

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Prestasi Akademik**

##### **2.1.1 Pengertian Prestasi**

Dalam proses pendidikan, prestasi dibatasi pada prestasi belajar atau prestasi akademik. Djamarah (2002) mendefinisikan prestasi akademik adalah hasil yang diperoleh berupa kesan-kesan yang mengakibatkan perubahan dalam diri individu sebagai hasil akhir dari aktivitas belajar. Sedangkan definisi prestasi akademik menurut Azwar (2002) adalah bukti peningkatan atau pencapaian yang diperoleh seorang siswa sebagai pernyataan ada tidaknya kemajuan atau keberhasilan dalam program pendidikan.

Selanjutnya menurut Suryabrata (2006) prestasi akademik adalah hasil belajar terakhir yang dicapai oleh siswa dalam jangka waktu tertentu, yang mana di sekolah prestasi akademik siswa biasanya dinyatakan dalam bentuk angka atau symbol tertentu. Kemudian dengan angka atau simbol tersebut, orang lain atau siswa sendiri akan dapat mengetahui sejauh mana prestasi akademik yang telah dicapai. Dengan demikian, prestasi akademik di sekolah merupakan bentuk lain dari besarnya penguasaan bahan pelajaran yang telah dicapai siswa, dan rapor bisa dijadikan hasil belajar terakhir dari penguasaan pelajaran tersebut.

Berdasarkan uraian dari beberapa pendapat tersebut, dapat disimpulkan bahwa prestasi akademik adalah hasil atau pencapaian yang diperoleh siswa dari aktivitas belajar, yang dinyatakan dalam bentuk angka atau simbol tertentu.

##### **2.1.2 Ukuran prestasi**

Menurut Azwar (1996) prestasi atau keberhasilan belajar dapat dioperasionalkan dalam bentuk-bentuk atau indikator-indikator berupa:

a. Nilai rapor

Dengan nilai rapor, kita dapat mengetahui prestasi belajar siswa. Siswa yang nilai rapornya baik dikatakan prestasinya tinggi, sedangkan yang nilainya jelek dikatakan prestasi belajarnya rendah.

b. Indeks prestasi akademik

Indeks prestasi akademik adalah hasil belajar yang dinyatakan dalam bentuk angka atau simbol. Indeks prestasi dapat digunakan sebagai tolak ukur prestasi belajar seseorang setelah menjalani proses belajar.

c. Angka kelulusan

Angka kelulusan merupakan suatu hasil yang diperoleh selama melaksanakan suatu pendidikan dalam institusi tertentu, dan hasil ini juga menjadi indikator penting prestasi belajar.

d. Predikat kelulusan

Predikat kelulusan merupakan status yang disandang oleh seseorang dalam menyelesaikan suatu pendidikan yang ditentukan oleh besarnya indeks prestasi yang dimiliki

e. Waktu tempuh pendidikan

Waktu tempuh pendidikan seseorang dalam menyelesaikan studinya menjadi salah satu ukuran prestasi, yang menyelesaikan studinya lebih awal menandakan prestasinya baik, sebaliknya waktu tempuh pendidikan yang melebihi waktu normal menandakan prestasi yang kurang baik.

### 2.1.3 Faktor-faktor yang Mempengaruhi Prestasi Akademik

Keberhasilan dalam proses belajar yang terjadi, dilatarbelakangi oleh adanya sumber atau penyebab yang mempengaruhi berlangsungnya proses belajar mengajar itu sendiri. Faktor tersebut dapat berupa penghambat maupun pendorong pencapaian prestasi. Soeryabrata (dalam Tjundjing, 2001) menggolongkan faktor-faktor yang mempengaruhi prestasi belajar menjadi dua faktor, yaitu:

## 1. Faktor Internal

Faktor ini merupakan hal-hal dalam diri individu yang mempengaruhi prestasi belajar yang dimiliki. Faktor ini dapat di golongkan ke dalam dua kelompok, yaitu:

### a) Faktor Fisiologis

Faktor fisiologis mengacu pada keadaan fisik, khususnya sistem penglihatan dan pendengaran, kedua sistem penginderaan tersebut dianggap sebagai factor yang paling bermanfaat di antara kelima indera yang dimiliki manusia. Untuk dapat menempuh pelajaran dengan baik seseorang perlu memperhatikan dan memelihara kesehatan tubuhnya. Keadaan fisik yang lemah merupakan suatu penghalang yang sangat besar bagi seseorang dalam menyelesaikan program studinya.

### a) Faktor Psikologis

Faktor psikologis meliputi faktor non fisik, seperti; motivasi, minat, intelegensi, perilaku dan sikap mental.

## 2. Faktor Eksternal

Selain faktor-faktor dalam diri inividu, masih ada hal-hal lain di luar diri yang dapat mempengaruhi prestasi yang diraih, yang di golongkan sebagai faktor eksternal, seperti lingkungan keluarga, sekolah dan masyarakat.

### a) Faktor lingkungan keluarga.

Faktor lingkungan keluarga dapat mempengaruhi prestasi siswa. Berikut ini dijelaskan faktor-faktor lingkungan keluarga tersebut:

- Sosial ekonomi keluarga
- Pendidikan orang tua
- Perhatian orang tua dan suasana hubungan antara anggota keluarga.

### b) Faktor lingkungan sekolah

- Sarana dan prasarana

- Kelengkapan fasilitas sekolah seperti OHP, kipas angin, pelantang (*microphone*) akan membantu kelancaran proses belajar mengajar di sekolah. Selain itu bentuk ruangan, sirkulasi udara dan lingkungan sekitar sekolah juga turut mempengaruhi proses belajar mengajar.
- Kompetensi guru dan siswa  
Kualitas guru dan siswa sangat penting dalam meraih prestasi. Kelengkapan sarana dan prasarana tanpa di sertai kinerja yang baik dari para penggunanya akan sia-sia belaka.
- Kurikulum dan metode mengajar.  
Hal ini meliputi materi dan bagaimana cara memberikan materi tersebut kepada siswa. Metode pengajaran yang lebih interaktif sangat di perlukan untuk menumbuhkan minat dan peran serta siswa dalam kegiatan pembelajaran.

c) Faktor lingkungan masyarakat

- Sosial budaya  
Pandangan masyarakat tentang pentingnya pendidikan akan mempengaruhi kesungguhan pendidik dan peserta didik. Masyarakat yang masih memandang rendah pendidikan akan enggan mengirim anaknya ke sekolah dan cenderung memandang rendah pekerjaan guru atau pengajar.
- Partisipasi terhadap pendidikan  
Bila semua pihak telah berpartisipasi dan mendukung kegiatan pendidikan, mulai dari pemerintah (berupa kebijakan dan anggaran) sampai pada masyarakat bawah (kesadaran akan pentingnya pendidikan), setiap orang akan lebih menghargai dan berusaha memajukan pendidikan dan ilmu pengetahuan. Hal ini akan memunculkan pendidik dan peserta didik yang lebih berkualitas.

Dari uraian tersebut dapat dikatakan bahwa faktor-faktor yang mempengaruhi prestasi belajar dapat bersifat individual dan kompleks. Bersifat individual

maksudnya adalah faktor penyebab tersebut pada setiap peserta didik selalu sama, bersifat kompleks maksudnya pengaruh tersebut merupakan interaksi dari beberapa faktor baik dari luar diri maupun dari dalam diri mahasiswa. Faktor-faktor tersebut secara langsung maupun tidak langsung saling berinteraksi.

## **2.2. Pengertian Data Mining**

Data mining merupakan suatu kegiatan yang meliputi pengumpulan, pemakaian dan historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dari data mining adalah pengelompokan clustering dimana data yang dikelompokkan belum mempunyai contoh kelompok. Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007).

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti database sistem, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, Image database, Signal processing (Han, 2006).

### **2.2.1 Tahap-tahap Data Mining**

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, memakai knowledge base. Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data diperoleh, baik dari

database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

## 2. Integrasi Data (data Integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

## 3. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

## 4. Transformasi Data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar

seperti analisis asosiasi dan clustering hanya bisa menerima input data katagorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

#### 5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

#### 6. Evaluasi Pola (Pattern evaluation)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

### **2.2.2. Teknik Data Mining**

Beberapa teknik data mining antara lain (Bala., et al, 2012) :

#### 1. Analisis asosiasi

Analisis asosiasi berupa penemuan aturan asosiasi yang menggambarkan kondisi atribut nilai yang sering terjadi bersamaan dalam sebuah satuan data tertentu. Analisis asosiasi secara luas digunakan untuk analisa data pasar dan transaksi

#### 2. Klasifikasi dan Prediksi

Klasifikasi adalah pemrosesan untuk menemukan sebuah model yang menjelaskan dan mincirikan konsep atau kelas data, untuk kepentingan tertentu, yang bisa menggunakan pemodelan untuk memprediksi kelas objek yang labelnya tidak diketahui. Model yang didapat mungkin diwakili dalam berbagai format

seperti aturan klasifikasi IF-THEN, pohon keputusan, formula matematika, atau jaringan syaraf tiruan pengklasifikasian bisa digunakan untuk memprediksi label kelas data objek data.

### 3. Analisis Clustering

Tidak seperti klasifikasi dan prediksi, yang menganalisa pelabelan objek data, clustering menganalisis objek data tanpa mengkonsultasikan label kelas yang dikenal. Secara umum label kelas bukan didapat dalam pengolahan data sederhana karena mereka tidak tahu bagaimana memulainya. Clustering dapat digunakan untuk me-generate label. Objek yang dicluster berdasarkan pada prinsip memaksimalkan persamaan dalam kelas dan meminimalkan kesamaan antar kelas. Sehingga cluster terhadap objek dibentuk sedemikian rupa sehingga objek dalam cluster mempunyai persamaan yang tinggi dalam perbandingan dengan objek lainnya, tapi sangat berlainan dengan objek dari cluster lain

### 4. Analisis Outlier

Sebuah database mungkin berisi objek data yang tidak sesuai dengan kebiasaan umumnya dari data yang disebut outlier. Analisa terhadap outlier mungkin membantu dalam pendeteksian kesalahan dan nilai-nilai abnormal.

#### **2.2.3. Clustering**

“Clustering atau analisis cluster adalah proses pengelompokan satu set benda – benda fisik atau abstrak ke dalam kelas objek yang sama” (Han and Kamber, 2006).

Baskoro (2010) menyatakan bahwa :

Clustering atau clusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan obyek–obyek ke dalam cluster–cluster. Cluster adalah sekelompok atau sekumpulan obyek–obyek data yang similiar satu sama lain dalam cluster yang sama dan disimiliar terhadap obyek–obyek yang berbeda cluster. Obyek akan dikelompokkan ke dalam satu atau lebih cluster sehingga obyek–obyek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Obyek–obyek dikelompokkan berdasarkan prinsip memaksimalkan

kesamaan obyek pada cluster yang sama dan memaksimalkan ketidaksamaan pada cluster yang berbeda. Kesamaan obyek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan obyek data, sedangkan obyek-obyek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi.

Dengan menggunakan clusterisasi, kita dapat mengidentifikasi daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut-atribut data. Dalam data mining usaha difokuskan pada metode-metode penemuan untuk cluster pada basis data berukuran besar secara selektif dan efisien. Beberapa kebutuhan clusterisasi dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda, mampu menangani dimensional yang tinggi, menangani data yang mempunyai noise, dan dapat diterjemahkan dengan mudah.

Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster.

## **2.3. Fuzzy Logic**

### **2.3.1 Pengertian Logika Fuzzy**

Logika fuzzy adalah salah satu cara yang tepat untuk memetakan suatu ruang input ke dalam suatu ruang output (Kusumadewi dan Purnomo,2004).sebagai contoh:

1. Manajer pergudangan mengatakan pada manajer produksi seberapa banyak persediaan barang pada akhir minggu ini.kemudian manajer produksi akan menetapkan jumlah barang yang harus di produksi esok hari
2. Pelayanan restoran memberikan pelayanan terhadap tamu, kemudian tamu Akan memberikan tip yang sesuai atas baik tidaknya pelayanan yang di berikan.
3. Anda mengatakan pada saya seberapa sejuk ruangan yang anda inginkan,saya akan mengatur putaran kipas yang ada pada ruangan ini.

### 2.3.2 Alasan Penggunaan Logika Fuzzy

Ada beberapa alasan orang menggunakan logika fuzzy, diantaranya :

- a. Konsep logika fuzzy mudah dimengerti. Konsep matematis yang mendasari penalaran fuzzy sangat sederhana dan mudah dimengerti.
- b. Logika fuzzy bersifat sangat fleksibel.
- c. Logika fuzzy juga memiliki toleransi terhadap data – data yang tidak tepat.
- d. Logika fuzzy mampu memodelkan fungsi – fungsi non–linear yang sangat kompleks.
- e. Logika fuzzy dapat membangun dan mengaplikasikan pengalaman - pengalaman para pakar secara langsung tanpa harus melalui proses pelatihan.
- f. Logika fuzzy dapat bekerjasama dengan teknik – teknik kendali secara konvensional.
- g. Logika fuzzy didasarkan pada bahasa alami fungsi keanggotaan adalah suatu kurva yang menunjukkan pemetaan titik – titik input ke dalam nilai keanggotaan. Fungsi keanggotaan sering juga disebut dengan derajat keanggotaan. Fungsi keanggotaan biasanya memiliki interval antara 0 sampai dengan 1. salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi.

### 2.4 Fuzzy Clustering Means

*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*. Metode *clustering* merupakan pengelompokan data beserta parameternya dalam kelompok – kelompok sesuai kecenderungan sifat dari masing-masing data tersebut (kesamaan sifat).

Ada beberapa algoritma *clustering* data, salah satu diantaranya adalah *Fuzzy C-Means*. *Fuzzy C-Means* adalah suatu teknik peng-*cluster*-an yang mana

keberadaannya tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

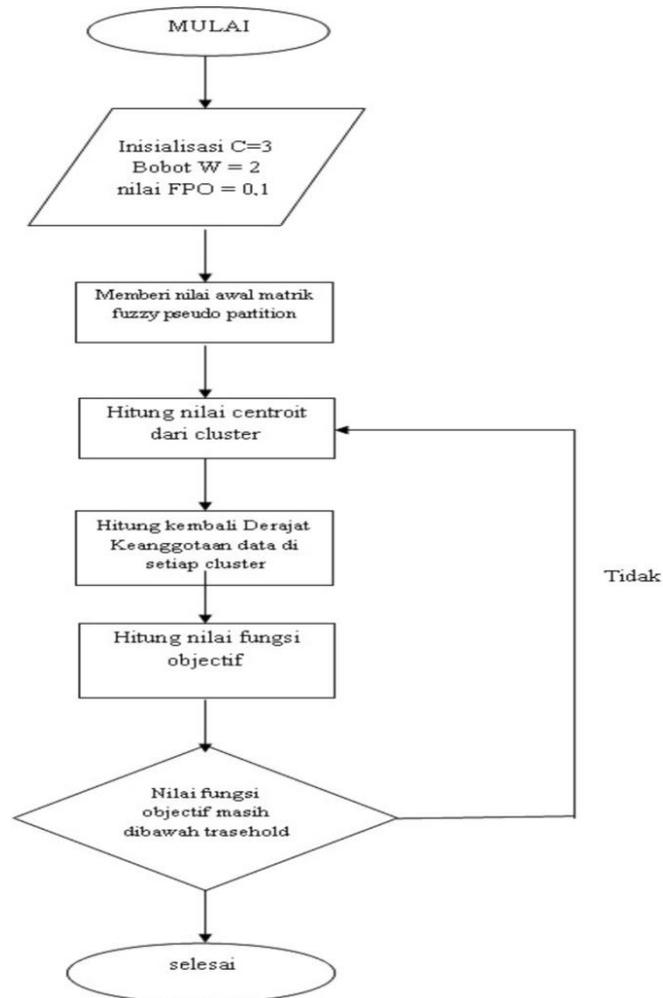
clustering dengan metode fuzzy C-means (FCM) didasarkan pada teori logika fuzzy, teorin ini pertama kali diperkenalkan oleh LotfiZadeh (1965) dengan nama himpunan fuzzy (fuzzy set). dalam teori fuzzy, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota) melainkan dengan suatu nilai derajat keanggotaanya yang jangkauan nilainya 0 sampai 1. nilai keanggotaan suatu dalam sebuah himpunan menjadi 0 ketika sama sekali tidak menjadi anggota, dan menjadi 1 ketika menjadi anggota secara penuh dalam himpunan. umumnya nilai keanggotaan antara 0 dan 1. semakin tinggi nilai keanggotaanya maka semakin tinggi derajat keanggotaanya, dan semakin kecil maka semakin rendah derajat keanggotaanya. kaitanya dengan K-means, Sebenarnya FCM merupakan versi fuzzy dari K-Means dengan beberapa modifikasi yang membedakanya dengan K-Means (Prasetyo, 2013).

Konsep dari *Fuzzy C-Means* 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 minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan kepusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut.

*Output* dari *Fuzzy C-Means* merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system*.

### 2.4.1 Algoritma *Fuzzy C-Means*

Diagram alir yang akan digunakan dalam penelitian ini secara umum dapat dilihat pada gambar 3.1



**Gambar 3.1** Flowchart Fuzzy clustering Means.

Pada gambar 3.1 digambarkan bahwa secara umum proses yang terjadi adalah :

1. Menginputkan data yang dicluster  $x$ , berupa matrik berukuran  $n \times m$  berfungsi untuk menentukan jumlah data dan atribut setiap data yang akan dipergunakan  
 $n$  = jumlah sampel data  
 $m$  = atribut setiap data

$X_{ij}$  = data sampel ke- $i$  ( $i=1,2,..n$ ) atribut ke- $j$  ( $j=1,2,..,m$ ).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \dots\dots\dots (2.1)$$

2. Menentukan :

- Jumlah cluster =  $c$ ;
- Bobot pangkat =  $w$ ;
- Maksimum Iterasi =  $\text{MaxIter}$ ;
- Error terkecil yang diharapkan =  $\epsilon$
- Fungsi Objektif awal =  $P_0 = 0$ ;
- Iterasi awal =  $t=1$ ;

Berfungsi untuk menentukan nilai awal dari persamaan, sebelum dilakukan pengolahan data.

3. Membangkitkan bilangan random  $u_{ik}$ ,  $i=1,2,..n; k = 1,2,..c$ ; berfungsi sebagai elemen-elemen matriks partisi awal  $U$ .

Menghitung jumlah tiap kolom (atribut) :

$$Q_j = \sum_{k=1}^c u_{ik} \dots\dots\dots (2.2)$$

Hitung

$$u_{ik} = \frac{u_{ik}}{Q_j} \dots\dots\dots (2.3)$$

4. Menghitung pusat cluster ke - $k$ :  $V_{kj}$ , dengan  $k=1,2,..c$ ; dan  $j=1,2,..,m$ ; penentuan pusat cluster digunakan untuk menandai lokasi rata-rata untuk tiap cluster dengan kondisi awal tidak akurat.

$$V_{kj} = \frac{\sum_{i=1}^n ((u_{ik})^w * X_{ij})}{\sum_{ij} (u_{ik})^w} \dots\dots\dots (2.4)$$

$$V = \begin{bmatrix} V_{11} & \cdots & V_{1m} \\ \vdots & \ddots & \vdots \\ V_{c1} & \cdots & V_{cm} \end{bmatrix} \dots\dots\dots (2.5)$$

5. Menghitung fungsi objektif pada iterasi ke =  $t$ ,  $P_t$  : perhitungan fungsi objektif digunakan untuk menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang berbobot oleh derajat keanggotaan titik data tersebut.

$$Pt = \sum_{i=1}^N \sum_{k=1}^c ([\sum_{j=1}^m (X_{ik} - V_{kj})^2] (\mu_{ik})^w) \dots\dots\dots (2.6)$$

6. Menghitung jarak data ke pusat cluster dengan menggunakan manhattan (Rectilinear) distance kemudian akan didapatkan matrik jarak sebagai berikut:

$$d_{ik} = d(x_k - v_i) = \sum_{j=1}^m |x_{kj} - v_{ij}| \dots\dots\dots (2.7)$$

7. Menghitung perubahan matriks partisi, penghitungan ini berfungsi sebagai nilai awal matriks jika mengalami perulangan dan agar lokasi cluster bisa berada pada posisi yang benar.

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ik} - V_{kj})^2]_{w-1}^{-1}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ik} - V_{kj})^2]_{w-1}^{-1}} \dots\dots\dots (2.8)$$

8. Cek Kondisi berhenti :

- Jika  $(|Pt - Pt_{-1}| < \epsilon)$  atau  $(t < \maxIter)$  maka berhenti;
- Jika tidak :  $t = t+1$ , ulangi langkah ke-4;

Langkah terakhir berfungsi sebagai pengkodisian perhitungan terhadap data, apakah suatu cluster yang telah dihasilkan, sudah memenuhi syarat atau perlu dilakukan iterasi selanjutnya agar lokasi cluster yang dihasilkan bisa berada pada posisi yang benar.

## 2.5 Jarak Manhattan

Jarak Rectilinear, sering disebut juga dengan jarak manhattan merupakan jarak yang diukur mengikuti jalur yang tegak lurus. disebut dengan jarak manhattan, mengingatkan jalan-jalan kot manhattan yang membentuk garis parallel dan saling tegak lurus antara satu jalan dengan jalan lainnya. pengukuran dengan jarak manhattan sering digunakan karena mudah perhitungannya, mudah dimengerti dan untuk beberapa masalah lebih sesuai, misalnya untuk menentukan jarak antar kota, jarak antar fasilitas dimana peralatan pemindahan bahan hanya dapat bergerak searah tegak lurus .

$$d_{ik} = d(x_k - v_i) = \sum_{j=1}^m |x_{kj} - v_{ij}|$$

$d(i,k)$  = *Manhattan distance* yaitu jarak antara data pada titik  $x$  dan titik  $y$  menggunakan kalkulasi matematika.

$i$  = Pusat Cluster

$k$  = Data

### 2.5.1 Kelebihan Manhattan Distance

Berikut ini adalah keunggulan menggunakan manhattan distance disbanding dengan algoritma pengukuran jarak lainnya adalah

1. Semua jalur – jalur dapat ditemukan (masalah dapat dipecahkan).
2. Hal ini disebabkan karena pada setiap penambahan nilai  $g(n)$ , pada perhitungan nilai heuristic-nya terjadi pula perubahan pada nilai  $d$  nya. Sehingga dengan penambahan nilai  $g(n)$ , tidak mempengaruhi pencarian jalur.
3. Dengan menggunakan fungsi heuristic manhattan distance, didapatkan nilai iterasi dan jumlah langkah yang paling kecil dibanding dengan menggunakan fungsi heuristic yang lain.

### 2.6 Indeks XB (Xie dan Beni)

Untuk menentukan banyak kelompok dapat dilakukan dengan menghitung Indeks XB (Xie dan Beni). Indeks ini ditemukan oleh Xie dan Beni dan pertama kali dikemukakan pada tahun 1991. Indeks XB dituliskan sebagai berikut :

$$XB = \frac{\sum_{i=1}^c \sum_{j=1}^m \mu_{ik}^w \|x_{kj} - v_{ij}\|^2}{n(\min \|x_{kj} - v_{ij}\|^2)}$$

dimana  $i$  = banyak kelompok,  $\mu_{ik}$  = derajat keanggotaan,  $j$  = jarak pengamatan dengan dan  $n$  = banyak objek yang akan dikelompokkan, serta kriteria banyak kelompok yang optimum ditunjukkan dengan nilai indeks XB yang minimum pada lembah pertama. Penggunaan indeks XB untuk menentukan jumlah kelompok yang optimum pada metode *fuzzy C-means* menyatakan bahwa indeks XB memiliki

ketepatan dan keandalan yang tinggi untuk memberikan jumlah kelompok yang optimum pada metode *fuzzy C-means*.

## 2.7 Penelitian sebelumnya

Sebagai bahan perbandingan dalam penelitian analisis prestasi akademik siswa SMK As'saadah Bunga Gresik tahun pertama sekolah, diantaranya dikemukakan oleh Wahyu Oktri Widyanto (2013) dengan judul "*Clustering Data Nilai Mahasiswa Untuk Pengelompokan Kosentrasi Jurusan Menggunakan Fuzzy C-Means*". Penelitian ini bertujuan untuk menganalisa data mahasiswa menurut bobot mata kuliah tertentu menggunakan konsep FCM. Pada penelitian ini menguji sebanyak 126 data. Hasil clustering ini menghasilkan sejumlah kelompok kosentrasi dengan jumlah masing-masing mahasiswa sesuai dengan cluster yang ada. Mahasiswa yang tercluster kedalam cluster 1 (Multimedia) sebanyak 28 mahasiswa, cluster 2 (WEB) sebanyak 70 mahasiswa dan cluster 3 (Pemogramanan) sebanyak 85 mahasiswa.

Penelitian kedua yaitu penelitian yang berjudul "*Sistem Inferensi Fuzzy untuk Memprediksi Prestasi Belajar Mahasiswa Berdasarkan Nilai Ujian Nasional, Tes Potensi Akademik, dan Motivasi Belajar*", dibuat oleh Hindayati Mustafidah dan Dwi Aryanto (Universitas Muhammadiyah Purwokerto, 2012). Tujuan dari penelitian tersebut adalah melakukan rancang bangun sistem yang bersifat interaktif yang bisa berfungsi sebagai sistem prediksi otomatis terhadap prestasi belajar mahasiswa dalam mengikuti perkuliahan di perguruan tinggi yang didasarkan atas nilai UN, nilai TPA, dan tingkat motivasi menggunakan *system inferensi fuzzy* metode *Mamdani*. Simpulan yang didapat dari penelitian tersebut adalah dengan menggunakan aplikasi logika fuzzy *system inferensi fuzzy* metode *Mamdani* dapat diprediksi prestasi belajar mahasiswa berdasarkan nilai TPA, NEM, dan tingkat motivasi belajar mahasiswa.

Penelitian ketiga yaitu "*Menentukan Nilai Akhir Kuliah Dengan Fuzzy C-Means*" Dibuat oleh Arwan Ahmad Khoiruddin, S.Kom. (*Universitas Islam Indonesia, 2007*). Dalam penelitian ini, akan diajukan alternatif yang bisa digunakan

untuk mempermudah perhitungan nilai akhir kuliah. Metode yang diajukan adalah dengan metode *clustering*, yaitu dengan menggunakan fuzzy c-means. Metode ini dipilih karena dengan metode ini, data-data beserta parameter-parameternya dapat dikelompokkan dalam *cluster-cluster* sesuai dengan kecenderungannya. Selain itu, metode FCM dipilih karena dengan metode ini, bisa ditentukan jumlah *cluster* yang akan dibentuk. Dengan penentuan jumlah *cluster* di awal, bisa diatur keragaman nilai akhir sesuai dengan *clusternya*. Dengan sistem berbasis FCM, bisa didapatkan nilai huruf yang labelnya A, B sampai E, atau juga bisa didapatkan yang labelnya A, A-, A/B, dan seterusnya.

Berdasarkan ketiga sumber penelitian tersebut di atas, tampak bahwa beberapa peneliti belum melakukan analisis secara spesifik terhadap suatu kasus tertentu dan melakukan proses pengelompokan terhadap mahasiswa yang mengikuti mata kuliah tersebut, sebagai upaya pengembangan akademik mahasiswa.