

## BAB II

### LANDASAN TEORI

#### 2.1 Pengertian Sistem

Sistem adalah suatu jaringan kerja prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran tertentu (Jogiyanto, 1999).

Dari pengertian diatas tadi penulis menyimpulkan bahwa pengertian sistem adalah suatu kumpulan komponen fisik maupun non-fisik yang membentuk suatu jaringan kerja untuk melakukan suatu kegiatan guna mencapai sasaran tertentu.

#### 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 salah satu teknik penelusuran data untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat menggali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga memanfaatkan teknik ini. Dalam *data mining*, pengelompokan data juga dilakukan. Tujuannya adalah agar penulis dapat mengetahui pola dan tindak lanjut yang diambil. Semua hal tersebut bertujuan untuk mendukung kegiatan evaluasi agar sesuai dengan yang diharapkan (Prasetyo, 2012).

*Data mining* merupakan salah satu cara untuk menemukan informasi yang terkandung pada suatu data (*knowledge discovery*). Teknik *Data mining*

dikembangkan untuk mencari data potensi serapan mahasiswa dan pola yang mungkin dapat digunakan pada database yang berskala besar (Pan-Ning Tan, 2006).

*Data mining* merupakan salah satu tahap pada proses *Knowledge Discovery in Database* (KDD). KDD adalah penyulingan informasi menarik yang tidak biasa, yang terkandung dalam basis data berukuran besar, yang sebelumnya tidak diketahui potensi dan manfaatnya (Han & Kamber 2001).

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).

Dari beberapa pernyataan tersebut dapat diambil garis besar bahwa *Data Mining* merupakan proses ekstraksi dari database yang berukuran besar untuk di jadikan sebuah informasi baru menggunakan metode yang sesuai, sehingga dapat memberikan pengetahuan baru dari data base yang sudah diolah.

### **2.3 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**

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

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 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 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

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. (Sunjana,2010)

## 2.4 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.

### 2.5 Clustering Data

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

*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 *similar* satu sama lain dalam *cluster* yang sama dan *dissimilar* 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 lainnya.

Dengan menggunakan *clusterisasi*, kita dapat mengidentifikasi ketrampilan mahasiswa yang mengambil mata kuliah pemrograman WEB berdasarkan potensi yang dimiliki mahasiswa dan mengelompokkannya berdasarkan potensi yang dimiliki mahasiswa tersebut.

## 2.6 Pengertian Metode SOM

SOM (*Self Organizing Feature Map*) pertama kali diperkenalkan oleh Kohonen (Kohonen,1989) dengan teknik pelatihan ANN yang menggunakan *basis winner takes all*. Di mana hanya neuron yang menjadi pemenang yang akan diperarui bobotnya. Meskipun menggunakan basis ANN, SOM tidak menggunakan target kelas, tidak ada kelas yang ditetapkan untuk setiap data. Karakteristik seperti inilah yang kemudian membuat SOM dapat digunakan untuk keperluan pengelompokan berbasis ANN (Prasetyo, 2012).

## 2.7 Algoritma SOM

Berikut ini adalah langkah-langkah yang perlu dilakukan dalam menerapkan metode SOM dalam pengolahan data (Prasetyo, 2012):

1. Inisialisasi Nilai bobot  $W_{ij}$  secara acak, tentukan parameter topologi ketetanggaan, tentukan parameter laju pembelajaran, tentukan jumlah iterasi pelatihan.
2. Selama jumlah maksimal iterasi belum tercapai, lakukan langkah 3 -7.
3. Untuk setiap data masukan  $X$  (matriks  $M \times N$ ), lakukan langkah 4 – 6.
4. Untuk setiap neuron  $j$ , hitung  $D_j = \sum_i (w_{ij} - x_i)^2$  ,  $i=1, \dots, N$  .....(2.1)  
 $N$  adalah dimensi data ( $N$ ).
5. Cari indeks dari sejumlah neuron, yaitu  $D_j$ , yang mempunyai nilai terkecil.
6. Untuk neuron  $j$  dan semua neuron yang menjadi tetangga  $J$  (yang sudah didefinisikan) dalam radius  $R$ , hitunglah perubahan bobot  
 $W_{ij}(\text{baru}) = W_{ij}(\text{lama}) + \eta (X_i - W_{ij}(\text{lama}))$ . .....(2.2)
7. Perbaharui laju pembelajaran dengan rumus :  
 $\eta(\text{baru}) = \text{fungsi\_pembelajaran} \times \eta(\text{lama})$  .....(2.3)

Pada algoritma tersebut dapat dijelaskan, parameter jarak untuk perbedaan atau kemiripan yang digunakan adalah Encludian kuadrat (square Euclidean). Hal ini dimaksudkan untuk mengurangi waktu komputasi dengan menyederhanakan kinerja algoritma, tetapi harus dibayar dengan penggunaan memori yang lebih besar untuk alokasi nilai jarak yang biasanya lebih besar. Nilai laju pembelajaran ( $\eta$ ) yang digunakan adalah jangkauan nilai 0 sampai 1. Tetapi, nilai ini akan terus diturunkan setiap kali ada kenaikan iterasi dengan sebuah fungsi pembelajaran.

Inisialisasi bobot awal bisa menggunakan nilai acak dengan jangkauan -0.5 sampai +0.5, atau menggunakan nilai acak dengan jangkauan nilai seperti pada data masukan.

Parameter yang bisa digunakan dalam pengelompokan set data  $X$  pada sejumlah  $C$ . untuk penjelasanya dapa dilihat pada tabel 2.1 :

**Tabel 2.1** : Parameter dan keterangan algoritma SOM.

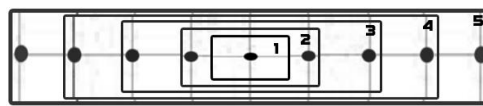
No	Parameter	Keterangan
1.	$X$	$X$ adalah matriks $M \times N$ . $M$ menyatakan jumlah data, dan $N$ menyatakan jumlah fitur.
2.	$C$	$C$ menyatakan jumlah kelompok (neuron pemroses)
3.	Iterasi	Jumlah maksimal iterasi pelatihan
4.	$W_{ij}$	Matriks $N \times C$ yang menyatakan bobot untuk setiap neuron (kelompok). $N$ menyatakan jumlah fitur, dan $C$ menyatakan jumlah neuron (kelompok).
5.	$D_j$	Matriks $M \times C$ . Baris menyatakan data, dan kolom menyatakan jarak ke neuron (kelompok).
6.	$J$	$J$ menyatakan neuron keluaran yang didapat dari data yang berdimensi tinggi. Neuron keluaran tersebut berfungsi jika dalam penelitian menggunakan topologi SOM.

## 2.8 Topologi Jaringan Neural Network Pada Metode SOM

Beberapa topologi SOM Kohonen yang digunakan pada data satu dimensi dan data dua dimensi. Topologi yang bisa digunakan yaitu Linear neighborhood, Rectangular neighborhood dan Hexagonal neighborhood.

### 1. Linear Neighborhood

Arsitektur SOM yang digambarkan secara topografis untuk memberikan visualisasi pengelompokan pada data satu dimensi dapat dilihat pada gambar 2.1:

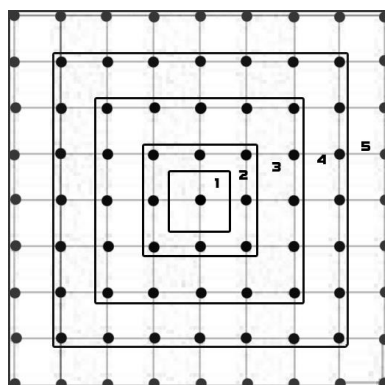


**Gambar 2.1** Linear Neighborhood

Untuk data dalam kotak bertuliskan angka 1, tidak ada tetangga yang berhubungan dengan setiap neuron yang diperbarui bobotnya. Untuk data dalam kotak bertuliskan angka 2 ada 2 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, Untuk data dalam kotak bertuliskan angka 3 4 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, begitu seterusnya.

### 2. Rectangular Neighborhood

Arsitektur SOM yang digambarkan secara topografis untuk memberikan visualisasi pengelompokan pada data dua dimensi berbentuk rectangular dapat dilihat pada gambar 2.2:



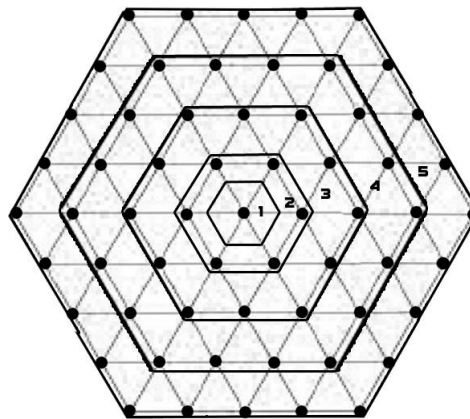
**Gambar 2.2** Rectangular Neighborhood



Untuk data dalam kotak bertuliskan angka 1, ketetanggaan didefinisikan dengan grid segi empat. Untuk data dalam kotak bertuliskan angka 1, tidak ada tetangga yang berhubungan dengan setiap neuron yang diperbarui bobotnya. Untuk data dalam kotak bertuliskan angka 2 ada 8 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, Untuk data dalam kotak bertuliskan angka 3 ada 24 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, begitu seterusnya.

### 3. Hexagonal Neighborhood

Arsitektur SOM yang digambarkan secara topografis untuk memberikan visualisasi pengelompokan pada data dua dimensi berbentuk hexagonal dapat dilihat pada gambar 2.3:



**Gambar 2.3** Hexagonal Neighborhood

Untuk data dalam grid bertuliskan angka 1, ketetanggaan didefinisikan dengan grid segi empat. Untuk data dalam grid bertuliskan angka 1, tidak ada tetangga yang berhubungan dengan setiap neuron yang diperbarui bobotnya. Untuk data dalam grid bertuliskan angka 2 ada 6 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, Untuk data dalam grid bertuliskan angka 3 ada 18 tetangga yang berhubungan dengan setiap neuron untuk diperharui bobotnya, begitu seterusnya.

## 2.9 Contoh Aplikasi SOM

Pada contoh ini akan dilakukan pengelompokan lima buah data dengan jumlah dimensi dua. Set data yang akan digunakan adalah data pada tabel 2.2 :

**Tabel 2.2 :** Set data pada contoh metode SOM.

Data	X	Y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

Fitur yang digunakan adalah X dan Y. Pada contoh ini akan dilakukan proses pengelompokan dengan metode SOM. Parameter yang digunakan jumlah kelompok ada 3, Jumlah iterasi pelatihan 5, laju pembelajaran 0,5 dan akan menjadi 0,6 (fungsi pembelajaran) darinya pada iterasi berikutnya.

Langkah pertama adalah inisialisasi bobot. Karna jumlah fitur ada 2 dan jumlah kelompok ada 3. Matriks bobot  $w$  berukuran  $2 \times 3$ . Secara acak penulis menggunakan bobot awal berikut :

$$w = \begin{matrix} & D_1 & D_2 & D_3 \\ \begin{bmatrix} 2 & 2 & 2 \\ 2 & 3 & 5 \end{bmatrix} \end{matrix}$$

### Iterasi 1 :

Untuk data 1 [1 1], hitung jarak ke setiap neuron:

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2-1)^2 + (2-1)^2 = 1^2 + 1^2 = 1+1 = 2 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (2-1)^2 + (3-1)^2 = 1^2 + 2^2 = 1+4 = 5 \end{aligned}$$

$$\begin{aligned}
 D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\
 &= (2-1)^2 + (5-1)^2 = 1^2 + 4^2 = 1+16 = 17
 \end{aligned}$$

Jarak terdekat adalah neuron ke-1 sehingga bobot neuron 1 akan diperbaharui.

$$\begin{aligned}
 w_{i1}(\text{baru}) &= w_{i1}(\text{lama}) + \eta(x_i - w_{i1}(\text{lama})) \\
 &= \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}
 \end{aligned}$$

Bobot baru menjadi :  $w = \begin{bmatrix} 1.5 & 2 & 2 \\ 1.5 & 3 & 5 \end{bmatrix}$

Untuk data 2 [4 1], hitung jarak ke setiap neuron

$$\begin{aligned}
 D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\
 &= (1.5-4)^2 + (1.5-1)^2 = (-2.5)^2 + 0.5^2 = 6.25 + 0.25 = 6.5
 \end{aligned}$$

$$\begin{aligned}
 D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\
 &= (2-4)^2 + (3-1)^2 = (-2)^2 + 2^2 = 4 + 4 = 8
 \end{aligned}$$

$$\begin{aligned}
 D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\
 &= (2-4)^2 + (5-1)^2 = (-2)^2 + 4^2 = 4 + 16 = 20
 \end{aligned}$$

Jarak terdekat adalah neuron ke-1 sehingga bobot neuron 1 akan diperbaharui

$$\begin{aligned}
 w_{i1}(\text{baru}) &= w_{i1}(\text{lama}) + \eta(x_i - w_{i1}(\text{lama})) = \\
 &= \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} \right) = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + 0.5 \begin{bmatrix} 2.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 1.25 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 2.75 \\ 1.25 \end{bmatrix}
 \end{aligned}$$

Bobot baru menjadi :  $w = \begin{bmatrix} 2.75 & 2 & 2 \\ 1.25 & 3 & 5 \end{bmatrix}$

Untuk data 3 [1 2], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2.75 - 1)^2 + (1.25 - 2)^2 = 1.75^2 + (-0.75)^2 \\ &= 3.0625 + 0.5625 = 3.625 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (2 - 1)^2 + (3 - 2)^2 = 1^2 + 1^2 = 1 + 1 = 2 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (2 - 1)^2 + (5 - 2)^2 = 1^2 + 3^2 = 1 + 9 = 10 \end{aligned}$$

Jarak terdekat adalah neuron ke-2 sehingga bobot neuron 2 akan diperbaharui

$$\begin{aligned} w_{i2}(\text{baru}) &= w_{i2}(\text{lama}) + \eta(x_i - w_{i2}(\text{lama})) = \\ &= \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 0.5 \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 2.75 & 1.5 & 2 \\ 1.25 & 2.5 & 5 \end{bmatrix}$$

Untuk data 4 [3 4], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2.75 - 3)^2 + (1.25 - 4)^2 = (-0.25)^2 + (-2.75)^2 \\ &= 0.0625 + 7.5625 = 7.625 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.5 - 3)^2 + (2.5 - 4)^2 = (-1.5)^2 + (-1.5)^2 \\ &= 2.25 + 2.25 = 4.5 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (2 - 3)^2 + (5 - 4)^2 = (-1)^2 + 1^2 = 1 + 1 = 2 \end{aligned}$$

Jarak terdekat adalah neuron ke-3 sehingga bobot neuron 3 akan diperbaharui

$$\begin{aligned} w_{i3}(\text{baru}) &= w_{i3}(\text{lama}) + \eta(x_i - w_{i3}(\text{lama})) = \\ &= \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 0.5 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} \end{aligned}$$

Bobot baru menjadi :  $w = \begin{bmatrix} 2.75 & 1.5 & 2.5 \\ 1.25 & 2.5 & 4.5 \end{bmatrix}$

Untuk data 5 [5 4], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2.75 - 5)^2 + (1.25 - 4)^2 = (-2.25)^2 + (-2.75)^2 \\ &= 5.0625 + 7.5625 = 12.625 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.5 - 5)^2 + (2.5 - 4)^2 = (-3.5)^2 + (-1.5)^2 \\ &= 12.25 + 2.25 = 14.5 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (2.5 - 5)^2 + (4.5 - 4)^2 = (-2.5)^2 + 0.5^2 \\ &= 6.25 + 0.25 = 6.5 \end{aligned}$$

Jarak terdekat adalah neuron ke-3 sehingga bobot neuron 3 akan diperbaharui

$$\begin{aligned} w_{i3}(\text{baru}) &= w_{i3}(\text{lama}) + \eta(x_i - w_{i3}(\text{lama})) = \\ &= \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} \right) = \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} + 0.5 \begin{bmatrix} 2.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} + \begin{bmatrix} 1.25 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 3.75 \\ 4.25 \end{bmatrix} \end{aligned}$$

Bobot baru menjadi :  $w = \begin{bmatrix} 2.75 & 1.5 & 3.75 \\ 1.25 & 2.5 & 4.25 \end{bmatrix}$

Semua data sudah diproses, langkah selanjutnya perbaharui laju pembelajaran :

$$\eta(\text{baru}) = \text{fungsi\_pembelajaran} \times \eta(\text{lama}) = 0.6 \times 0.5 = 0.3$$

Data 1 : Neuron 1 menjadi pemenang

Data 2 : Neuron 1 menjadi pemenang

Data 3 : Neuron 2 menjadi pemenang

Data 4 : Neuron 3 menjadi pemenang

Data 5 : Neuron 3 menjadi pemenang

Indeks kelompok yang di ikuti data adalah:[1 1 2 3 3]

### Iterasi 2 :

Untuk data 1 [1 1], hitung jarak ke setiap neuron:

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2.75 - 1)^2 + (1.25 - 1)^2 = 1.75^2 + 0.25^2 = 3.0625 + 0.0625 = 3.125 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.5 - 1)^2 + (2.5 - 1)^2 = 0.5^2 + 1.5^2 = 0.25 + 2.25 = 2.55 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (3.75 - 1)^2 + (4.25 - 1)^2 = 2.75^2 + 3.25^2 = 7.5625 + 10.5625 = 18.125 \end{aligned}$$

Jarak terdekat adalah neuron ke-2 sehingga bobot neuron 2 akan diperbaharui.

$$\begin{aligned} w_{i2}(\text{baru}) &= w_{i2}(\text{lama}) + \eta(x_i - w_{i2}(\text{lama})) \\ &= \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} + 0.3 \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} \right) = \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} + 0.3 \begin{bmatrix} -0.5 \\ -1.5 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + \begin{bmatrix} -0.15 \\ -0.45 \end{bmatrix} = \begin{bmatrix} 1.35 \\ 2.05 \end{bmatrix} \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 2.75 & 1.35 & 3.75 \\ 1.25 & 2.05 & 4.25 \end{bmatrix}$$

Untuk data 2 [4 1], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (2.75 - 4)^2 + (1.25 - 1)^2 = (1.25)^2 + 0.25^2 = 1.5625 + 0.0625 = 1.625 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.35 - 4)^2 + (2.05 - 1)^2 = (-2.65)^2 + 1.05^2 = 7.0225 + 1.1025 = 8.125 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (3.75 - 4)^2 + (4.25 - 1)^2 = (-0.25)^2 + 3.25^2 = 0.0625 + 10.5625 = 10.625 \end{aligned}$$

Jarak terdekat adalah neuron ke-1 sehingga bobot neuron 1 akan diperbaharui

$$\begin{aligned} w_{i1}(\text{baru}) &= w_{i1}(\text{lama}) + \eta(x_i - w_{i1}(\text{lama})) = \\ &= \begin{bmatrix} 2.75 \\ 1.25 \end{bmatrix} + 0.3 \left( \begin{bmatrix} 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 2.75 \\ 1.25 \end{bmatrix} \right) = \begin{bmatrix} 2.75 \\ 1.25 \end{bmatrix} + 0.3 \begin{bmatrix} 1.25 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 2.75 \\ 1.25 \end{bmatrix} + \begin{bmatrix} 0.375 \\ -0.275 \end{bmatrix} = \begin{bmatrix} 3.125 \\ 1.175 \end{bmatrix} \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 3.125 & 1.35 & 3.75 \\ 1.175 & 2.05 & 4.25 \end{bmatrix}$$

Untuk data 3 [1 2], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (3.125 - 1)^2 + (1.75 - 2)^2 = 2.125^2 + (-0.825)^2 \\ &= 4.5156 + 0.6806 = 5.1962 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.35 - 1)^2 + (2.05 - 2)^2 = 0.35^2 + 0.05^2 = 0.1225 + 0.0025 = 0.125 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (3.75 - 1)^2 + (4.25 - 2)^2 = 2.75^2 + 2.25^2 = 7.5625 + 5.0625 = 12.625 \end{aligned}$$

Jarak terdekat adalah neuron ke-2 sehingga bobot neuron 2 akan diperbaharui

$$\begin{aligned} w_{i2}(\text{baru}) &= w_{i2}(\text{lama}) + \eta(x_i - w_{i2}(\text{lama})) = \\ &= \begin{bmatrix} 1.35 \\ 2.05 \end{bmatrix} + 0.3 \left( \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1.35 \\ 2.05 \end{bmatrix} \right) = \begin{bmatrix} 1.35 \\ 2.05 \end{bmatrix} + 0.3 \begin{bmatrix} -0.35 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 1.35 \\ 2.05 \end{bmatrix} + \begin{bmatrix} -0.105 \\ -0.015 \end{bmatrix} = \begin{bmatrix} 1.245 \\ 2.035 \end{bmatrix} \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 3.125 & 1.245 & 3.75 \\ 1.175 & 2.035 & 4.25 \end{bmatrix}$$

Untuk data 4 [3 4], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (3.125 - 3)^2 + (1.75 - 4)^2 = (-0.125)^2 + (-2.825)^2 \\ &= 0.0156 + 7.9806 = 7.9962 \end{aligned}$$

$$\begin{aligned} D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\ &= (1.245 - 3)^2 + (2.035 - 4)^2 = 1.755^2 + 1.965^2 \\ &= 3.0800 + 3.8612 = 6.9412 \end{aligned}$$

$$\begin{aligned} D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\ &= (3.75 - 3)^2 + (4.25 - 4)^2 = 0.75^2 + 0.25^2 = 0.5625 + 0.0625 = 0.625 \end{aligned}$$

Jarak terdekat adalah neuron ke-3 sehingga bobot neuron 3 akan diperbaharui

$$\begin{aligned} w_{i3}(\text{baru}) &= w_{i3}(\text{lama}) + \eta(x_i - w_{i3}(\text{lama})) = \\ &= \begin{bmatrix} 3.75 \\ 4.25 \end{bmatrix} + 0.3 \left( \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 3.75 \\ 4.25 \end{bmatrix} \right) = \begin{bmatrix} 3.75 \\ 4.25 \end{bmatrix} + 0.3 \begin{bmatrix} -0.75 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 3.75 \\ 4.25 \end{bmatrix} + \begin{bmatrix} -0.225 \\ -0.075 \end{bmatrix} = \begin{bmatrix} 3.525 \\ 4.175 \end{bmatrix} \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 3.125 & 1.245 & 3.525 \\ 1.175 & 2.035 & 4.175 \end{bmatrix}$$

Untuk data 5 [5 4], hitung jarak ke setiap neuron

$$\begin{aligned} D_1 &= \sum_i (w_{i1} - x_i)^2 = (w_{11} - x_1)^2 + (w_{21} - x_2)^2 \\ &= (3.125 - 5)^2 + (1.75 - 4)^2 = 1.875^2 + (-2.825)^2 \\ &= 3.5156 + 7.9806 = 11.4962 \end{aligned}$$



$$\begin{aligned}
 D_2 &= \sum_i (w_{i2} - x_i)^2 = (w_{12} - x_1)^2 + (w_{22} - x_2)^2 \\
 &= (1.245 - 5)^2 + (2.035 - 4)^2 = 3.755^2 + 1.965^2 \\
 &= 14.1000 + 3.8612 = 17.9612
 \end{aligned}$$

$$\begin{aligned}
 D_3 &= \sum_i (w_{i3} - x_i)^2 = (w_{13} - x_1)^2 + (w_{23} - x_2)^2 \\
 &= (3.525 - 5)^2 + (4.175 - 4)^2 = 1.475^2 + 0.175^2 \\
 &= 2.1756 + 0.0306 = 2.2062
 \end{aligned}$$

Jarak terdekat adalah neuron ke-3 sehingga bobot neuron 3 akan diperbaharui

$$\begin{aligned}
 w_{i3}(\text{baru}) &= w_{i3}(\text{lama}) + \eta(x_i - w_{i3}(\text{lama})) = \\
 &= \begin{bmatrix} 3.525 \\ 4.175 \end{bmatrix} + 0.3 \left( \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 3.525 \\ 4.175 \end{bmatrix} \right) = \begin{bmatrix} 3.525 \\ 4.175 \end{bmatrix} + 0.3 \begin{bmatrix} 1.475 \\ -0.175 \end{bmatrix}
 \end{aligned}$$

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 3.125 & 1.245 & 3.9675 \\ 1.175 & 2.035 & 4.1225 \end{bmatrix}$$

Semua data sudah diproses, langkah selanjutnya perbaharui laju pembelajaran :

$$\eta(\text{baru}) = 0.6 \times \eta(\text{lama}) = 0.6 \times 0.3 = 0.18$$

Data 1 : Neuron 2 menjadi pemenang

Data 2 : Neuron 1 menjadi pemenang

Data 3 : Neuron 2 menjadi pemenang

Data 4 : Neuron 3 menjadi pemenang

Data 5 : Neuron 3 menjadi pemenang

Indeks kelompok yang di ikuti data adalah:[2 1 2 3 3]

**Iterasi 3 :**

$$\text{Bobot baru menjadi : } w = \begin{bmatrix} 3.2825 & 1.1647 & 4.0105 \\ 1.1435 & 1.8759 & 4.0824 \end{bmatrix}$$

Laju pembelajaran menjadi :

$$\eta(\text{baru}) = \text{fungsi\_pembelajaran} \times \eta(\text{lama}) = 0.6 \times 0.18 = 0.108$$

Indeks kelompok yang di ikuti data adalah:[2 1 2 3 3]

**Iterasi 4 :**

Bobot baru menjadi :  $w = \begin{bmatrix} 3.36 & 1.1311 & 4.0201 \\ 1.1280 & 1.8049 & 4.0655 \end{bmatrix}$

Laju pembelajaran menjadi :

$$\eta(\text{baru}) = \text{fungsi\_pembelajaran} \times \eta(\text{lama}) = 0.6 \times 0.108 = 0.0648$$

Indeks kelompok yang di ikuti data adalah:[2 1 2 3 3]

**Iterasi 5 :**

Bobot baru menjadi :  $w = \begin{bmatrix} 3.4015 & 1.1146 & 4.0217 \\ 1.1197 & 1.7688 & 4.0573 \end{bmatrix}$

Laju pembelajaran menjadi :

$$\eta(\text{baru}) = \text{fungsi\_pembelajaran} \times \eta(\text{lama}) = 0.6 \times 0.0648 = 0.0389$$

Indeks kelompok yang di ikuti data adalah:[2 1 2 3 3]

Jadi, setelah melalui 5 kali iterasi yang ditetapkan, hasil pengelompokan menunjukkan bahwa data 1 dan 3 masuk ke kelompok 2, data 2 ,asuk ke kelompo 1, sedangkan 4 dan 5 masuk ke kelompok 3. Seperi pada tabel 2.3 :

**Tabel 2.3 :** Hasil pengelompokan sesuai cluster.

Data Ke-	Cluster 1	Cluster 2	Cluster 3
1		x	
2	X		
3		x	
4			X
5			X

## 2.10 Penelitian Sebelumnya

Penelitian sebelumnya yang dilakukan oleh M. habibulloh (2014) adalah sistem pengelompokan tingkat serapan materi kuliah pengenalan pola menggunakan metode Self Organizing Map. Pada penelitian ini algoritma yang digunakan yaitu algoritma Self Organizing Map. Setiap percobaan yang

dilakukan pada sistem pengelompokan ini menghasilkan cluster tetap. Perubahan hanya terjadi pada nilai pengujian validitas menggunakan silhouette index yang berbeda. Validitas terbaik ada pada pengclusteran keempat dengan tingkat akurasi rata-rata 0.1135. Sedangkan untuk pengelompokan terbaik pada metode SOM ada di percobaan 2 cluster, 3 cluster dan 4 cluster. Sedangkan pada 5 cluster mahasiswa yang tergabung pada kelompok mengalami perubahan. Sedangkan pada metode Fuzzy C-Means cluster terbaik hanya pada percobaan 3 cluster. Untuk percobaan 2 dan 4 cluster ada mahasiswa yang mengalami perubahan kelompok. Hasil pengujian untuk grade nilai tinggi, sedang dan rendah pada sistem ini mempunyai tingkat kesesuaian 52.491% dari 85 Mahasiswa dibandingkan dengan nilai yang didapat dari dosen pengampuh mata kuliah pengenalan pola.

Penelitian sebelumnya yang dilakukan oleh M. Hamdi Asyrofi (2014) adalah pengelompokan ketrampilan pemrograman WEB mahasiswa Teknik Informatika UMG menggunakan metode K – MEANS. Pada penelitian ini data diambil dari kuisioner mahasiswa jurusan teknik informatika angkatan 2010 dengan menggunakan algoritma K – MEANS sesuai dengan evaluasi *cluster internal Davies-Bouldin Index* adalah dijadikan 7 *cluster* karena memiliki nilai DBI yang paling minimal dari pada dikelompokkan menjadi 2,3, atau 5 *cluster*. Dari hasil pengujian sistem bisa berubah – ubah karna pembangkitan centroid awal dilakukan secara acak dan pengelompokan 3 cluster yang paling baik pada percobaan ke-7, hingga mencapai mahasiswa yang mempunyai kompetensi ketrampilan pemrograman WEB tinggi 63.53% (54 mahasiswa), sedang 30.59% ( 26 mahasiswa), rendah 5.88% (5 mahasiswa).

Penelitian sebelumnya yang dilakukan oleh Irman Hermadi, Imas S. Sitanggang, Edward (2007) adalah clustering menggunakan Self Organizing Maps (studi kasus : Data PPMB IPB). Pada penelitian ini data yang digunakan adalah rata-rata nilai Biologi, Fisika, Matematika, dan Kimia (cawu 1 sampai

cawu 7) dari pelamar tahun 2004 dengan pilihan pertama program sarjana di Fakultas Pertanian, IPB. Data (sebanyak 1899 baris dan 4 field yaitu: Biologi, Fisika, Kimia, dan Matematika), Penentuan bobot pemenang dalam algoritma SOM menggunakan Jarak Mahalanobis, dan inisialisasi nilai bobot awal dengan nilai midpoint. Kriteria pemberhentian algoritma SOM dalam penelitian ini adalah iterasi, dengan banyak iterasi: 1, 5, dan 10. Hasil clustering dari SOM divalidasi menggunakan Indeks Davies-Bouldin. Hasil clustering data yang memiliki DBI minimal (53.472) dari penelitian adalah ukuran vektor bobot 9 dengan learning rate 0.9, penurunan learning rate 0.1, dan 5 iterasi. Pelamar dari Sumatera banyak berada pada cluster yang memiliki rata-rata nilai Biologi, Fisika, Kimia, dan Matematika lebih tinggi (81.12, 77.50, dan 74.16). Berbeda dengan daerah asal Jawa, yang banyak berada di cluster yang memiliki rata-rata lebih rendah (74.08, 73.09, 71.91, 70.04, 68.59, dan 67.93). Pelamar dari Luar Negeri tergolong pelamar dengan nilai rendah, hanya berada di cluster dengan rata-rata 68.59.