

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. Komponen

Suatu sistem terdiri 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.

2. Batasan sistem

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

3. Lingkungan Luar Sistem.

Lingkungan luar (*environment*) dari suatu sistem adalah apapun diluar batas sistem yang mempengaruhi operasi. Lingkungan luar sistem dapat bersifat menguntungkan dan dapat juga bersifat merugikan sistem tersebut.

4. Penghubung Sistem

Penghubung (*interface*) merupakan media penghubung antara satu subsistem dengan subsistem yang lainnya. melalui penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke

subsistem yang lainnya. Dengan penghubung satu subsistem dapat berintegrasi dengan subsistem yang lainnya membentuk satu kesatuan.

2.3 Pengertian *Data Mining*

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). *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 (Santoso, 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 system*, *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). *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar (Witten, 2005).

Karakteristik *data mining* sebagai berikut :

- *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

- *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering*, *classification*, *association rules mining*, *neural network*, *genetic algorithm* dan lain-lain (Pramudiono, 2007).

2.4 Tahap-Tahap *Data mining*

Tahap-tahap *data mining* ada 7 yaitu :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang 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 juga 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-atribut yang mengidentifikasi entitas-entitas yang unik seperti 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 analysis*, tidak perlu 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 kategorikal. 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 kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi 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.

7. Presentasi pengetahuan (*knowledge presentation*),

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining* (Han, 2006).

2.5 Fungsi *Data Mining*

Fungsi *data mining* dan macam-macam pola yang dapat ditemukan menurut Han dan 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 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. *Classification and Prediction*

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksi kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data (yaitu objek data yang memiliki label kelas yang diketahui). Model yang diturunkan dapat direpresentasikan 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. Aturan pengelompokan tersebut akan digunakan untuk 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.

3. *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 *Custer* adalah proses pengelompokan satu set benda-benda fisik atau abstrak kedalam 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.

4. *Outlier analysis*

Outlier merupakan objek data yang tidak mengikuti perilaku umum dari data. *Outlier* dianggap sebagai noise atau pengecualian. Analisis *data outlier* dapat dianggap sebagai *noise* atau pengecualian. Analisis *data*

outlier dinamakan *Outlier Mining*. Teknik ini berguna dalam *fraud detection* dan *rare events analysis*.

5. 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.

6. Association rules

Association rules (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *market basket analysis*.

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 banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan syarat minimum untuk confidence (*minimum confidence*) (Pramudiono, 2007).

2.6 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 yang 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 konsekuensi 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 rasionalnya, 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.

1. *Support* (dukungan)

Support 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]:

Sebagai contoh, support dari aturan ”A=>B” adalah probabilitas kemunculan barang A dan B secara bersamaan.

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

2. *Confidence* (tingkat kepercayaan)

Probabilitas kejadian beberapa barang dibeli bersamaan dimana salah satu produk sudah pasti dibeli. Sebagai contohnya, jika terdapat n transaksi dimana barang A dibeli dan ada m transaksi dimana barang A dan B dibeli secara bersamaan, maka confidence dari aturan *if A then B* adalah m/n .

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

$$\text{Support } (A \rightarrow B) = \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Jumlah transaksi mengandung } A} \dots\dots(2.3)$$

b. Membentuk aturan asosiatif (*Associaton Rule Mining*)

Association rule mining adalah suatu prosedur yang mencari assosiasi antar item dalam suatu data set yang ditentukan[KAM03]. Interestingness measure yang dapat digunakan dalam pembentukan aturan asosiatif adalah

Aturan assosiatif untuk 2-item biasanya dinyatakan dalam bentuk :

{roti, mentega} (support = 50%, confidence = 62%)

Keterangan : "62% dari transaksi yang memuat item roti juga memuat mentega. Sedangkan 50% dari seluruh transaksi yang ada di database memuat kedua item itu (roti dan mentega)".

Aturan assosiatif untuk 3-item dinyatakan dalam bentuk :

$\{\text{roti, mentega}\} \rightarrow \{\text{susu}\}$ (support = 40%, confidence = 50%)

Keterangan : "50% dari transaksi yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu (roti, mentega dan susu)".

Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini mewakili 40% dari catatan transaksi selama ini."

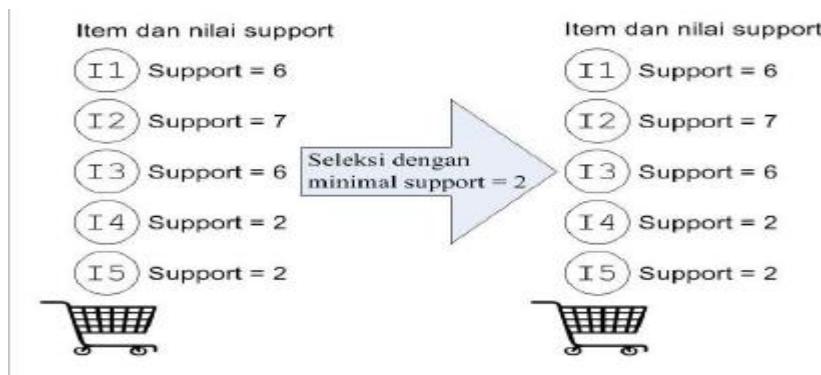
Ilustrasi algoritma Apriori dapat dilihat pada proses berikut :

Tahap I : Analisa pola frekuensi

- Tentukan nilai *minimum support* dan *minimum confidence*

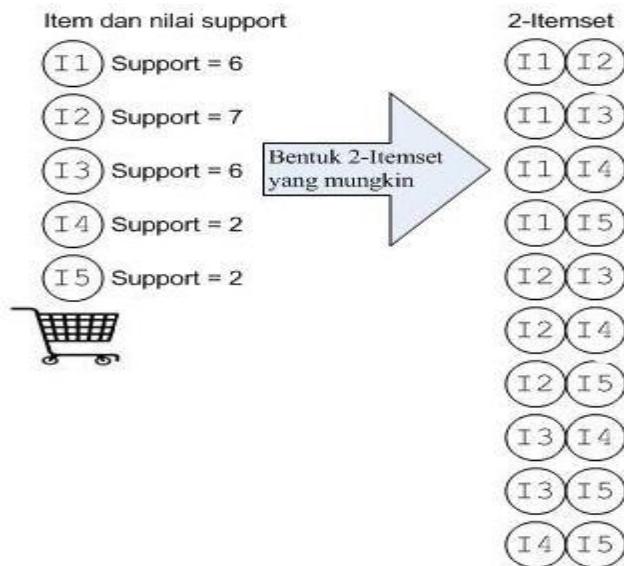
Apabila diinginkan nilai *minimum support* adalah 2, atau $\text{minimum support} = 2/9 = 22\%$ dan *minimum confidence* yang diinginkan sebesar 60%.

- Mencari jumlah dari masing-masing item pada tabel transaksi dan menyeleksi sesuai *minimum support* terlihat pada **Gambar 2.1**.



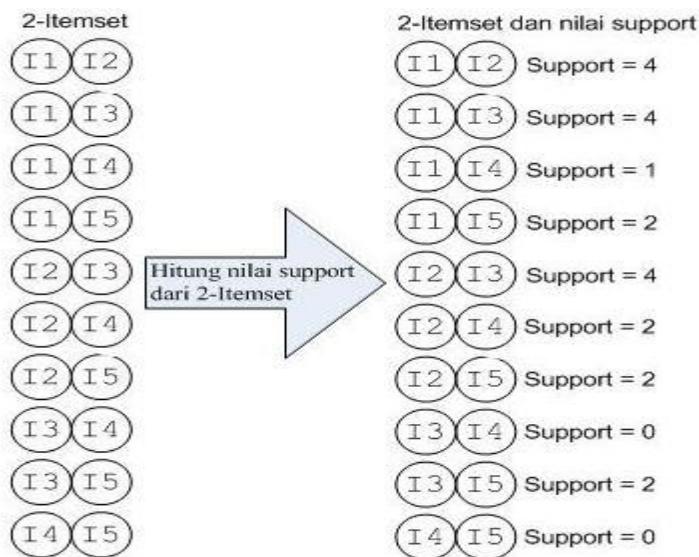
Gambar 2.1 Menghitung dan menyeleksi nilai support barang

- Membentuk kombinasi *2-itemset* dari hasil penyeleksian minimum support pada masing-masing *item* terlihat pada **Gambar 2.2**.



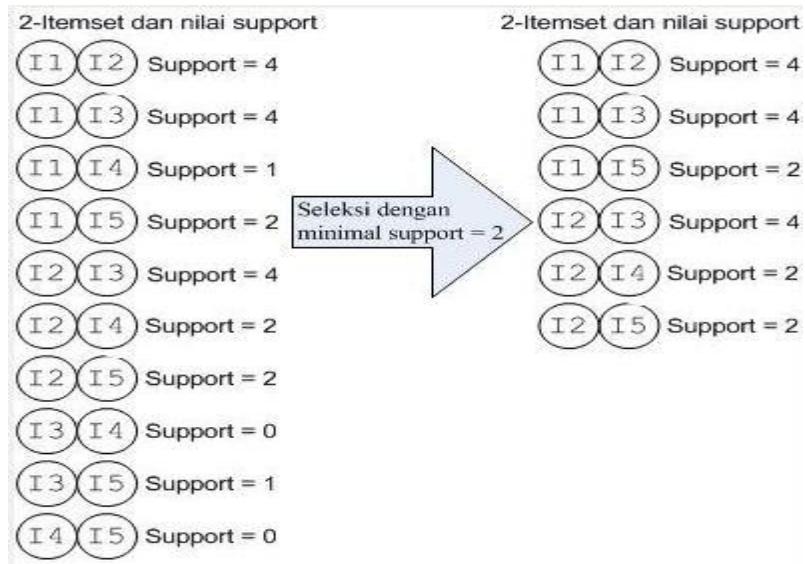
Gambar 2.2 Membentuk 2-itemset dari tabel candidate 1

- d) Mencari nilai support dari kombinasi 2-itemset sesuai dengan tabel transaksi terlihat pada **Gambar 2.3**.



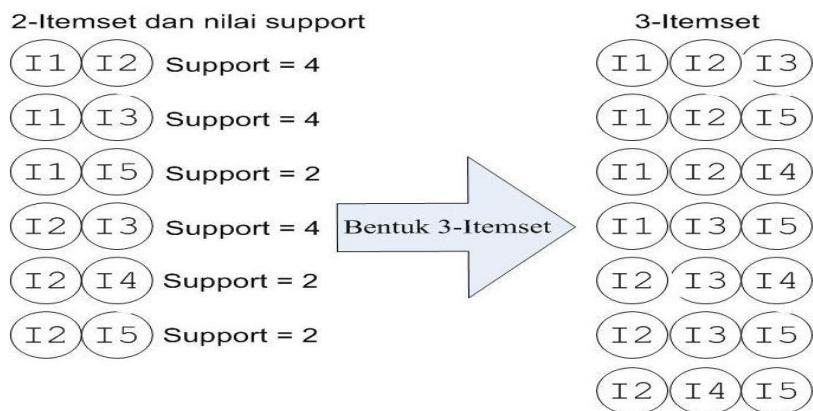
Gambar 2.3 Mencari nilai support dari kombinasi 2-itemset

- e) Menyeleksi nilai support *2-itemset* dengan nilai minimum support terlihat pada **Gambar 2.4**.



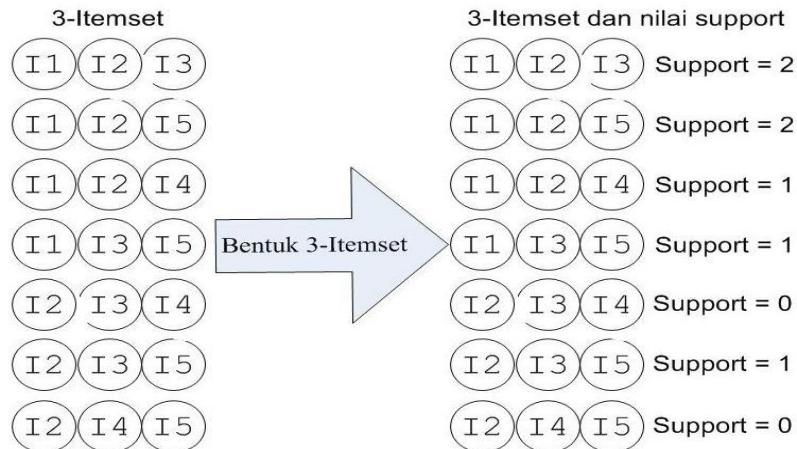
Gambar 2.4 Menyeleksi nilai support *2-itemset* dengan nilai minimum support

- f) Membentuk kombinasi *3-itemset* dari *2-itemset* yang memiliki nilai minimum support yang ditentukan terlihat pada **Gambar 2.5**.



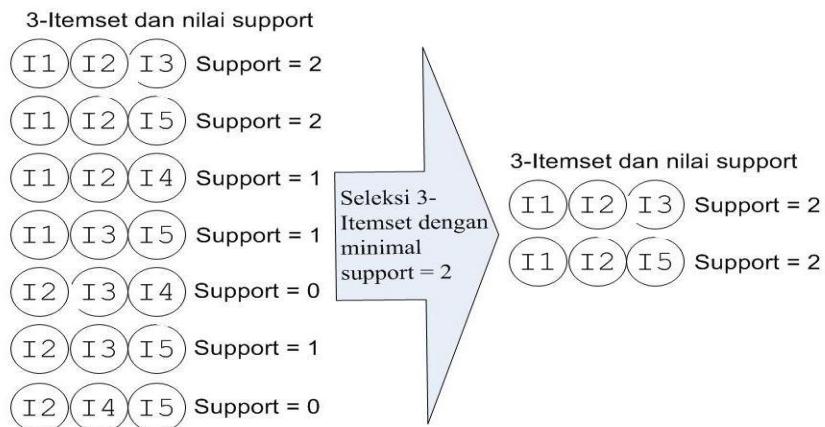
Gambar 2.5 Membentuk kombinasi *3-itemset* dari *2-itemset*

- g) Mencari nilai support dari *3-itemset* sesuai tabel transaksi terlihat pada **Gambar 2.6**.



Gambar 2.6 Mencari nilai support dari *3-itemset*

- h) Menyeleksi nilai *support* dari *3-itemset* dengan *minimum support* terlihat pada **Gambar 2.7**.



Gambar 2.7 Menyeleksi nilai *support* dari *3-itemset* dengan *minimum support*

- i) Hasil dari *2-itemset* dan *3-itemset* yang memenuhi nilai minimum support sebesar 22% terlihat pada **Gambar 2.8**.

2-Itemset dan nilai support	3-Itemset dan nilai support
(I1) (I2) Support = 4	(I1) (I2) (I3) Support = 2
(I1) (I3) Support = 4	(I1) (I2) (I5) Support = 2
(I1) (I5) Support = 2	
(I2) (I3) Support = 4	
(I2) (I4) Support = 2	
(I2) (I5) Support = 2	

Gambar 2.8 Data *2-itemset* dan *3-itemset* yang memenuhi nilai minimum support

Tahap II : Membentuk aturan asosiasi

Aturan asosiasi dibentuk dari data *2-itemset* dan *3-itemset* yang memenuhi nilai minimum support yaitu terlihat pada **Gambar 2.8**. Data *2-itemset* dan *3-itemset* akan dibentuk kombinasi yang mungkin dengan memperhatikan urutannya (tidak berulang). Misalkan terdapat *2-itemset* {A, B}, maka dihasilkan kombinasi {A, B} dan {B,A}. Dari data pada **Gambar 2.8** dapat dibentuk aturan asosiatif seperti terlihat pada **Tabel 2.1**.

Tabel 2.1 menunjukkan pada kolom pertama merupakan aturan asosiasi, diikuti dengan kolom kedua yaitu kolom nilai support dari itemset dari aturan asosiasi yang merupakan prosentasi kemunculan dari itemset dibandingkan dengan jumlah transaksi yang ada. Pada kolom ketiga merupakan nilai prosentase frekuensi kemunculan variabel X yang merupakan item atau itemset yang berada disebelah kiri dari aturan asosiasi. Pada kolom keempat merupakan nilai prosentase kekuatan hubungan antara variabel X dengan variabel Y, dimana variabel X merupakan item atau itemset yang berada disebelah kiri dari aturan asosiasi. Sedangkan variabel Y merupakan item yang berada disebelah kiri dari aturan asosiasi.

Setelah nilai confidence diperoleh, aturan asosiasi akan diseleksi dengan nilai minimal confidence yang ditentukan yaitu 60%. Sehingga akan dihasilkan tabel aturan asosiasi yang memenuhi kriteria minimum support 22% dan minimum confidence 60% yang dapat dilihat pada **Tabel 2.2**

Tabel 2.1 Tabel Aturan Asosiasi

Aturan Asosiasi ($X \rightarrow Y$)	Prosentase Sup(XUY)	Prosentase Sup (X)	Prosentase Confidence
I1, I2 → I3	22%	25%	25%
I1, I3 → I2	22%	25%	25%
I2, I3 → I1	22%	25%	25%
I1, I2 → I5	22%	25%	25%
I1, I5 → I2	22%	100%	100%
I2, I5 → I3	22%	100%	100%
I1, → I2	44%	66%	66%
I2 → I1	44%	57%	57%
I3 → I1	44%	66%	66%
I1 → I3	44%	66%	66%
I1 → I5	22%	33%	33%
I5 → I1	22%	100%	100%

Tabel 2.2 Tabel aturan asosiasi yang memenuhi kriteria minimum support 22% dan minimum confidence 60%

Aturan Asosiasi ($X \rightarrow Y$)	Prosentase Sup(XUY)	Prosentase Sup (X)	Prosentase Confidence
I1, I5 → I2	22%	100%	100%
I2, I5 → I1	22%	100%	100%
I1 → I2	44%	66%	66%
I1 → I3	44%	66%	66%
I3 → I1	44%	66%	66%
I5 → I1	22%	100%	100%
I3 → I2	44%	66%	66%

2.7 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 2.3**.

Tabel 2.3 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

2.8 Tinjauan Pustaka

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 assosiatif 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 “*Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)*” oleh Nuqson Masykur Huda. Adapun data yang diambil dalam penelitian ini adalah data induk mahasiswa dan data kelulusan mahasiswa, dapat menghasilkan informasi tentang tingkat kelulusan dengan data induk mahasiswa melalui teknik *data mining*. Kategori tingkat kelulusan di ukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma *apriori*, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan.

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 northwide SQL Server yang telah diubah ke database MySQL, nilai support tertinggi yang dapat digunakan untuk membentuk 2-itemset sebesar 1% (satu persen) dengan kemunculan sebanyak 8 transaksi berbanding terhadap seluruh transaksi yang berjumlah 830 transaksi.

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. Pengujian aplikasi ini dilakukan secara external dan internal, dilakukan percobaan sebanyak 20 kali, 10 kali menggunakan 2 cluster dan 10 kali menggunakan 3 cluster.