

BAB II

LANDASAN TEORI

2.1 Definisi sistem

Sistem secara fisik adalah kumpulan dari elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (Gordon, 1991). Jogianto (2005 :2) mengemukakan bahwa sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.2 Karakteristik sistem

Jogianto (2005 ;3) mengemukakan sistem mempunyai karakteristik atau sifat-sifat tertentu yakni :

1. Batasan sistem

Batasan sistem (bondary) merupakan daerah yang membatasi antara suatu sistem dengan sistem yang lainnya atau dengan lingkungan luarnya. Batasan suatu sistem menunjukkan ruang lingkup dari sistem tersebut.

2. Komponen

Suatu sistem dari sejumlah komponen yang saling berinteraksi yang artinya saling bekerja sama membentuk satu kesatuan komponen-komponen sistem atau elemen-elemen sistem dapat berupa suatu subsistem atau bagian-bagian dari sistem. Setiap subsistem mempunyai sifat-sifat dari sistem untuk menjalankan suatu fungsi tertentu mempengaruhi proses sistem secara keseluruhan.

3. Lingkungan luar sistem

Lingkungan luar (*evinronment*) dari suatu sistem adalah apapun diluar batas sistem yang mempengaruhi operasi. Lingkungan luar

sistem dapat bersifat menguntungkan dan dapat juga bersifat merugikan sistem tersebut.

4. Penghubung sistem

Penghubung (interface) merupakan media penghubung antara satu subsistem dengan subsistem yang lainnya. Melalui penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke subsistem yang lainnya. Dengan penghubung satu subsistem dapat berintegrasi dengan subsistem yang lainnya membentuk satu kesatuan.

2.3 Data Mining

2.3.1 Pengertian Data mining

Secara sederhana, data mining merupakan ekstraksi informasi yang tersirat dalam sekumpulan data. Data mining merupakan sebuah proses untuk menggali kumpulan data dan menemukan informasi di dalamnya. (Turban, dkk, 2005). Data mining merupakan pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan teknik gambar dari statistik, mesin pembelajaran dan sistem manajemen database. Penggalan data ini dilakukan pada sekumpulan data yang besar untuk menemukan pola atau hubungan yang ada dalam kumpulan data tersebut (Kusrini & Luthfi, 2009). Hasil penemuan yang diperoleh setelah proses penggalan data ini, kemudian dapat digunakan untuk analisis yang lebih lanjut.

Data mining yang disebut juga dengan Knowledge-Discovery in Database (KDD). Istilah *data mining* sering dipakai, mungkin karena istilah ini lebih pendek dari Knowledge Discovery in Database. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Data mining dianggap hanya sebagai suatu langkah terpenting dalam KDD. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Han dan Kamber, 2006)

1. Pembersihan data, untuk menghilangkan *noise* dan data yang tidak konsisten.
2. Integrasi data, dimana beberapa sumber data dapat dikombinasikan. Sebuah tren populer di industri informasi adalah untuk melakukan pembersihan dan integrasi data sebagai langkah preprocessing, dimana data yang dihasilkan akan disimpan dalam *data warehouse*.
3. Seleksi data, dimana data yang relevan dengan tugas analisis yang di ambil dari database.
4. Data transformasi (dimana data diubah dan digabung ke dalam bentuk yang sesuai untuk pertambangan dengan melakukan ringkasan atau agregasi operasi) terkadang transformasi data dilakukan sebelum proses seleksi data, khususnya dalam kasus *data warehouse*.
5. *Data mining* merupakan proses esensial dimana metode cerdas diaplikasikan untuk mengekstrak data pola.
6. Evaluasi pola, untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan.
7. Presentasi pengetahuan, dimana visualisaasi dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan hasil *data mining* kepada pengguna.

Langkah 1 sampai 4 merupakan berbagai bentuk preprocessing data, dimana data dipersiapkan untuk data mining. Hal ini menunjukkan bahwa data mining sebagai salah satu langkah dalam proses KDD, karena dapat mengungkap pola-pola tersembunyi yang digunakan untuk evaluasi.

2.3.2 Tugas Data Mining

Secara Umum, tugas data mining dapat diklasifikasikan kedalam 2 kategori (Han dan Kamber, 2006), yaitu :

a. Prediktif

Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variable tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai *explanatory* atau *variable* bebas.

b. Deskriptif

Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, teritori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas *data mining* deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

2.3.3 Fungsi Data Mining

Fungsi data mining dan macam-macam pola yang dapat ditemukan menurut Han Kamber (2006), yaitu :

1. *Concept / class Description : Characterization and Discrimination*

Data characterization adalah ringkasan dari semua karakteristik atau fitur dari data yang telah diperoleh dari target kelas. Data yang sesuai dengan kelas yang telah ditentukan oleh pengguna biasanya dikumpulkan di dalam *database*. Misalnya, untuk mempelajari karakteristik produk perangkat lunak dimana pada tahun lalu seluruh penjualan telah meningkat sebesar 10% data yang terkait dengan produk-produk tersebut dapat dikumpulkan dengan menjalankan sebuah query SQL.

Datadiscrimination adalah perbandingan antara fitur umum objek data target kelas dengan fitur umum objek dari satu atau satu set kelas lainnya. Target diambil melalui *query database*. Misalnya, pengguna

mungkin ingin membandingkan fitur umum dari produk perangkat lunak yang pada tahun lalu penjualannya meningkat sebesar 10% tetapi selama periode yang sama seluruh penjualan juga menurun setidaknya 30%.

2. *Mining Frequent Patterns, Associations, and Correlations*

Frequent Pattern adalah pola yang sering terjadi di dalam data. Ada banyak jenis dari *Frequent Patterns*, termasuk di dalamnya pola, sekelompok *item set*, *sub-sequence*, dan *sub-struktur*. Sebuah frequent patterns biasanya mengacu pada satu set item yang sering muncul bersama-sama dalam suatu kumpulan data transaksional, misalnya seperti susu dan roti. *Frequent patterns* sering mengarah pada penemuan asosiasi yang menarik dan korelasi data.

Associations Analysis adalah pencarian aturan-aturan asosiasi yang menunjukkan kondisi-kondisi nilai atribut yang sering terjadi bersama-sama dalam sekumpulan data. Analisis asosiasi sering digunakan untuk menganalisa *Market Basket Analysis* dan data transaksi.

3. *Classification and prediction*

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data (aitu objek data yang memiliki label kelas yang diketahui). Model yang diturunkan dapat diinterpretasikan dalam berbagai bentuk seperti *If-then* klasifikasi, decision tree, naïve bayes, dan sebagainya.

Teknik *classification* bekerja dengan mengelompokkan data berdasarkan *data training* dan nilai atribut klasifikasi data baru ke dalam kelompok yang ada.

Dalam banyak kasus, pengguna ingin memprediksi nilai-nilai data yang tidak tersedia atau hilang (bukan label dari kelas). Dalam kasus ini nilai data yang akan diprediksi merupakan data *numeric*. Disamping itu, prediksi lebih menekankan pada identifikasi *trend* dari distribusi berdasarkan data yang tersedia.

4. Cluster Analysis

Cluster adalah kumpulan objek data yang mirip satu sama lain dalam kelompok yang sama dan berbeda dengan objek data di kelompok lain. Sedangkan, *Clustering* atau Analisis *Cluster* adalah proses pengelompokkan satu set benda-benda fisik atau abstrak ke dalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan pengelompokkan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu *cluster* dan semakin besar perbedaan tiap *cluster* maka kualitas analisis *cluster* semakin baik.

Dari tugas-tugas data mining yang telah dijelaskan, perbandingan antara *Classification* dan *Clustering* menurut Han dan Kamber (2006) lebih spesifik seperti tabel 2.1

Tabel 2.1 Perbandingan *Classification* dan *Clustering*

Classification	Clustering
1. Menganalisis label kelas dari data objek.	Menganalisis data objek tanpa ada label kelas.
2. Label kelas ada atau terlihat jelas pada training data.	Label kelas tidak ada atau tidak terlihat pada training data.
3. Bertujuan untuk mengelompokkan pada kelas-kelas yang telah ditentukan	Bertujuan untuk mengelompokkan dan menentukan label kelas dari tiap cluster yang telah terbentuk.
4. Proses klasifikasi berdasarkan pada menemukan sebuah model atau fungsi yang menggambarkan dan membedakan data kelas atau konsep, dengan tujuan untuk dapat menggunakan model untuk memprediksi objek kelas yang	Proses clustering berdasarkan pada prinsip ; objek yang ada di dalam satu cluster memiliki kemiripan yang tinggi daripada yang lainnya, tetapi sangat berbeda dengan objek yang ada pada cluster lainnya.

<p>kelas label nya belum diketahui. Model tersebut berdasarkan pada analisis dari <i>training data</i> (data objek yang kelas label nya telah diketahui).</p>	
---	--

5. *Outlier analysis*

Outlier merupakan objek data yang tidak mengikuti perilaku umum dari data. *Outlier* dianggap sebagai noise atau pengecualian. Analisis data outlier dapat di anggap sebagai noise atau pengecualian. Analisis data outlier dinamakan *Outlier Mining*. Teknik ini berguna dalam fraud detection dan rare event analysis

6. *Evolution Analysis*

Analisis evolusi data menjelaskan dan memodelkan *trend* dari objek yang memiliki perilaku yang berubah setiap waktu. Teknik ini dapat meliputi karakterisasi, diskriminasi, asosiasi, klasifikasi, atau *clustering* dari data yang berkaitan dengan waktu.

Dari buku Data Mining Technique yang dikarang oleh Berry and Linoff proses terjadinya data mining dapat di deskripsikan sebagai virtuous cycle. Didasari oleh pengembangan berkelanjutan dari proses bisnis serta didorong oleh penemuan knowledge ditindak lanjuti dengan pengambilan tindakan dari penemuan.

2.3.4 Langkah-langkah Data mining

1. *Identity The Bussiness Problem*

Yang pertama dan juga dasar dari virtuous cyycle adalah mengetahui masalah bisnis ang kita hadapi. Karena kita tidak bisa mengolah data jika kita tidak tau yang sedang kita hadapi. Kita harus mengetahui masalah-masalah apa ang sedang dihadapi. Dengan mengetahui masalah ang dihadapi kita dapat menentukan data-data mana saja ang ita butuhkan untuk dapat dilakukan tahap analisa.

2. *Mine The Data For Actonable Information*

Setelah mengetahui identifikasi masalah, kita memperoleh data-data mana saja yang diperlukan untuk analisa. Barulah kita melakukan analisa terhadap data-data tersebut. Dan dari analisa tersebut analisis akan dapat memperoleh sebuah knowledge baru dan barulah dapat diambil suatu keputusan atau kebijaksanaan.

3. *Take The Action*

Dan dari keputusan / kebiaksanaan yang didapat dari proses data mining itu barulah kita terapkan dengan aksi berupa tindakan-tindakan yang kongkrit/nata dalam proses bisnis

4. *Measure Result*

Setelah diambil tindakan-tindakan dan keputusan kita memonitori hasil tersebut. Apakah sudah sesuai (memuaskan) dengan target yang ingin kita capai, apakah bisa mengatasi masalah-masalah yang dihadapi.

2.3.5 Jenis - jenis Data Mining

1. *Market Basket Analysis*

Himpunan data yang dijadikan sebagai objek penelitian pada area data mining. Market basket analysis adalah proses untuk menganalisis kebiasaan pelanggan dalam menyimpan item-item yang akan dibeli ke dalam keranjang belanjanya. Market basket analysis memanfaatkan data transaksi penjualan untuk dianalisis sehingga dapat ditemukan pola berupa item-item yang sering dibeli bersama ke dalam sebuah area yang berdekatan, merancang tampilan item-item di katalog, merancang kupon diskon (untuk diberikan kepada pelanggan yang membeli item tertentu), merancang penjualan item-item dalam bentuk paket, dan sebagainya. Dengan menggunakan teknologi data mining. Analisis data secara manual tidak diperlukan lagi.

2. *Memory-based Reasoning*

Metode klasifikasi yang digabungkan dengan penalaran berbasis memori proses menggunakan satu set data untuk membuat model dari prediksi atau asumsi-asumsi yang dapat dibuat tentang objek baru yang diperkenalkan. Ada dua komponen dasar untuk metode MBR. Yang pertama adalah kesamaan fungsi, yang mengukur bagaimana anggota yang sama dari setiap pasangan object satu sama lain. Yang kedua adalah fungsi kombinasi yang digunakan untuk menggabungkan hasil dari himpunan ptetangga untuk sampai pada keputusan.

3. *Cluster Detection*

Ada dua pendekatan untuk clustering. Pendekatan pertama adalah dengan mengasumsikan bahwa sejumlah cluster sudah tersimpan dalam data tujuannya adalah untuk memecah data ke dalam cluster. Pendekatan lain disebut clustering agglomerative, dengan asumsi keberadaan setiap jumlah yang telah ditetapkan cluster tertentu, setiap item keluar di cluster sendiri dan proses terjadi berulang-ulang yang berupaya untuk menggabungkan cluster, meskipun proses komputasi sama.

4. *Link Analysis*

Proses mencari dan membangun hubungan antara object dalam kumpulan data juga mencirikan sifat yang terkait dengan hubungan antara dua object. Link Analysis berguna untuk aplikasi analisis yang mengandalkan teori grafik untuk mengambil kesimpulan. Selain itu Link Analysis berguna untuk proses optimasi.

5. *Rule Induction*

Ekstraksi aturan sebab-akibat dari data secara static, identifikasi aturan bisnis yang tersimpan di dalam data. Metode berhubungan dengan induksi aturan yang digunakan untuk proses penemuan. Salah satu pendekatan untuk penemuan aturan adalah menggunakan pohon keputusan.

6. *Neural Networks*

Model prediksi non linier yang melakukan pembelajaran melalui latihan dan menyerupai struktur jaringan neural yang terdapat pada makhluk hidup. Mampu menurunkan pengertian dari data yang kompleks dan tidak jelas dan dapat digunakan pula untuk mengekstrak pola dan mendeteksi tren yang sangat kompleks untuk dibicarakan baik oleh manusia maupun teknik komputer lainnya.

2.3.6 Tugas Data Mining

Classification

- Proses untuk menempatkan beberapa nilai numerik secara terus suatu objek, estimasi juga dapat digunakan sebagai bagian dari proses klasifikasi.

Estimation

- Berbeda dengan Estimation dan Classification, prediction adalah upaya-upaya untuk mengklasifikasikan suatu objek berdasarkan dari behaviour yang akan ditentukan (diharapkan) dari candidate behavior.

Prediction

- Berbeda dengan Estimation dan Classification, prediction adalah upaya-upaya untuk mengklasifikasikan suatu objek berdasarkan dari behaviour yang akan ditentukan (diharapkan) dari candidate behavior.

Affinity Grouping

- Proses yang mengevaluasi hubungan atau asosiasi antara unsur-unsur data berupa atribut atau behaviour data yang menunjukkan beberapa tingkat afinitas antara objek.

Clustering

- Sama seperti klasifikasi tetapi kelompok yang tidak/belum ditentukan standarnya, sehingga secara algoritma data

tersebut akan dikelompokkan berdasarkan data yang serupa dengan data yang di submit

Description

- Proses yang menggambarkan apa yang telah terjadi dan diidentifikasi atau proses yang menjelaskan hasil akhir dari jalannya proses data mining.

2.4 Fuzzy C-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.

Ada beberapa algoritma *clustering* data, salah satu diantaranya adalah Fuzzy C-Means (FCM). Fuzzy C-Means (FCM) adalah suatu teknik pengclusteran 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 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 minimisasi fungsi objektif 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-Menas (FCM)

Algoritma Fuzzy C-Means adalah sebagai berikut:

1. Inisialisasi : Tentukan jumlah cluster ($k \geq 2$), tentukan bobot pangkat ($w > 1$), tentukan threshold perubahan nilai fungsi objektif.
2. Berikan nilai awal pada matrik fuzzy pseudo-partition, sesuai dengan persamaan (2.2).

$$\sum_{j=1}^k u_{ij} = 1 \quad \dots\dots\dots (2.2)$$

Keterangan :

- k = jumlah cluster
- j = menyatakan cluster
- u_{ij} = derajat keanggotaan pada setiap cluster

3. Lakukan langkah 4 sampai 7 selama perubahan pada nilai fungsi objektif masih di atas nilai threshold yang ditentukan.
4. Hitung nilai centroid dari masing-masing cluster menggunakan persamaan (2.3).

$$c_{lj} = \frac{\sum_{i=1}^N (u_{il})^w x_{ij}}{\sum_{i=1}^N (u_{il})^w} \quad \dots\dots\dots (2.3)$$

Keterangan :

- c_l = centroid pada cluster
- l = fitur
- N = jumlah data
- w = bobot pangkat
- u_{il} = nilai derajat keanggotaan
- x_i = data ke- i

5. Hitung nilai jarak centroid menggunakan persamaan eucliden.

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^k |x_{2j} - x_{1j}|^2} \quad \dots\dots\dots (2.4)$$

Keterangan :

- D = jarak antara data dan centroid
- x_1 = data ke- 1
- x_2 = data ke- 2

6. Hitung kembali matrik fuzzy pseudo-partition (derajat keanggotaan setiap data pada cluster) menggunakan persamaan (2.5).

$$u_{ij} = \frac{D(x_i, c_j)^{\frac{-2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{\frac{-2}{w-1}}} \dots\dots\dots (2.5)$$

7. Hitung nilai fungsi objektif menggunakan persamaan (2.6).

$$J = \sum_{i=1}^N \sum_{l=1}^k (u_{ij})^w D(x_i, c_l)^2 \dots\dots\dots (2.6)$$

2.4.2 Rumus Normalisasi Data

1. Normalisasi data : menggunakan rumus non linear, sesuai dengan persamaan (2.1)

$$\sigma_k^2 = \frac{1}{N-1} \sum_{i=1}^N (x_{ik} - \bar{x}_k)^2$$

$$y = \frac{x_{ik} - \bar{x}_k}{r\sigma_k}$$

$$\hat{x}_{ik} = \frac{1}{1 + e^{-y}} \dots\dots\dots (2.1)$$

Keterangan

- y = perkiraan fungsi linear
- r = konstanta yang ditentukan sendiri
- σ_k = standar deviasi fitur k
- e = konstanta = 2,728281828459

2.4.3 Contoh Perhitungan Kasus

Terdapat 10 data dengan 3 fitur yaitu fitur harta, utang dan gaji

Tabel 2.2 Tabel Data set

No	Harta	Utang	Gaji
1	3.2321	2.5000	2.0000
2	4.7321	2.5000	1.6000
3	7.7321	3.0000	4.0000

No	Harta	Utang	Gaji
4	9.2321	3.0000	2.3000
5	3.2321	3.5000	1.0000
6	4.7321	4.5000	1.4000
7	8.7321	4.5000	5.4000
8	2.2321	6.5000	2.5000
9	2.7321	8.5000	1.1000
10	3.2321	8.5000	3.0000

Tabel 2.2 Lanjutan Tabel Data Set

1. Proses Inisialisasi

Pada tabel kasus diatas di tentukan jumlah cluster sebanyak 3 cluster, Bobot pangkat (w) sebesar 2 dan nilai threshold adalah 0,1

2. Proses Pembangkitan Pseudo Awal

Proses ini adalah membangkitkan nilai matrik secara acak.

Tabel 2.3 Tabel Matrix pseudo awal

Data ke-i	U1	U2	U3
1	0.1000	0.3000	0.6000
2	0.7000	0.2000	0.1000
3	0.8000	0.0500	0.1500
4	0.4000	0.5000	0.1000
5	0.5500	0.1000	0.3500
6	0.2500	0.2500	0.5000
7	0.4000	0.1500	0.4500
8	0.5000	0.2000	0.3000
9	0.3000	0.1500	0.5500
10	0.4000	0.2000	0.4000

Dari **tabel 2.3** jumlah dari baris u_1 , u_2 dan u_3 apabila di jumlahkan berjumlah 1.

3. Proses Perhitungan Centroid dari masing – masing Cluster

Berikut ini adalah contoh cara perhitungan nilai centroid dari masing – masing cluster. Menggunakan persamaan (2.4).

- Perhitungan $(u_1)^w$

Nilai pseudo awal (u_1) di pangkat w .

Contoh : $0.1000^2 = 0.0100$

- Perhitungan $(u_1)^w \cdot x_{i1}$

Nilai perhitungan u_1 dikalikan dengan fitur 1 data ke-1

Contoh : $0.0100 * 3.2321 = 0.0323$

- Perhitungan $(u_1)^w \cdot x_{i2}$

Nilai perhitungan u_1 dikalikan dengan fitur 2 data ke-1

Contoh : $0.0100 * 2.5000 = 0.0250$

- Perhitungan $(u_1)^w \cdot x_{i3}$

Nilai perhitungan u_1 dikalikan dengan fitur 3 data ke-1

Contoh : $0.0100 * 2.0000 = 0.0200$

Tabel 2.4 Tabel Iterasi 1, cluster 1.

Data ke-i	$(u_1)^w$	$(u_1)^w \cdot x_{i1}$	$(u_1)^w \cdot x_{i2}$	$(u_1)^w \cdot x_{i3}$
1	0.0100	0.0323	0.0250	0.0200
2	0.4900	2.3187	1.2250	0.7840
3	0.6400	4.9485	1.9200	2.5600
4	0.1600	1.4771	0.4800	0.3680
5	0.3025	0.9777	1.0588	0.3025
6	0.0625	0.2958	0.2813	0.0875
7	0.1600	1.3971	0.7200	0.8640
8	0.2500	0.5580	1.6250	0.6250
9	0.0900	0.2459	0.7650	0.0990

10	0.1600	0.5171	1.3600	0.4800
Jumlah	2.3250	12.7684	9.4600	6.1900

Tabel 2.5 Tabel Iterasi 1, Cluster 2

Data ke-	$(ui_2)^w$	$(ui_2)^w xi_1$	$(ui_2)^w xi_2$	$(ui_2)^w xi_3$
1	0.0900	0.2909	0.2250	0.1800
2	0.0400	0.1893	0.1000	0.0640
3	0.0025	0.0193	0.0075	0.0100
4	0.2500	2.3080	0.7500	0.5750
5	0.0100	0.0323	0.0350	0.0100
6	0.0625	0.2958	0.2813	0.0875
7	0.0225	0.1965	0.1013	0.1215
8	0.0400	0.0893	0.2600	0.1000
9	0.0225	0.0615	0.1913	0.0248
10	0.0400	0.1293	0.3400	0.1200
Jumlah	0.5800	3.6121	2.2913	1.2928

Tabel 2.6 Tabel Iterasi 1, Cluster 3

Data ke-	$(ui_3)^w$	$(ui_3)^w xi_1$	$(ui_3)^w xi_2$	$(ui_3)^w xi_3$
1	0.3600	1.1636	0.9000	0.7200
2	0.0100	0.0473	0.0250	0.0160
3	0.0225	0.1740	0.0675	0.0900
4	0.0100	0.0923	0.0300	0.0230
5	0.1225	0.3959	0.4288	0.1225
6	0.2500	1.1830	1.1250	0.3500
7	0.2025	1.7683	0.9113	1.0935
8	0.0900	0.2009	0.5850	0.2250
9	0.3025	0.8265	2.5713	0.3328
10	0.1600	0.5171	1.3600	0.4800
Jumlah	1.5300	6.3689	8.0038	3.4528

Setelah menghitung centroid tiap cluster pada iterasi 1, maka didapatkan nilai centroid sebagai berikut :

- Perhitungan fitur x dicentroid 1

Di dapat dari perhitungan cluster 1

$$c_{1j} = \frac{\sum_{i=1}^N (u_{i1})^w x_{i1}}{\sum_{i=1}^N (u_{i1})^w} = \frac{12.7684}{2.3250} = 5.4918$$

- Perhitungan fitur y dicentroid 1

$$c_{1j} = \frac{\sum_{i=1}^N (u_{i1})^w x_{i2}}{\sum_{i=1}^N (u_{i1})^w} = \frac{9.4600}{2.3250} = 4.0688$$

Tabel 2.7 Tabel Centroid 1

centroid 1		
Fitur x	Fitur y	Fitur z
5.4918	4.0688	2.6624

Tabel 2.8 Tabel Centroid 2

centroid 2		
Fitur x	Fitur y	Fitur z
6.2278	3.9504	2.2289

Tabel 2.9 Tabel Centroid 3

centroid 3		
Fitur x	Fitur y	Fitur z
4.1627	5.2312	2.2567

4. Proses Menghitung Jarak Antara data dan Centroid

Menghitung Jarak antara data dan centroid dengan rumus ecludien sesuai dengan persamaan (2.4). Berikut ini adalah contoh perhitungan jarak centroid.

- Perhitungan jarak antara data dan centroid

$$\begin{aligned}
D(x_1, c_1) &= \sqrt{(x_{11} - c_{11})^2 + (x_{12} - c_{12})^2 + (x_{13} - c_{13})^2} \\
&= \sqrt{(3.2321 - 5.4918)^2 + (2.5000 - 4.0688)^2 + (2.0000 - 2.6624)^2} \\
&= \sqrt{5.1062 + 2.4611 + 0.4388} = 2.8295
\end{aligned}$$

Tabel 2.10 Tabel Perhitungan Jarak centroid

Data ke-	centroid 1	centroid 2	centroid 3
1	2.8295	3.3362	2.8968
2	2.0413	2.1763	2.8662
3	2.8197	2.5106	4.5561
4	3.9069	3.1519	5.5389
5	2.8624	3.2691	2.3329
6	1.5351	1.7961	1.2621
7	4.2638	4.0779	5.5942
8	4.0697	4.7476	2.3229
9	5.4491	5.8475	3.7509
10	4.9855	5.5016	3.4790

5. Proses Menghitung Nilai Keanggotaan Pseudo Baru

Menghitung nilai keanggotaan pseudo baru sesuai dengan persamaan (2.5). Berikut ini adalah contoh perhitungan nilai keanggotaan pseudo baru.

- Perhitungan u_{11} anggota pseudo baru

$$\begin{aligned}
u_{11} &= \frac{D(x_1, c_1)^{\frac{-2}{w-1}}}{D(x_1, c_1)^{\frac{-2}{w-1}} + D(x_1, c_2)^{\frac{-2}{w-1}} + D(x_1, c_3)^{\frac{-2}{w-1}}} \\
&= \frac{2.8295^{\frac{-2}{2-1}}}{2.8295^{\frac{-2}{2-1}} + 3.3362^{\frac{-2}{2-1}} + 2.8968^{\frac{-2}{2-1}}} = 0.3741
\end{aligned}$$

- Perhitungan u_{12} anggota pseudo baru

$$u_{12} = \frac{D(x_1, c_2)^{\frac{-2}{w-1}}}{D(x_1, c_1)^{\frac{-2}{w-1}} + D(x_1, c_2)^{\frac{-2}{w-1}} + D(x_1, c_3)^{\frac{-2}{w-1}}}$$

$$= \frac{3.3362^{\frac{-2}{2-1}}}{2.8295^{\frac{-2}{2-1}} + 3.3362^{\frac{-2}{2-1}} + 2.8968^{\frac{-2}{2-1}}} = 0.2691$$

- Perhitungan u_{13} anggota pseudo baru

$$u_{13} = \frac{D(x_1, c_3)^{\frac{-2}{w-1}}}{D(x_1, c_1)^{\frac{-2}{w-1}} + D(x_1, c_2)^{\frac{-2}{w-1}} + D(x_1, c_3)^{\frac{-2}{w-1}}}$$

$$= \frac{2.8968^{\frac{-2}{2-1}}}{2.8295^{\frac{-2}{2-1}} + 3.3362^{\frac{-2}{2-1}} + 2.8968^{\frac{-2}{2-1}}} = 0.3569$$

Tabel 2.11 Tabel Nilai Keanggotaan Pseudo Baru

Data ke-	u_{11}	u_{12}	u_{13}
1	0.3741	0.2691	0.3569
2	0.4189	0.3686	0.2125
3	0.3782	0.4770	0.1448
4	0.3296	0.5064	0.1640
5	0.3056	0.2343	0.4601
6	0.3115	0.2276	0.4609
7	0.3739	0.4088	0.2172
8	0.2082	0.1530	0.6389
9	0.2513	0.2183	0.5304
10	0.2581	0.2119	0.5300

6. Proses Menghitung Fungsi Objektif

Menghitung Fungsi Objektif sesuai dengan persamaan (2.6).

Berikut ini adalah contoh perhitungan fungsi objektif.

- Perhitungan fungsi objektif cluster 1

$$J = 2.8295^2 * 0.3741^2 = 1.1202$$

- Perhitungan fungsi objektif cluster 2

$$J = 3.3362^2 * 0.2691^2 = 0.8058$$

Tabel 2.12 Tabel Fungsi Objektif

data ke-	cluster 1	cluster 2	cluster 3
1	1.1202	0.8058	1.0688
2	0.7313	0.6434	0.3709
3	1.1370	1.4342	0.4355
4	1.6582	2.5477	0.8250
5	0.7652	0.5867	1.1520
6	0.2287	0.1671	0.3384
7	2.5423	2.7793	1.4769
8	0.7176	0.5273	2.2026
9	1.8756	1.6287	3.9583
10	1.6555	1.3595	3.3997

Kemudian di lakukan perhitungan nilai fungsi objektive (j), dengan cara menjumlah seluruh data cluster 1 + cluster 2 + cluster 3. Pada contoh kasus diatas didapatkan Nilai Fungsi Objektive (j) sebesar 40.1393.

Kemudian dilakukan perhitungan nilai perubahan Fungsi Objektive dengan cara mengurangi nilai fungsi objektive awal dengan nilai fungsi objektif (j). Pada contoh kasus diatas didapatkan perubahan fungsi objektivenya adalah 959.8607.

2.5 Penelitian Terkait

Penelitian sebelumnya dilakukan oleh Dimas Anggoro. Judul “PENGELOMPOKAN KETERAMPILAN DATABASE MAHASISWA TEKNIK INFORMATIKA UMG MENGGUNAKAN FUZZY C-MENAS”. Pada penelitian ini dibangun dengan tujuan membuat suatu sistem yang berbasis data mining dengan metode Fuzzy C-means untuk mengelompokkan mahasiswa berdasarkan keterampilan database. Variabel yang digunakan sebanyak 9 variabel yang terdiri atas X1 (mahasiswa), X2

(NIM), X3 (tempat lahir), X4 (tanggal lahir), X5 (Alamat Mahasiswa), X6 (Instansi sekolah), X7 (Sekolah asal), X8 (jurusan), X9 (IPK), Data bobot nilai dari masing-masing variabel di atas kemudian diolah menggunakan Fuzzy C-Means untuk menentukan data clustering yang nantinya akan digunakan untuk penentuan konsentrasi jurusan. Dengan adanya sistem fuzzy c-means ini, dapat diharapkan mampu menjadi salah satu acuan untuk menentukan konsentrasi jurusan mahasiswa.

Penelitian kedua dilakukan oleh Syahriyatur Rahmawati. Judul "PENERAPAN METODE FUZZY C-MEANS UNTUK PENGELOMPOKAN KELUARGA PENERIMA BERAS MISKIN DI KELURAHAN PEKAUMAN KECAMATAN GRESIK". Pada penelitian ini, dibangun sebuah sistem fuzzy yang bertujuan untuk memberikan pelabelan pada data pengelompokan keluarga dengan kategori prasejahtera, sejahtera 1, sejahtera tahap II, sejahtera tahap III, sejahtera tahap III plus dan yang kedua bertujuan untuk menganalisa penerimaan beras miskin sesuai dengan hasil pengelompokan. Variabel yang digunakan sebanyak 5 cluster yang terdiri atas kategori prasejahtera, sejahtera 1, sejahtera tahap II, sejahtera tahap III, sejahtera tahap III plus. Data bobot nilai dari masing-masing variabel di atas kemudian diolah menggunakan Fuzzy C-Means untuk menentukan data clustering yang nantinya akan digunakan untuk mengklasifikasikan keluarga penerima beras miskin di Kelurahan Pekauman kecamatan Gresik ke dalam cluster-cluster (prasejahtera, sejahtera 1, sejahtera tahap II, sejahtera tahap III, sejahtera tahap III plus).

Penelitian ketiga dilakukan oleh Arwan Ahmad Khoiruddin. Dengan judul "MENENTUKAN NILAI AKHIR KULIAH DENGAN FUZZY C-MEANS". 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 cluster-nya. 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.

Dalam penelitian ini digunakan sebanyak 5 variabel penunjang yaitu tugas kuliah, Nilai kuis, Persentase kehadiran, Nilai ujian tengah semester dan Nilai ujian akhir semester. Masing-masing parameter itu diberi bobot sendiri. Nilai akhir kuliah akan dihitung dari perhitungan parameter-parameter tersebut dikalikan dengan bobotnya.