

BAB II

LANDASAN TEORI

Landasan teori merupakan hal yang cukup penting dalam suatu kerangka kerja dari penelitian. Dengan memanfaatkan teori-teori, metode-metode yang ada dan melakukan pengkajian dari penelitian yang pernah dilakukan akan lebih memudahkan langkah pengerjaan penelitian. Dibawah ini landasan teori pendukung dari penelitian yang dilakukan.

2.1 *Funding*

Berdasarkan undang-undang perbankan Republik Indonesia (RI) No. 10 tertanggal 10 November 1998, pengertian bank adalah *"badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak"*.

Funding merupakan kegiatan penghimpunan dana pihak ketiga yang dilakukan oleh pihak bank. Kegiatan penghimpunan dana pihak ketiga ini memerlukan suatu usaha dan strategi, sehingga masyarakat mau menyimpan atau mempercayakan dananya kepada pihak bank. Ada beberapa cara untuk menghimpun dana masyarakat yaitu dengan menawarkan berbagai jenis produk seperti rekening tabungan, giro, dan deposito (Kasmir, 2000).

Secara sederhana dalam kehidupan sehari-hari kata simpanan dapat juga diartikan dengan nama rekening atau *account*. Seseorang, perusahaan atau badan

hukum yang memiliki simpanan atau rekening berarti memiliki sejumlah uang atau dana yang dipercayakan untuk dititipkan di bank.

2.1.1 Simpanan Giro (*Demand Deposit*)

Pengertian giro adalah simpanan yang penarikannya dapat dilakukan setiap saat dengan menggunakan cek, bilyet, giro atau sarana perintah pembayaran lainnya, atau dengan pemindahbukuan (Undang-undang perbankan RI, 1998).

Penjelasan mengenai simpanan yang penarikannya dapat dilakukan setiap saat maksudnya adalah nasabah dapat melakukan penarikan dana yang disimpan dalam rekening gironya berkali-kali dalam jangka waktu satu hari dengan persyaratan dana dalam giro masih tersedia, dan memenuhi persyaratan lain yang telah ditentukan oleh bank. Sedangkan pengertian dari penarikan adalah pengambilan sejumlah dana/uang yang mengakibatkan berkurang jumlah dana/uang yang ada pada rekening giro tersebut. Ada beberapa cara dalam penarikan yaitu dengan menggunakan cek untuk penarikan tunai dan bilyet untuk penarikan non tunai. (Kasmir, 2000), (Hasibuan, 2006).

Transaksi dalam rekening giro terbagi dua, pertama transaksi penyetoran/kredit dimana dengan adanya penyetoran ini akan menambah jumlah saldo nasabah pada rekening gironya. Kedua transaksi penarikan/debet yaitu transaksi yang mengakibatkan berkurangnya sejumlah saldo pada rekening giro nasabah.

Ada beberapa jenis kepemilikan rekening giro diantaranya rekening perorangan atau atas nama pribadi dan rekening atas nama suatu badan seperti

badan usaha yang merupakan badan hukum (PT, Koperasi, Yayasan) dan bukan badan hukum (CV, Firma).

Dibawah ini produk simpanan giro yang dimiliki oleh PT. Bank X:

- *Giro X Rupiah*

Produk ini ditujukan pada nasabah sebagai pilihan yang tepat dalam menjalankan bisnisnya yang memerlukan rekening giro dalam mata uang rupiah, fasilitas yang diberikan dalam penarikan dana menggunakan *cek, bilyet giro*, surat perintah pembayaran atau pemindahbukuan. Produk ini juga dilengkapi dengan layanan *ATM, Internet banking*, dan *phone banking*.

- *Giro X Valas*

Perbedaan produk Giro X Valas dengan Giro X Rupiah adalah penggunaan mata uang asing. Didalam produk ini transaksi dapat berupa jual beli valuta asing. Dimana pihak bank menjual valuta asing dan menerima *counter value* dari pihak nasabah dalam bentuk rupiah atau mata uang asing lainnya.

- *Syariah Giro X*

Produk syariah giro yang dimiliki oleh PT. Bank X ada dua yaitu: giro dengan bentuk mata uang rupiah dan mata uang asing. Pengelolaan produk syariah dikelola oleh manajemen khusus yang disesuaikan dengan undang-undang perbankan syariah yang telah ditetapkan oleh pemerintah Republik Indonesia, namun manajemen tersebut masih berada dalam bagian organisasi PT. Bank X.

Nasabah dinyatakan memiliki rekening giro aktif bila saldo dalam rekening gironya tidak bernilai negatif dalam jangka waktu yang telah ditentukan, dan dalam kondisi masih adanya aktifitas transaksi yang dilakukan oleh nasabah selama jangka waktu tertentu.

2.1.2 Simpanan Tabungan (*Saving Deposit*)

Tabungan adalah simpanan yang penarikannya dapat dilakukan menurut syarat-syarat tertentu yang disepakati, tetapi tidak dapat ditarik dengan cek, bilyet giro dan atau alat lainnya yang dipersamakan dengan itu (Undang-undang perbankan RI, 1998).

Prinsip transaksi tabungan ada dua yaitu: debit dan kredit. Kepemilikan rekening tabungan bersifat individual atau atas nama perorangan. Dibawah ini produk simpanan tabungan yang dimiliki oleh PT. Bank X:

- *Tabungan X Rupiah*

Ada dua jenis tabungan X rupiah yaitu: tabungan berbasis buku tabungan (*passbook*) dan tabungan berbasis *statement* dengan fasilitas yang lebih leluasa karena memiliki batas transaksi lebih tinggi daripada *passbook*. Tabungan berbasis *statement* tidak perlu repot dengan buku tabungan, karena laporan keuangan (*statement*) akan dikirimkan setiap bulan langsung ke alamat.

- *Tabungan X Foreign Currency*

Produk tabungan *foreign currency* diterbitkan PT. Bank X untuk memenuhi kebutuhan masyarakat yang ingin menabung dalam bentuk mata uang asing. Sampai saat ini rekening tabungan dalam bentuk mata uang asing tersedia dalam 3 jenis pilihan mata uang asing, yaitu: Dollar America (USD), Dollar Singapore (SGD), dan EURO (EUR).

- Syariah tabungan X

Produk syariah tabungan diterbitkan PT. Bank X, untuk pemenuhan keinginan masyarakat yang ingin menabung dengan prinsip syariah. Produk syariah tabungan dikelola oleh manajemen khusus yang disesuaikan dengan

undang-undang perbankan syariah yang telah ditetapkan oleh pemerintah Republik Indonesia, namun manajemen tersebut masih berada dalam bagian organisasi PT. Bank X.

Semua jenis rekening tabungan yang diterbitkan oleh PT. Bank X mempunyai fasilitas layanan *ATM*, *Internet banking* dan *phone banking*.

2.1.3 Simpanan Deposito (*Time Deposit*)

Simpanan deposito memiliki perbedaan dengan dua jenis simpanan sebelumnya, pada simpanan deposito jangka waktu penarikan saldo uang dapat ditarik setelah jatuh tempo. Pengertian dari jatuh tempo disini adalah masa berakhirnya simpanan deposito.

Deposito adalah simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah dengan bank (Undang-undang perbankan RI, 1998).

Dibawah ini produk simpanan deposito yang dimiliki oleh PT. Bank X:

- *Deposito X berjangka Rupiah*

Deposito berjangka ini diterbitkan dengan jangka waktu yang bervariasi antara 1, 3, 6, 12, dan 24 bulan. Deposito berjangka diterbitkan atas nama, baik perorangan maupun lembaga. Simpanan deposito ini dalam bentuk mata uang rupiah.

- *Deposito X berjangka Valas*

Perbedaan produk ini adalah dengan bentuk mata uang asing. Jangka waktu yang disediakan antara 1, 3, 6, 12, dan 24 bulan.

- *Deposito X On Call*

Simpanan ini dapat berupa rupiah dan dollar Amerika. Produk ini merupakan penyimpanan dana sekitar 1 minggu, berbeda dengan deposito lainnya penarikan deposito ini dapat dilakukan kapan saja dengan minimal selama 3 hari dana telah berada di bank.

- Syariah deposito X

Penarikan dana yang disimpan di produk syariah deposito x ini sama dengan deposito lainnya. Produk syariah deposito x yang diterbitkan PT. Bank X ada dua, yaitu: syariah deposito x dengan bentuk mata uang rupiah dan mata uang asing. Pengelolaan dana yang disimpan di produk syariah deposito ini dikelola oleh manajemen khusus yang disesuaikan dengan undang-undang perbankan syariah yang telah ditetapkan oleh pemerintah Republik Indonesia, namun manajemen tersebut masih berada dalam bagian organisasi PT. Bank X.

2.1.4 Loyalitas Nasabah

Ketatnya persaingan di industri perbankan khususnya di bidang *funding* dengan tingkat suku bunga yang cenderung hampir sama dari tiap kategori produk *funding*, membuat persaingan dalam penghimpunan dana pihak ketiga menjadi lebih ketat lagi, tidak hanya antar bank tetapi juga persaingan antara jenis produk *funding* perbankan.

Di bidang *funding* perbankan nasabah, dikatakan sebagai nasabah yang loyal bila memiliki kepercayaan untuk menyimpan dananya terhadap suatu produk *funding* dan menggunakan produk tersebut dalam transaksi keseharian nasabah sesuai fasilitas dan manfaat yang diberikan produk dalam interval waktu yang

cukup (lebih dari satu tahun). Indikator loyalitas ini dapat diukur dari jumlah dana yang disimpan oleh nasabah, dimana semakin besar dana yang disimpan dalam interval waktu yang semakin lama, menandakan semakin tinggi tingkat loyalitas nasabah (Vibiznews, 2007), (Markplusinc, 2007).

Tingkat loyalitas nasabah menjadi faktor yang sangat penting bagi kelangsungan sebuah produk *funding* bank karena nasabah yang loyal akan aktif berpromosi dan memberikan rekomendasi kepada pihak lain, menjadikan produk sebagai pilihan utama dan tidak mudah untuk berpindah ke kompetitor lain. Kotler, Hayes dan Bloom (2002) menyebutkan ada beberapa alasan mengapa suatu bank perlu mendapatkan loyalitas nasabah:

1. Nasabah yang loyal memberikan keuntungan besar bagi bank
2. Biaya yang dikeluarkan untuk mendapatkan nasabah baru jauh lebih besar dibandingkan mempertahankan nasabah yang ada
3. Nasabah loyal akan senantiasa membela bank dan bahkan mampu menarik orang lain untuk menjadi nasabah

Untuk mendapatkan dan mempertahankan nasabah yang loyal tidaklah mudah, diperlukan strategi yang tepat. Langkah awal untuk mempertahankan nasabah loyal adalah dengan melakukan identifikasi terhadap seluruh nasabah yang dimiliki oleh pihak bank khususnya nasabah *funding* bank. Dimana dimensi yang paling menentukan tingkat loyalitas nasabah terhadap produk *funding* perbankan adalah *transaction* (kebiasaan transaksi) dari nasabah (Pqm, 2007), (Subkhan, et.al, 2007).

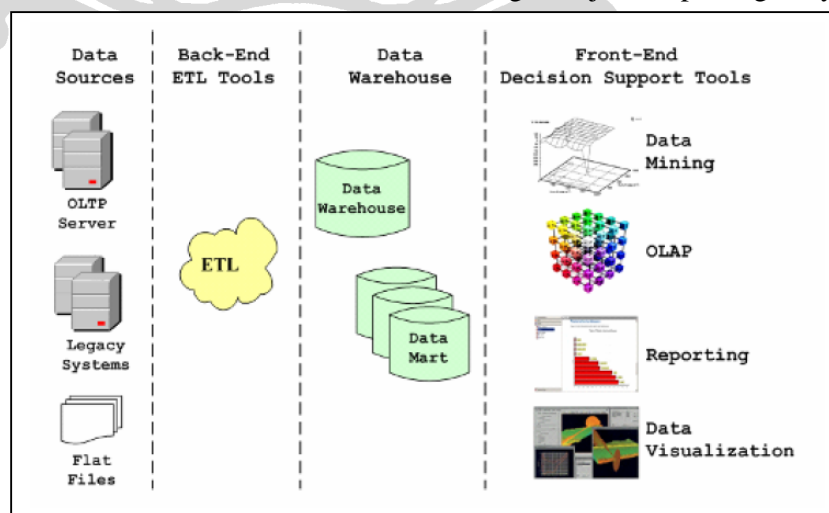
2.2 Data Warehouse

Perkembangan *data warehouse* saat ini telah dipahami peranannya sebagai sumber informasi dalam membuat keputusan dan berkaitan erat dengan *Decisions Support Systems (DSS)* (Wells, 2000). Data yang terdapat di *data warehouse* juga dapat dimanfaatkan dalam penyusunan perencanaan dan pengambilan keputusan strategis dengan menggunakan teknologi seperti: *On-Line Analytical Processing (OLAP)* dan *data mining* (Kontio, 2006).

2.2.1 Arsitektur Data Warehouse

Agar didapat pemahaman yang cukup jelas mengenai arsitektur *data warehouse* langkah awal adalah dengan mengetahui apa definisi dari *data warehouse*. *Data warehouse* merupakan himpunan data yang berorientasi pada subjek, terintegrasi, *non-volatile* yang dapat mendukung manajemen untuk mengambil kebijakan (Inmon, 1993).

Bentuk dari arsitektur *data warehouse* ditunjukkan oleh gambar 2.1, dalam gambar tersebut arsitektur *data warehouse* terbagi menjadi empat bagian, yaitu:



Gambar 2.1 Arsitektur *data warehouse* (Simistsis, Theodoratos, 2006).

1. *Data Sources*

Data source diperoleh dari berbagai data transaksi dan produksi perusahaan hasil operasional aplikasi yang dijalankan tiap harinya. Data ini bisa tersimpan oleh *tools database* dan format struktur data yang mungkin berbeda, untuk itu diperlukan suatu bentuk pemodelan untuk pemahaman sesuai kebutuhan bisnis yang nantinya dilanjutkan ke proses berikutnya.

2. *Extraction, Transformation and Loading (ETL)*

Bagian ini terdiri dari tiga proses yang digabungkan menjadi satu. Proses *extraction* adalah proses yang dikerjakan untuk mendapatkan data dari sumber data. Proses *transformation* adalah proses perubahan format struktur *field* data dari format sebelumnya. Proses *loading* adalah proses memasukkan data yang telah mengalami proses *extraction* dan *transformation* ke dalam tabel target yang sudah dibuat.

Di proses *ETL* juga dilakukan pengerjaan seperti: pembersihan data, pengecekan duplikasi data, standarisasi data, dan penamaan konvensi data sesuai kebutuhan *data warehouse*. Dimana tingkat kualitas *data warehouse* yang dihasilkan sangat ditentukan diproses ini (Winkler, 2006). Proses pengerjaan *ETL* biasanya dibantu oleh *tools* yang merupakan bagian dari arsitektur *data warehouse* (Simistis, Theodoratos, 2006).

3. *Data Storage/Data Warehouse.*

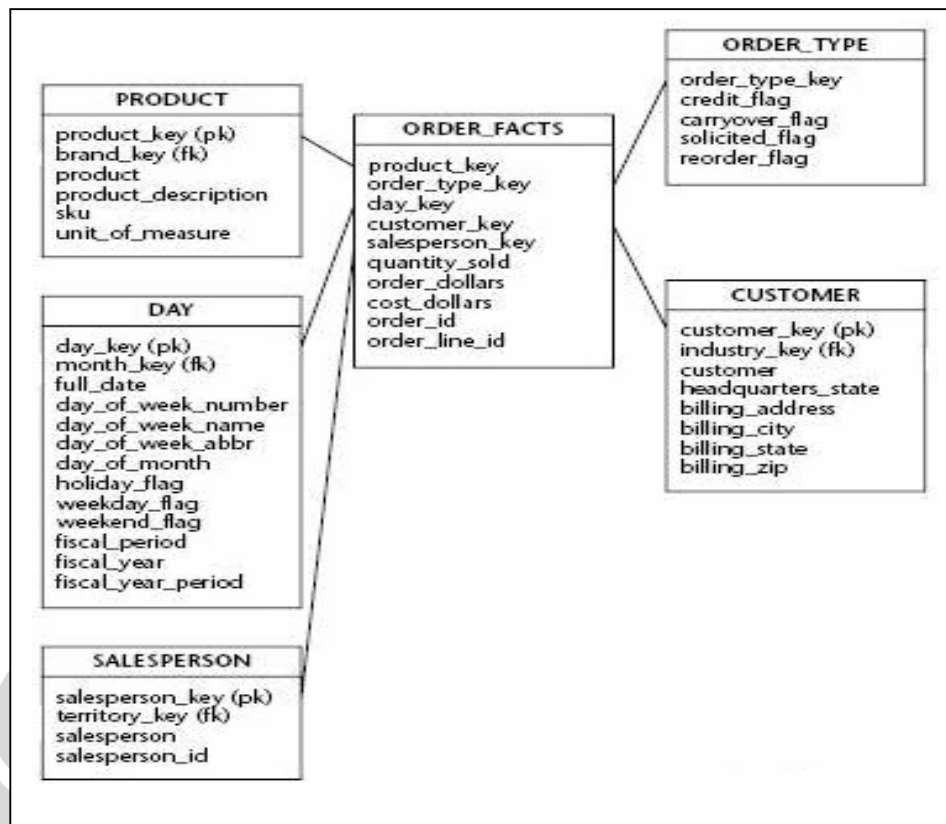
Data warehouse sudah menjadi bagian terakhir pembebasan data, dimana data yang berada di *data warehouse* siap untuk digunakan untuk pemenuhan kebutuhan dalam mendukung pengambilan kebijakan dan keputusan strategis. *Data warehouse* bisa terdiri dari kumpulan beberapa *data mart* yang bersifat *dependence mart* dan *independence mart*.

4. *Front-End Decision Support Tools*

Front-End Decision Support Tools adalah bagian pemanfaatan dari *data warehouse* yang telah dibuat, antara lainnya menggunakan teknologi *OLAP* dan *data mining*.

2.2.2 *Skema Data Warehouse*

Tabel-tabel didalam *data warehouse* dirancang menggunakan pemodelan dimensi atau yang biasa disebut *star schema*. Adamson, 2006 meyakini dengan pemodelan *star schema* dapat meningkatkan kinerja *data warehouse* yang dibuat. Komponen *star schema* di *data warehouse* dideskripsikan dengan menggunakan teknik pemodelan tabel fakta dan tabel dimensi. Keterkaitan pemodelan antara tabel fakta dan tabel dimensi hampir mirip dengan penggunaan konsep dari *Entity-Relationship (ER)* dengan memberikan beberapa pembatasan. Setiap model *star schema* terdiri dari satu tabel fakta dengan komposisi *primary key*, dan beberapa tabel dimensi dengan komposisi *foreign key* yang berrelasi dengan *primary key* yang ada di tabel fakta (Connolly, 2005). Contoh sederhana dari *star schema* dapat dilihat pada gambar 2.2 dibawah ini.



Gambar 2.2 *Star schema* untuk proses pemesanan (Adamson, 2006).

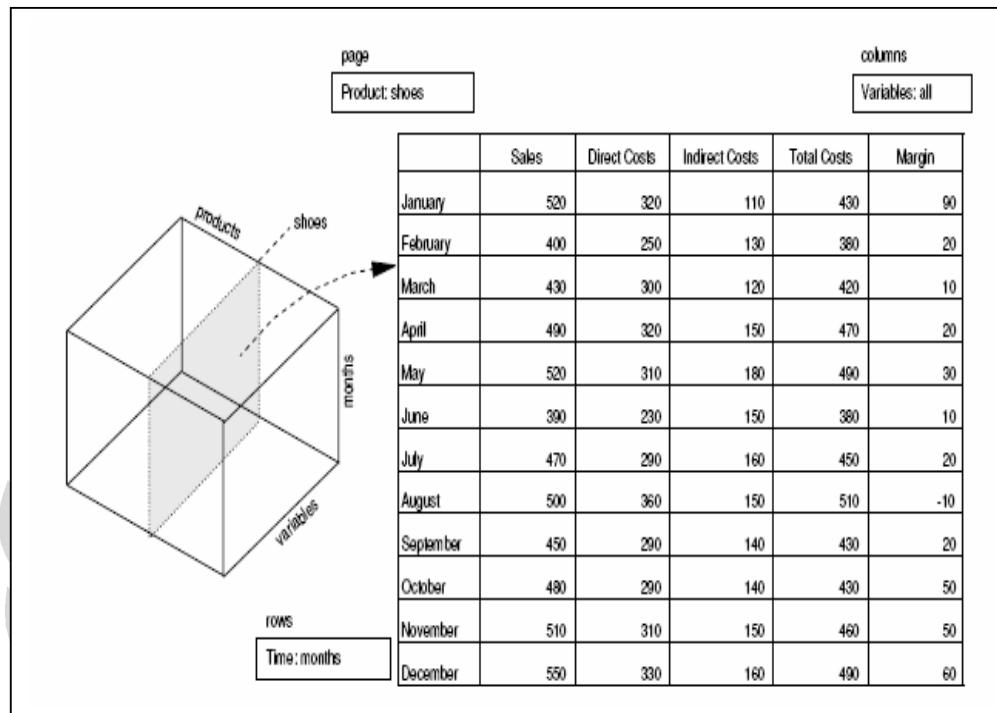
2.3 *On-Line Analytical Processing (OLAP)*.

Kemampuan *On-Line Analytical Processing (OLAP)* dalam mendukung *queries* dan analisa data pada *aggregation database* yang ada di *data warehouse*, menjadikan *OLAP* sebagai sistem dengan kemampuan mengumpulkan, mengelola, memproses dan menghadirkan multidimensional data. Penerapan teknologi *OLAP* dapat memberikan informasi untuk penyusunan perencanaan dan pembuatan kebijakan bisnis perusahaan. (Kontio, 2006), (Peter, 2006).

Ada dua metode utama *OLAP* yang banyak diterapkan yaitu: *Relational OLAP (ROLAP)* dan *Multidimensional OLAP (MOLAP)* (Boon-NoiTan, 2006).

Kedua metode ini memanfaatkan pembentukan data secara *cube* atau

multidimensional (Boon-NoiTan, 2006), dimana karakter data multidimensional ini dipercaya sebagai penyelesaian *OLAP* (Thomsen, 2002). Contoh sederhana dari tampilan pemodelan data *cube* dan multidimensional dapat dilihat pada gambar 2.3



Gambar 2.3 Bentuk pemodelan *data cube* dan tampilan data sisi dari *cube* (Thomsen, 2002).

Bagian kiri gambar menunjukkan pemodelan data yang terbentuk dalam sebuah *cube*. *Cube* ini terdiri dari data multidimensional *product*, *months (time)*, dan variabel. Di bagian tengah *cube* dibuat suatu irisan (sisi bidang) untuk *product shoes* yang kemudian ditampilkan menjadi halaman data yang terlihat di bagian kanan gambar.

2.4 Data Mining

Teknologi *data mining* memiliki kemampuan memberikan informasi dan data yang dibutuhkan untuk mendukung pengambil kebijakan strategis. Tingginya kebutuhan pemanfaatan *data warehouse* yang telah tersedia untuk mendukung keputusan strategis menjadikan *data mining* merupakan suatu kelanjutan yang diyakini mampu menjawab berbagai permasalahan bisnis. Oleh karena itu faktor terpenting dalam keberhasilan proses *data mining* adalah tersedianya *data warehouse* yang terintegrasi secara baik dan tepat dalam mengidentifikasi kebutuhan bisnis yang ada.

2.4.1 Konsep Data Mining

Data Mining (penambangan data) dapat dikatakan sebagai teknologi baru dalam pencarian informasi yang sangat bermanfaat bagi kepentingan bisnis perusahaan. Informasi yang diperoleh dari proses pencarian yang dilakukan dalam *data mining* harus memiliki arti yang sangat penting dan bernilai tinggi. Dibawah ini beberapa definisi sebagai konsep dari *data mining*:

1. *Data mining* adalah suatu proses pencarian pengetahuan. *Data mining* dapat membantu kita memahami substansi dari data secara khusus yang tadinya tidak diketahui seperti pola-pola atau *trend* yang ada dalam data yang sebelumnya tidak pernah diketahui (Connolly, 2005).
2. *Data mining* juga bisa dikatakan proses ekstraksi data yang valid, dimana sebelumnya ada informasi yang sangat berharga, lebih menyeluruh namun belum diketahui, hal ini diperlukan usaha yang lebih besar untuk menemukannya dari suatu *database* dimana informasi yang

diperoleh dapat digunakan untuk membuat kebijakan bisnis yang penting (Ponniah, 2001).

3. *Data mining* merupakan suatu proses penyelidikan yang berlangsung secara besar-besaran dari suatu data untuk menemukan pola-pola yang tersembunyi dan menemukan pengetahuan baru (Oracle Corporation, 2006).
4. *Data Mining* adalah suatu proses penggunaan berbagai variasi dari *tools* analisa data untuk menemukan pola-pola dan keterkaitan dalam suatu data dimana hasilnya dapat digunakan untuk membuat prediksi yang valid (Two Crows Corporation, 2005).
5. *Data Mining* adalah suatu proses pencarian berarti mengenai hubungan, pola, dan *trend* yang baru dengan melakukan penyelidikan secara terus-menerus dari data yang berjumlah sangat besar di suatu media penyimpanan data. Dalam pencarian dan penggunaan pola-pola dengan memadukan teknologi yang sudah ada seperti teknologi statistik (Gatner Group, 2007).

Pendekatan yang dilakukan dalam *data mining* menggunakan *discovery based* dengan bantuan algoritma-algoritma untuk menentukan keterkaitan dari data yang di eksplorasi sehingga didapat pola-pola informasi.

2.4.2 Knowledge Discovery in Database (KDD).

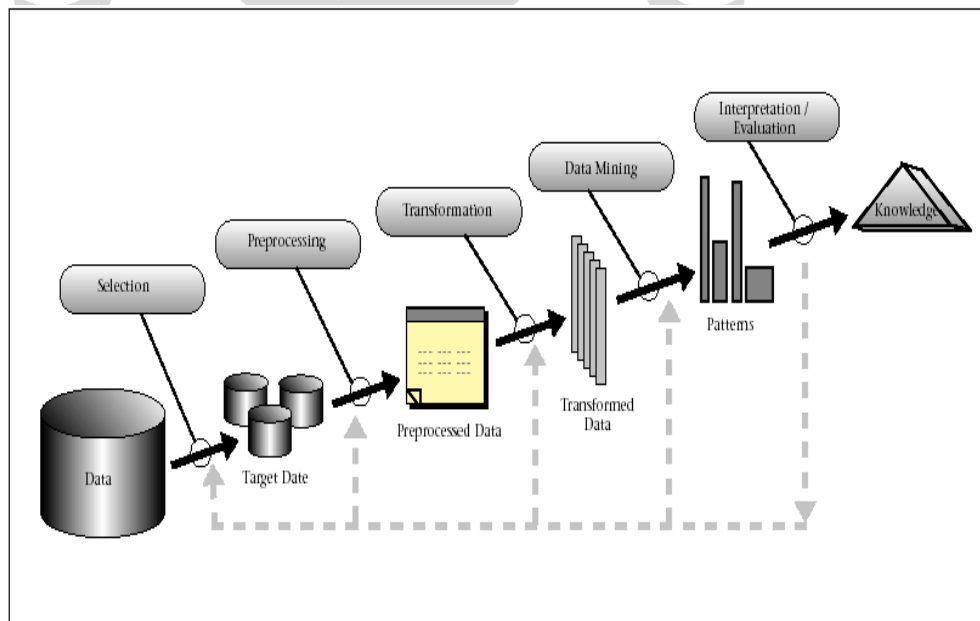
Data mining bukanlah suatu teknologi yang berdiri sendiri melainkan suatu tahapan dari proses pencarian pengetahuan dalam *database* yang biasa dikenal dengan sebutan *Knowledge Discovery in Database (KDD)*. Proses

berulang *data mining* yang dilakukan dari tahapan-tahapan *KDD* bisa disebut juga suatu metodologi *data mining*.

Tahapan yang ada diproses *KDD* sampai saat ini banyak dikeluarkan oleh peneliti *KDD*. Berdasarkan kajian beberapa dari proses *KDD* yang dilakukan, dapat diketahui bahwasannya tahapan yang ada antara satu proses *KDD* dengan lainnya tidak terjadi pertentangan yang prinsip. Perbedaan yang ditemukan adalah jumlah tahapan diproses *KDD*. Beberapa definisi dan tahapan dari proses *KDD* akan dibahas berikut ini.

2.4.2.1 Proses *KDD* (Fayyad et.al, 1996)

KDD merupakan suatu proses interaktif dan iteratif, meliputi banyak tahapan dengan berbagai keputusan yang dibuat oleh pemakai (Fayyad et.al, 1996). Tahapan-tahapan yang dilakukan dapat dilihat pada gambar.2.4.



Gambar 2.4. Tahapan-tahapan dalam proses *KDD* (Fayyad et.al, 1996).

1. Pemahaman mengenai domain aplikasi dan relevansi terhadap pengetahuan yang ada serta mengidentifikasi tujuan atau sasaran yang ingin dicapai.
2. Membuat *database* sebagai data target: melakukan pemilihan data untuk membuat *database* atau memfokuskan penyediaan *database* untuk proses *discovery* yang akan dilakukan dengan menggunakan subset variabel atau sampel data.
3. Pembersihan data: melakukan pembersihan data seperti menghilangkan *noise*, menjaga data dari ketidak konsistenan atau tidak relevan serta melengkapi relasi *field* yang ada di *database*.
4. Memproyeksikan pengurangan data: disini sudah dimulai dilakukan pencarian fitur-fitur yang bermanfaat untuk dipresentasikan kedalam tujuan atau sasaran yang telah ditentukan.
5. Pemilihan teknik dari metode *data mining* untuk melakukan pencocokkan dari tujuan yang sudah ditentukan ditahap satu seperti: *clustering*, *classification*, *regresi*, dsb.
6. Melakukan analisa eksplorasi dan pemodelan serta pemilihan hipotesis: pemilihan algoritma *data mining* dan metode atau teknik untuk digunakan dalam pencarian pola-pola data.
7. *Data Mining*: proses pencarian pola-pola data berharga yang diinginkan dari kumpulan data yang direpresentasikan.
8. Mengevaluasi dan penterjemahan pola-pola yang diperoleh dari *data mining* di tahapan ketujuh, ditahapan ini ada kemungkinan dilakukan pengulangan dari tahap satu sampai tujuh.

9. Menindaklanjuti hasil pencarian pola-pola sebagai pengetahuan baru dengan melakukan pengujian pada data berbeda, juga dilakukan penyelesaian apabila terjadi konflik antara pengetahuan yang diperoleh dengan pengetahuan yang dipercaya sebelumnya.

2.4.2.2 Proses *KDD* (Baritchi, 2004)

Baritchi (2004), mendefinisikan *KDD* adalah suatu proses ketekunan dalam pencarian pengetahuan di *database* yang besar untuk tujuan menggali pola-pola tersembunyi atau pengetahuan didalam data. Berikut tahapan-tahapan yang ditetapkan:

1. *Domain Analysis*
2. *Selection data*
3. *Preprocessing data*
4. *Transformation data*
5. *Data Mining*
6. *Interpretation and evaluation*
7. *Consolidation of Discovered Knowledge*

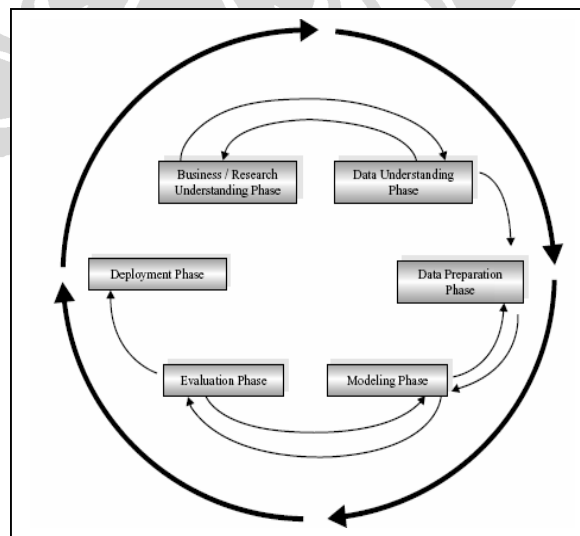
2.4.2.3 Proses *KDD* Cross-Industry Standard Process for Data Mining (CRISP-DM V.1.0, 2002)

Metodologi ini dibuat oleh beberapa orang dari sebuah konsorsium perusahaan yaitu: Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (Daimler Chrysler), Colin Shearer (SPSS) dan Rudiger Wirth (Daimler Chrysler). Proses tahapan *KDD*

dalam metodologi ini dinamakan *Cross-Industry Standard Process for Data Mining (CRISP-DM)*, (*CRISP-DM V.1.0*, 2002). Dibawah ini tahapan-tahapan dari *CRISP-DM*:

1. *Business understanding phase*
2. *Data understanding phase*
3. *Data Preparation phase*
4. *Modeling phase*
5. *Evaluation phase*
6. *Deployment phase*

Metodologi *CRISP-DM* saat ini banyak digunakan sebagai referensi dalam proses *KDD*, Ini dapat dilihat dari beberapa referensi yang menggunakan metodologi *CRISP-DM* (Larose, 2005, p.5-9), (Ye, 2003, p.490), (Connolly, Begg, 2005, p.1239-1240), (Tang, MacLennan, p28-30). Tanuwijaya, 2007 dalam penelitiannya juga menggunakan metodologi *CRISP-DM*. Dibawah ini proses iterasi dari tahapan *KDD CRISP-DM*, yang ditunjukkan oleh gambar 2.5.



Gambar 2.5 Proses tahapan *KDD CRISP-DM (CRISP-DM V.1.0, 2002)*

2.4.2.4 Proses *KDD Two Crows Corporation* (*Two Crows Corporation, 2005*)

Metodologi *KDD* dari *Two Crows Corporation* dibuat dengan mengembangkan metodologi *CRISP-DM*. Terdapat satu tahapan di metodologi *KDD Two Crows Corporation* yang tidak ada di metodologi *CRISP-DM*, yaitu tahapan penyelidikan data sebelum dilakukan proses *data mining* dengan menggunakan teknologi *OLAP*. Dikarenakan adanya kombinasi teknologi *OLAP* diproses tahapan *KDD* maka penelitian ini menggunakan metodologi *KDD* dari *Two Crows Corporation* yang diyakini sesuai dengan pengerjaan tesis ini. Penggunaan teknologi *OLAP* di tahapan proses *KDD* juga digunakan oleh peneliti *KDD* antara lain; oleh Silver (Silver 2005) dan Payton (Payton 2005). Penjelasan lebih lanjut mengenai metodologi *KDD* dari *Two Crows Corporation* dibahas pada bab 3.

2.4.3 Teknik–Teknik *Data Mining*

Saat ini kategori *data mining* dibagi dalam dua kegiatan utama yaitu: deskripsi dan prediksi (Baricthi, 2004), (*Two Crows Corporation, 2005*).

- Deskripsi

Deskripsi merupakan kegiatan yang paling banyak dilakukan di *data mining* untuk memperoleh pola data. Proses ini berfokus pada pencarian pola-pola berharga yang tersembunyi dalam *database* yang dieksplorasi. Teknik-teknik *data mining* yang termasuk dalam kategori ini antara lain: *association, clustering, summaries and visualization*, dan sebagainya.

- Prediksi

Sasaran dari tugas prediksi adalah melakukan prediksi dengan menggunakan beberapa variabel atau *field-field* yang ada didalam *database* untuk menentukan nilai-nilai variabel dimasa akan datang yang dibutuhkan, dimana informasi belum diketahui saat ini. Dalam konteks *KDD*, prediksi dianggap kurang penting dibandingkan dengan deskripsi. Teknik-teknik *data mining* yang termasuk dalam kategori prediksi adalah: *classification*, *regression*, *time series*, dan lain-lain.

2.4.3.1 Association Rule

Association rule adalah teknik *data mining* untuk menemukan aturan *associative* antara suatu kombinasi item yang ada didalam *database* antara satu item dengan item lainnya. *Association rule* terdiri dari dua komponen yang biasa disebut dengan *antecedent* dan *consequent* (Webb, 2003). *Association* dapat di tulis seperti $A \Rightarrow B$, dimana A disebut *antecedent or left hand side* (LHS), dan B disebut *consequent or right hand side* (RHS). Contoh sederhana dari *association rule* adalah sebagai berikut “Jika seseorang membeli palu maka mereka membeli juga paku” *antecedent* disini membeli palu dan *consequent* adalah membeli paku (Two Crows Corporation, 2005). Analisa mengenai adanya keterkaitan dari suatu item dengan item lainnya yang terjadi pada suatu transaksi dari satu keranjang pembelian yang dilakukan algoritma *association rule*, menyebabkan teknik *association rule* disebut juga dengan *market basket analysis*.

Denton dan Besemann, berpendapat bahwa penggunaan *database data mining* yang menerapkan *Relational Database Management System (RDBMS)*

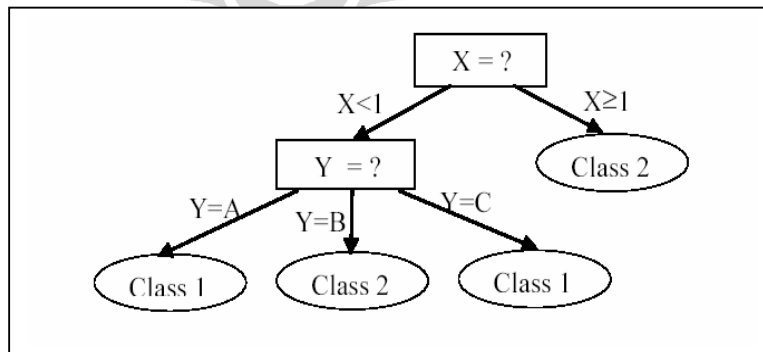
akan cukup membantu dalam proses *association rule* (2006). Pendapat ini juga didukung oleh Martine, dimana data yang dihasilkan dapat ditampilkan dengan menggunakan metode statistik melalui perhitungan (et.al, 2006).

2.4.3.2 Classification

Classification adalah proses pembelajaran fungsi, dimana bila diberikan suatu *input* sebagai variabel kedalam barisan data yang direpresentasikan maka akan dapat memetakan (mengklasifikasi) sebuah item data kedalam salah satu dari beberapa kelas yang sudah terdefiniskan (Baricthi, 2004), (Larose, 2005). *Classification* dapat dikatakan sebagai pembelajaran dalam statistik dan *machine learning* (An, 2006). Dibawah ini beberapa teknik algoritma *classification*:

- *Decision Tree*

Decision tree adalah salah satu algoritma *classification* yang paling populer. *Decision tree* adalah model prediksi menggunakan struktur *tree*, dimana setiap simpul internal menandakan suatu tes pada *attribute*, setiap cabang merepresentasikan hasil tes dan simpul hasil merupakan dari pemetaan kelas atau distribusi kelas. Contoh sederhana dari *decision tree* dapat dilihat pada gambar 2.6.

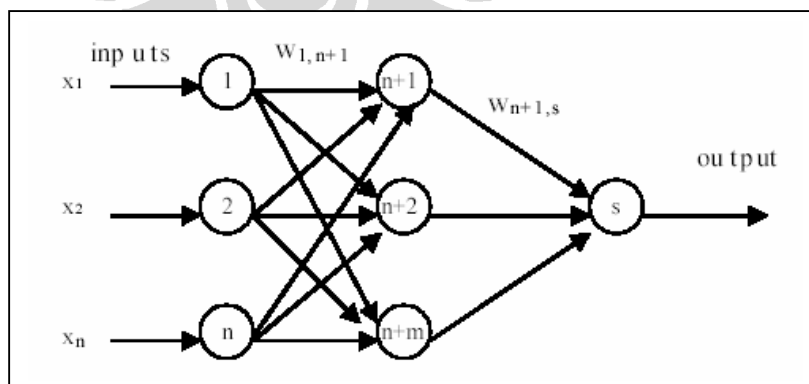


Gambar 2.6 *Decision tree* dengan pengujian pada atribut X dan Y (An, 2006).

Dijelaskan bahwa awal *decision tree* dimulai dengan menginputkan variabel terhadap X, dimana bila pencocokkan variabel tersebut lebih besar atau sama dengan satu maka hasil dari pengujian ini akan menjadi kelas data dua dan nilai yang lebih kecil dari satu masih dapat didefinisikan kembali melalui input variabel terhadap Y, dimana bila hasil pengujian tersebut hasilnya adalah B maka dapat dimasukkan kedalam kelas data dua, dan bila hasil pengujian hasilnya A atau C maka dimasukkan kedalam kelas data satu.

- *Neural Network*

Neural network adalah penggunaan komputasi sederhana untuk menguji data dan mengembangkan model untuk membantu identifikasi pola-pola atau struktur yang bernilai (Smith, 2006). *Neural network* juga mereferensi pada *artificial neural network* (Marti, 2006). *Neural network* digunakan untuk mempelajari pola-pola yang ada sebagai referensi dalam memprediksi kejadian di masa depan dan mengklasifikasi data yang sebelumnya tidak terlihat kedalam basis *group* yang terdefinisi dari observasi karakteristik berdasarkan informasi yang diperoleh dari sejarah pengujian data. Contoh diagram dari *neural network* ditunjukkan oleh gambar 2.7



Gambar 2.7 *Neural network* diagram.(Marti, 2006)

2.4.3.3 Clustering

Clustering (pengelompokkan) merupakan teknik *data mining* deskripsi yang melakukan proses pengelompokkan sejumlah kumpulan pola-pola kedalam bagian atau *cluster* didasari pada kesesuaian perkiraan kedekatan atau kesamaan terbesar yang ada di pola-pola tersebut (Ghosh, 2003). Prinsip dari *Clustering* adalah memaksimalkan kesamaan yang ada antara anggota satu kelas dan meminimalkan kesamaan antar *cluster* (Two Crows Corporation, 2006), (Baricthi, 2004). Berbeda dengan *assosication rule* dan *clasification* yang telah ditentukan definisi kelas data sebelum proses pemodelan, *clustering* tidak didasari kelas data terdahulu dalam melakukan pengelompokkan, bahkan dengan *cluster* dapat ditemukan kelas data yang sebelumnya tidak diketahui (Larose, 2005).

Penerapan *clustering* dalam teknik *data mining* cukup luas diberbagai bidang, dengan bantuan algoritma *clustering* (Huang, 2006), dapat dipahami *natural cluster* atau struktur yang terdapat dalam *database*. Penggunaan metode algoritma *cluster* seperti algoritma *K-Means* (Ghosh, 2003), algoritma *center-based clustering* (Zhang, 2006), merupakan hal utama dari kegiatan proses teknik *clustering*.