

BAB II

LANDASAN TEORI

2.1. 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. 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)

2.3. 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.

2.4. 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.

2.5. Algoritma K-Means

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Agusta (2007)

Algoritma *K-means*

1. Menentukan nilai K sebagai jumlah *cluster* yang ingin di bentuk.

Jumlah *cluster* yang akan dibentuk ditentukan sendiri oleh pengguna system.

2. Membangkitkan K *centroid* (titik pusat *cluster*) secara random.

Dalam menentukan n buah pusat *cluster* awal dilakukan pembangkitan secara random yang mempresentasikan urutan data input. Pusat awal *cluster* di dapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data.

3. Menghitung jarak setiap data ke masing – masing *centroid*.

Untuk mengukur jarak antar data dengan pusat *cluster* digunakan jarak kosinus.

Algoritma perhitungan jarak data dengan pusat *cluster*.

1. Ambil nilai data dan nilai titik pusat *cluster*.
2. Hitung *cosinus* data dengan tiap pusat *cluster*.

$$\cos(x,c) = \frac{\sum_{j=1}^N X_j \cdot c_j}{\|x\| \|c\|} \dots\dots\dots (2.1)$$

3. Hitung nilai jarak dari *cosinus* dengan menggunakan rumus sebagai berikut : $d = S - 1 \dots\dots\dots (2.2)$

Dimana S adalah $\cos(x,c)$

4. Cari jarak terdekat dan masukkan X kedalam *cluster* sesuai dengan *centroid* tersebut.

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat (terkecil) antara data dan pusat cluster. Jarak ini menunjukkan bahwa data tersebut berada dalam satu keompok dengan pusat *cluster* terdekat.

Algoritma pengelompokan data.

1. Ambil nilai jarak tiap pusat *cluster* dengan data.
2. Cari nilai jarak yang terkecil.
3. Kelompokkan data dengan pusat *cluster* yang memiliki jarak terkecil

5. Menentukan posisi *centroid* baru dengan cara menghitung rata – rata dari data – data yang terpilih pada *centroid* yang sama.

Untuk mendapatkan pusat *cluster* baru, bisa di hitung dengan rata – rata nilai anggota *cluster* yang baru.

Algoritma penentuan pusat *cluster* baru :

1. Cari jumlah anggota tiap *cluster*.
2. Hitung pusat baru dengan rumus :
3. $\bar{v}_{ij} = \frac{1}{N} \sum_{k=0}^{N_i} X_{kj} \dots\dots\dots (2.3)$

Dimana :

v_{ij} = centroid/rata-rata *cluster* ke-i untuk variable ke-j

N_i = jumlah data yang menjadi anggota *cluster* ke-i

i, k = indeks dari *cluster*

j = indeks dari variable

X_{kj} = nilai data ke-k yang ada di dalam *cluster* tersebut untuk variable ke-j

6. Lakukan langkah 3 – 5 hingga posisi anggota *cluster* baru dengan anggota *cluster* lama tidak berubah.

Pusat *cluster* yang baru digunakan untuk melakukan perhitungan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen, dan akan berhenti jika hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama) atau apabila ada perubahan nilai centroid diatas nilai ambang atau nilai pada fungsi objektif yang telah ditentukan. Dimana nilai ambang (threshold) adalah $0.0000 < 0.01$.

K-means merupakan algoritma clustering yang bersifat partitional yaitu membagi himpunan objek kata ke dalam sub himpunan (*cluster*) yang tidak overlap, sehingga setiap objek data berada tepat dalam satu *cluster*. Strategi partitional- clustering yang paling sering digunakan adalah berdasarkan kriteria *square error*. Secara umum, tujuan kriteria *square error* adalah untuk memperoleh partisi (jumlah *cluster* tetap) yang meminimalkan total *square error*.

Contoh perhitungan :

Misalnya : Sebuah set data memiliki 4 objek sebagai titik data pelatihan dan setiap objek memiliki 2 atribut. Setiap atribut mewakili koordinat dari objek, yaitu :

Objek atribut 1 (x) : bobot indeks

Objek atribut 2 (y) : pH

Tabel 2.1 Data kasus

Objek	Atribut 1 (x) : Bobot indeks	Atribut 2 (y) : Ph
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Untuk menyelesaikan permasalahan tersebut, kita dapat melakukan beberapa tahap, yaitu :

1. Menentukan jumlah *cluster*

Dengan melihat data yang ada, maka kita dapat mengelompokkan objek menjadi dua cluster (cluster 1 dan cluster 2) sesuai atributnya. Masalahnya adalah bagaimana menentukan medicine tersebut merupakan anggota cluster 1 atau cluster 2. Dari data yang diperoleh, dapat ditentukan bahwa 4 objek tersebut memiliki 2 atribut (bobot indeks dan pH), dimana tiap-tiap medicine mewakili satu titik dengan 2 atribut (x,y).

2. Menentukan nilai awal centroid

Untuk menentukan nilai awal centroid dilakukan secara acak, dalam contoh kasus ini dimisalkan titik koordinat medicine A adalah cluster 1 (C1) dan medicine B (C2) sebagai nilai centroid awal.

- C1 = (1,1)
- C2 = (1,2)

3. Menghitung jarak antara titik centroid dengan tiap titik objek. Pada contoh kasus ini menggunakan jarak Euclidean. Berikut adalah cara untuk menghitung jarak dari tiap objek :

- Medicine A= (1,1) dengan C1 = (1,1)

$$\rightarrow = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

dengan C2 = (2,1)

$$\rightarrow = \sqrt{(1 - 2)^2 + (1 - 1)^2} = 1$$

- Medicine B = (2,1) dengan C1= (1,1)

$$\rightarrow = \sqrt{(2-1)^2 + (1-1)^2} = 1$$

dengan C2 = (2,1)

$$\rightarrow = \sqrt{(2-2)^2 + (1-1)^2} = 0$$

- Medicine C = (4,3) dengan C1= (1,1)

$$\rightarrow = \sqrt{(4-1)^2 + (3-1)^2} = 3.61$$

dengan C2 = (2,1)

$$\rightarrow = \sqrt{(4-2)^2 + (3-1)^2} = 2.83$$

- Medicine D = (5,4) dengan C1 = (1,1)

$$\rightarrow = \sqrt{(5-1)^2 + (4-1)^2} = 5$$

dengan C2 = (2,1)

$$\rightarrow = \sqrt{(5-2)^2 + (4-1)^2} = 4.24$$

Dari perhitungan diatas, diperoleh jarak matriksnya, yaitu :

$$D^0 = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} C1 \\ C2 \end{matrix} & \begin{pmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{pmatrix} \end{matrix} \rightarrow \begin{matrix} C1 = (1,1) \\ C2 = (2,1) \end{matrix}$$

4. Pengelompokan objek

Setelah menghitung jarak matriks, kita menentukan anggota cluster menurut jarak minimum dari centroid. Dengan melihat lagi pada jarak matriks, medicine A termasuk cluster 1, sedangkan medicine B, C dan D termasuk cluster 2. Hal ini dapat dilihat pada perolehan nilai sebagai berikut :

$$G^0 = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} Cluster 1 \\ Cluster 2 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \end{matrix} \rightarrow \begin{matrix} Cluster 1 \\ Cluster 2 \end{matrix}$$

5. Pada iterasi 1

a. Menentukan centroid baru.

Himpunan yang terbentuk pada iterasi sebelumnya, telah diketahui anggota tiap cluster. Untuk cluster 1 mempunyai anggota medicine A saja, sedangkan cluster 2 mempunyai anggota medicine B, C dan D. Dari data

tersebut hitung kembali centroid untuk menentukan centroid baru. Karena pada cluster 1 hanya mempunyai 1 anggota, maka untuk centroid baru masih berada di $C1 = (C1)$. Sedangkan pada C2 dengan menghitung nilai rata-ratanya dapat diperoleh nilai centroid barunya, yaitu :

$$C2 = \left(\frac{2+4+5}{3}, \frac{1+3+4}{3} \right) \quad C2 = \left(\frac{11}{3}, \frac{8}{3} \right)$$

- b. Menghitung jarak antara titik centroid baru dengan tiap titik objek. Pada tahap menghitung jarak antara objek dengan centroid baru. Hal ini hampir sama dengan tahap 3, yaitu menghitung jarak dengan C2

$$C2 = \left(\frac{11}{3}, \frac{8}{3} \right)$$

Dengan cara perhitungan yang sama pada tahap 3, maka diperoleh jarak matriksnya, yaitu

$$\begin{array}{cccc} & A & B & C & D \\ D^1 = & \begin{pmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.89 \end{pmatrix} & \rightarrow & C1 = (1,1) \\ & & & & \rightarrow & C2 = \left(\frac{11}{3}, \frac{8}{3} \right) \end{array}$$

- c. Melakukan pengelompokan objek

Hampir sama dengan tahap 4, yaitu menentukan anggota cluster dengan menghitung jarak minimum tiap objek dengan centroid baru. Hasil yang diperoleh adalah :

$$\begin{array}{cccc} & A & B & C & D \\ G^1 = & \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} & \rightarrow & \text{Cluster 1} \\ & & & & \rightarrow & \text{Cluster 2} \end{array}$$

6. Pada iterasi 2

- a. Menentukan centroid baru

Tahap ini mengulang kembali tahap 5, yaitu menghitung centroid baru. Dari cluster 1 yang mempunyai 2 anggota yaitu medicine A dan B, dan cluster 2 mempunyai anggota yaitu medicine C dan D, maka hasil centroid baru yang diperoleh adalah :

$$C1 = \left(\frac{1+2}{2}, \frac{1+1}{2} \right) \quad C1 = \left(\frac{3}{2}, 1 \right)$$

$$C2 = \left(\frac{4+5}{2}, \frac{3+4}{2} \right) \quad C2 = \left(\frac{9}{2}, \frac{7}{2} \right)$$

- b. Menghitung jarak antara titik centroid baru dengan tiap titik objek

Tahap ini juga hampir sama dengan tahap 3, yaitu menghitung jarak dengan centroid baru

$$C1 = \left(\frac{3}{2}, 1 \right) \quad C2 = \left(\frac{9}{2}, \frac{7}{2} \right)$$

Dengan cara perhitungan yang sama pada tahap 3, maka diperoleh jarak matriksnya, yaitu

$$D^2 = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A & B & C & D \end{matrix} & \begin{pmatrix} 0.5 & 0.5 & 3.20 & 4.61 \\ 4.30 & 3.54 & 0.71 & 0.71 \end{pmatrix} \end{matrix} \rightarrow \begin{matrix} C1 = \left(\frac{3}{2}, 1 \right) \\ C2 = \left(\frac{9}{2}, \frac{7}{2} \right) \end{matrix}$$

- c. Melakukan pengelompokan objek

Hampir sama dengan tahap 4, yaitu menentukan anggota cluster dengan menghitung jarak minimum tiap objek dengan centroid baru yang telah dihasilkan. Hasil yang diperoleh adalah :

$$G^2 = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A & B & C & D \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix} \rightarrow \begin{matrix} \text{Cluster 1} \\ \text{Cluster 2} \end{matrix}$$

Berdasarkan hasil anggota cluster yang diperoleh tetap sama antara $G^1 = G^2$, maka iterasi dihentikan.

Tabel 2.2 Hasil clustering

Objek	Atribut 1 (x) : bobot indeks	Atribut 2 (y) : pH	Cluster (result)
Medicine A	1	1	1
Medicine B	2	1	1
Medicine C	4	3	2
Medicine D	5	4	2

2.7 Davies-Bouldin Index

Davies-Bouldin Index diperkenalkan oleh David L. Davies and Donald W. Bouldin pada tahun 1979. Sum-of square within cluster (SSW) sebagai metrik kohesi dalam sebuah cluster. Pendekatan perhitungan validitas Davies Bouldin Index ini untuk memaksimalkan jarak *inter-cluster* di antara *Cluster* C_i dan C_j atau *Sum-of-square-between-cluster* (SSB) dan pada waktu yang sama mencoba untuk meminimalkan jarak antara titik dalam sebuah *cluster* atau *Sum-of-square-within-cluster* (SSW). Rumus SSW adalah :

$$SSW = \frac{1}{N} \sum_{i=1}^N \|x_i - C_{p_i}\|^2 \dots\dots\dots (2.5)$$

Rumus SSB :

$$SSB(C_i, C_j) = \sum_{j=1, j \neq i}^M \|C_i - C_j\|^2 \dots\dots\dots (2.6)$$

Rumus R dan DBI

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots\dots\dots (2.7)$$

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \dots\dots\dots (2.8)$$

2.9 Penelitian sebelumnya

Penelitian sebelumnya yang dilakukan oleh Dwi Novianti Nango (2012) adalah clustering data anggaran pendapatan belanja daerah di Kabupaten xyz dengan menggunakan metode K-Means. Pada penelitian ini digunakan algoritma K-Means untuk melakukan clustering data. Pembentukan cluster diuji dengan 3 nilai centroid dan 2 nilai centroid, yang dilanjutkan dengan menghitung nilai SSE. Hasil cluster dengan nilai SSE terkecil dijadikan sebagai parameter untuk memperkirakan data anggaran belanja yang akan datang. Berdasarkan penelitian yang dilakukan dapat diketahui bahwa pembentukan cluster dengan 3 nilai centroid adalah cluster yang terbaik, karena memiliki SSE terkecil.

Penelitian yang dilakukan oleh Najmatun Nabilah(2013) adalah pengklasifikasian jenis hadits dengan menggunakan metode *Fuzzy K-Nearest Neighbor In Every Class*. Perhitungan jarak menggunakan jarak

cosinus. Pengklasifikasian ini dilakukan dengan menggunakan atribut-atribut nilai persambungan sanad, jarh ta'dil atas dan jarh ta'dil bawah antara nilai data uji dengan nilai data training yang ada pada database. Pengujian sistem dilakukan dengan membandingkan hasil pengklasifikasian sistem dengan hasil pengklasifikasian manual. Hasil pengujian sistem pengklasifikasian jenis hadits dengan menggunakan metode FKNNEC menghasilkan nilai akurasi sebesar 89%.

Penelitian yang dilakukan oleh M. Nizam Al Anshori (2014) adalah pengelompokan kompetensi *database* mahasiswa teknik informatika Universitas Muhammadiyah Gresik menggunakan metode *K-Harmonic Means*. Perhitungan menggunakan jarak *Euclidean*, pengelompokan dilakukan dengan menggunakan atribut – atribut penilaian kompetensi *database* dasar, kompetensi SQL DDL, kompetensi SQL DML, kompetensi *database agregasi* dan *database advance*. Hasil uji system dengan evaluasi *cluster* internal Devies-Bouldin Index yang paling bagus yakni 0,167 dengan mengelompokkan mahasiswa menjadi 2 dan 3 klaster. Dengan jumlah prosentase 23,53% mahasiswa kompetensi *database* rendah, 34,12 % mahasiswa kompetensi *database* sedang, dan 42,35 % mahasiswa kompetensi *database* tinggi