

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 *hard c-means*..

3.1.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 clustering 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.

3.2. Hard C-Means (HCM)

Algoritma clustering HCM merupakan satu dari beberapa teknik *clustering* yang banyak digunakan. Metode tersebut berusaha untuk menghasilkan *hard C-partitions* yang optimal dari nilai X dengan meminimalkan fungsi obyektif. Untuk menghitung fungsi obyektif pada hard partition digunakan persamaan (3.1).

$$J_w(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik} \|X_k - V_i\|^2 \quad (3.1)$$

Dalam metode pengelompokan *hard partition*, setiap objek dari kumpulan data harus ditugaskan secara tepat pada satu klaster. Analisis cluster adalah teknik untuk memecah data ke dalam komponen terkait sedemikian rupa sehingga pola dapat dilihat (Cox, 2005). Sebuah *cluster* mempunyai titik pusat yang disebut *centroid* yang didasarkan dari rata-rata bobot tiap data yang terdapat pada *cluster* tersebut. Keanggotaan pada *cluster* ditentukan dengan mengukur jarak data

dengan titik pusat *cluster*. Tiap data memiliki derajat keanggotaan pada tiap *cluster*, data dianggap anggota dari sebuah cluster dimana derajat keanggotaannya yang paling tinggi. *Flowchart* algoritma *Hard C-Means* dapat dilihat pada Gambar 3.1.

Algoritma *Hard C-Means* menurut Wang (1997) :

Langkah 1.

Dataset $X = \{X_1, X_2, \dots, X_n\}, X_i \in R^p$,

dengan $2 \leq c < n$

dan inisialisasi $U^{(0)} \in M_c$

Keterangan :

n = banyaknya data

Tentukan :

c = banyaknya cluster

MaxIter = Maksimum iterasi

ϵ = Error terkecil yang diharapkan

Langkah 2.

Pada iterasi $l, l = 0, 1, 2, \dots$, hitung pusat cluster dengan persamaan (3.2)

$$V_i^{(l)} = \frac{\sum_{k=1}^n (\mu_{ik}^{(l)} * X_k)}{\sum_{k=1}^n \mu_{ik}^{(l)}}, 1 \leq i \leq c \quad (3.2)$$

Dimana $[\mu_{ik}^{(l)}] = U^{(l)}$

Keterangan :

$V_i^{(l)}$ = Pusat kluster ke i pada iterasi ke l
 i = index cluster, (1,2,.....,c)
 k = index data, (1,2,.....,n)
 X_k = data sampel ke- k ($k=1,2,\dots,n$)
 μ_{ik} = nilai keanggotaan cluster ke i dan data ke k

Langkah 3.

Perbaiki nilai U yang baru

$U^{(l)} = [u_{ik}^{(l)}]$ menjadi $U^{(l+1)} = [u_{ik}^{(l+1)}]$ menggunakan persamaan (3.3)

$$U^{(l+1)} = \begin{cases} 1 & \|X_k - V_i^{(l)}\| = \min_{1 \leq j \leq c} (\|X_k - V_j^{(l)}\|) \\ 0 & \text{lainya} \end{cases} \quad (3.3)$$

Keterangan :

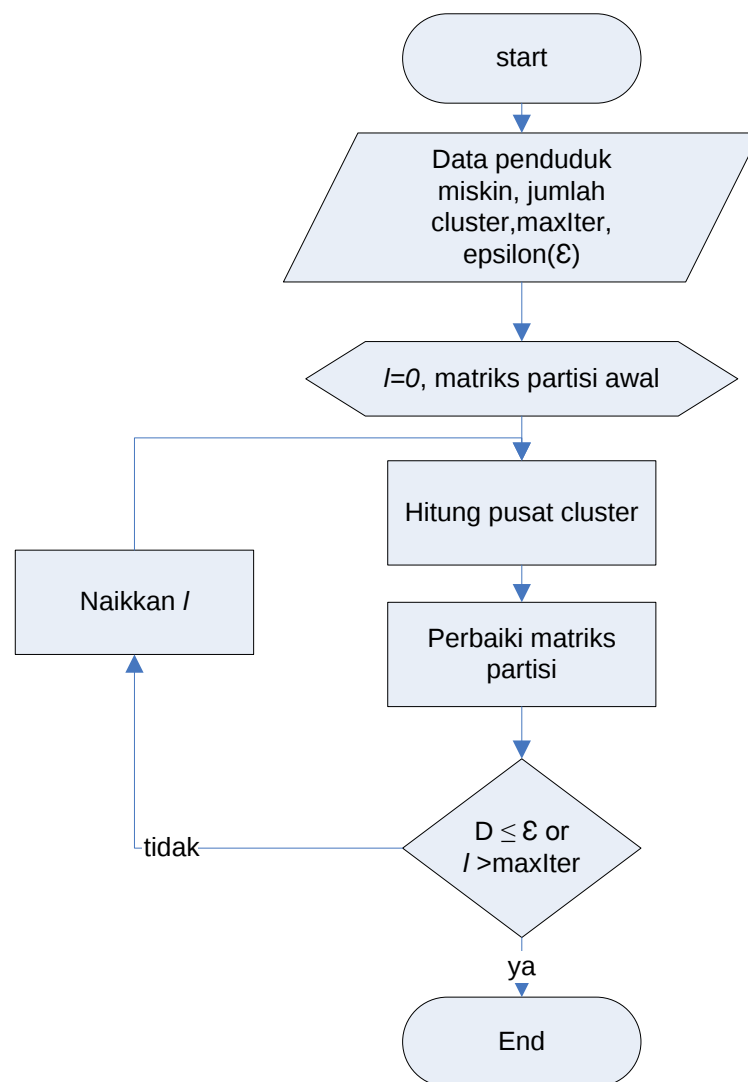
$V_i^{(l)}$ = Pusat kluster ke i pada iterasi ke l
 i = index cluster, (1,2,.....,c)
 k = index data, (1,2,.....,n)
 X_k = data sampel ke- k ($k=1,2,\dots,n$)
 μ_{ik} = nilai keanggotaan cluster ke i dan data ke k

Langkah 4.

Jika $\|U^{(l+1)} - U^{(l)}\| < \epsilon$, maka iterasi dihentikan, namun apabila tidak, maka naikkan iterasi ($l = l+1$) dan kembali kelangkah 2.

Keterangan :

Pencarian nilai D dapat dilakukan dengan mengambil elemen terbesar dari nilai mutlak selisih antara $\mu_{ik}(l)$ dengan $\mu_{ik}(l-1)$.



Gambar 3.1 Flowchart hard c-means

3.3.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. Pembandingan ini penting mengingat bahwa kemiskinan tidak hanya terjadi di pedesaan, melainkan juga diperkotaan. 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).

3.4.Model Garis Kemiskinan

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

3.4.1. 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).

3.4.2. 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.

3.4.3. 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 tangga	3
11	Jumlah Jiwa	Jiwa dalam KK (termasuk kepala keluarga) 5 jiwa atau lebih	5