

BAB II

LANDASAN TEORI

2.1 Perpustakaan

Menurut Undang-undang Perpustakaan (UU nomor 43 tahun 2007) disebutkan bahwa perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan/ atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka. Sedangkan menurut Sulistyobasuki (1991: 3) perpustakaan adalah: sebuah ruangan, bagian sebuah gedung, ataupun gedung itu sendiri yang digunakan untuk menyimpan buku dan terbitan lainnya yang biasanya disimpan menurut tata susunan tertentu untuk digunakan pembaca, bukan untuk dijual.

Institusi merupakan struktur dan mekanisme aturan dan kerjasama sosial yang mengawal perlakuan dua atau lebih individu. Institusi bisa juga berarti lembaga yaitu badan (organisasi) yang bermaksud melakukan suatu penyelidikan keilmuan atau melakukan suatu usaha. Pengelola berasal dari kata *to manage* yang berarti mengurus, mengatur, melaksanakan, mengelola. Jadi pengelola adalah seseorang yang mengurus, mengatur, melaksanakan, mengelola. Koleksi berarti kumpulan benda yang digemari. Dengan demikian maka koleksi karya tulis, karya cetak dan/ atau karya rekam adalah kumpulan informasi yang berbentuk tulisan tangan, buku cetakan maupun yang direkam dalam berbagai media

termasuk media elektronik dan digital. Profesional berarti memerlukan kepandaian khusus untuk menjalankan. Dengan demikian “mengelola koleksi karya tulis, karya cetak dan atau karya rekam secara profesional” berarti mengurus, mengatur, melaksanakan, mengelola kumpulan informasi dalam berbagai bentuk atau format dimana dalam melakukan pengelolaannya tersebut diperlukan keahlian khusus. Baku berarti sesuatu yang dipakai dasar ukuran (nilai, harga, dsb) standar. Jadi sistem baku merupakan sistem yang digunakan sebagai dasar dalam melakukan pengelolaan koleksi karya tulis, karya cetak dan atau karya rekam. Pemustaka menurut UU 43 tahun 2007 adalah pengguna perpustakaan, yaitu perseorangan, kelompok orang, masyarakat, atau lembaga yang memanfaatkan fasilitas layanan perpustakaan.

Dengan demikian maka makna dari kedua definisi yang dikutip pada awal tulisan ini adalah: perpustakaan merupakan institusi atau lembaga tempat menyimpan informasi dalam bentuk buku dan bentuk-bentuk lain yang disimpan menurut aturan tertentu yang baku untuk digunakan oleh orang lain (bukan hanya digunakan oleh pribadi) secara gratis untuk bermacam-macam tujuan atau kebutuhan seperti untuk pendidikan, penelitian, pelestarian, informasi, dan rekreasi. Mari kita bandingkan dengan definisinya Wikipedia yang mendefinikan perpustakaan sebagai berikut: “*A library is a collection of sources, resources, and services, and the structure in which it is housed; it is organized for use and maintained by a public body, an institution, or a private individual. In the more traditional sense, a library is a collection*

of books. It can mean the collection, the building or room that houses such a collection, or both.” Jadi makna beberapa definisi tersebut memiliki pengertian yang sama yakni: (1) merupakan kumpulan bahan perpustakaan; (2) dikelola secara profesional dengan sistem tertentu (baku); (3) dikelola oleh lembaga atau institusi dan atau individu; (4) diselenggarakan untuk kebutuhan pemustaka.

2.2. Dewey Decimal Classification

Dewey Decimal Classification adalah merupakan salah satu sistem pengklasifikasian koleksi buku yang ditemukan oleh Melvil Dewey. Nama lengkapnya Melville Louis Kassuth Dewey (1851-1931). Pada 1874 Dewey sebagai pustakawan di Amhers College, Massachuseetts, Tahun 1876 ia menerbitkan DDC edisi pertama dengan judul “*A classification and subject index for a library*”. Terbit pertama kali hanya sebanyak 42 halaman yang berisi 12 halaman pendahuluan, 12 halaman bagan dan 18 halaman indeks. Sejak edisi pertama diterbitkan, DDC terus menerus mengikuti perkembangan ilmu pengetahuan. Banyak subyek-subyek baru yang ditambahkan. Adakalanya notasi mengalami perluasan dan perubahan lokasi karena perkembangan subyek tersebut. Kelestarian DDC sampai dapat mencapai umur lebih seabad dan banyak pemakainya di dunia, disebabkan karena DDC secara berkala ditinjau kembali dan diterbitkan edisi barunya. Lembaga yang mengawasi dan mendukung penerbitan DDC ialah “*The Lake Placed Education Foundation*” dan “*The Library of Congress*” di Amerika Serikat sarana komunikasi

diterbitkan "*Decimal Classification, addition, notes, decisions*" (disingkat DC). DDC dalam pengembangannya menggunakan sistem desimal angka arab sebagai simbol notasinya.

2.2.1 Komponen DDC

Dalam klasifikasi Persepuluhan Dewey ini terdapat 3 komponen, yaitu Bagan, indeks Relatif, dan Tabel-tabel. Untuk lebih jelasnya dapat diperhatikan pada uraian berikut ini.

A. Bagan (Schedules)

Klasifikasi Dewey adalah bagan klasifikasi sistem hirarki yang menganut prinsip "desimal" untuk membagi semua bidang ilmu pengetahuan. Ilmu pengetahuan dibagi ke dalam 9 kelas utama, yang diberi kode/lambang angka (selanjutnya disebut notasi). Seperti telah dijelaskan pada halaman sebelumnya. Dalam DDC ini semakin khusus suatu subyek, semakin panjang notasinya. Karena banyak angka yang ditambahkan pada notasi dasarnya. Pembagiannya dari umum ke khusus.

B. Indeks Relatif (Relative Index)

Untuk membantu mencari notasi suatu subyek dalam DDC terdapat 'Indeks Relatif'. Pada indeks relatif ini terdaftar sejumlah istilah yang disusun berabjad. Istilah-istilah tersebut mengacu ke notasi yang terdapat dalam bagan. Dalam indeks ini didaftar sinonim untuk suatu istilah, hubungan-hubungan dengan subyek lainnya. Bila suatu subyek telah ditemukan dalam indeks relatif, hendaklah ditentukan lebih lanjut aspek dari subyek yang bersangkutan. Cara yang paling cepat untuk menentukan notasi suatu subyek adalah melalui indeks relatif. Tetapi

menentukan notasi hanya melalui dan berdasarkan indeks relatif saja tidak dapat dibenarkan. Setelah suatu subyek diperoleh notasinya dalam indeks relatif, harus diadakan pengecekan dengan notasi yang terdapat dalam bagan. Dengan demikian dapat diketahui apakah notasi tersebut betul-betul sesuai dengan karya yang sedang diklasifikasikan.

C. Tabel-Tabel

Kecuali pembagian kelas secara desimal dengan notasi yang terdaftar dalam bagan, DDC juga mempunyai sarana lain. Untuk membagi/ memperluas subyek lebih lanjut, yaitu dengan menyediakan sejumlah tabel pembantu atau *auxiliary tables*. Notasi pada table-table tersebut hanya dapat digunakan dalam rangkaian dengan notasi yang terdapat dalam bagan. Dengan kata lain, notasi yang terdapat dalam tabel tidak pernah berdiri sendiri, selalu dirangkaikan dengan notasi dalam bagan.

2.2.2 Cara Kerja Sistem DDC

Dewey membagi berbagai disiplin pengetahuan yang ada ke dalam sepuluh kelas utama (*main class*) dengan satu “*Generalities*”. Selanjutnya, kelas-kelas utama tersebut dibagi lagi ke dalam sepuluh divisi, dan setiap divisi dibagi lagi ke dalam sepuluh *section*. Ke-sepuluh kelas utama tersebut adalah :

1. 000 *Generalities*.
2. 100 *Philosophy, psychology*.
3. 200 *Religion*.
4. 300 *Social Science (incl. economics)*.

5. 400 *Language*.
6. 500 *Natural Science*.
7. 600 *Technology (incl. medicine, management)*.
8. 700 *Art (incl. architecture, paintings, photography)*.
9. 800 *Literature*.
10. 900 *History geography, biography*.

Kelas utama 000 digunakan untuk karya-karya yang tidak terbatas pada satu disiplin ilmu saja, misalnya ensiklopedia. Kelas ini juga digunakan untuk bidang yang berhubungan dengan pengetahuan dan informasi, misalnya ilmu komputer, ilmu perpustakaan. Angka pertama pada nomor-nomor tersebut menunjukkan *main class*. Masing-masing *main class* terdiri dari 10 divisi, juga menggunakan nomor 0-9. Angka yang menunjukkan divisi adalah angka kedua. Misalnya, 600 digunakan untuk buku-buku yang membahas tentang teknologi/ ilmu terapan secara umum, 610 untuk ilmu kedokteran, 620 untuk ilmu teknik, 630 untuk pertanian.

Masing-masing divisi dibagi lagi menjadi 10 section, juga menggunakan nomor 0-9. Angka ketiga dalam nomor DDC menunjukkan section. Misal, 610 digunakan untuk karya umum di bidang kedokteran, 611 untuk anatomi manusia, 612 untuk fisiologi manusia, 613 untuk bidang promosi kesehatan. Selanjutnya, setelah tiga nomor utama tersebut, angka desimal dapat digunakan sejauh diperlukan. Misalnya, 611.1 untuk buku yang membahas tentang organ-organ kardiovaskular, 611.2 untuk buku yang membahas tentang organ-organ pernafasan.

2.2.3 Keunggulan Dan Kelemahan DDC

a. keunggulan

1. DDC merupakan sistem yang praktis dan paling banyak digunakan di dunia karena kehandalannya.
2. Indeks relatif menyatukan subjek yang sama dengan aspek berlainan yang tersebar dalam berbagai disiplin ilmu.
3. Notasi dengan angka Arab dikenal secara universal. Pustakawan dengan latar belakang budaya dan bahasa yang berbeda dengan mudah dapat menyesuaikan sistem tersebut.
4. Urutan numeric kasat mata memudahkan penjajaran dan penempatan buku-buku di rak.
5. Penggunaan notasi desimal memungkinkan perluasan pembagian sub divisi tanpa batas.
6. Revisi berkala dengan interval teratur menjamin kemutakhiran bagan klasifikasi DDC.

b. Kelemahan

1. Disiplin ilmu yang berakitan sering kali terpecah. Misalnya 300 ilmu-ilmu sosial terpisah dari 900 Geografi dan Sejarah. Pada bidang lain, kelas 400 Bahasa terpisah dari 800 Sastra.
2. Penempatan beberapa subjek tertentu dipermasalahkan. Misalnya ilmu perpustakaan pada kelas karya umum 000, psikologi sebagai sub divisi dari filsafat 100 dan olahraga serta hiburan dan kesenian 700.

3. Pada kelas 800, karya literer oleh pengarang yang sama disebarakan berdasarkan bentuk literer. Padahal para pakar menginginkannya menjadi satu.
4. Basis sepuluh dalam DDC membatasi kemampuan perluasan sistem notasi. Karena, dari sepuluh divisi hanya Sembilan yang dapat diperluas untuk member tempat subjek yang bertingkat sama dalam hirarki. Bila sebuah subjek dibagi dalam 10 subdivisi, perluasan hanya mampu menampung 9 subdivisi yang setara. Padahal dalam kenyataan subdivisi yang ada lebih dari 9 sehingga beberapa subdivisi terpaksa mengalah turun lebih rendah menjadi subdivisi-subdivisi.
5. Laju pertumbuhan ilmu pengetahuan tidak sama sehingga membuat struktur ilmu pengetahuan tidak seimbang. Ada kelas yang dianggap statis seperti Agama dan Filsafat, ada pula yang tumbuh cepat seperti kelas 300 ilmu-ilmu sosial, 500 Sains dan 600 Teknologi. Pada kelas 300, 500 dan 600 ada kesan terlalu padat.
6. Perluasan sebuah subjek dapat dilakukan dengan sistem desimal, tetapi anehnya angka baru untuk subjek baru tidak dobel disisipkan antara nomor koordinat (yaitu nomor antara divisi, misalnya 610 dan 620) walaupun memang tempat yang layak ada diantara nomor koordinat. DDC melakukan penambahan subjek baru dengan memasukkannya pada subdivisi dari subjek yang telah ada.

7. Karena kemampuan perluasan tidak terbatas berkat sistem desimal, hasilnya ialah angka yang cukup panjang untuk beberapa subjek. Angka yang panjang menyulitkan penempatan buku di rak.

2.3. Sistem Pendukung Keputusan

Sistem Pendukung Keputusan atau yang biasa disebut *Decision Support System* (DSS) adalah sebuah sistem yang ditujukan untuk mendukung para pengambil keputusan manajerial untuk masalah semiterstruktur. Scott Morton mendefinisikan DSS sebagai “sistem berbasis komputer interaktif, yang membantu para pengambil keputusan untuk menggunakan data dan berbagai model untuk memecahkan masalah-masalah tidak terstruktur” (Gory dan Scott Morton, 1971). Seperti yang disebutkan oleh Turban (2005: 136) yaitu DSS dimaksudkan untuk menjadi alat bantu bagi para pengambil keputusan untuk memperluas kapabilitas mereka, namun tidak untuk menggantikan penilaian mereka. DSS ditujukan untuk keputusan-keputusan yang memerlukan penilaian atau pada keputusan-keputusan yang sama sekali tidak dapat didukung oleh algoritma. Sebagai istilah umum DSS digunakan untuk menggambarkan semua sistem terkomputerisasi yang mendukung pengambilan keputusan pada suatu organisasi. Tujuan utama dari DSS yaitu untuk mendukung dan meningkatkan pengambilan keputusan (Turban, 2005: 138).

Sesuai dengan konsep DSS diatas, maka menurut Turban (2005: 20) yang membedakan DSS dengan Sistem Informasi Manajemen adalah

“Organisasi bisa saja memiliki suatu sistem manajemen pengetahuan untuk memandu seluruh personelya dalam memecahkan masalah, ia dapat memiliki DSS tersendiri untuk pemasaran, keuangan, dan akuntansi, sistem SCM untuk produksi, dan beberapa sistem pakar untuk membuat diagnosis dan *help desk* perbaikan”. Jadi bisa dikatakan perbedaan antara Sistem Pendukung Keputusan (SPK) dengan Sistem Informasi Manajemen (SIM) dapat dilihat pada Tabel 2.1.

Tabel 2.1 Perbedaan Antara Sistem Pendukung Keputusan dengan Sistem Informasi Manajemen (Laudon, 2005)

Sistem Pendukung Keputusan (SPK)	Sistem Informasi Manajemen (SIM)
Memberikan serangkaian kemampuan baru untuk keputusan-keputusan non rutin dan kendali pengguna.	Menghasilkan laporan regular terjadwal dan baku berdasarkan data yang diambil dan dirangkum dari sistem pemrosesan transaksi (SPT) milik organisasi. Sebuah laporan SIM pada umumnya menunjukkan ringkasan penjualan perbulan untuk masing-masing wilayah penjualan utama.
Menekankan perubahan, fleksibilitas, dan respon cepat.	Memberikan kepada manajer laporan berdasarkan aliran rutin

Sistem Pendukung Keputusan (SPK)	Sistem Informasi Manajemen (SIM)
Dengan adanya SPK tidak perlu susah payah menghubungkan pengguna dengan alur informasi terstruktur, menghasilkan model-model, asumsi dan pertanyaan khusus yang semakin diperluas serta menampilkan grafik.	data dan membantu kendali umum perusahaan.
SPK menyelesaikan semi terstruktur dan menganalisis masalah tak terstruktur.	Menyelesaikan masalah-masalah terstruktur.

2.4. Pengertian Logika Fuzzy

Menurut Kusuma Dewi [2006] Logika Fuzzy adalah peningkatan dari logika *Boolean* yang berkaitan dengan konsep kebenaran sebagian. Logika klasik menyatakan bahwa segala hal dapat diekspresikan dalam istilah binary (0 atau 1, hitam atau putih, ya atau tidak), logika fuzzy menggantikan kebenaran *boolean* dengan tingkat kebenaran.

Logika Fuzzy memungkinkan nilai keanggotaan antara 0 dan 1, tingkat keabuan dan juga hitam dan putih, dan dalam bentuk linguistik, konsep tidak pasti seperti "sedikit", "sedang", dan "banyak". Pada akhir

tahun 1965, Dr. Lotfi A. Zadeh memperkenalkan teori himpunan fuzzy yang dapat digunakan untuk mempresentasikan masalah ketidakpastian.

2.4.1. Himpunan Fuzzy

Pada himpunan tegas (*crisp set*), nilai keanggotaan suatu item x dalam suatu himpunan A (ditulis $\mu_A[x]$) memiliki 2 kemungkinan :

1. Satu (1), artinya x adalah anggota A
2. Nol (0), artinya x bukan anggota A

Contoh 1 :

Jika diketahui :

$S = \{1, 2, 3, 4, 5, 6\}$ adalah semesta pembicaraan

$A = \{1, 2, 3\}$

$B = \{3, 4, 5\}$

maka :

Nilai keanggotaan 2 pada A , $\mu_A[2] = 1$, karena $2 \in A$

Nilai keanggotaan 4 pada A , $\mu_A[4] = 0$, karena $4 \notin A$

Contoh 2:

Misal variable umur dibagi menjadi 3 katagori :

MUDA, umur < 35 tahun

PARUHBAYA, $35 \leq \text{umur} \leq 55$ tahun

TUA, umur > 55 tahun



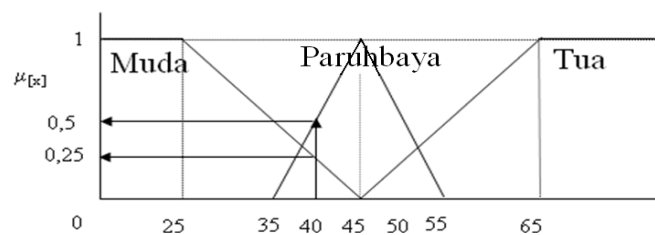
Gambar 2.1 Keanggotaan Himpunan Klasik (*Crisp*) Umur Muda, Paruhbaya dan Tua

- Apabila seseorang berusia 34 tahun, maka ia dikatakan MUDA
- Apabila seseorang berusia 35 tahun, maka ia dikatakan TIDAK MUDA
- Apabila seseorang berusia 35 tahun, maka ia dikatakan PARUHBAYA
- Apabila seseorang berusia 35 tahun kurang 1 hari, maka ia dikatakan TIDAK PARUHBAYA
- Apabila seseorang berusia 55 tahun, maka ia dikatakan TIDAK TUA
- Apabila seseorang berusia 55 tahun lebih $\frac{1}{2}$ hari, maka ia dikatakan TUA

Dari sini bisa dikatakan bahwa pemakaian himpunan *crisp* untuk menyatakan umur sangat tidak adil, adanya perubahan kecil saja pada suatu nilai mengakibatkan perbedaan katagori yang cukup signifikan.

Himpunan fuzzy digunakan untuk mengantisipasi hal tersebut. Seseorang dapat masuk dalam 2 himpunan yang berbeda. MUDA dan PARUHBAYA, PARUHBAYA dan TUA, dsb. Seberapa besar eksistensinya dapat dilihat pada nilai/ derajat keanggotaannya.

Gambar berikut menunjukkan himpunan fuzzy untuk variabel umur :



Gambar 2.2 Himpunan Fuzzy Untuk Variabel Umur

2.4.2. Operasi Logika (Operasi Himpunan Fuzzy)

Operasi logika adalah operasi yang mengkombinasikan dan memodifikasi 2 atau lebih himpunan fuzzy. Nilai keanggotaan baru hasil operasi dua himpunan disebut *firing strength* atau α predikat. Terdapat 3 operasi dasar pada himpunan fuzzy yang diciptakan oleh Zadeh :

1. *AND* (*Intersection*)
2. *OR* (*Union*)
3. *NOT* (*Complement*)

1. *AND* (*Intersection*)

Fuzzy *intersection* (\cap): irisan dari 2 himpunan fuzzy adalah minimum dari tiap pasang elemen pada kedua himpunan.

Contoh :

$$\begin{aligned}
 A \cap B &= \{\text{MIN}(1.0, 0.2), \text{MIN}(0.20, 0.45), \text{MIN}(0.75, 0.50)\} \\
 &= \{0.2, 0.20, 0.50\}
 \end{aligned}$$

Misal nilai keanggotaan umur 27 pada himpunan muda adalah $\mu_{\text{MUDA}}[27] = 0,6$ dan nilai keanggotaan 2 juta pada himpunan penghasilan TINGGI adalah $\mu_{\text{GAJITINGGI}}[2\text{juta}] = 0,8$.

Maka α -predikat untuk usia MUDA dan berpenghasilan TINGGI adalah nilai keanggotaan minimum :

$$\begin{aligned} & \mu_{\text{MUDA} \cap \text{GAJITINGGI}} \\ &= \min(\mu_{\text{MUDA}}[27], \mu_{\text{GAJITINGGI}}[2\text{juta}]) \\ &= \min(0,6; 0,8) \\ &= 0,6 \end{aligned}$$

2. OR (Union)

Fuzzy union (\cup): union dari 2 himpunan adalah maksimum dari tiap pasang elemen pada kedua himpunan.

Contoh :

$$A = \{1.0, 0.20, 0.75\}$$

$$B = \{0.2, 0.45, 0.50\}$$

$$\begin{aligned} A \cup B &= \{\text{MAX}(1.0, 0.2), \text{MAX}(0.20, 0.45), \text{MAX}(0.75, 0.50)\} \\ &= \{1.0, 0.45, 0.75\} \end{aligned}$$

Misal nilai keanggotaan umur 27 pada himpunan muda adalah $\mu_{\text{MUDA}}[27] = 0,6$ dan nilai keanggotaan 2 juta pada himpunan penghasilan TINGGI adalah $\mu_{\text{GAJITINGGI}}[2\text{juta}] = 0,8$.

Maka α -predikat untuk usia MUDA atau berpenghasilan TINGGI adalah nilai keanggotaan maksimum :

$$\mu_{\text{MUDA} \cup \text{GAJITINGGI}}$$

$$\begin{aligned}
 &= \max(\text{MUDA}[27], \text{GAJITINGGI}[2\text{juta}]) \\
 &= \max(0,6 ; 0,8) \\
 &= 0,8
 \end{aligned}$$

3. NOT (Complement)

Komplemen dari variabel fuzzy dengan derajat keanggotaan x adalah $(1-x)$.

Komplemen ($_c$): komplemen dari himpunan fuzzy terdiri dari semua komplemen elemen.

Contoh :

$$\begin{aligned}
 A^c &= \{1 - 1.0, 1 - 0.2, 1 - 0.75\} \\
 &= \{0.0, 0.8, 0.25\}
 \end{aligned}$$

Misal nilai keanggotaan umur 27 pada himpunan muda adalah $\mu_{\text{MUDA}}[27] = 0,6$.

Maka α -predikat untuk usia TIDAK MUDA adalah :

$$\begin{aligned}
 \mu_{\text{MUDA}'}[27] &= 1 - \text{MUDA}[27] \\
 &= 1 - 0,6 \\
 &= 0,4
 \end{aligned}$$

2.5. Clustering

Metode data *mining* dapat diklasifikasikan berdasarkan fungsi atau aplikasi dimana mereka digunakan. Teknik data *mining* yang digunakan dalam menyelesaikan tugas akhir ini adalah menggunakan teknik *cluster* analisis. *Clustering* adalah sekumpulan data yang memiliki kesamaan terhadap data lain yang ada dalam satu *cluster* dan tidak memiliki

kesamaan dengan objek di *cluster* yang berbeda (Han, 2007:383). *Clustering* atau yang biasa disebut *data segmentation* di dalam sebuah aplikasi karena *clustering* membagi data yang sangat besar ke dalam kelompok-kelompok berdasarkan kepada kesamaan yang ada.

Clustering juga dapat digunakan untuk *outlier detection*, dimana jarak terluar lebih menarik dari kasus-kasus yang biasanya. Sebagai cabang dari statistika, analisis *cluster* telah lebih luas dipelajari dalam beberapa tahun, mengutamakan pada *distance-based cluster analysis* (Han, 2007:384). Perangkat analisis *cluster* berdasarkan pada *k-means*, *k-medoids*, dan beberapa metode lainnya yang juga dibangun ke dalam banyak paket software analisis statistik, seperti S-Plus, SPSS, dan SAS.

Metode *clustering* pada dasarnya melakukan segmentasi atau pengelompokan suatu populasi data yang heterogen menjadi beberapa sub group atau *cluster*. Metode ini dikategorikan ke dalam teknik *undirect knowledge* atau *unsupervised learning* karena tidak membutuhkan proses pelatihan untuk klasifikasi awal data dalam masing-masing group atau *cluster*. Ada beberapa kategori pendekatan *clustering* (Gunadarma, 2008), diantaranya :

- a. Algoritma Partisi : mempartisi objek-objek ke dalam k *cluster* dan realokasi objek-objek secara iteratif untuk memperbaiki *clustering*.
- b. Algoritma Hirarkis : *Agglomerative* dimana setiap objek merupakan *cluster*, gabungan dari *cluster-cluster* membentuk *cluster* yang besar dan *Divisive* dimana semua objek berada

dalam suatu *cluster*, pembagian *cluster* tersebut membentuk *cluster-cluster* yang kecil.

- c. Metode berbasis densitas : berbasis koneksitas dan fungsi densitas dan *noise* disaring, kemudian temukan *cluster-cluster* dalam bentuk sembarang.
- d. Metode berbasis *grid* : kuantitas ruang objek ke dalam struktur *grid*.
- e. Berbasis Model : menggunakan model untuk menemukan keadaan data yang baik.

2.6. Fuzzy C-Means Cluster (FCM)

Fuzzy *Clustering* adalah salah satu teknik untuk menentukan *cluster* optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar vektor. 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 (FCM) adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

Konsep dasar Fuzzy C-Means *Cluster* (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 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.

Output dari FCM bukan merupakan fuzzy *inference system*, namun merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu fuzzy *inference system*. [SRI&HARI2010]

2.6.1 Algoritma Fuzzy C-Means Cluster (FCM)

Algoritma Fuzzy C-Means adalah sebagai berikut:

1. *Input* data yang akan di*cluster* X , berupa matriks berukuran $n \times m$ (n =jumlah sample data, m =atribut setiap data). X_{ij} =data sample ke- i ($i=1,2,\dots,n$), atribut ke- j ($j=1,2,\dots,m$).
2. Tentukan:
 - o Jumlah *cluster* (c)
 - o Pangkat (w)
 - o Maksimum iterasi (MaxIter)
 - o Error terkecil yang diharapkan (ϵ)
 - o Fungsi obyektif awal ($P_0 = 0$)
 - o Iterasi awal ($t=1$)

3. Bangkitkan nilai acak μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$; sebagai elemen-elemen

matriks partisi awal U (derajat keanggotaan dalam *cluster*).

μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu *cluster*.

Posisi dan nilai matriks dibangun secara random. Dimana nilai keanggotaan

terletak pada interval 0 sampai dengan 1. Pada posisi awal matriks partisi

U masih belum akurat begitu juga pusat *clusternya*. Sehingga

kecendrungan data untuk masuk suatu *cluster* juga belum akurat.

Hitung jumlah setiap kolom (atribut)

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad \dots \dots \dots (2.1)$$

Q_j adalah jumlah nilai derajat keanggotaan perkolom = 1

dengan $j=1,2,\dots,m$

Hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad \dots \dots \dots (2.2)$$

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

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w * X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \dots\dots\dots (2.3)$$

5. Hitung fungsi obyektif pada iterasi ke- t , P_t .

Fungsi obyektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kecendrungan data untuk masuk ke *cluster* mana pada *step* akhir.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \dots\dots\dots (2.4)$$

6. Hitung perubahan matriks partisi:

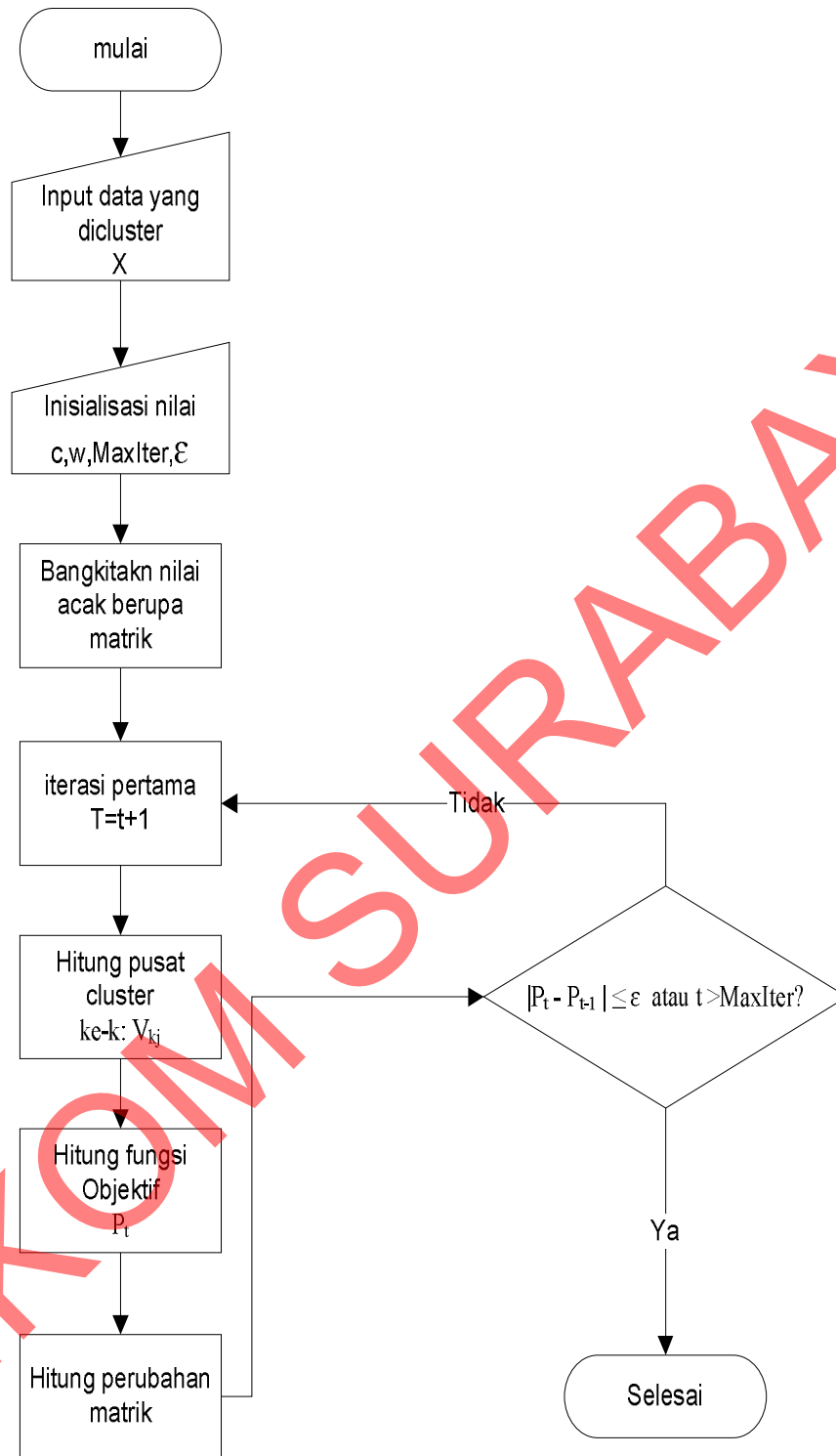
$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \dots\dots\dots (2.5)$$

dengan: $i=1,2,\dots,n$; dan $k=1,2,\dots,c$.

7. Cek kondisi berhenti:

o jika: $(|P_t - P_{t-1}| \leq \epsilon)$ atau $(t > \maxIter)$ maka berhenti;

o jika tidak: $t=t+1$, ulangi langkah ke-4.



Gambar 2.3 Flowchart Metode Fuzzy C-means Clustering