

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 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 :

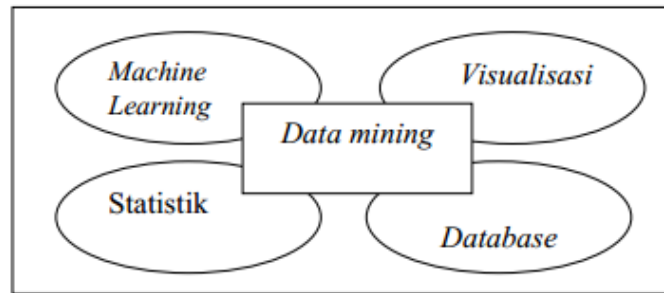
- Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data mining berguna untuk membuat keputusan yang 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, lassification, association rules mining, neural network, genetic algorithm dan lain-lain (Pramudiono, 2007).

## **2.2 Pengenalan Pola, Data Mining, dan Machine Learning**

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohon kredit, image atau signal atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya (Santoso, 2007). Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition jarang digunakan karena termasuk bagian dari data mining (Santoso, 2007).

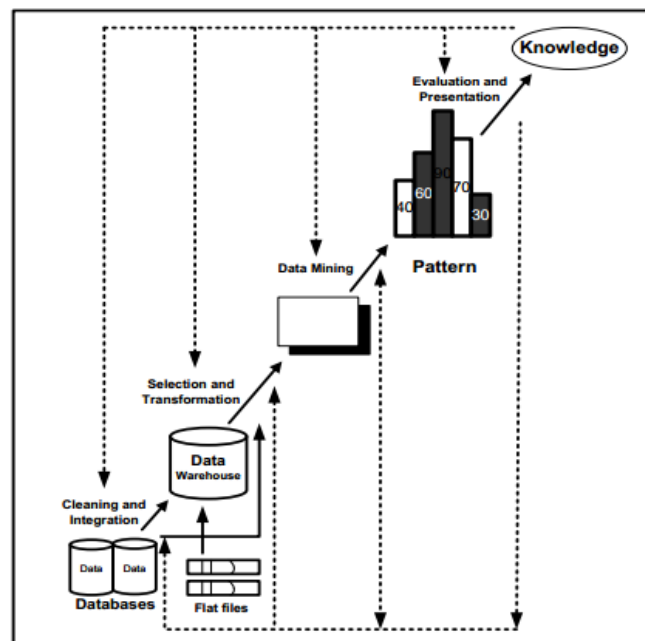
Machine Learning adalah suatu area dalam artificial intelligence atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, data mining dan machine learning sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi. Machine learning menjadi alat analisis dalam data mining. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam gambar 2.4 (Santoso, 2007).



**Gambar 2.1** Hubungan Bidang Ilmu Data mining, Manchine learning dan Pengenalan Pola

### 2.2.1 Tahap-Tahap Data mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 2.5. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base



**Gambar 2.2** Tahap – tahap Data Mining (Han,2006)

Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data 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 performansi 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)

### **2.3 Pengertian Data Warehouse**

Pengertian Data Warehouse dapat bermacam-macam namun mempunyai inti yang sama, seperti pendapat beberapa ahli berikut ini :

Menurut W.H. Inmon dan Richard D.H., data warehouse adalah koleksi data yang mempunyai sifat berorientasi subjek, terintegrasi, time-variant, dan bersifat tetap dari koleksi data dalam mendukung proses pengambilan keputusan management.

Menurut Vidette Poe, data warehouse merupakan database yang bersifat analisis dan read only yang digunakan sebagai fondasi dari sistem penunjang keputusan.

Menurut Paul Lane, data warehouse merupakan database relasional yang didesain lebih kepada query dan analisa dari pada proses transaksi, biasanya mengandung history data dari proses transaksi dan bisa juga data dari sumber lainnya. Data warehouse memisahkan beban kerja analisis dari beban kerja transaksi dan memungkinkan organisasi menggabung/konsolidasi data dari berbagai macam sumber.

Jadi, data warehouse merupakan metode dalam perancangan database, yang menunjang DSS (Decision Support System) dan EIS (Executive Information System). Secara fisik data warehouse adalah database, tapi perancangan data warehouse dan database sangat berbeda. Dalam perancangan database tradisional menggunakan normalisasi, sedangkan pada data warehouse normalisasi bukanlah cara yang terbaik.

Dari definisi-definisi yang dijelaskan tadi, dapat disimpulkan data warehouse adalah database yang saling bereaksi yang dapat digunakan untuk query dan analisis, bersifat orientasi subjek, terintegrasi, time-variant, tidak berubah yang digunakan untuk membantu para pengambil keputusan.

### **2.3.1 Istilah-Istilah dalam Data Warehouse**

Beberapa istilah-istilah yang berhubungan dengan data warehouse antara lain:

#### **1. Decision Support System(DSS)**

Menurut O'Brien(2003, p21), "Decision support system is the new role for *information systems* was provide managerial end-user with *ad hoc and interactive support of their decision-making process*", yang artinya DSS merupakan role baru untuk sistem informasi yang menyediakan sistem manajerial end-user secara harian dan merupakan dukungan interaktif dalam proses pembuatan keputusan. Juga merupakan sistem yang menyediakan informasi kepada user yang menjelaskan bagaimana sistem ini dapat menganalisa situasi dan mendukung suatu keputusan yang baik.

## 2. Data Mart

Menurut Connolly-Begg (2002, p1067) “Data Mart is a subset of data warehouse that support the requirements of a particular department of business function”, yang artinya data mart adalah subset dari data warehouse yang mendukung kebutuhan informasi dari suatu departemen atau fungsi bisnis tertentu. data mart merupakan suatu bagian dari data warehouse yang dapat mendukung pembuatan laporan dan analisis data pada suatu unit, bagian atau operasi perusahaan.

Perbedaan antara data mart dan data warehouse adalah:

- Data Mart hanya berfokus pada kebutuhan user yang berkaitan dengan suatu departemen atau fungsi bisnis.
- Data Mart tidak mengandung data operasional secara detil, tidak seperti *data warehouse*.
- Data yang ada dalam data mart lebih sedikit daripada yang ada dalam *data warehouse*, data mart juga lebih mudah dimengerti karena lebih sederhana.

## 3. OLAP (Online Analytical Processing)

Menurut Mallach (2000, p531), “OLAP is a category of software that enables analyst, managers, and executive to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user”, yang artinya OLAP adalah kategori teknologi software yang dapat memungkinkan penganalisa, manager, dan eksekutif untuk melihat data yang ada dengan akses yang cepat, konsisten dan interaktif sehingga dapat melihat informasi yang sudah di transformasi dari data mentah menjadi dimensi keadaan nyata yang dapat dimengerti dengan mudah oleh user. OLAP juga merupakan suatu pemrosesan database yang menggunakan tabel fakta dan dimensi untuk dapat menampilkan berbagai bentuk laporan, query dari data yang berukuran besar.

#### 4. OLTP (Online Transaction Processing)

Menurut O'Brien (2003, p224), "OLTP is a real time transaction processing systems", yang artinya sebuah Transaction Processing Systems saat ini. "Transaction Processing Systems (TPS) are cross-functional information systems that process data resulting from the occurrence of business transaction", yang artinya TPS adalah sistem informasi yang cross-functional yang memproses data sehari-hari dari sebuah kejadian dari transaksi bisnis. OLTP dirancang untuk memungkinkan terjadinya pengaksesan secara bersamaan oleh beberapa user terhadap sumber data yang sama dan mengatur proses yang diperlukan.

#### 5. Dimensional Table (Tabel Dimensi)

Tabel yang berisikan kategori dengan ringkasan data detil yang dapat dilaporkan seperti laporan keuntungan pada tabel fakta dapat dilaporkan sebagai dimensi waktu (yang berupa perbulan, perkuartal, dan pertahun).

#### 6. Fact Table (Tabel Fakta)

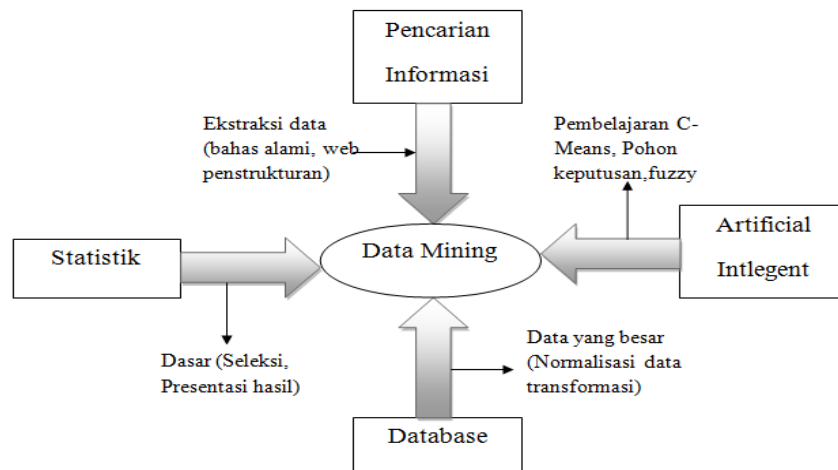
Merupakan tabel yang umumnya mengandung angka dan data historis dimana key (kunci) yang dihasilkan sangat unik karena key nya merupakan kumpulan foreign key dan primary key yang ada pada masing-masing tabel dimensi yang berhubungan atau merupakan tabel terpusat dari skema bintang. Tabel Fakta menyimpan tipe-tipe measure yang berbeda, seperti measure, yang secara langsung terhubung dengan tabel dimensi dan measure yang tidak berhubungan dengan tabel dimensi.

#### 7. Data Mining

Menurut Aldeman (2000, p145), "Data Mining is a discovery process of *unknown unsuspected pattern of data*", yang artinya data mining adalah proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya, sedangkan menurut Hui Tang-MacLennan 17 (2005, p2), "Data mining is about analyzing data and finding hidden pattern using automatic and semi-automatic means", yang artinya data mining adalah penganalisisan data dan penemuan pola yang tersembunyi menggunakan arti yang otomatis dan semi-otomatis. Tujuan utama dari data mining adalah untuk mengekstraksi pola



dari data yang ada, menambah nilai intrinsik dari data serta mengubahnya menjadi pengetahuan. Nama lain data mining adalah *Knowledge discovery from Databases (KDD)*.



**Gambar 2.3** Bidang Ilmu data mining

#### 2.4 Analisis Asosiasi

Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. 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*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan *confidence* (nilai

kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Aturan assosiatif biasanya dinyatakan dalam bentuk : {roti, mentega} \_ {susu} (support = 40%, confidence = 50%). Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu." Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

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*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

#### 2.4.1 Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut:

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

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

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

#### 2.4.2 Pembentukan aturan assosiatif

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

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

Untuk memahami proses algoritma apriori maka berikut ini akan diberikan ilustrasi penggunaan algoritma apriori. Dengan menggunakan database pada Gambar 2.4 dan mengasumsikan minimum support 0,5 atau 2 transaksi .

Sebagai contoh suatu data transaksi yang didapat dari paper penerapan data mining untuk strategi promosi pendidikan (Gunadi Widi Nurcahyo) dengan Tabel 2.1 data transaksi sebagai berikut :

Transaksi 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

**Tabel 2.1** Data Transaksi

Misalkan diinginkan *minimum support* : 50% atau 0,5 (2 dari 4 transaksi)

- ✓ Langkah 1: Mencari nilai *support* untuk masing – masing *itemset*. Seperti yang ditunjukkan pada Tabel 2.2  $L_1 = \{\text{large 1-itemset}\}$

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

**Tabel 2.2** Nilai Support 1 itemset

- ✓ Langkah 2 : Mencari kandidat itemset untuk  $L_2$ , gabungkan itemset pada  $L_1$  (algoritma apriori-gen) { A B, A C, A E, B C, B E, C E}. Item D tidak masuk dalam kombinasi karena nilai support tidak memenuhi minimum support.

- ✓ Langkah 3 : Hitung nilai Support untuk masing- masing itemset. Hasilnya dapat ditunjukkan pada Tabel 2.3.

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

**Tabel 2.3** Nilai Support 2 itemset

- ✓ Langkah 4: tentukan item set yang memenuhi minimum support. Hasilnya dapat ditunjukkan pada tabel 2.4.  $L_2$  { large 2-itemset }

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

**Tabel 2.4** item set yang memenuhi minimum support

- ✓ Langkah 5 : Ulangi langkah 2 - 4
- ✓ Selanjutnya gabungkan itemset pada  $L_1$  &  $L_2$ . Hasilnya dapat ditunjukkan pada Tabel 2.5.

Itemset	Gabungan 3 itemset
A C + B C	ABC
A C + B E	ACB, ACE, ABE
A C + C E	ACE
B C + B E	BCE
B C + C E	BCE
B E + C E	BCE

**Tabel 2.5** Hasil Gabungan 3 itemset

- ✓ Langkah 6 : Hitung support dari setiap kandidat itemset  $L_3$   
Hasilnya dapat ditunjukkan pada Tabel 2.6

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

**Tabel 2.6** Nilai Support 3 itemset

- ✓ Langkah 7 : L3 { large 3-itemset } { B C E} dan 3 itemset yang lainnya dipangkas karena nilai support tidak memenuhi nilai minimum support.
- ✓ Langkah 8 : Untuk mencari aturan asosiasi diperlukan juga minconfidence. Misal minconfidence : 75 %, aturan asosiasi yang mungkin terbentuk dapat ditunjukkan pada Tabel 2.7.

Aturan (X → Y)	Support ( X ∪ Y)	Support 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%	66.67%

**Tabel 2.7** Nilai Confidence Untuk Tiap Itemset

## 2.5 Algoritma Apriori

*Apriori* adalah suatu algoritma yang sudah sangat dikenal dalam melakukan pencarian frequent itemset dengan menggunakan teknik association rule . Algoritma *Apriori* menggunakan knowledge mengenai frequent itemset yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pada algoritma *Apriori* untuk menentukan kandidat-kandidat yang mungkin muncul dengan cara memperhatikan minimum support.

Adapun dua proses utama yang dilakukan dalam Algoritma Apriori, yaitu :

➤ **Join (penggabungan).**

Pada proses ini setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi. Algoritma Apriori dibagi menjadi beberapa tahap yang disebut iterasi atau pass. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari pass pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu.

Di iterasi pertama, support dari setiap item dihitung dengan men-scan database. Setelah support dari setiap item didapat, item yang memiliki support diatas minimum support dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.

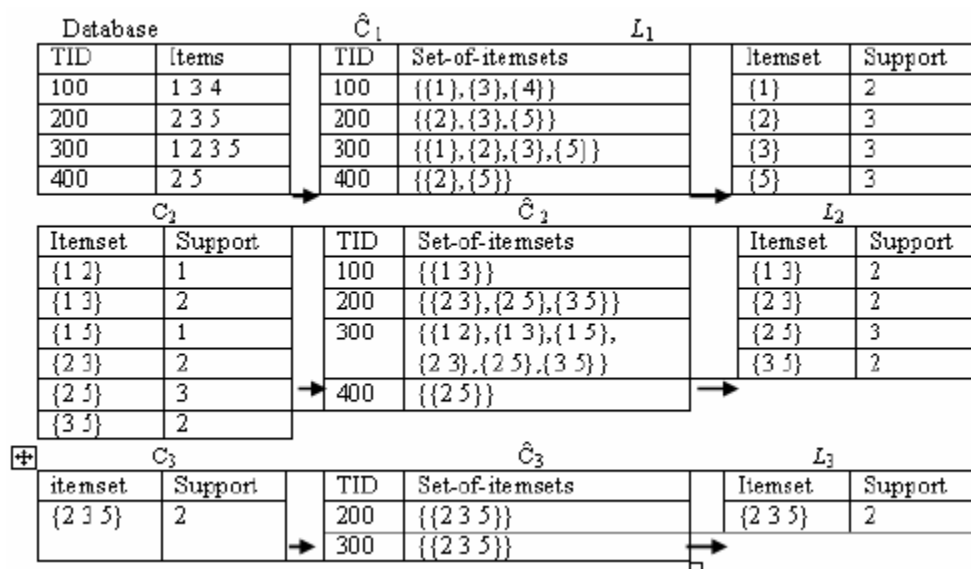
Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung support-nya dengan men-scan database.

➤ **Prune (pemangkasan).**

Pada proses ini, hasil dari item yang telah dikombinasikan tadi lalu dipangkas dengan menggunakan minimum support dan minimum confidence yang telah ditentukan oleh user. Nilai minimum support disini artinya nilai yang digunakan untuk menyeleksi lolos dan tidak lolosnya kandidat itemset. Penilaian untuk minimum confidence yaitu nilai yang digunakan untuk menentukan batas lolos dan tidak lolos aturan asosiasi dari kandidat 2 itemset dan 3 itemset yang direlasikan. Setelah support dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum support dan minimum confidence dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2 item. Untuk selanjutnya pada iterasi ke-k dapat dibagi lagi menjadi beberapa bagian (Sumber : Agrawal , R., Srikant, 1996) :

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 subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.

2. Penghitungan support dari tiap kandidat k-itemset, support dari tiap kandidat k-itemset didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tsb. Ini adalah juga ciri dari algoritma Apriori dimana diperlukan penghitungan 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 support-nya lebih besar dari minimum support.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian 1.



**Gambar 2.4** Ilustrasi Algoritma Apriori

Walaupun algoritma Apriori mudah untuk dipahami dan diimplementasikan dibandingkan dengan algoritma yang lainnya yang memang diterapkan untuk proses association rule, akan tetapi algoritma Apriori juga memiliki kekurangan yaitu, untuk melakukan pencarian frequent itemset, algoritma Apriori harus melakukan scanning database berulang kali untuk setiap kombinasi item.

Hal tersebut menyebabkan banyaknya waktu yang dibutuhkan untuk melakukan scanning database. Selain itu, dibutuhkan generate Analisis Market

Basket Dengan Algoritma Apriori dan FP-Growth Erwin candidate yang besar untuk mendapatkan kombinasi item dari database.

## 2.6 DBMS (Database Management System) dan MySQL

DBMS merupakan suatu sistem perangkat lunak yang memungkinkan pengguna (user) untuk membuat, memelihara, mengontrol, dan mengakses database secara praktis dan efisien. Hampir semua DBMS dapat diintegrasikan dengan web application salah satunya adalah MySQL. Konektivitas MySQL dengan web application sangat baik, sehingga lebih sering digunakan dalam aplikasi-aplikasi web application dari pada DBMS lainnya.

MySQL adalah DBMS yang didistribusikan secara gratis dibawah lisensi dari General Public License (GPL) dimana setiap orang bebas untuk menggunakannya tetapi tidak boleh menggunakan MySQL untuk dijadikan induk turunan yang bersifat close source (komersial). MySQL merupakan turunan dari salah satu konsep utama dalam database yaitu SQL (Structured Query Language).

Keunggulan dari MySQL adalah sebagai berikut :

### 1. Portability

Dapat berjalan stabil pada berbagai system operasi, diantaranya Windows, Linux, FreeBSD, Mac OS X Server, Solaris, Asigma.

### 2. Open source

Didistribusikan secara gratis dibawah lisesnsi dari General Public License (GPL) dimana setiap orang bebas untuk menggunakannya tetapi tidak boleh menggunakan MySQL untuk dijadikan induk turunan yang bersifat close source (komersil)

### 3. Multi user

Dapat digunakan oleh beberapa user dalam waktu yang bersamaan.

### 4. Performance Tuning

Mempunyai kecepatan yang tinggi dalam menangani query.

### 5. Column types

Memiliki tipe kolom yang sangat kompleks, seperti : signed/unsigned integer, float, double, char, varchar, text, blob, date, time, datetime, timestamp, year dan enum.



6. Command dan function

Memiliki operator dan fungsi secara penuh yang mendukung select dan where dalam sebuah query.

7. Security

Memiliki beberapa lapisan sekuritas seperti tingkat subnetmask, hostname, privilege user dengan system perijinan yang mendetail serta password yang terenkripsi.

8. Scalability dan limits

Mampu menangani basis data dalam jumlah besar, dengan jumlah field lebih dari 50 juta, 60 ribu table dan 5 milyar record. Batas indeks yang dimiliki mencapai 32 buah per table.

9. Localication

Dapat mendeteksi pesan kesalahan (error code) pada klien dengan menggunakan lebih dari 20 bahasa.

10. Connnectivity

Dapat dilakukan koneksi dengan klien menggunakan protocol TCP/IP, Unix Socket, Named Pipes.

11. Interface

Memiliki antarmuka terhadap berbagai aplikasi dan bahasa pemrograman dengan menggunakan fungsi API.

12. Clients dan tools

Dilengkapi dengan berbagai tool yang dapat digunakan untuk administrasi basis data sekaligus dokumen petunjuk *online*.

13. Struktur table

Memiliki struktur table yang lebih fleksibel dalam menangani alter table dibandingkan dengan PostgreSQL dan Oracle.

### **2.6.1 SQL (*Structured Query Language*)**

MySQL sebenarnya merupakan turunan dari salah satu konsep utama dalam basis data sejak lama, yaitu SQL (Structured Query Language). SQL adalah sebuah konsep pengoperasian basis data terutama untuk proses seleksi,

pemasukan, perubahan dan penghapusan data yang dimungkinkan dapat dikerjakan dengan mudah dan otomatis.

Didalam SQL terdapat 3 sub bahasa, yaitu: Data Definition Language (DDL) yang digunakan untuk membangun objek – objek dalam database seperti tabel dan indeks; Data manipulation Language (DML) yang digunakan untuk menambah, mencari, mengubah, dan menghapus aris dalam tabel; dan Data Control Language (DCL) yang digunakan untuk menangani masalah sekuriti dalam database. Ketiga sub bahasa ini dapat kita akses setelah database dipanggil.

#### **a. DDL (Data Definition Language)**

DDL adalah perintah – perintah yang digunakan untuk membangun isi dari database. DDL bertugas untuk membuat objek SQL dan menyimpan definisinya dalam tabel. Contoh dari objek yang dimaksud adalah tabel, view, dan index. Pembuatan tabel, perubahan struktur tabel, perubahan nama tabel, serta perintah untuk menghapus tabel, dilakukan dengan sub bahasa yang tergolong dalam DDL Perintah – perintah yang digolongkan dalam DDL adalah :

1. **Create** : Perintah ini digunakan untuk membuat databs, tabel, dan objek lain dalam database
2. **Alter** : Perintah ini digunakan untuk memodifikasi tabel, seperti mengganti nama tabel, mengubah jenis/tipe field yang digunakan, mengubah nama field baru atau mengurangi field tertentu
3. **Drop** : Perintah ini digunakan untuk menghapus database, tabel dan objek lain dalam dabase.

#### **b. DML (Data Manipulation Language)**

DML digunakan untuk menampilkan, mengubah, menambah dan menghapus baris dalam tabel. Perintah – perintah yang dihubungkan digolongkan dalam DML antara lain:

1. **Select** : Perintah ini digunakan untuk menampilkan isi tabel baik semua isi tabel atau sebagian isi tabel. Bahkan bisa pula menampilkan isi tabel yang dihubungkan dengan tabel – tabel yang lain.

2. **Insert** : Perintah ini untuk mengisi atau menambahkan isi tabel. Pengisian dapat dilakukan untuk seluruh field atau hanya sebagian field saja.
3. **Update** : Seperti terlihat pada perintahnya. Perintah ini digunakan untuk mengubah item – item tertentu dalam satu tabel.
4. **Delete** : Perintah ini digunakan untuk menghapus record – record dengan kriteria tertentu.

## 2.7 Pengertian Puskesmas

Puskesmas adalah organisasi kesehatan fungsional yang merupakan pusat pengembangan kesehatan masyarakat yang juga membina peran serta masyarakat dan memberikan pelayanan secara menyeluruh dan terpadu kepada masyarakat di wilayah kerjanya dalam bentuk kegiatan pokok (Depkes RI, 1991).

Dengan kata lain puskesmas mempunyai wewenang dan tanggungjawab atas pemeliharaan kesehatan masyarakat dalam wilayah kerjanya.

Menurut Kepmenkes RI No. 128/Menkes/SK/II/2004 puskesmas merupakan Unit Pelayanan Teknis Dinas kesehatan kabupaten/kota yang bertanggung jawab menyelenggarakan pembangunan kesehatan di suatu wilayah kerja.

Pembangunan Kesehatan merupakan program unggulan diantara 45 program penting yang harus dijalankan diseluruh tanah air berkaitan dengan pembangunan sektoral dan regional, dimana program kesehatan masuk di urutan ke 12. Untuk Kemenkes telah menetapkan rencana jangka menengah 2010 -2014 yang disusun dalam sebuah rencana strategis dengan mengangkat 4 isu penting yaitu :

1. Peningkatan Pembiayaan kesehatan untuk memberikan jaminan kesehatan masyarakat.
2. Peningkatan Kesehatan Masyarakat untuk mempercepat pencapaian target MDGs.
3. Peningkatan Ketersediaan, Pemerataan, dan kualitas tenaga kesehatan terutama di daerah terpencil, tertinggal, perbatasan dan kepulauan (DTPK).
4. Pengendalian Penyakit dan penanggulangan masalah kesehatan akibat bencana.

Pembangunan Kesehatan telah mencapai berbagai macam kemajuan, namun saat ini kita tengah menghadapi tantangan baru, yaitu meningkatnya masalah *penyakit tidak menular*. Data Riskesdas 2010 menunjukkan 59% kematian di Indonesia disebabkan penyakit tidak menular, yang membutuhkan biaya pengobatan sangat besar seperti stroke, kanker, diabetes, gagal ginjal, dan penyakit jantung. Penyakit tidak menular dapat dicegah dengan gaya hidup sehat yaitu : pola makan, gizi seimbang, memantau berat badan secara teratur, melakukan olah raga teratur, mengendalikan stress, dan tidak merokok.

Pemerintah berupaya meningkatkan akses masyarakat terhadap pelayanan kesehatan yang bermutu dan menitik beratkan upaya promotif-preventif dengan tetap memperhatikan upaya upaya kuratif-rehabilitatif. Untuk meningkatkan akses dan kualitas pelayanan, maka jangkauan program jaminan kesehatan masyarakat diperluas, sedangkan pelayanan kesehatan ibu dan anak diperkuat dengan Program jaminan persalinan atau program jampersa; sejak tahun 2011,.

Sejak tahun 2011, pemerintah menyediakan dana bantuan operasional kesehatan (BOK) yang berkisar 75 juta sampai 250 juta per puskesmas per tahun. Dana BOK dimaksudkan untuk mendukung kegiatan operasional puskesmas, termasuk pembinaan posyandu, percepatan penurunan angka kematian ibu dan anak, pemulihan gizi kurang dan gizi buruk di masyarakat.

Sesuai amanat Undang Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN) dan Undang Undang Nomor 24 Tahun 2011 tentang Badan Penyelenggaraan Jaminan Sosial (BPJS), pada saat ini pemerintah sedang melakukan penyiapan pelaksanaan SJSN bidang kesehatan untuk mewujudkan jaminan kesehatan semesta atau universal health coverage. Pelaksanaan SJSN bidang kesehatan tersebut dimulai pada tahun 2014 secara bertahap ditingkatkan cakupan kepesertaan hingga mencapai jaminan kesehatan semesta pada tahun 2019. Dengan Jaminan Kesehatan Semesta 2019 ini diharapkan seluruh penduduk Indonesia mempunyai jaminan kesehatan sehingga dapat memperoleh pelayanan kesehatan yang komprehensif dan bermutu tanpa kendala pembiayaan. Pelayanan kesehatan ini mencakup aspek preventif dan kuratif rehabilitative.

Dalam pembangunan kesehatan, upaya promotif-preventif harus diutamakan karena selain akan menurunkan jumlah orang yang sakit, juga berdampak pada efisiensi biaya kesehatan. Upaya promotif-preventif juga mencakup upaya perilaku hidup bersih dan sehat, sebagai perilaku sehari-hari dan menciptakan kemandirian masyarakat untuk hidup sehat dengan menjaga dirinya agar tetap sehat.

Untuk dapat melaksanakan pembangunan kesehatan tersebut, perlu dukungan, peranserta dan komitmen seluruh jajaran kesehatan dan juga peran serta dari pihak terkait serta seluruh lapisan masyarakat. Oleh karena itu kerjasama dari semua pihak sangat diharapkan untuk mensukseskan pembangunan kesehatan di Indonesia pada umumnya dan di Kabupaten 'X' pada khususnya.

## **2.8 Riset Sebelumnya**

Sebagai bahan perbandingan dalam penilitan Analisis asosiasi menggunakan metode algoritma apriori diantaranya dikemukakan oleh Rohman Dijaya (2013) Dengan judul **APLIKASI ANALISA POLA DIAGNOSA PENYAKIT PADA REKAM MEDIS ELEKTRONIK DENGAN METODE APRIORI** Pembentukan Confidence pada asosiasi dari pasangan obat dapat dibentuk dengan metode apriori melalui query mysql yang dapat membentuk pasangan obat 2 maupun 3 item set. Pasangan obat yang sering muncul secara bersamaan adalah pasangan obat 85, 888 dan 1720 atau Gliseril Guayakalot tablet 100 mg, Klorfen M 4 mg dan Pamset Tab 500 mg. Hasil keluaran (output) perangkat lunak berupa kaidah asosiasi yang didasarkan pada nilai minimum support, yaitu nilai prosentase minimal yang menunjukkan seberapa besar tingkat dominasi suatu obat (item) atau kumpulan obat (item set) terhadap keseluruhan data data transaksi yang dianalisa dari nilai minimum confidence.

Penelitian selanjutnya oleh Muchammad Iljas (2011) dengan judul **RANCANG BANGUN PERANGKAT LUNAK ANALISA KERANJANG PASAR DENGAN METODE APRIORI**. Penggunaan SQL Query dapat langsung melakukan pembentukan 2-itemset dari data transaksi vertical tanpa melalui Candidate 1 sehingga metode apriori dapat dipangkas pada proses pembentukan 2-itemset dari Candidate 1. Penggunaan SQL query dapat melakukan perhitungan

nilai support dari 3-itemset dengan menggunakan relasi terhadap tabel data 3-itemset yang memenuhi minimum support dengan tiga duplikat tabel data transaksi *order\_detail*. Penggunaan SQL query dapat melakukan pembentukan kaidah asosiasi dengan teknik permutasi, serta dapat menghitung nilai confidence dari kaidah asosiasi. Hasil keluaran (*ouput*) perangkat lunak berupa informasi kaidah asosiasi yang didasarkan pada nilai *minimum support*, yaitu nilai prosentase minimal yang menunjukkan seberapa besar tingkat dominasi suatu barang (*item*) atau kumpulan barang (*itemset*) terhadap keseluruhan data transaksi yang dianalisa dan nilai *minimum confidence*, yaitu nilai prosentase minimal yang menunjukkan hubungan dari kaidah asosiasi.

Berdasarkan kedua sumber penelitian tersebut di atas, tampak bahwa beberapa peneliti yang menggunakan analisa dengan metode apriori terlihat Hasil keluaran (*ouput*) perangkat lunak berupa informasi kaidah asosiasi yang didasarkan pada nilai minimum support, yaitu nilai prosentase minimal yang menunjukkan seberapa besar tingkat dominasi suatu barang (*item*) atau kumpulan barang (*itemset*) terhadap keseluruhan data transaksi yang dianalisa dan nilai minimum confidence, yaitu nilai prosentase minimal yang menunjukkan hubungan dari kaidah asosiasi dapat menemukan suatu relasi istimewa pada data set yang dianalisa dan dapat memberikan informasi baru . Algoritma apriori dapat diterapkan pada data set puskesmas kabupaten 'X' untuk mendapatkan informasi baru yang belum didapatkan sebelumnya seperti 2 penelitian sebelumnya.