

BAB II

LANDASAN TEORI

2.1 Definisi Sistem

Sistem secara fisik adalah kumpulan elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (gordon,1991)[2].

Jogianto (2005 :2)[4] mengemukakan bahwa sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.2 Karakteristik Sistem

Jogianto (2005 :3)[4] 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 untuk menjalankan suatu fungsi tertentu mempengaruhi proses sistem secara keseluruhan.

3. Lingkungan Luar Sistem

Lingkungan luar sistem (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 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 Definisi Produk Bayi

Bayi merupakan individu yang berusia 0 – 12 bulan yang ditandai dengan pertumbuhan dan perkembangan yang cepat disertai dengan perubahan dalam kebutuhan zat gizi (Wong,2003). Sedangkan dalam rentang usia 1- 4 tahun maka dikatakan sebagai balita, jika usia 6 – 12 tahun dinamakan later childhood (anak besar).

Dalam marketing produk adalah sesuatu yang dapat ditawarkan ke pasar untuk diperhatikan, dipakai, dimiliki atau dikonsumsi sehingga dapat memuaskan keinginan atau kebutuhan.

Menurut kotler dan Armstrong (2008) beberapa atribut yang menyertai dan melengkapi produk adalah:

1. Merek (branding)

Merek adalah nama, istilah, tanda, simbol, atau rancangan, atau kombinasi dari semua ini yang dimaksudkan untuk mengidentifikasi produk dari satu atau kelompok penjualan dan membedakannya dari produk pesaing. Pemberian merek merupakan masalah pokok dalam strategi produk.

2. Pengemasan (Packing)

Pengemasan adalah kegiatan merancang dan membuat wadah atau pembungkusan suatu produk.

3. Kualitas Produk (Produk Quality)

Kualitas produk adalah kemampuan suatu produk untuk melaksanakan fungsinya meliputi, daya tahan keandalan, ketepatan kemudahan operasi dan perbaikan, serta atribut bernilai lainnya.

Demikian pula dengan produk bayi yang ada di toko mamiko, disana banyak kebutuhan yang dijual mulai dari perlengkapan mandi, perlengkapan tidur, dan lain - lain. Dari berbagai jenis merek yang ada menawarkan kepada konsumen merek apa yang terbaik dan menarik konsumen untuk membeli sesuai umur bayi dan balita masing-masing. Dengan pertumbuhan bayi yang selalu berkembang seiring waktu, maka kebutuhannya selalu berbeda dan memungkinkan para konsumen memilih produk yang berbeda pula sesuai kebutuhan perkembangan, misalnya dalam hal perlengkapan mandi maka konsumen harus menyesuaikan dengan apa yang dibutuhkan, dalam hal makanan juga, sehingga bayi atau balita tersebut akan berkembang dengan baik dan nyaman dengan adanya produk bayi yang melengkapi pertumbuhan dan perkembangannya.

Penjelasan dari beberapa jenis produk bayi adalah sebagai berikut:

1. Shampoo adalah sejenis cairan, yang berfungsi untuk meningkatkan tegangan permukaan kulit kepala sehingga dapat meluruhkan kotoran. Contoh produk shampo antara lain : zwitsal, johson, pigeon, cussions, dan lain-lain.
2. Sabun adalah pembersih yang digunakan untuk membersihkan anggota tubuh. Contoh produk sabun antara lain: johnson, zwitsal, cussions, dan lain-lain.
3. Minyak rambut adalah salah satu produk penata rambut yang berbentuk cair yang membuat rambut rapi dan wangi. Contoh produk minyak rambut antara lain : cussions, johnson, dan lain-lain.
4. Popok adalah alas bayi yang terbuat dari bahan dimana dapat menyerap air dengan baik. Contoh produk popok antara lain : mamypoko, pampers, sweety, dan lain-lain.
5. Susu formula adalah susu yang dibuat dari susu sapi atau susu buatan yang diubah komposisi hingga dapat dipakai sebagai pengganti ASI. Contoh produk susu formula antara lain: bebelove, SGM, chil mil, nutribaby, dan lain-lain.

2.4 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)[9]. Data mining merupakan pengestrakan 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 & lutfi, 2009)[5]. Hasil penemuan yang 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)[3].

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 visualisasi 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.

2.5 Pekerjaan Dalam Data Mining

Pekerjaan yang berkaitan dengan data mining dapat dibagi menjadi empat kelompok (Eko Prasetyo, 2012)[7], yaitu :

1. Model Prediksi

Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang di dapat. Ada dua jenis model prediksi, yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk variabel target diskret, sedangkan regresi untuk variabel target kontinu.

2. Analisis Kelompok

Contoh pekerjaan yang berkaitan dengan analisis kelompok (cluster analysis) adalah bagaimana caranya mengetahui pola pembelian barang oleh para konsumen pada waktu-waktu tertentu. Dengan mengetahui pola kelompok pembelian tersebut, perusahaan/pengecer dapat menentukan jadwal promosi yang dapat diberikan sehingga omzet penjualan bisa ditingkatkan.

Analisis kelompok melakukan pengelompokan data-data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data-data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan dengan kelompok tersebut.

3. Analisis Asosiasi

Analisis asosiasi (*association analysis*) digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efisien.

Penerapan yang paling dekat dengan kehidupan sehari-hari adalah analisis data keranjang belanja. Sebagai contoh, pembeli adalah ibu rumah tangga yang akan membeli kebutuhan rumah tangga di sebuah supermarket. Jika ibu tersebut membeli beras, sangat besar kemungkinannya bahwa ibu itu juga akan membeli barang lain, misalnya minyak, telur, dan tidak mungkin dan jarang membeli barang lain seperti topi atau buku. Dengan mengetahui hubungan yang lebih kuat antara beras dengan telur daripada beras dengan topi, pengecer dapat menentukan barang-barang yang sebaiknya disediakan dalam jumlah yang cukup banyak.

4. Deteksi Anomali

Pekerjaan deteksi anomali (*anomaly detection*) berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristiknya menyimpang (berbeda) dari data yang lain disebut outlier. Algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju eror yang rendah. Deteksi anomali dapat diterapkan pada sistem jaringan untuk mengetahui pola data yang memasuki jaringan sehingga penyusup bisa ditemukan jika pola kerja data

yang datang berbeda. Perilaku kondisi cuaca yang mengalami anomali juga dapat dideteksi dengan algoritma ini.

2.6 Fungsi Data Mining

Fungsi data mining dan macam-macam pola yang dapat ditemukan menurut Ham Kamber (2006)[3], yaitu :

1. *Concept /class Description : Characterization and Discrimination*

Data characterization adalah ringkasan dari semua karakteristik atau fitur dari data yang 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.

Association Analisis 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 memprediksi kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data (yaitu objek data yang dipresentasikan dalam berbagai bentuk seperti *if-then* klasifikasi, decision tree, naive 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 dikelompokkan lain. Sedangkan, *clustering* atau Analisis cluster adalah proses pengelompokkan satu set benda-benda fisik atau abstrak ke dalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan pengelompokkan 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)[3] lebih spesifik dijelaskan pada **tabel 2.1** :

Tabel 2.1 Perbandingan antara Classification dan Clustering

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 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 Training data (data objek yang kelas label nya telah diketahui).	Proses clustering berdasarkan pada prinsip objek yang ada dalam satu cluster memiliki kemiripan yang tinggi daripada yang lainnya, tetapi sangat berbeda dengan objek yang ada pada cluster lainnya.

5. *Outlier Analysis*

Outlier merupakan objek data yang tidak mengikuti perilaku umum dari data. *Outlier* dianggap sebagai noise atau pengecualian. *Analysis data outlier* dapat dianggap sebagai noise atau pengecualian. *Analysis data outlier* dinamakan *Outlier Mining*. Teknik ini berguna dalam *fraud detection* dan *rare event analysis*.

6. *Evolution Analysis*

Analisis evolusi adalah data menjelaskan dan memodelkan *trnd* dari objek yang memiliki karakteristik, 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

di dorong oleh penemu knowledge ditindak lanjuti dengan pengambilan tindakan dari penemuan.

7. *Association Rules*

Association rules (aturan asosiasi) atau *affinit analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa studi transaksi di supermarket misalnya seorang yang membeli susu baik juga membeli 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)[8].

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 asosiatif dapat diketahui dengan dua parameter *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antara item dalam aturan asosiatif. Analisis asosiasi di definisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat untuk *confidence* (*minimum confidence*) (pramudiono, 2007).

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 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 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 tetangga 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 aglomerative, 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 Link Analysis berguna untuk proses optimasi.

5. *Rule Induction*

Ekstraksi aturan sebab-akibat dari data secara static, identifikasi aturan yang tersimpan di dalam data. Metode berhubungan dengan induksi antara 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 dapat digunakan pula untuk mengekstark pola dan mendeteksi tren yang sangat kompleks untuk dibicarakan baik oleh manusia maupun teknik komputer lainnya.

2.8 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)[7].

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)[7].

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).

Confidence = jumlah transaksi dengan item dalam antedent (A) dan concequent (B) dibagi jumlah transaksi dengan item dalam antedent (A) dan biasanya pada istilah antecedent untuk mewakili bagian “jika” dan concequent untuk mewakili bagian “maka”. (Budi Santoso,2007:228)[8].

2.9 Tahapan Association Rules

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik 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. (Muhammad Ikhsan et al, 2007).

Metodologi dasar analisis terbagi menjadi dua tahap :

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 dari rumus berikut :

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

Nilai support 2 *item* diperoleh dari rumus berikut :

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

b. Pembentukan Aturan Asosiatif

Setelah semua pola frekuensi tinggi ditemukan barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung aturan asosiatif $A \rightarrow B$ diperoleh dari rumus berikut :

$$\text{Confidence} = P(B | A) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \dots\dots\dots(2.3)$$

c. Contoh Algoritma Apriori untuk Pencarian Association Rule

Misalkan terdapat data transaksi seperti yang ditunjukkan pada **tabel 2.2**

Tabel 2.2 Data Item Transaksi

ID Transaksi	Itemset
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 2.3 Nilai *Support* untuk 1 itemset

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

Hapus itemset D karena nilai support tidak memenuhi minsup yang telah ditetapkan.

Langkah 2: mencari kandidat itemset untuk L2:

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

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

Hasilnya dapat ditunjukkan pada **Tabel 2.4**.

Tabel 2.4 Nilai Support untuk 2 Itemset

Itemset	Support
AB	25%
AC	50%
AE	25%
BC	50%
BE	75%
CE	50%

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

Hasilnya dapat ditunjukkan pada **tabel 2.5**.

Tabel 2.5 frequent 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 2.5**.

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

Kombinasi 3 itemset tersebut diperoleh dari frequent 2 itemset dikombinasikan dengan itemset selanjutnya.

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
ACB	25%
ACE	25%
ABE	25%
ACE	25%
BCE	50%

Langkah 7: L3 {large 3-itemset}

Hasil yang lolos adalah BCE,BEC karena memenuhi minsup yaitu 50%.

Dapat di lihat pada **tabel 2.7**.

Tabel 2.7 frequent 3- *itemset* yang memenuhi Miinimum *Support*

Itemset	Support
BCE	50%

Untuk mencari aturan asosiasi diperlukan juga minconf

Misal minconf 75%, aturan asosiasi yang mungkin terbentuk dapat ditunjukkan pada **tabel 2.8**.

Tabel 2.8 aturan asosiasi

Aturan ($X \rightarrow Y$)	Sup ($X \cup Y$)	Sup (X)	Confidence
A \rightarrow C	50%	50%	100%
B \rightarrow C	50%	75%	66.67%
B \rightarrow E	75%	75%	100%
C \rightarrow E	50%	75%	66.67%
C \rightarrow A	50%	75%	66.67%
C \rightarrow B	50%	75%	66.67%
E \rightarrow B	75%	75%	100%

Lanjutan **Tabel 2.8** aturan asosiasi

Aturan ($X \rightarrow Y$)	Sup ($X \cup Y$)	Sup (X)	Confidence
E \rightarrow C	50%	75%	66.67%
B \rightarrow CE	50%	75%	66.67%
C \rightarrow BE	50%	75%	66.67%
E \rightarrow BC	50%	75%	66.67%
B \rightarrow EC	50%	75%	66.67%
E \rightarrow BC	50%	75%	66.67%

Tabel 2.9 tabel aturan asosiasi yang memenuhi kriteria minimum support 50% dan minimum confidence 75%

Aturan ($X \rightarrow Y$)	Sup ($X \cup Y$)	Sup (X)	Confidence
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)^[1]. Untuk contoh dasar proses diskritisasi dapat dilihat pada **Tabel 2.10**.

Tabel 2.10 Proses Dasar Diskritisasi

Gender	Male	Female
Male	1	0
Female	0	1
Female	0	1
Male	1	0
Female	0	1
Male	1	0

2.11 Penelitian Sebelumnya

Algoritma Apriori adalah algoritma paling terkenal untuk menentukan 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 “ *Implementasi Data Mining Dengan Metode Algoritma Apriori Dalam menentukan Pola Pembelian Obat*” oleh Roby Yanto Dan Riri Khoiriah. Penelitian ini tujuannya untuk menemukan suatu strategi dimana akan menambah omset penjualan obat. Salah satunya dengan tetap tersedianya berbagai jenis obat yang dibutuhkan oleh konsumen. Penerapan algoritma apriori dapat membantu dalam membentuk kandidat kombinasi *item*, kemudian dilakukan pengujian apakah kombinasi tersebut memenuhi parameter *support* dan *confidence* minimum yang merupakan nilai ambang yang diberikan oleh pengguna.

Jika memenuhi parameter *support* dan *confidence* maka hasil tersebut dapat membantu dalam penentuan pola pembelian obat dan membantu tata letak obat berdasarkan kecenderungan konsumen membeli obat. Pengujian dilakukan berdasarkan nama obat bukan taksonomi obat, kombinasi 2 *itemset* yang akan dijadikan tolak ukur dalam pencarian nilai *support* dan *confidence* pada data transaksi pembelian obat, pengolahan data dilakukan terhadap data penjualan obat pada periode januari obat pada periode januari dan februari 2014 dengan sampel 20 data transaksi.

Penelitian selanjutnya oleh Eka Novita Sari yang berjudul “ *Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan*”. Penjualan pakaian wanita mengalami peningkatan setiap bulannya, produk yang ditawarkan bermacam merek, untuk mengetahui merek yang paling diminati diperlukan algoritma apriori.

dan untuk mendapatkan hasil yang lebih baik digunakan Tools Tanagra. Data mining adalah metode untuk mencari data terlaris atau paling banyak dibutuhkan customer. Data mining merupakan analisis data menggunakan tool untuk menemukan pola dan aturan dalam himpunan data. Perangkat lunak bertugas untuk menemukan pola dengan mengidentifikasi aturan dan fitur pada data dan diharapkan mampu mengenal pola ini dalam data dengan input minimal dari user. Salah satunya bentuk pola yang dapat dihasilkan data mining adalah association rule. Dengan menggunakan nilai minimum *support* 50% dan minimum *confidence* 70%. Mendapatkan nilai terbaik yaitu nilai *confidence* 100%.

Berikutnya penelitian yang berjudul “*Implementasi Data Mining Algoritma Apriori Pada Sistem Penjualan*” oleh Muhammad Afif Syaifullah. Dalam penelitian ini dilakukan untuk menentukan barang yang banyak terjual sehingga dapat meningkatkan penjualan barang tersebut. Sama dengan penelitian yang lain, Data mining adalah metode untuk mencari data terlaris atau paling banyak dibutuhkan customer. Data mining merupakan analisis data menggunakan tool untuk menemukan pola dan aturan dalam himpunan data. Perangkat lunak bertugas untuk menemukan pola dengan mengidentifikasi aturan dan fitur pada data dan diharapkan mampu mengenal pola ini dalam data dengan input minimal dari user. Salah satunya bentuk pola yang dapat dihasilkan data mining adalah association rule.