumlah data/obyek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang semirip mungkin. (Santosa, 2007). Dalam clustering kita berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu cluster dan membuat jarak antar cluster sejauh mungkin. Ini berarti obyek dalam satu cluster sangat mirip satu sama lain dan berbeda dengan obyek dalam cluster-cluster yang lain.

Beberapa tipe penting dari klastering adalah :

1. Partitional dan hierarchical

Partitional clustering adalah pembagian obyek data ke dalam sub himpunan(cluster) yang tidak overlap sedemikian rupa sehingga tiap obyek data berada dalam tepat satu sub-himpunan.

Hierarchi calclustering merupakan sebuah himpunan cluster bersarang yang diatur sebagai suatu pohon hirarki. Tiap simpul (cluster) dalam pohon (kecuali simpul daun) merupakan gabungan dari anaknya (subcluster) dan simpul akar berisi semua obyek.

## 2. Exclusive dan non exclusive

Semua bentuk clustering merupakan exclusive clustering, karena setiap obyek berada tepat di satu cluster. Sebaliknya dalam overlapping atau non exclusive clustering, sebuah obyek dapat berada di lebih dari satu cluster secara bersamaan

## 3. Fuzzy dan non-fuzzy

Dalam fuzzy clustering, sebuah titik termasuk dalam setiap cluster dengan suatu nilai bobot antara 0 dan 1. Jumlah dari bobot-bobot tersebut sama dengan 1. Clustering probabilitas mempunyai karakteristik yang sama.

## 4. Partial dan complete

Dalam complete clustering, setiap obyek ditempatkan dalam sebuah cluster. Tetapi dalam partial clustering, tidak semua obyek ditempatkan dalam sebuah cluster. Kemungkinan ada obyek yang tidak tepat untuk ditempatkan di salah satu cluster, misalkan berupa outlier atau noise.

Jenis-jenis clustering berdasarkan kesamaan datanya, yaitu :

### 1. Distance based clustering

Jika dua atau lebih data memiliki jarak yang berdekatan berdasarkan sebuah jarak yang ditentukan (dalam hal ini jarak geometris), maka hal ini disebut dengan distance based clustering.

## 2. Conceptual clustering

Jika dua atau lebih data yang dimiliki oleh sebuah cluster memiliki kesamaan umum dalam konsep tertentu dan bukan berdasarkan kesamaan pengukuran maka hal ini disebut dengan conceptual clustering.

Data clustering yang secara umum digunakan untuk pengolahan data matematis adalah distance-based clustering, karena clustering jenis ini dilakukan secara matematis dan fuzzy clustering termasuk di dalamnya.

### **1.2.Fuzzy C-Means (FCM)**

“Fuzzy clustering adalah salah satu teknik untuk menentukan cluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidian untuk jarak antar vector” (Kusumadewi dan Purnomo, 2010). Fuzzy clustering sangat berguna bagi pemodelan fuzzy terutama dalam mengidentifikasi aturan-aturan fuzzy.

Ada beberapa algoritma clustering data, salah satu diantaranya adalah Fuzzy C-Means. Fuzzy C-Means merupakan salah satu metode fuzzy clustering yang paling umum digunakan dalam proses clustering. Fuzzy C-means pertama kali dikemukakan oleh Dunn (1973) dan kemudian dikembangkan oleh Bezdek (1981) yang banyak digunakan dalam pattern recognition.

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 memiliki 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 objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot oleh derajat keanggotaan titik data tersebut.

Dengan tingkat keanggotaan ini bisa dilihat data mana yang sebenarnya berada dalam daerah abu-abu. Dengan nilai tingkat keanggotaan harus diambil keputusan ke klaster mana suatu data harus dimasukkan. Hasil fuzzy c-means ini juga sangat bergantung pada nilai awal pusat klaster yang kita berikan. Fuzzy C-Means baik digunakan untuk mengelompokkan obyek yang tersebar secara tidak teratur dan memiliki nilai-nilai ekstrim di dalamnya. Maksud ketidakteraturan disini adalah obyek-obyek tersebut tidak memiliki pola yang jelas dan tidak memiliki kecenderungan yang pasti bahwa obyek tersebut akan mengelompok secara jelas pula.

Algoritma pengelompokan FCM menurut (Yan, 1994) dijelaskan sebagai berikut :

1. Input data yang akan di cluster  $X$ , berupa matriks berukuran  $n \times m$  ( $n$ =jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  = data sampel ke- $i$  ( $i=1,2,\dots,n$ ), atribut ke- $j$  ( $j=1,2,\dots,m$ ).

2. Tentukan :

Jumlah cluster =  $c$

Pangkat =  $w$

Maksimum iterasi =  $\text{MaxIter}$

Error terkecil yang diharapkan =  $\epsilon$

Fungsi obyektif awal  $= P_0 = 0$

Iterasi awal  $= t = 1$

3. Bangkitkan bilangan random  $\mu_{ik}$  sebagai elemen-elemen matriks partisi awal U.

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (1)$$

Keterangan :

$Q_i$  = matriks partisi awal

$\mu_{ik}$  = bilangan random

$i$  = banyaknya data, 1,2,,.....,n

$k$  = jumlah kluster, 1,2,,.....,c

Hitung :

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

4. Hitung pusat cluster ke-k :  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,m$ .

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

Keterangan :

$V_{kj}$  = Pusat kluster  $kj$

$k$  = jumlah kluster, (1,2,,.....,c)

$j$  = jumlah atribut, (1,2,,.....,m)

w = pangkat

X<sub>ij</sub> = data sampel ke-i (i=1,2,...,n), atribut ke-j (j=1,2,...,m)

5. Hitung fungsi objektif pada iterasi ke-t, P<sub>t</sub>

$$P_t = \sum_{k=0}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (x_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (4)$$

6. Hitung perubahan matriks partisi

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (x_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left( \left[ \sum_{j=1}^m (x_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}} \right)} \quad (5)$$

Keterangan :

i = banyaknya data, (1,2,...,n)

k = jumlah kluster, (1,2,...,c)

7. Cek kondisi berhenti :

Jika : (|P<sub>t</sub> - P<sub>t-1</sub>| < ε ) atau (t > MaxIter) maka berhenti;

Jika tidak : t = t+1, ulangi langkah 4

### 1.3.Konsep Kemiskinan

“Kemiskinan merupakan masalah deprivasi atau problematika kekurangan. Kemiskinan adalah suatu keadaan seseorang atau keluarga yang serba kekurangan” (Sen dan Foster, 1997). Dari definisi tersebut, kemiskinan dapat dibedakan menjadi dua yakni absolut dan relative,

- a. Kemiskinan absolut didefinisikan sebagai ketidakmampuan mencapai standar minimal dari kehidupan. Pengertian kebutuhan standar minimal berbeda-beda di setiap negara.
- b. Kemiskinan relatif pada sisi lain didefinisikan sebagai ketidakmampuan mencapai standar kebutuhan kontemporer, yang dihubungkan dengan kesejahteraan rata-rata atau pendapatan rata-rata masyarakat pada saat itu.

Berdasarkan data, dibedakan faktor-faktor yang mempengaruhi kemiskinan di pedesaan dan yang berada di perkotaan. Perbandingan ini penting mengingat bahwa kemiskinan tidak hanya terjadi di pedesaan, melainkan juga di perkotaan. Berdasarkan pendekatan geografis kemiskinan dapat dibedakan kemiskinan di pedesaan dan perkotaan.

- a. Kemiskinan pedesaan (rural poverty) memiliki karakteristik antara lain : i) terbatasnya akses ke fasilitas tanah dan irigasi, ii) lambatnya adaptasi terhadap teknologi modern, iii) terlalu besar beban yang ditanggung, iv) terbatasnya modal manusia, v) terkonsentrasi di daerah pinggiran dan vi) terkonsentrasi pada minoritas etnis (Pernia dan Quibra, 1999).
- b. Kemiskinan perkotaan (urban poverty) memiliki karakteristik antara lain : i) mempunyai keterbatasan terhadap akses sumber daya dan pelayanan, ii) keterbatasan sumber daya manusia yang bermutu, iii) terlalu besar beban yang ditanggung, iv) rendahnya upah yang didapat, v) banyaknya usaha kecil yang tidak terorganisasi dan vi) banyaknya kelompok-kelompok yang tidak mempunyai kapabilitas (Pernia dan Quibra, 1999).

#### **1.4. Model Garis Kemiskinan**

Kemiskinan berdasarkan pada model garis kemiskinan, dapat dibedakan menjadi tiga macam antara lain adalah :

##### **1.4.1. Model Garis Kemiskinan (BPS)**

BPS mengartikan kemiskinan sebagai ketidakmampuan untuk memenuhi standar minimum kebutuhan dasar yang meliputi kebutuhan makanan maupun non makanan. Standar minimum itu diwujudkan dalam bentuk garis kemiskinan (poverty line). Berdasarkan hal tersebut dapat diketahui berapa banyak masyarakat miskin atau orang yang memiliki tingkat konsumsi di bawah garis kemiskinan. Proses untuk mengeluarkan angka kemiskinan BPS dilakukan melakukan Survei Sosial Ekonomi Nasional (Susenas).

##### **1.4.2. Model Keluarga Sejahtera (BKKBN)**

Berbeda dengan BPS, Badan Koordinasi Keluarga Berencana Nasional (BKKBN) lebih melihat dari sisi kesejahteraan dari pada sudut kemiskinan. Unit survei yang dipergunakan oleh BKKBN adalah keluarga, sementara BPS menggunakan rumah tangga seperti telah disebutkan di atas. Pendataan dimulai dari unit terkecil yakni Rukun Tetangga atau Dusun yang kemudian dapat diagregasikan pada tingkat Desa, Kecamatan, Kabupaten, Propinsi sampai dengan tingkat nasional.

##### **1.4.3. Indikator keluarga miskin Kabupaten Bantul**

Indikator yang digunakan untuk menentukan keluarga miskin di Kabupaten Bantul diantaranya meliputi Aspek Pangan, Sandang, Papan, Penghasilan, Kesehatan, Pendidikan, Kekayaan 1, Kekayaan 2, Aspek Air Bersih, Aspek Listrik dan Aspek

Jumlah Jiwa. Berikut ini adalah penjelasan secara rinci mengenai keempat indikator tersebut :

Tabel 3.1. Indikator Kemiskinan BKKBN

No	Aspek	Keterangan	Skor
1	Pangan	Seluruh anggota keluarga tidak mampu makan dengan layak atau senilai Rp. 1.500,- minimal 2 kali dalam sehari	12
2	Sandang	Lebih dari sebagian anggota keluarga tidak memiliki pakaian pantas pakai minimal 6 stel	9
3	Papan	Tempat tinggal/ rumah berlantai tanah/ berdinding bambu/berataprumbia	9
4	Penghasilan	Jumlah penghasilan yang diterima seluruh anggota keluarga yang berusia 16 tahun keatas < Rp. 666.788	35
5	Kesehatan	Bila ada anggota keluarga yang sakit tidak mampu berobat ke fasilitas kesehatan dasar	6
6	Pendidikan	Keluarga tidak mampu menyekolahkan anak yang berumur 7 – 15 tahun	6
7	Kekayaan 1	Jumlah kekayaan/aset milik keluarga kurang dari Rp.2.500.000,-	5
8	Kekayaan 2	Tanah bangunan yang ditempati bukan milik sendiri	6
9	Air Bersih	Tidak menggunakan air bersih untuk keperluan makan, minum & MCK	4
10	Listrik	Tidak menggunakan listrik untuk keperluan rumah	3

		tangga	
11	Jumlah Jiwa	Jiwa dalam KK ( termasuk kepala keluarga ) 5 jiwa atau lebih	5