

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 ang 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

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 ang 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.

2.4 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.

2.5 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.

Data discrimination 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 dipresentasikan 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 pengelompokan satu set benda-benda fisik atau abstrak ke dalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan pengelompokan 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 digambarkan sebagai berikut :

| 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 | Proses clustering berdasarkan pada prinsip ; objek yang ada di dalam |

| | |
|--|---|
| <p>atau fungsi yang menggambarkan dan membedakan data kelas atau konsep, dengan tujuan untuk dapat menggunakan model untuk memprediksi objek kelas yang kelas label nya belum diketahui. Model tersebut berdasarkan pada analisis dari <i>training data</i> (data objek yang kelas label nya telah diketahui).</p> | <p>satu cluster memiliki kemiripan yang tinggi daripada yang lainnya, tetapi sangat berbeda dengan objek yang ada pada cluster lainnya.</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.

7. Association rules

Association rules (aturan asosiasi) atau *affinit analysis* (analisis afinitas) bernan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa studi transaksi di supermarket misalnya seseorang ang membeli susu bai juga membeli sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa maka aturan asosiaso juga sering dinamakan *market basket analyysis*.

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “if-then” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistik (Santoso 2007).

Analisis asosiasi dikenal juga sebagai salah satu metode data mining yang menjadi dasar dari berbagai metode data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter support (nilai panunjang) yaitu prosentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antara item dalam aturan assosiatif. Analisis asosiasi di definisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan sarat minimum untuk confidence (*minimum confidence*) (pramudiono, 2007).

2.6 Langkah-langkah Data mining

1. *Identity The Bussiness Problem*

Yang pertama dan juga dasar dari virtous 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 yang dihadapi kita dapat menentukan data-data mana saja yang kita butuhkan untuk dapat dilakukan tahap analisa.

2. *Mine The Data For Actionable 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 / kebijaksanaan 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.

2.7 Teknik-teknik/ 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 psetangga 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.

2.8 Tugas Data Mining (Six Tax 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 di tentukan 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 di identifikasi atau proses yang menjelaskan hasil akhir dari jalanya proses data mining.

2.9 Analisis Asosiasi Apriori

Analisis asosiasi berguna untuk menemukan hubungan penting yang tersembunyi di antara set data yang sangat besar. Hubungan yang sudah terbuka dipresentasikan dalam bentuk aturan asosiasi atau set aturan item ang sering muncul.

Isu penting dalam analisis asosiasi adalah bagaimana cara menemukan pola tertentu dari data yang berjumlah sangat besar, yang membuat biaya komputasi menjadi sangat mahal. Isu penting lainnya adalah bahwa diantara pola-pola yang harus dipelajari dalam teknik analisis asosiasi, ada banyak pola yang akhirnya dibuang karena tidak termasuk dalam kriteria pola yang dicari. (Prasetyo, 2012:311).

Algoritma Apriori menggunakan pendekatan level yang baik dalam pembangkitan asosiasi, dimana setiap level berhubungan dengan sejumlah item yang dimiliki konsekuen aturan. Awalnya, aturan dengan *confidence* yang tinggilah yang mempunyai satu item diekstrak. Aturan ini kemudian digunakan untuk membangkitkan kandidat aturan yang baru. (Prasetyo, 2012:331).

Dalam algoritma apriori semakin kecil nilai minimum support, semakin banyak rule yang dihasilkan tetapi tidak semua rule yang dihasilkan valid, yang menunjukkan bahwa nilai minimum support berpengaruh pada pembentukan rule tetapi tidak banyak berpengaruh

terhadap kevalidan rule. Semakin kecil nilai minimum confidence, semakin banyak rule yang dihasilkan dan besar kemungkinan rule tersebut valid, yang menunjukkan bahwa nilai minimum confidence berpengaruh pada kevalidan rule. Dari rule-rule yang memenuhi nilai confidence dihitung nilai lift atau interest factor rasionya, rule yang memiliki nilai lift rasio atau interest factor lebih dari 1 maka rule tersebut valid dan bisa dijadikan rekomendasi. (Rizqi, 2010).

a. **Analisa pola frekuensi tinggi**

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support item diperoleh dengan persamaan *Support* (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau itemset dari keseluruhan transaksi.

Nilai support 1 *item* diperoleh dengan rumus 1 berikut[LAR05]:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Jumlah transaksi}} \dots\dots(2.1)$$

Nilai support 2 *item* diperoleh dengan rumus 2 berikut[LAR05]:

$$\text{Support (A, B)} = p(A \cap B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah transaksi}} \dots\dots(2.2)$$

b. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk

confidence dengan menghitung confidence aturan asosiatif $A \rightarrow B$ diperoleh dari persamaan 3.

Nilai confident 2 item diperoleh dengan rumus 1 berikut[LAR05]:

$$\text{Confidence} = P(A|B) = \frac{\text{Nilai Support A dan B}}{\text{Nilai Support A}} \dots\dots (2.3)$$

c. Contoh algoritma Apriori untuk Pencarian Association Rule

Misalkan terdapat data transaksi seperti ang ditunjukkan pada Tabel 1

Tabel 2.1 Data Item Transaksi

| Transaction ID | Item Set |
|-----------------------|--------------------------------|
| 1 | Item A, Item C, Item D |
| 2 | Item B, Item C, Item E |
| 3 | Item A, Item B, Item C, Item E |
| 4 | Item B, Item E |

Misalkan diinginkan minimum support : 50% (2 dari 4 transaksi)

Langkah 1 : mencari nilai *support* untuk masing-masing itemset.

Seperti yang ditunjukkan pada tabel $L1 = \{\text{large 1-itemset}\}$

Tabel 2.2 Nilai *Support* untuk 1 Item Set

| Itemset | Support |
|----------------|----------------|
| A | 50 % |
| B | 75 % |
| C | 75 % |
| D | 25 % |
| E | 75 % |

Langkah 2: mencari kandidat itemset untuk L2:

2.1 : Gabungkan itemset pada L1 (algoritma apriori-gen)

{AB, AC, AD, AE, BC, BD, BE, CD, CE, DE}

2.2 : Hapus yang tidak ada dalam itemset

Itemset {BD, DE} dihapus karena tidak ada dalam itemset

Langkah 3 : Hitung nilai *support* untuk masing-masing *itemset*.

Hasilnya dapat ditunjukkan pada Tabel 3.

Tabel 2.3 Nilai *Support* untuk 2 *Itemset*

| Itemset | Support |
|----------------|----------------|
| AB | 25 % |
| AC | 50 % |
| AD | 25 % |
| AE | 25 % |
| BC | 50 % |
| BE | 75 % |
| CD | 25 % |
| CE | 50 % |

Langkah 4: tentukan *itemset* yang memenuhi minimum *support*.

Hasilnya dapat ditunjukkan pada Tabel 4.

L2 {large 2-itemset}

Tabel 2.4 anggota 2 *itemset* yang Memenuhi Minimum *Support*

| Itemset | Support |
|---------|---------|
| AC | 50 % |
| BC | 50 % |
| BE | 75 % |
| CE | 50 % |

Langkah 5 :Ulangi langkah 2-4

Langkah 5.1, Gabungkan itemset pada L2 dan L2. Hasilnya dapat ditunjukkan pada Tabel 5.

Tabel 2.5 Anggota 3 *Itemset*

| Itemset | Hasil Gabungan (3 itemset) |
|---------|----------------------------|
| AC + BC | A C B |
| AC + BE | A C B, A C E , A B E |
| AC + CE | A C E |
| BC + BE | B C E |
| BC + CE | B C E |
| BE + CE | B C E |

Langkah, 5.2, Hapus yang tidak ada dalam itemset : {A C E}

Langkah 6 :Hitung *support* dari setiap kandidat *itemset* L3

Hasilnya dapat ditunjukkan pada Tabel 2.6

Tabel 2.6 Nilai Support untuk 3 *Itemset*

| Itemset | Support |
|----------------|----------------|
| A B C | 25 % |
| A B E | 25 % |
| B C E | 50% |

Langkah 7 :L3 {large 3-itemset} {B C E}

Langkah 8 : STOP karena sudah tidak ada lagi kandidat untuk 4-itemset.

Untuk mencari aturan asosiasi diperlukan juga minconf

Misal minconf: 75 %, aturan asosiasi yang mungkin terbentuk dapat ditunjukkan pada **Tabel 2.7**

| Aturan (X → Y) | Sup (X∪Y) | Sup (X) | Confidence |
|-----------------------|------------------|----------------|-------------------|
| B C → E | 50 % | 50 % | 100 % |
| B E → C | 50 % | 75 % | 66,67 % |
| C E → B | 50 % | 50 % | 100 % |
| A → C | 50 % | 50 % | 100 % |
| C → A | 50 % | 75 % | 66,67 % |
| B → C | 50 % | 75 % | 66,67 % |
| C → B | 50 % | 75 % | 66,67 % |
| B → E | 75 % | 75 % | 100 % |
| E → B | 75 % | 75 % | 100 % |
| C → E | 50 % | 75 % | 66,67 % |
| E → C | 50 % | 75 % | 66,6 % |

Tabel 2.8 tabel aturan asosiasi yang memenuhi kriteria minimum support 50 % dan minimum confidence 75 %

| Aturan ($X \rightarrow Y$) | Sup ($X \cup Y$) | Sup (X) | Confidence |
|---------------------------------------|--------------------|-------------|--------------|
| B C \rightarrow E | 50 % | 50 % | 100 % |
| C E \rightarrow B | 50 % | 50 % | 100 % |
| A \rightarrow C | 50 % | 50 % | 100 % |
| B \rightarrow E | 75 % | 75 % | 100 % |
| E \rightarrow B | 75 % | 75 % | 100 % |

2.10 Analisis Dasar Diskritisasi

Diskritisasi adalah metode dasar pendekatan untuk mengatur keberlanjutan sebuah atribut. Pendekatan tersebut dilakukan dengan cara mengubah representasi atribut kategorikal menjadi biner pada setiap interval. Nilai untuk sebuah item adalah 1 jika item tersebut ada dalam atribut, atau 0 jika tidak ada dalam atribut. (Kumar, Tan & Steinbach, 2006). Untuk contoh dasar proses diskritisasi dapat dilihat pada **Tabel 2.9**.

Tabel 2.9 Proses Dasar Diskritisasi

| Gender | Male | Female |
|--------|------|--------|
| Male | 1 | 0 |
| Female | 0 | 1 |
| Female | 0 | 1 |
| Male | 1 | 0 |
| Female | 0 | 1 |
| Male | 1 | 0 |
| Female | 0 | 1 |

2.1.2 Penelitian Sebelumnya

Algoritma Apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola – pola item di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum support. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik data mining lainnya. Beberapa riset yang telah dilakukan berkaitan dengan kasus asosiasi yang menggunakan metode apriori antara lain :

Penelitian yang berjudul “*Analisis Profil Mahasiswa Berdasarkan Latar Belakang Atribut Mahasiswa Baru menggunakan Metode Apriori (Association Rules) di Universitas Muhammadiyah Gresik*” oleh Muhammad Abdi Setyawan. Adapun data yang diambil dalam penelitian ini adalah status sekolah jenis kelamin alamat, instansi sekolah, umur mahasiswa baru dengan Menggunakan Metode Apriori. Penelitian ini bertujuan untuk mendapatkan informasi yang berguna tentang asosiasi latar belakang mahasiswa baru untuk strategi promosi mahasiswa baru. Informasi yang ditampilkan berupa nilai support dan confidence hubungan antara profil mahasiswa baru dengan atribut latar belakang mahasiswa. Rule yang mempunyai tingkat confidence tertinggi pada masing-masing periode dengan rule 3 itemset adalah {kebomas, SMK} → {Swasta} dengan nilai *confidence* sebesar 96,96% dan 95,99%. Sedangkan rule 2 itemset tertinggi adalah {SMK} → {L} dan {Cerme} → {L} dengan nilai confidence sebesar 83,15% dan 90,90%. Ada 18 rule pada periode 2012/2013 dan 15 rule pada periode 2013/2014 dan memiliki korelasi atau hubungan yang positif dari pasangan variabel mahasiswa dengan nilai korelasi ≥ 1 .

Penelitian lain dilakukan oleh Muchammad iljas mengenai “*Rancang Bangun Perangkat Lunak Analisa Keranjang Pasar Dengan Metode Apriori*”. Dalam penelitiannya analisa keranjang pasar dengan metode Apriori sebagai salah satu teknik data mining dapat digunakan untuk menggali pola kecenderungan kemunculan barang secara bersamaan yang dilakukan oleh para pembeli. Dengan data transaksi yang besar maka bahasa SQL (Structured Query Language) dapat digunakan sebagai solusi yang tepat untuk meningkatkan kecepatan proses pembentukan frequent itemset pembentukan 3 itemset dari 2 itemset perhitungan nilai support dan confidence serta dapat melakukan pembentukan kaidah asosiasi dari 2-itemset dan 3-itemset. Berdasarkan hasil grafik analisa data transaksi order detail pada database northwise SQL Server yang telah diubah ke database MySQL nilai support tertinggi yang dapat digunakan untuk membentuk 2-itemset sebesar 1% (satu persen) dengan kemunculan sebanak 8 transaksi berbanding terhadap seluruh transaksi yang berjumlah 830 transaksi.

Penelitian yang berjudul “*Penerapan Data Mining dengan Algoritma Apriori untuk mendukung Strategi promosi pendidikan*” oleh Gunadi Widi Nurcahyo. Algoritma Apriori dapat diterapkan untuk mendukung strategi promosi pendidikan pada perguruan tinggi. Informasi yang berkaitan dengan pelaksanaan promosi dapat tersedia dengan cepat, sehingga pihak manajemen dapat melakukan pengambilan keputusan dengan cepat. Pelaksanaak promosi pendidikan pada perguruan tinggi sangat dibantu dengan adanya penerapan algoritma Apriori ini sehingga diharapkan efektifitas pelaksanaan promosi pendidikan akan semakin ditingkatkan. Penerapan algoritma Apriori yang dilakukan melalui perangkat lunak yang dirancang terbukti menunjukkan hasil yang sama dibandingkan dengan melakukan perhitungan secara manual ataupun dengan perangkat lunak sejenis. Hal

ini dibuktikan melalui perhitungan nilai *support* dan *confidence* yang menunjukkan hasil yang sama. SMA A menjadi target promosi yang sangat potensial karena memiliki nilai minimum *support* yang cukup besar yaitu 25 % artinya adalah bahwa dari seluruh mahasiswa perguruan tinggi, maka 25 % nya adalah berasal dari SMA. Bila fokus promosi mau lebih dipertajam lagi maka kita dapat melihat bahwa konsentrasi yang cukup dominan di SMA adalah di bidang IPS yang memiliki nilai minimum *confidence* sebesar 75%. Artinya dari keseluruhan siswa SMA A yang masuk ke suatu perguruan tinggi maka 75 % nya adalah berasal dari jurusan IPS sedangkan Jurusan IPA hanya mencapai 25% dari keseluruhan siswa SMA A yang masuk ke perguruan tinggi.

Selain itu penelitian-penelitian yang terkait mengenai analisis data mahasiswa salah satunya adalah penelitian yang dilakukan oleh Muhammad Fathur Rozi dalam risetnya mengenai “*pencarian Distribusi Potensi Daerah Untuk Promosi Universitas Muhammadiyah Gresik Teknik Informatika Menggunakan Metode K-Modes*”. Penulis menggunakan kriteria yaitu asal sekolah jurusan alamat instansi sekolah dengan menggunakan metode K-Modes. Penelitian ini bertujuan untuk memudahkan panitia PMB untuk mengetahui potensi daerah yang terbaik untuk dilakukanya promosi kampus teknik informatika. Penguji aplikasi ini dilakukan secara external dan internal dilakukan percobaan sebanyak 20 kali 10 kali menggunakan 2 cluster dan 10 kali menggunakan 3 cluster.