

## BAB II

### TINJAUAN STUDI DAN LANDASAN TEORI

#### 2.1 Tinjauan Studi

Penelitian mengenai penerapan metode *market basket analysis* bukan merupakan hal asing, sebab telah banyak dilakukan oleh peneliti sebelumnya. Penulis telah mengamati beberapa penelitian yang relevan terhadap topik penelitian ini.

Andreas Handoyo [8]. Judul penelitian “Aplikasi *Data Mining* untuk meneliti Asosiasi Pembelian Item Barang di Supermarket dengan Metode *Market Basket Analysis*”. Kesimpulan dari penelitian ini sebagai berikut:

- a. Aplikasi mampu mengolah data transaksi yang disediakan oleh *user*, untuk menemukan *frequent itemset* dan *association rule* yang memenuhi syarat minimum support, berdasarkan item yang ada dalam transaksi dan mampu menampilkan *rules* dalam bentuk grafik dan teks.
- b. Dengan aplikasi ini dapat diketahui asosiasi barang apa saja yang sering dibeli bersamaan oleh konsumen di supermarket yang nantinya informasi ini dapat memberikan pertimbangan tambahan bagi manajer dalam pengambilan keputusan guna pembelian barang dan pengaturan barang pada rak supermarket.
- c. Pada analisa terhadap sejumlah data, ditemukan bahwa semakin kecil minimum *support* dan *confidence* yang ditentukan, semakin banyak pula *rules* yang dapat dihasilkan oleh aplikasi, dengan konsekuensi waktu proses akan lebih lama dibandingkan minimum *support* yang lebih besar.

Gregorius Satia Budhi [9]. Judul Penelitian “Aplikasi *Data Mining* Market Basket Analysis Pada Tabel Data Absensi Elektronik Untuk Mendeteksi Kecurangan Absensi (Check-Lock) Karyawan Di Perusahaan”. Kesimpulan dari penelitian ini sebagai berikut:

- a. Dari hasil pengujian perangkat lunak yang mendapatkan hasil sesuai dengan harapan, dapat disimpulkan bahwa metode *market basket*

*analysis* dapat pula dimanfaatkan untuk menggali *pattern* kebiasaan absensi(*check-lock*) pegawai sebuah perusahaan. Dari sini kemungkinan terjadinya kecurangan saat melakukan absensi masuk / pulang dapat dideteksi.

- b. Dari hasil pengujian kecepatan perangkat lunak dapat disimpulkan bahwa penggunaan algoritma *Pincer Search* untuk menggenerasi *frequent itemset* masih belum dapat mengatasi terjadinya „*bottleneck*“ pada proses tersebut.
- c. Mengingat bahwa proses *generate frequent itemset* dilakukan paling cepat 1 minggu sekali, dapat disimpulkan bahwa perangkat lunak telah dapat diaplikasikan pada dunia nyata, terutama pada PT. Mulia Batara Semesta Surabaya.

Pada penelitian dengan judul “*Implementasi Data Mining Algoritma Apriori pada Sistem Persediaan Alat-Alat Kesehatan [4]*” dianalisa sejumlah data dengan 30 jenis *item*, serta ambang batas yang ditemukan adalah minimum *support* 16% dan minimum *confidence* 70%. Dari batas-batas yang diberikan, terbentuk dua aturan asosiasi. Kesimpulan dari penelitian tersebut adalah :

algoritma apriori cukup efisien dan dapat mempercepat proses pembentukan kecenderungan pola kombinasi *itemset*, namun juga memiliki kelemahan yaitu memerlukan waktu yang lama untuk mengolah data dengan skala besar karena algoritma ini harus melakukan scan menyeluruh pada *database* dalam setiap kali iterasi.

Penelitian berikutnya pada tahun 2013 yang dilakukan oleh Tomi Listiawan [10] dengan judul “*Pembuatan Prototype Perangkat Lunak Data Mining Berbasis Web Untuk Penggalian Kaidah Asosiasi Berdasarkan Algoritma Apriori Menggunakan PHP*” melakukan uji coba dengan algoritma apriori terhadap 3 *database* berbeda dengan beberapa nilai minimum *support* yang berbeda pula. Dari percobaan analisa terhadap

ketiga *database* tersebut dengan masing-masing diperiksa dengan nilai minimum *support* 25%, 35%, dan 40%, dapat disimpulkan bahwa :

- a. Jumlah transaksi bukanlah satu-satunya faktor yang mempengaruhi waktu proses penggalian aturan asosiasi.
- b. Jumlah transaksi yang lebih besar belum tentu memerlukan waktu komputasi yang besar pula. Waktu proses komputasi dipengaruhi oleh beberapa faktor, yaitu : kemunculan setiap *item* pada transaksi, jumlah transaksi, batas minimum *support* dan minimum *confidence*.
- c. Nilai minimum *support* berbanding terbalik dengan jumlah aturan yang ditemukan dan waktu komputasi. Artinya semakin tinggi nilai minimum *support* yang diberikan, jumlah aturan yang ditemukan semakin kecil begitu juga dengan waktu komputasi yang semakin berkurang, begitu juga sebaliknya.

Tabel 2.1 *State of The Art*

No	Judul	Penulis	Tahun	Metode	Hasil
1	Pembuatan Prototype Perangkat Lunak <i>Data Mining</i> Berbasis Web Untuk Penggalian Kaidah Asosiasi Berdasarkan Algoritma Apriori Menggunakan PHP	Tomi Listiawan	2013	<i>Market Basket Analysis</i> dengan Algoritma Apriori	nilai minimum <i>support</i> berbanding terbalik dengan jumlah aturan yang ditemukan dan waktu komputasi.
2	Implementasi <i>Data Mining</i> Algoritma Apriori pada Sistem Persediaan Alat-alat	Kennedi Tampubolon, Hoga Saragih, Bobby Reza	2013	<i>Market Basket Analysis</i> dengan Algoritma Apriori	Algoritma Apriori memiliki kelemahan karena harus melakukan <i>scan database</i> setiap kali iterasi, sehingga untuk <i>database</i> yang

	Kesehatan				sangat besar membutuhkan waktu yang lama.
3	Aplikasi Data Mining Market Basket Analysis Pada Tabel Data Absensi Elektronik Untuk Mendeteksi Kecurangan Absensi (Check-Lock) Karyawan Di Perusahaan	Gregorius Satia Budhi	2011	<i>Market Basket Analysis</i>	metode <i>market basket analysis</i> dapat pula dimanfaatkan untuk menggali <i>pattern</i> kebiasaan absensi( <i>check-lock</i> ) pegawai sebuah perusahaan
4	Aplikasi Data Mining untuk meneliti Asosiasi Pembelian Item Barang di Supermarket dengan Metode <i>Market Basket Analysis</i>	Andreas Handojo	2004	<i>Market Basket Analysis</i>	semakin kecil minimum <i>support</i> dan <i>confidence</i> yang ditentukan, semakin banyak pula rules yang dapat dihasilkan

## 2.2 Landasan Teori

### 2.2.1 Apotek

Definisi apotek menurut PP 51 Tahun 2009. Apotek merupakan suatu tempat atau terminal distribusi obat perbekalan farmasi yang dikelola oleh apoteker sesuai standar dan etika kefarmasian. Sedangkan menurut Keputusan Menkes RI No.1332/Menkes/SK/X/2002, Apotek merupakan suatu tempat tertentu untuk melakukan pekerjaan kefarmasian dan penyaluran obat kepada masyarakat.

Apotek memiliki tugas dan fungsi sebagai berikut :

1. Tempat pengabdian profesi seorang apoteker yang telah mengucapkan sumpah jabatan.

2. Sarana farmasi untuk melaksanakan peracikan, pengubahan bentuk, pencampuran dan penyerahan obat atau bahan obat.
3. Sarana penyaluran perbekalan farmasi dalam menyebarkan obat - obatan yang diperlukan masyarakat secara luas dan merata.

### 2.2.2 Inventori

Berdasarkan jenis operasi perusahaan, inventori dapat diklasifikasikan menjadi 2 (dua):

1. Pada perusahaan manufaktur yang memproses *input* menjadi *output*, persediaan adalah simpanan bahan baku dan barang setengah jadi (*work in process*) untuk diproses menjadi barang jadi (*finished goods*) yang mempunyai nilai tambah lebih besar secara ekonomis, untuk selanjutnya dijual kepada pihak ketiga (konsumen).
2. Pada perusahaan dagang, persediaan adalah simpanan sejumlah barang jadi yang siap untuk dijual kepada pihak ketiga (konsumen).

Dari kedua jenis persediaan tersebut, titik fokus bahasan adalah persediaan dalam arti untuk perusahaan manufaktur. Walaupun pada beberapa hal lain berlaku pula untuk perusahaan nonmanufaktur.

### 2.2.3 Data Mining

*Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data, dimana proses-nya harus otomatis atau semi-otomatis. Pola-pola yang ditemukan harus berarti dan menghasilkan keuntungan, terutama keuntungan ekonomi. *Data Mining*, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian dan *historis* untuk menemukan keteraturan, pola atau hubungan dalam suatu data berukuran besar [12]. Definisi sederhana dari *data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang

ada di *database* yang besar. Menurut beberapa sumber, *data mining* juga dikenal dengan nama *Knowledge Discovery in Databases* (KDD).

Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor, antara lain :

- a. Pertumbuhan yang cepat dalam kumpulan data.
- b. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang handal.
- c. Adanya peningkatan akses data melalui navigasi web dan intranet.
- d. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
- e. Perkembangan teknologi perangkat lunak untuk data mining.
- f. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Data mining terbagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

a) Deskripsi

Deskripsi pola dan *trend* seringkali memberikan penjelasan yang masuk akal untuk pola dan *trend*. Model *data mining* harus dibuat sejelas (transparan) mungkin, yang berarti hasil dari model *data mining* harus mendeskripsikan pola jelas yang sesuai dengan interpretasi dan penjelasan intuitif. Metode *data mining* tertentu lebih sesuai dari metode lain dalam hal interpretasi transparan. Deskripsi yang berkualitas tinggi seringkali diperoleh melalui *exploratory data analysis*, metode grafis dalam eksplorasi data dalam pencarian pola dan trend.

b) Klasifikasi

Dalam klasifikasi terdapat sebuah target variabel kategori, misalnya *income bracket*, dimana misalnya dapat dipartisi

menjadi 3 kelas atau kategori: *high income*, *middle income*, dan *low income*. Model *data mining* meneliti set *record* dalam jumlah besar, dimana tiap *record* berisi informasi mengenai variabel target serta satu set input. Metode *data mining* yang umum untuk klasifikasi adalah *k-nearest neighbor*, *decision tree*, dan *neural network*.

c) Estimasi

Estimasi hampir sama dengan klasifikasi kecuali bahwa variabel targetnya berupa numerik bukan kategori. Salah satu contoh pekerjaan estimasi adalah mengestimasi GPA dari seorang mahasiswa S2 berdasarkan GPA S1 mahasiswa tersebut. Metode estimasi pada umumnya menggunakan analisis statistik termasuk *point estimation* dan *confidence interval estimation*, *simple linear regression and correlation* dan *multiple regression*.

d) Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi. Perbedaan mendasar yaitu, hasil dari prediksi adalah di masa depan. Contoh dari prediksi adalah memprediksi harga saham selama 3 bulan mendatang. Semua metode dan teknik yang digunakan untuk klasifikasi dan estimasi dapat pula digunakan untuk prediksi dalam situasi yang sesuai.

e) *Clustering*

*Clustering* merupakan pengelompokan *record*, observasi, atau kasus ke dalam kelas-kelas dengan objek yang serupa. Sebuah *cluster* adalah koleksi *record* yang sama satu sama lain, dan tidak sama dengan *record* di *cluster* lain. *Clustering* berbeda dengan *classification* karena tidak ada variabel target dalam *clustering*. *Clustering* tidak mengklasifikasi, estimasi ataupun prediksi nilai dari variabel target. Akan tetapi algoritma *clustering* mencari segmen dari keseluruhan set data ke dalam subgrup yang relatif homogen atau *cluster* di mana keserupaan (*similarity*) *record*

dalam *cluster* adalah maksimal dan keserupaan *record* di luar *cluster* adalah minimal. Contoh *clustering* adalah target pemasaran produk dari bisnis kecil dengan *budget marketing* yang terbatas.

f) Asosiasi

Teknik asosiasi dalam *data mining* adalah teknik untuk menemukan atribut yang muncul bersamaan dalam satu waktu. Dalam dunia bisnis juga dikenal dengan nama *Market Basket Analysis*/Analisis Keranjang Belanja dan digunakan untuk menemukan aturan asosiatif antara suatu kombinasi item atau barang. Asosiasi merupakan sebuah teknik *data mining* yaitu melakukan pencarian atribut mana yang digabungkan bersama. Contoh yang paling umum adalah *affinity analysis* atau *market basket analysis*, yaitu mencari aturan yang tersirat untuk mengkuantifikasi hubungan antara dua atribut atau lebih

#### 2.2.4 Tahapan-tahapan Data Mining

*Data mining*, sering juga disebut sebagai *knowledge discovery in database* (KDD), karena kegiatan yang dilakukan meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Sebenarnya, *Data mining* dan *knowledge discovery in database* memiliki konsep yang berbeda, namun berkaitan. *Data Mining* merupakan salah satu tahapan dalam keseluruhan proses KDD, yang secara garis besar dapat diuraikan sebagai berikut :

a. *Data cleaning* (pembersihan data)

Pada kenyataannya, data yang didapat dari suatu *database* belum tentu memiliki kualitas yang cukup baik. Misalnya data tersebut tidak lengkap atau ada informasi yang hilang, maupun data tidak valid, juga terdapat atribut-atribut data yang tidak relevan



terhadap teknik *data mining* yang digunakan. *Data cleaning* bertujuan untuk membuang data-data yang tidak konsisten, menghilangkan *noise* dan melengkapi data yang kehilangan informasi, sehingga performansi dari *data mining* dapat meningkat.

b. *Data integration* (integrasi data)

Data yang akan diproses dalam *data mining* dapat berasal dari berbagai *database*, dan bukan hanya dari satu *database*. Integrasi data diperlukan untuk menggabungkan data dari berbagai sumber data kedalam satu *database* baru. Integrasi yang teliti dapat mengurangi dan menolak redundansi data, sehingga dapat meningkatkan akurasi dan kecepatan dari proses *data mining*.

c. *Data selection* (pemilihan data)

Sering kali terdapat data yang tidak terpakai dalam *database*. Hanya data yang sesuai untuk dianalisis yang diperlukan. Sebagai contoh, untuk meneliti kebiasaan beli konsumen, tidak perlu mengambil data nama konsumen, cukup dengan id konsumen saja. Dalam kasus *market basket analysis*, kuantitas barang dan harga kurang begitu diperlukan.

d. *Data transformation* (transformasi data)

Pada tahapan ini, data diubah atau ditransformasikan menjadi format data yang sesuai untuk diproses dalam *data mining*, sebab ada metode-metode *data mining* yang memerlukan format data tertentu untuk diolah. Proses mentransformasikan data yang telah dipilih sehingga sesuai untuk *data mining* adalah *coding*. Proses *coding* dalam KDD sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

e. *Data mining* (penggalian data)

Metode dan algoritma yang telah ditentukan mulai diterapkan untuk mencari pola dan menemukan informasi berharga yang

tersembunyi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

f. *Pattern Evaluation* (evaluasi pola)

Pola informasi yang dihasilkan dari proses *data mining* mungkin berbeda dan tidak sesuai dengan hipotesa. Bila hal ini terjadi, hasil tersebut dapat dijadikan umpan balik untuk memperbaiki proses *data mining*. Solusi lain adalah dengan mengubah metode yang digunakan, atau menerima hasil yang ada sebagai pengetahuan baru yang mungkin dapat bermanfaat.

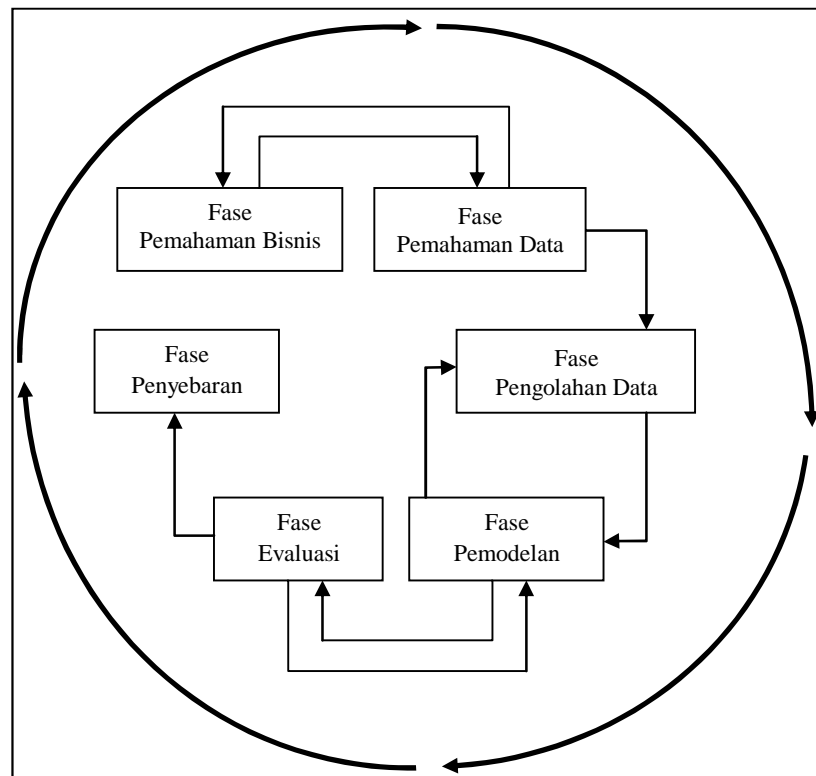
g. *Knowledge presentation*.

*Knowledge presentation* merupakan tahapan akhir dalam proses *data mining*. Bagaimana pengetahuan yang telah ditemukan akan disajikan kepada *user*. Tidak semua *user* memahami *data mining*, karenanya penting untuk menyusun dengan baik penyajian hasil *data mining* dalam bentuk yang dapat dipahami oleh *user*. Dalam hal ini, visualisasi juga dapat digunakan untuk membantu menyampaikan hasil *data mining*.

### 2.2.5 CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM yang dikembangkan pada tahun 1996 oleh para analis dari beberapa industri seperti DaimlerChrysler, SPSS, dan NCR. CRISP menyediakan standar proses *data mining* sebagai strategi pemecahan masalah umum dari bisnis atau unit penelitian.

Dalam CRISP-DM, sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase, dimana keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung pada keluaran dari fase sebelumnya. Hubungan penting antarfase digambarkan dengan panah.



Gambar 2.1 CRISP-DM

Andaikan proses yang sedang berlangsung adalah pada fase pemodelan, tergantung pada sifat dan karakteristik model, proses mungkin dapat kembali ke fase pengolahan data untuk perbaikan lebih lanjut sebelum menuju fase evaluasi.

Enam dalam CRISP-DM adalah :

- a. Fase Pemahaman Bisnis (*Business Understanding Phase*)
  - i. Menentukan tujuan dan kebutuhan secara jelas dalam lingkup bisnis atau unit penelitian secara keseluruhan.
  - ii. Menerjemahkan tujuan-tujuan dan batasan-batasan dari rumusan masalah.
  - iii. Mempersiapkan strategi awal untuk mencapai tujuan.
- b. Fase Pemahaman Data (*Data Understanding Phase*)
  - i. Mengumpulkan data.

- ii. Menggunakan analisis data untuk mengenali data dan menemukan pengetahuan awal.
  - iii. Mengevaluasi kualitas data.
  - iv. Memilih sebagian kecil kumpulan data yang mungkin mengandung pola-pola yang dapat diamati lebih lanjut.
- c. Fase Pengolahan Data (*Data Preparation Phase*)
- i. Mempersiapkan dataset akhir dari data mentah awal yang akan digunakan untuk keseluruhan fase berikutnya, yakni fase pemodelan.
  - ii. Memilih kasus-kasus dan variabel-variabel yang ingin dianalisa dan sesuai dengan tujuan penelitian.
  - iii. Melakukan perubahan pada beberapa variabel jika diperlukan.
  - iv. Membersihkan data sehingga siap untuk perangkat pemodelan.
- d. Fase Pemodelan (*Modeling Phase*)
- i. Memilih dan menerapkan teknik pemodelan yang sesuai dengan kasus yang akan dianalisa.
  - ii. Menyesuaikan pengaturan model untuk mengoptimalkan hasil.
  - iii. Perlu diingat bahwa beberapa teknik mungkin dapat digunakan untuk permasalahan *data mining* yang sama.
  - iv. Jika diperlukan, kembali ke fase pengolahan data untuk mengubah bentuk data sesuai dengan kebutuhan teknik *data mining* tertentu.
- e. Fase Evaluasi (*Evaluation Phase*)
- i. Mengevaluasi satu atau lebih model-model yang digunakan pada fase pemodelan untuk kualitas dan efektifitas sebelum disebarkan untuk digunakan.
  - ii. Menetapkan apakah terdapat model sudah memenuhi tujuan yang ditentukan pada fase pertama.

- iii. Menentukan apakah terdapat permasalahan dari bisnis atau penelitian yang tidak tertangani dengan baik.
  - iv. Menentukan keputusan mengenai penggunaan hasil dari model *data mining*.
- f. Fase Penyebaran (*Deployment Phase*)
- i. Menggunakan model yang dihasilkan : hasil model tidak menandakan selesainya proyek.
  - ii. Contoh sederhana dari penyebaran : Pembuatan laporan.
  - iii. Contoh dari penyebaran yang lebih kompleks : Menerapkan proses *data mining* secara paralel pada perusahaan lain.

### 2.2.6 Association Rule Analysis

Analisis asosiasi atau *association rule mining* adalah teknik data *mining* untuk menemukan aturan *asosiatif* antara suatu kombinasi *item*. Contoh dari aturan *asosiatif* dari analisis pembelian obat di Apotek Sehat adalah mengetahui besarnya kemungkinan seorang pelanggan untuk membeli panadol bersamaan dengan laserin. Analisis asosiasi juga sering disebut dengan istilah *market basket analysis*. 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*) yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan *asosiatif* bisa diketahui menggunakan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi *item* tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar *item* dalam aturan asosiasi.

Salah satu teknik *data mining* yaitu teknik asosiasi sering dikenal juga dengan *Market Basket Analysis* karena sering diaplikasikan untuk menganalisis isi keranjang belanja. Market basket analisis atau analisis keranjang pasar adalah analisis keranjang data

untuk mengidentifikasi kombinasi *item* dengan afinitas yang satu sama lain. Artinya, analisis keranjang pasar bertujuan untuk mengetahui kombinasi dari *item*, kehadiran yang ada dalam suatu transaksi mempengaruhi kemungkinan keberadaan sebuah *item* atau kombinasi *item*. *Market Basket Analysis* merupakan suatu teknik matematis untuk mengungkap hubungan keterikatan antar produk maupun antar kumpulan produk. Sesuai dengan namanya, teknik ini menganalisa keranjang belanja untuk mengetahui kebiasaan beli konsumen. Pola beli konsumen dapat dianalisa melalui suatu transaksi pembelian. Dalam dunia bisnis, *Market Basket Analysis* sering digunakan dalam bidang pemasaran untuk meningkatkan performa perusahaan.

Tujuan dari *Market Basket Analysis* adalah menemukan *frequent itemset* dan membentuk aturan asosiasi dari *frequent itemset* yang ditemukan sesuai dengan parameter minimal *support* dan minimal *confidence* yang telah ditentukan. *Market basket Analysis* dapat mengetahui kumpulan produk yang sering dibeli konsumen. Informasi ini dapat dimanfaatkan oleh pedagang untuk mengatur tata letak produk dalam toko maupun mengatur susunan layout produk dalam katalog. Produk yang kurang laku juga dapat ikut terjual jika dijual bersama dengan produk lain.

Perusahaan dapat lebih efektif menentukan target pemasaran dan strategi pemasaran selanjutnya, sementara konsumen juga mendapatkan kenyamanan karena produk-produk yang sering dibeli bersamaan ditempatkan berdekatan, yang tentu saja akan meningkatkan kepuasan konsumen.

### 2.2.7 Algoritma Apriori

*Apriori* adalah suatu algoritma yang sudah sangat dikenal dalam melakukan pencarian frequent itemset dengan menggunakan teknik association rule. Algoritma Apriori diperkenalkan oleh Agrawal dan Srikant pada tahun 1994. Sampai saat ini algoritma tersebut

merupakan algoritma asosiasi yang telah banyak digunakan dan dikembangkan oleh para peneliti. Algoritma Apriori dalam menangani masalah asosiasi adalah dengan mengurangi jumlah *set item* yang dipertimbangkan.

Algoritma Apriori adalah salah satu algoritma yang dapat digunakan pada penerapan *market basket analysis* untuk mencari aturan-aturan asosiasi yang memenuhi batas *support* dan *confidence*. Algoritma apriori merupakan metode yang efisien untuk mencari aturan kuat yang terkandung dalam satu set transaksi. Selama proses tahap pertama, algoritma menghasilkan penggalan secara sistematis tanpa menjelajahi semua kandidat, sedangkan pada tahap kedua dilakukan ekstraksi terhadap aturan yang kuat. *Frequent itemset* biasanya mengacu pada kumpulan *item* yang sering muncul bersamaan dalam sebuah data transaksional. Contohnya jika *item* A dan B sering dibeli bersamaan dalam suatu toko. Setelah menemukan *frequent itemset*, algoritma ini kemudian meneliti *knowledge* dari *frequent item* sebelumnya untuk menggali informasi selanjutnya. Apriori menggunakan pendekatan iteratif dengan *level-wise search* dimana *k-itemset* dipakai untuk mencari  $(k+1)$ -*itemset*.

Iterasi  $i$  menghitung semua kumpulan data  $i$  (kumpulan yang mengandung elemen  $i$ ) yang sering muncul. Setiap iterasi terdiri dari dua langkah yaitu *candidate generation* (penentuan kandidat) dan *candidate counting and selection* (pemilihan serta penghitungan kandidat).

Pencarian aturan asosiasi harus menggunakan parameter sehingga aturan yang didapat akurat. Parameter yang digunakan untuk pembentukan *rules* yaitu :

a. *Support*

*Support* adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu barang atau itemset dari keseluruhan Transaksi.

$$Support(A) = \frac{ju_{ti} \quad m}{ti \quad m} \quad A \quad (1)$$

Persamaan (1) merupakan rumus umum untuk menghitung nilai *support* suatu item.

Sedangkan rumus untuk menghitung prosentasi *support* dari suatu item ditunjukkan pada persamaan (3), yaitu :

$$Support(A) = \frac{ju_{ti} \quad m}{ti \quad m} \quad A \times 100\% \quad (2)$$

*b. Confidence*

*Confidence* adalah suatu ukuran yang menunjukkan hubungan kondisional antar dua barang (misal seberapa sering obat B dibeli jika orang membeli obat A).

Rumus untuk menghitung nilai *confidence* tersebut yaitu :

$$Confidence(A \rightarrow B) = \frac{ju_{tr} \quad m \quad Ad \quad B}{ju_{ti} \quad m \quad A} \quad (3)$$

Atau,

$$Confidence(A \rightarrow B) = \frac{s \quad (A,B)}{s \quad (A)} \quad (4)$$

Sedangkan rumus untuk menghitung nilai prosentase *confidence* tersebut yaitu :



$$Confidence(A \rightarrow B) = \frac{S_{(A,B)}}{S_{(A)}} \times 100\% \quad (5)$$

Prinsip kerja dasar dari algoritma ini yaitu dengan mengembangkan *frequent itemset*. Mulai dari satu *item* dan secara rekursif mengembangkan *frequent itemset* menjadi dua *item*, tiga *item*, dan seterusnya hingga *frequent itemset* tidak dapat dikembangkan lagi.

Untuk mengembangkan *frequent itemset* dengan dua *item*, dapat menggunakan satu *item*, dengan alasan bila set satu *item* tidak dapat mencapai minimum *support*, maka setiap *itemset* dengan ukuran yang lebih besar juga tidak akan melebihi minimum *support*.

Terdapat dua proses utama dalam algoritma apriori yaitu :

a. *Join* (penggabungan)

Dalam proses ini, setiap *item* dikombinasikan dengan *item* lain sampai tidak dapat terbentuk kombinasi lagi.

b. *Pruning* (pemangkasan)

Pada proses ini, hasil kombinasi *item* akan dipangkas berdasarkan minimum *support* yang telah ditentukan.

Langkah-langkah dari proses algoritma apriori adalah :

1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi (*k-1*)-*itemset* yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma Apriori adalah adanya pemangkasan kandidat *k-itemset* yang subsetnya yang berisi *k-1* *item* tidak termasuk dalam pola frekuensi tinggi dengan panjang *k-1*.
2. Perhitungan *support* dari tiap kandidat *k-itemset*. *Support* dari tiap kandidat *k-itemset* didapat dengan menscan database untuk menghitung jumlah transaksi yang memuat semua *item* di dalam kandidat *k-itemset* tersebut. Ini juga merupakan ciri dari

algoritma apriori dimana diperlukan perhitungan dengan scan seluruh database sebanyak k-itemset terpanjang.

3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k item atau k-itemset ditetapkan dari kandidat k-itemset yang supportnya lebih besar dari minimum support.
4. Bila tidak didapat pola frekuensi tinggi, maka proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian 1.

### 2.2.8 Association Rules

Aturan keterkaitan (*association rules*) merupakan suatu prosedur untuk mencari hubungan antar *item* dalam suatu set data yang ditentukan [5]. Dengan menggunakan teknik pengolahan informasi Data Mining yaitu *Association rule mining* [14] yang digunakan untuk menemukan hubungan antara data atau bagaimana suatu kelompok data yang mempengaruhi suatu keberadaan data yang lain [4]. Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar.

Aturan asosiasi terdiri dari dua himpunan item disebut antecedent dan consequent. Consequent biasanya dibatasi untuk berisi suatu item tunggal. Aturan asosiasi berbentuk “If antecedent, then consequent”, dilengkapi dengan tingkat support dan confidence aturan tersebut. Kuat tidaknya sebuah aturan asosiasi ditentukan oleh dua parameter yaitu support dan confidence.

Berdasarkan penjelasan di atas maka pencarian pola kaidah asosiasi menggunakan dua buah parameter nilai yaitu dukungan (*support*) dan keterpercayaan (*confidence*) yang memiliki nilai antara 0% - 100 %.

*Association Rules* (aturan asosiasi) merupakan sebuah *rule* atau aturan tertentu yang menyatakan hubungan atau korelasi tingkat kemunculan beberapa atribut / *item* dalam sebuah *database*. Aturan ini bersifat satu arah (implikasi). Istilah *Analysis Unit* dan *Associated*

*Unit* dipakai untuk membedakan posisi *item* dalam *association rules*. *Analysis unit* adalah *item* yang menjadi pokok aturan, sedangkan *associated unit* adalah *item* yang dipengaruhi oleh pokok aturan. Aturan asosiasi memberikan informasi dalam bentuk hubungan “*if-then*” atau “jika-maka.” Bentuk umum dari *association rules* yaitu :

$$X_1 \dots X_n \rightarrow Y \quad (6)$$

Yang artinya bahwa, konsumen yang membeli *item* X juga berpeluang membeli *item* Y dengan mengabaikan kuantitas dan harga dari *item-item* tersebut. Jika I adalah kumpulan *item* yang terdapat dalam transaksi dan T adalah setiap transaksi yang memenuhi  $T \subseteq I$ , maka dari setiap transaksi T dikatakan mengandung *item* X jika dan hanya jika  $X \subseteq T$ . Secara tidak langsung ketentuan tersebut mempengaruhi  $X \rightarrow Y$  dimana  $X \subseteq I$  dan  $Y \subseteq I$  dan  $X \cap Y = \emptyset$ . Bentuk  $X \rightarrow Y$  menjelaskan asosiasi antara X dan Y. Contoh dari *association rule* yaitu :

$$\text{panadol} \rightarrow \text{vitacimin, laserin} \quad (7)$$

Dimana berarti bahwa, konsumen yang membeli *panadol* juga cenderung membeli *vitacimin* dan *laserin*, namun bukan berarti bahwa konsumen yang membeli *vitacimin* dan *laserin* juga cenderung membeli *panadol*. Menurut posisi dalam aturan, *panadol* adalah *analysis unit* atau biasa disebut *antecedent*, sedangkan *vitacimin* dan *laserin* adalah *associated unit*, atau biasa disebut dengan *consequent*.

Aturan asosiasi biasanya dinyatakan dalam bentuk :

$$\text{panadol} \rightarrow \text{vitacimin, laserin} \quad (8)$$

*support* = 40% dan *confidence* = 50%

Artinya, 50% dari transaksi di *database* yang memuat *item* *panadol* juga memuat *item* *vitacimin* dan *laserin*. Sementara 40% dari seluruh transaksi yang ada di *database* memuat ketiga *item* tersebut.

Bisa juga diartikan seorang konsumen yang membeli panadol memiliki kemungkinan sebesar 50% untuk juga membeli vitacimin dan laserin. Aturan tersebut cukup signifikan, karena mewakili 40% dari catatan transaksi selama ini. Analisis asosiasi didefinisikan sebagai suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat *minimum* untuk *support* (*minimum support*) dan syarat *minimum* untuk *confidence* (*minimum confidence*).

### **2.2.9 PHP**

PHP (*Hyper Text Preprocessor*) adalah bahasa skrip yang dapat ditanamkan atau disisipkan ke dalam HTML (*Hyper Text Markup Language*). PHP banyak dipakai untuk memrogram situs *web* dinamis. PHP dapat digunakan untuk membangun sebuah CMS. Pada awalnya PHP merupakan kependekan dari *Personal Home Page* (Situs personal). PHP pertama kali dibuat oleh Rasmus Lerdorf pada tahun 1995. Pada waktu itu PHP masih bernama *Form Interpreted* (FI), yang wujudnya berupa sekumpulan skrip yang digunakan untuk mengolah data formulir dari *web*. Selanjutnya Rasmus merilis kode sumber tersebut untuk umum dan menamakannya PHP/FI. Dengan perilsan kode sumber ini menjadi sumber terbuka, maka banyak *programmer* yang tertarik untuk ikut mengembangkan PHP.