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, pemakai 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 performasi 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 perli 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.

2.3. 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 Custering

   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)

Langkah-langkah dalam algoritma K-means clustering adalah :

1. Menentukan jumlah cluster
2. Menentukan nilai centroid

   Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut.

$$\bar{v}_{ij} = \frac{1}{N}\sum_{k=0}^{N_i} X_{kj} \quad \text{.....} \quad (2.1)$$

   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

3. Menghitung jarak antara titik centroid dengan titik tiap objek. Pada kasus ini untuk menghitung jarak menggunakan cosinus. Rumus cosinus:

   $D(x,y) = \cos(x,y) = \frac{\sum(x.y)}{||x||||y||}$

4. Pengelompokan objek

   Untuk menentukan anggota cluster bisa menggunakan similarity atau dissimilarity. Hubungan nilai similarity dan dissimilarity bisa dilihat dari d=1-s. d adalah dissimilarity, sedangkan s adalah similarity. Jika menggunakan similarity maka di cari jarak yang terbesar, jika menggunakan dissimilarity maka di cari jarak yang terkecil. Nilai yang diperoleh dalam keanggotaan data pada distance matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke cluster yang sama dan nilai 0 untuk data yang dialokasikan ke cluster yang lain.

5. Kembali ke tahap 2, lakukan perulangan hingga nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain.

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 kita 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 = (2,1)

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

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

     ➔ $= \frac{(1x1)+(1x1)}{\sqrt{(1^2+1^2)}x\sqrt{(1^2+1^2)}} = 1 \longrightarrow$ d = 1-1 = 0

     dengan C2 = (2,1)

     ➔ $= \frac{(1x2)+(1x1)}{\sqrt{(1^2+1^2)}x\sqrt{(2^2+1^2)}} = 0.948683 \longrightarrow$ d = 1-0.948683 = 0.0513167

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

     ➔ $= \frac{(2x1)+(1x1)}{\sqrt{(2^2+1^2)}x\sqrt{(1^2+1^2)}} = 0.948683 \longrightarrow$ d = 1- 0.948683 = 0.051316

     dengan C2 = (2,1)

     ➔ $= \frac{(2x2)+(1x1)}{\sqrt{(2^2+1^2)}x\sqrt{(2^2+1^2)}} = 1 \longrightarrow$ d = 1-1 = 0

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

  ➔ $= \frac{(4x1)+(3x1)}{\sqrt{(4^2+3^2)}x\sqrt{(1^2+1^2)}} = 0.989949 \longrightarrow d = 1\text{-}0.989949 = 0.0100505$

  dengan C2 = (2,1)

  ➔ $= \frac{(4x2)+(3x1)}{\sqrt{(4^2+3^2)}x\sqrt{(2^2+1^2)}} = 0.98387 \longrightarrow d = 1\text{-}0.98387 = 0.01613009$

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

  ➔ $= \frac{(5x1)+(4x1)}{\sqrt{(5+4^2)}x\sqrt{(1^2+1^2)}} = 0.993884 \longrightarrow d = 1\text{-}0.993884 = 0.0061163$

  dengan C2 = (2,1)

  ➔ $= \frac{(5x2)+(4x1)}{\sqrt{(5+4^2)}x\sqrt{(2^2+1^2)}} = 0.977802 \longrightarrow d = 1 - 0.977802 = 0.02219759$

Dari perhitungan diatas, diperoleh jarak matriksnya, yaitu :

$$D^0 = \begin{matrix} A & B & C & D \\ \end{matrix}$$

$$\begin{pmatrix} 0 & 0.051316 & 0.0100505 & 0.0061163 \\ 0.0513167 & 0 & 0.01613009 & 0.02219759 \end{pmatrix} \begin{matrix} \rightarrow C1 = (1,1) \\ \rightarrow C2 = (2,1) \end{matrix}$$

4. Pengelompokan objek

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

$$G^0 = \begin{matrix} A & B & C & D \\ \end{matrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} \rightarrow \text{Cluster 1} \\ \rightarrow \text{Cluster 2} \end{matrix}$$

5. Iterasi 1, menentukan centroid baru.

   Himpunan yang terbentuk pada iterasi sebelumnya, telah diketahui anggota tiap cluster. Untuk cluster 1 mempunyai anggota medicine A, C dan D, sedangkan cluster 2 mempunyai anggota medicine B saja. Dari data tersebut hitung kembali centroid untuk menentukan centroid baru. Karena pada cluster 2 hanya mempunyai 1 anggota, maka untuk centroid baru masih berada di C2 = (C2). Sedangkan pada C1 dengan menghitung nilai rata-ratanya dapat diperoleh nilai centroid barunya, yaitu :

C1 = $\left(\frac{1+4+5}{3}, \frac{1+3+4}{3}\right)$ C1 = $\left(\frac{10}{3}, \frac{8}{3}\right)$

6. Iterasi 1, 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 C1 dan C2

   C1 = $\left(\frac{10}{3}, \frac{8}{3}\right)$ C2 = (2, 1)

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

   $$D^1 = \begin{matrix} A & B & C & D \\ \begin{pmatrix} 0.0061163 & 0.0221976 & 0.0004879 & 0 \\ 0.0513167 & 0 & 0.01613009 & 0.02219759 \end{pmatrix} & \begin{matrix} \rightarrow C1 = (\frac{10}{3}, \frac{8}{3}) \\ \rightarrow C2 = (2,1) \end{matrix} \end{matrix}$$

7. Iterasi 1, 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 :

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

8. Iterasi 2, menentukan centroid baru

   Tahap ini mengulang kembali tahap 5, yaitu menghitung centroid baru. Dari cluster 1 yang mempunyai 3 anggota yaitu medicine A, C dan D, dan cluster 2 mempunyai anggota yaitu medicine B, maka hasil centroid baru yang diperoleh adalah :

   C1 = $\left(\frac{1+4+5}{3}, \frac{1+3+4}{3}\right)$ C1 = $\left(\frac{10}{3}, \frac{8}{3}\right)$

   Iterasi 2, 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 = $\left(\frac{10}{3}, \frac{8}{3}\right)$ C2 = (2, 1)

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

$$D^1 = \begin{matrix} \text{A} \\ \\ \end{matrix}\begin{matrix} \text{B} \\ \\ \end{matrix}\begin{matrix} \text{C} \\ \\ \end{matrix}\begin{matrix} \text{D} \\ \\ \end{matrix}\begin{pmatrix} 0.0061163 & 0.0221976 & 0.0004879 & 0 \\ 0.0513167 & 0 & 0.01613009 & 0.02219759 \end{pmatrix}\begin{matrix} \rightarrow \text{C1} = (\frac{10}{3}, \frac{8}{3}) \\ \rightarrow \text{C2} = (2,1) \end{matrix}$$

9. Iterasi 2, 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 :

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

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	2
Medicine C	4	3	1
Medicine D	5	4	1

2.6. Jarak Cosinus

Cosinus merupakan sebuah ukuran kemiripan dua vektor dengan mencari cosine diantara dua vector tersebut. Semakin dekat nilai yang dihasilkan dengan angka 1, maka data tersebut semakin similarity. Hubungan nilai similarity dan dissimilarity bisa dilihat dari d=1-s. d adalah dissimilarity, sedangkan s adalah similarity. Sehingga nilai similarity bisa dikatakan 1 dan nilai dissimilarity adalah 0. Mencari nilai similarity bisa menggunakan rumus cosinus

Mencari jarak kemiripan

$$D(x,y) = \cos(x,y) = \frac{\sum x.y}{||x||||y||} \quad \text{.....} \quad (2.2)$$

Atau

x.y = $\sum_{j=1}^{N} Xj, Yj$ atau ||x|| = $\sqrt{\sum_{j=1}^{N} x^2{}_j}$ =$\sqrt{x.x}$ ..... (2.3)

Dengan x dan y adalah sebagai berikut :

X = nilai data uji Y = nilai data training

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. Separasi dengan Sum-of-square-between-cluster (SSWB) dengan mengukur jarak antara centroid $C_i$dan C_j. $R_{i,j}$ adalah ukuran rasio seberapa baik nilai perbandingan antara cluster ke-i dan cluster ke-j. Rumus SSW adalah :

$$SSW = \frac{1}{N}\sum_{i=1}^{N} \|x_i - C_{Pi}\|^2$$

Rumus SSB :

$$SSB = \frac{2}{M(M-1)}\sum_{I=1}^{M}\sum_{j=1,j\neq 1}^{M} \left\|C_i - C_j\right\|^2$$

Rumus R dan DBI

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \qquad DBI = \frac{1}{K}\sum_{i=1}^{K} \max_{i\neq j}\left(R_{i,j}\right)$$

2.8. Penelitian sebelumnya

Penelitian sebelumnya yang dilakukan oleh Dwi Noviati 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 Jonh Fredrik Ulysses (2012) adalah Prediksi lama masa studi mahasiswa berdasarkan jalur penerimaan menggunakan metode naive bayes. Penelitian ini menggunakan data pendidikan berupa sampel data alumni dari mahasiswa lulusan STMIK Palangkaraya jurusan Manajemen Informatika dengan atribut-atribut yaitu NIM, nama, alamat, tempat tanggal lahir, lulus tahun, IPK, lama studi/semester, model penerimaan dan asal daerah. Untuk model penerimaan dibagi menjadi dua kategori yaitu SPMB dan jalur khusus. Berdasarkan pengujian, dapat dianalisa hasil bahwa mahasiswa yang masuk melalui jalur khusus memiliki kecenderungan untuk lulus lebih cepat daripada mahasiswa yang masuk melalui SPMB.

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 FK-NNC menghasilkan nilai akurasi sebesar 89%.