

BAB 2

LANDASAN TEORI

2.1 Data Mining

2.1.1 Definisi Data Mining

Data mining menyediakan teknologi yang pintar dan kemampuan untuk menjelajahi kemungkinan pengetahuan atau informasi yang tersimpan di sebuah data (Berson, 2000, p33).

Data mining adalah proses untuk menemukan *pattern* yang bernilai dan *relationship* yang tersembunyi dalam *database* yang sangat besar (Seidman, 2000, p3).

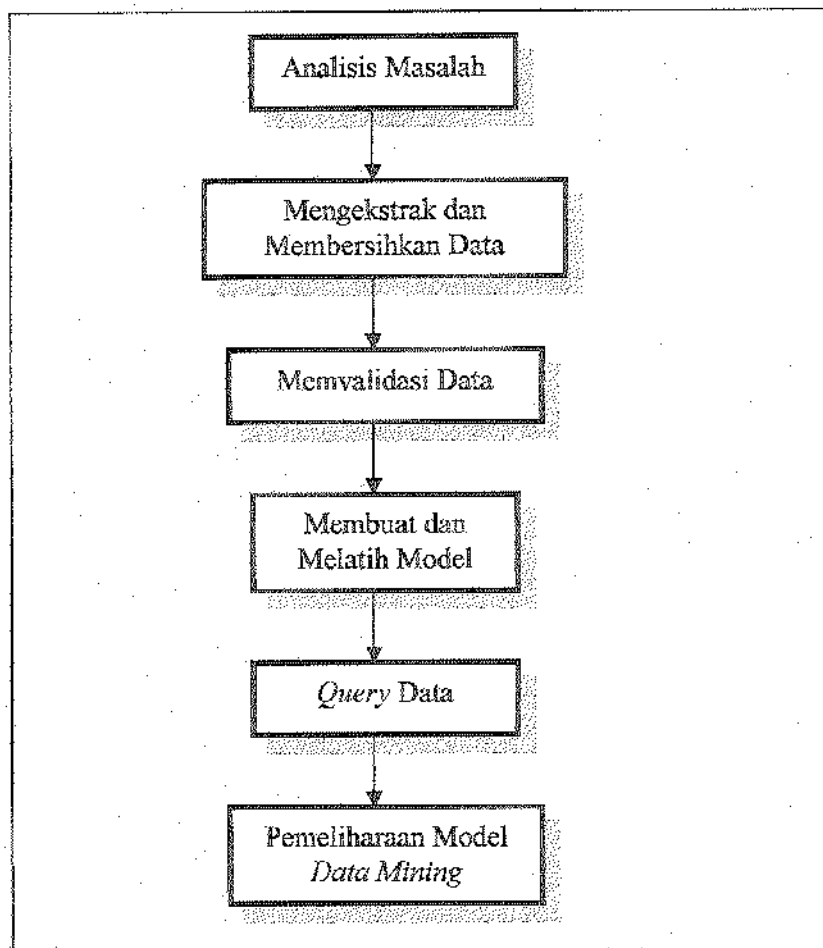
Definisi umum dari *data mining* adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2003, p1).

Data mining adalah beberapa cara pengembangan dari ilmu statistik dengan sedikit *Artificial Intelligence* dan seperti sebuah mesin yang mempelajari data untuk mengatasi masalah dengan menghasilkan informasi yang tidak kelihatan atau tidak disadari oleh pengguna informasi tersebut (Thearling, 1995).

2.1.2 Metodologi Data Mining

Sebagai salah satu bagian dari sistem informasi, *data mining* menyediakan perencanaan dari ide hingga implementasi akhir. Komponen-komponen perancangan *data mining* (Seidman, 2000, pp9-11) adalah sebagai berikut :

1. Analisis masalah.
2. Mengekstrak dan membersihkan data.
3. Memvalidasi data.
4. Membuat dan melatih model.
5. *Query* data.
6. Pemeliharaan model *data mining*.



Gambar 2.1 Diagram Alir Komponen-Komponen Perancangan *Data Mining*

Sumber : Seidman, 2000, p9

2.1.3 Teknik Pembelajaran dalam *Data Mining*

Teknik yang digunakan dalam *data mining* erat kaitannya dengan “penemuan” (*discovery*) dan “pembelajaran” (*learning*) yang terbagi dalam tiga kelompok utama pembelajaran yaitu : *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Berson, 2000, p36).

2.1.3.1 *Supervised Learning*

Supervised learning adalah teknik yang paling banyak digunakan. Teknik ini sama dengan “*programming by example*”. Teknik ini melibatkan fase pelatihan dimana data pelatihan historis yang karakter-karakternya dipetakan ke hasil-hasil yang telah diketabui diolah dalam algoritma *data mining*. Proses ini melatih algoritma untuk mengenali variabel-variabel dan nilai-nilai kunci yang nantinya akan digunakan sebagai dasar dalam membuat perkiraan-perkiraan ketika diberikan data baru.

2.1.3.2 *Unsupervised Learning*

Teknik pembelajaran ini tidak melibatkan fase pelatihan seperti yang terdapat pada *supervised learning*. Teknik ini bergantung pada penggunaan algoritma yang mendeteksi semua pola, seperti *associations* dan *sequences*, yang muncul dari kriteria penting yang spesifik dalam data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan (*rules*) yang mengkarakterisasikan penemuan *associations*, *clusters*, dan *segments*. Aturan-aturan ini kemudian dianalisis untuk menemukan hal-hal yang penting.

2.1.3.3 Reinforcement Learning

Walaupun teknik pembelajaran ini jarang digunakan dibandingkan dengan dua teknik lainnya, namun memiliki penerapan-penerapan yang terus dioptimalkan dari waktu ke waktu dan memiliki kontrol adaptif. Teknik ini sangat menyerupai kehidupan nyata yaitu seperti "*on-job-training*", dimana seorang pekerja diberikan sekumpulan tugas yang membutuhkan keputusan-keputusan. Pada beberapa titik waktu kelak diberikan penilaian atas *performance* pekerja tersebut kemudian pekerja diminta mengevaluasi keputusan-keputusan yang telah dibuatnya sehubungan dengan hasil *performance* pekerja tersebut. *Reinforcement learning* sangat tepat digunakan untuk menyelesaikan masalah-masalah yang sulit yang bergantung pada waktu.

2.1.4 Teknik Data Mining

Berson (2000) dalam bukunya mengelompokkan teknik *data mining* menjadi dua kelompok utama yaitu teknik klasik dan teknik generasi berikutnya. Teknik klasik terdiri dari *statistik*, *nearest neighbor* dan *clustering*. Sedangkan teknik generasi berikutnya terdiri dari *decision trees*, *neural networks* dan *rule induction*.

2.1.4.1 Teknik Klasik

2.1.4.1.1 Statistik

Teknik statistik digerakkan oleh data dan digunakan untuk menemukan pola-pola dan membuat model perkiraan (*predictive model*). Statistik sangat membantu dalam menjawab beberapa pertanyaan penting mengenai :

1. Pola-pola apa yang terdapat dalam *database* ?
2. Berapa besar kemungkinan kejadian akan muncul ?

3. Pola-pola manakah yang penting ?
4. Rangkuman data yang mana yang dapat memberikan ide mengenai apa yang terkandung dalam *database* ?

Salah satu cara terbaik untuk meringkas data adalah dengan menggunakan diagram batang (*histogram*). Beberapa ringkasan statistik yang sering digunakan adalah nilai maksimum, nilai minimum, rata-rata, nilai tengah, modus dan varian. Dalam statistik, prediksi biasanya sinonim dengan beberapa bentuk regresi.

2.1.4.1.2 Nearest Neighbor

Nearest neighbor adalah teknik perkiraan yang hampir sama dengan *clustering*, untuk memperkirakan suatu nilai dalam satu *record* yaitu dengan melihat *record-record* lain yang mempunyai nilai *predictor* serupa dalam *database* historis dan menggunakan nilai perkiraan dari *record* yang “terdekat” (*nearest*). Algoritma prediksi *nearest neighbor* didefinisikan sebagai berikut : “*Object that are ‘near’ each other will also have similar prediction values. Thus, if you know the prediction value of one of the object, you can predict it for its nearest neighbor*” (Berson, 2000, p136).

2.1.4.1.3 Clustering

Clustering adalah suatu metode dimana *record-record* dikumpulkan bersama agar dapat memberikan pandangan terhadap apa yang ada dalam *database*. *Clustering* juga dapat berarti segmentasi yang sangat berguna dalam prediksi masalah bisnis. Sebagai contoh sistem *clustering* adalah sistem PRIZM dari perusahaan Claritas dan MicroVision dari perusahaan Equifax. Sistem ini mengelompokkan populasi

berdasarkan informasi demografi ke dalam segmen-segmen yang dipercayai sangat bermanfaat untuk pemasaran cara langsung (*direct marketing*) dan penjualan.

2.1.4.2 Teknik Generasi Berikutnya

2.1.4.2.1 Decision Trees

Seperti namanya *decision trees* adalah model prediksi yang dapat dilihat sebagai pohon. Setiap cabangnya merupakan hasil klasifikasi dari pertanyaan dan daunnya merupakan hasil partisi dari kumpulan data sesuai dengan klasifikasinya.

Jika dilihat dari sudut pandang bisnis, pohon keputusan dapat dilihat sebagai segmentasi dari data, dimana setiap segmen merupakan data yang mempunyai sifat yang sama, sehingga memudahkan pengguna bisnis dalam memahami informasi yang terkandung di dalamnya. Pohon keputusan dapat digunakan dalam menangani berbagai masalah bisnis yaitu untuk eksplorasi, preproses data dan prediksi.

Teknik ini lebih banyak digunakan untuk eksplorasi dan preproses data. Eksplorasi dilakukan dengan melihat *predictor* dan nilai yang dipilih pada setiap segmen dari pohon. Preproses data yaitu dimana pohon keputusan digunakan untuk mencari *predictor* pada tahap pertama proses *data mining*, kemudian hasilnya dapat digunakan oleh teknik *neural networks*, *nearest neighbor* atau statistik.

2.1.4.2.2 Neural Networks

Neural networks yang sesungguhnya adalah jaringan saraf dalam otak manusia yang mengenali pola-pola, membuat perkiraan dan pembelajaran. Sedangkan yang dimaksud disini adalah *artificial neural networks* yaitu program komputer yang

menerapkan pencarian pola dan algoritma pembelajaran mesin komputer untuk membuat perkiraan dari data historis.

Neural networks pertama kali diperkenalkan oleh McCulloch dan Pits dalam seminarnya semasa Perang Dunia ke II yang mulanya hanya menyampaikan sebuah ide tentang unit pemrosesan yang sederhana (seperti sebuah neuron dalam otak manusia). Ide tersebut dapat dihubungkan bersama-sama dalam jaringan yang besar sehingga membangun sebuah sistem dengan kemampuan untuk menyelesaikan masalah-masalah yang sulit dan menampilkan pola-pola yang kompleks dari yang dapat dibuat oleh sebuah unit saja. (Berson, 2000, p167)

2.1.4.2.3 Rule Induction

Rule induction adalah salah satu teknik dalam *data mining* yang paling sering digunakan untuk menemukan pengetahuan dalam sistem *unsupervised learning*. *Rule* (aturan) adalah bentuk sederhana dari "jika ini maka ini dan ini dan kemudian ini". Sebagai contoh : jika seseorang membeli roti maka orang tersebut juga cenderung membeli susu.

Agar aturan-aturan tersebut bermanfaat maka harus ditambahkan dua informasi tambahan sesuai dengan keadaan sebenarnya yaitu :

1. Keakuratan (*accuracy/confidence*) yang menunjukkan seberapa sering aturan tersebut benar.
2. Penerapan (*coverage/support*) yaitu angka yang menunjukkan seberapa sering aturan tersebut dipakai.

2.2 Regression Sebagai Salah Satu Teknik Data Mining

Teknik regresi (*regression*) merupakan salah satu teknik statistik yang digunakan dalam *data mining*. Persamaan regresi adalah persamaan matematika yang menyatakan hubungan fungsional antara satu peubah tak bebas (*dependent variable*) atau respon dengan satu atau lebih peubah bebas (*independent variable*) atau deterministik berdasarkan sekumpulan data percobaan (Walpole dan Myers, 2002, p340). Persamaan regresi ditemukan oleh Sir Francis Galton (1822-1911).

Regresi digunakan untuk mendapat hubungan sebab akibat antara peubah bebas dan peubah tak bebas dalam suatu model matematika untuk melakukan peramalan dan perkiraan.

Menurut Walpole (2002), bentuk-bentuk regresi terdiri atas regresi linier (*linear regression*), dan regresi polinomial (*polynomial regression*).

2.2.1 Linear Regression

Regresi linier (*linear regression*) menggambarkan hubungan linier antara peubah tak bebas dengan peubah bebas. Berdasarkan jumlah peubah bebas yang digunakan dalam persamaan regresi linier dapat dibagi menjadi dua yaitu regresi linier sederhana (*simple linear regression*) dan regresi linear ganda (*multiple linear regression*).

2.2.1.1 Simple Linear Regression

Regresi linier sederhana (*simple linear regression*) menggambarkan hubungan linier antara satu peubah tak bebas dengan satu peubah bebas. Hubungan tersebut dinyatakan dalam bentuk persamaan sebagai berikut :

$$y = a + bx$$

Persamaan 2.1 Persamaan *Simple Linear Regression*

Keterangan:

y = nilai peubah tak bebas

x = nilai peubah bebas

a, b = penduga parameter koefisien regresi

2.2.1.2 Multiple Linear Regression

2.2.1.2.1 Persamaan Matematika *Multiple Linear Regression*

Regresi linier ganda (*multiple linear regression*) menggambarkan hubungan linier antara satu peubah tak bebas dengan beberapa peubah bebas. Persamaan regresi ganda dinyatakan sebagai berikut :

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

Persamaan 2.2 Persamaan *Multiple Linear Regression*

Keterangan :

y = nilai peubah tak bebas

x = nilai peubah bebas

b_0, b_1, \dots, b_p = penduga parameter koefisien regresi

p = jumlah peubah bebas

Jika terdapat n sampel acak maka susunan data dapat dinyatakan sebagai berikut:

$$y_1 = b_0 + b_1 x_{11} + b_2 x_{21} + \dots + b_p x_{p1}$$

$$y_2 = b_0 + b_1 x_{12} + b_2 x_{22} + \dots + b_p x_{p2}$$

.....

.....

$$y_n = b_0 + b_1 x_{1n} + b_2 x_{2n} + \dots + b_p x_{pn}$$

Bentuk di atas dapat dinyatakan dalam bentuk matriks sebagai berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & & & \dots & \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

Persamaan 2.3 Persamaan *Multiple Linear Regression* dalam Bentuk Matriks

Nilai parameter koefisien regresi ditentukan dengan menggunakan metode kuadrat terkecil (*least squares*). Dalam metode kuadrat terkecil akan dicari b_0, b_1, \dots, b_p sehingga jumlah kuadrat sisa minimum. Jumlah kuadrat sisa sering pula disebut Jumlah Kuadrat Galat terhadap garis regresi dan dinyatakan dalam JKG (Walpole dan Myers, 2002, p408)

Nilai parameter regresi menggunakan metode kuadrat terkecil dapat dihitung dengan teori matriks yang dijabarkan sebagai berikut :

$$y = X b$$

$$X'y = X'X b$$

Sehingga :

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Persamaan 2.4 Persamaan Koefisien Regresi dalam Bentuk Matriks

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \dots & \sum x_{pi} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \dots & \sum x_{1i}x_{pi} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{pi} & \sum x_{1i}x_{pi} & \sum x_{2i}x_{pi} & \dots & \sum x_{pi}^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \vdots \\ \sum x_{pi}y_i \end{bmatrix}$$

2.2.1.2.2 Penaksiran Selang Kepercayaan

Selang kepercayaan $(1-\alpha)$ 100% untuk y_0 diberikan oleh :

$$\hat{y}_0 - t_{\alpha/2} s \sqrt{1 + x_0 (X'X)^{-1} x_0} < y_0 < \hat{y}_0 + t_{\alpha/2} s \sqrt{1 + x_0 (X'X)^{-1} x_0}$$

Persamaan 2.5 Persamaan Selang Kepercayaan

Dimana :

$$s^2 = \frac{JKT - JKR}{n - k - 1} = \frac{JKG}{dbGalat}$$

$$JKT = \mathbf{y}'\mathbf{y} - n\bar{y}^2$$

$$JKR = \mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2$$

2.2.1.2.3 Perhitungan Koefisien Determinasi

Koefisien determinasi ($R^2 \times 100\%$) menyatakan persentase variasi yang dapat diterangkan oleh model yang dipostulasikan. R^2 dihitung dengan :

$$R^2 = \frac{JKR}{JKT}$$

Persamaan 2.6 Persamaan Koefisien Determinasi

2.2.2 Polynomial Regression

Persamaan regresi polinomial (*polynomial regression*) dinyatakan sebagai berikut (Walpole dan Myers, 2002, p414) :

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + \dots + b_r x_i^r$$

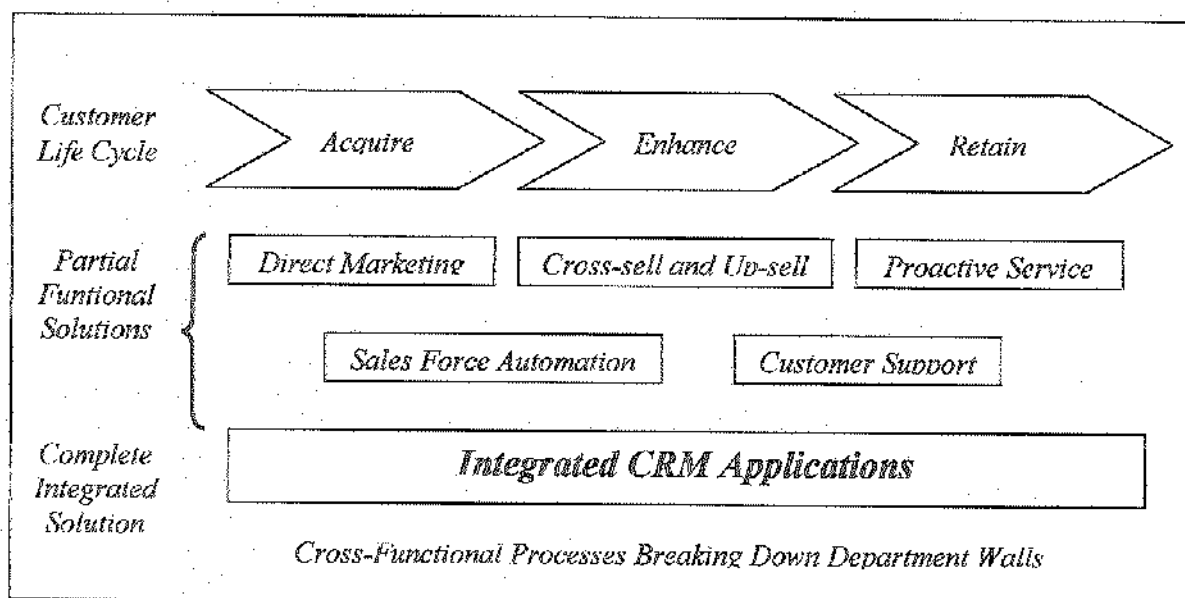
Persamaan 2.7 Persamaan *Polynomial Regression*

2.3 Customer Relationship Management (CRM)

2.3.1 Definisi CRM

Dalam buku *E-Business Roadmap for Success* (Kalakota dan Robinson, 1999, p117) menyatakan *Customer Relationship Management* (CRM) adalah kombinasi dari proses bisnis dan teknologi untuk memperoleh informasi agar dapat memahami pelanggan-pelanggan perusahaan dari berbagai sudut pandang : siapa mereka, apa yang mereka lakukan dan apa yang mereka suka.

CRM dideskripsikan sebagai suatu integrasi antara strategi penjualan, pemasaran dan pelayanan yang tidak melihat kepentingan masing-masing departemen dalam perusahaan dan bergantung pada tindakan-tindakan terkoordinasi perusahaan (lihat Gambar 2.2).



Gambar 2.2 CRM Terintegrasi
 Sumber : Kalakota dan Robinson, 1999, p117

2.3.2 Tujuan dan Manfaat CRM

Tujuan CRM adalah untuk mengoptimalkan keuntungan perusahaan dengan membuat perusahaan lebih efisien dalam menggunakan sumber daya yang ada untuk melayani pelanggan yang diinginkan dan memelihara hubungan dengan pelanggan (Kalakota dan Robinson, 1999, p117).

Manfaat CRM adalah dijabarkan sebagai berikut :

1. Meningkatkan pendapatan.
2. Mendorong loyalitas pelanggan.
3. Mengurangi biaya.
4. Meningkatkan efisiensi operasional.
5. Meningkatkan *time to market*.

2.3.3 Arsitektur Aplikasi CRM

CRM dibagi menjadi 3 komponen penting (Berson, 2000, p45), yaitu :

1. *Operational CRM*. Automatisasi proses bisnis yang terintegrasi, meliputi *customer touchpoint, channels* dan *front-back office integration*.
2. *Analytical CRM*. Analisis data yang dihasilkan oleh *operational CRM*.
3. *Collaborative CRM*. Aplikasi dari pelayanan bersama yang meliputi *e-mail, personalized publishing, e-communities*, dan fasilitas sejenisnya yang dirancang untuk interaksi antara pelanggan dan perusahaan.

2.3.4 Teknologi CRM

Aplikasi CRM secara umum terbagi ke dalam tiga jenis (Seybold, 2000) yaitu :

1. *Customer Facing Application*. Merupakan aplikasi yang mendukung *staff* bagian penjualan dan pelayanan berinteraksi langsung dengan pelanggan. Contohnya adalah aplikasi *contact center* yang mendukung *customer service* dalam melayani pelanggan secara 24 jam serta aplikasi pendukung *staff* penjualan dan pelayanan.
2. *Customer Touching Application*. Dalam aplikasi ini pelanggan berinteraksi secara langsung dengan aplikasi. Aplikasi ini juga dapat terhubung dengan *staff* perusahaan untuk menyeimbangkan kebutuhan pelayanan. Contohnya adalah pelayanan *customer service* secara otomatis melalui *e-mail* serta layanan informasi melalui *website*.
3. *Customer Centric Intelligence Application*. Merupakan aplikasi yang melakukan analisis terhadap data-data hasil dari proses operasional untuk memberikan dukungan keputusan yang bertujuan untuk meningkatkan efisiensi dan efektifitas perusahaan dalam hubungan dengan pelanggan. Contohnya adalah aplikasi *data*

warehouse, reporting dan *analytic application* seperti analisis pola belanja pelanggan dan analisis trend produk.

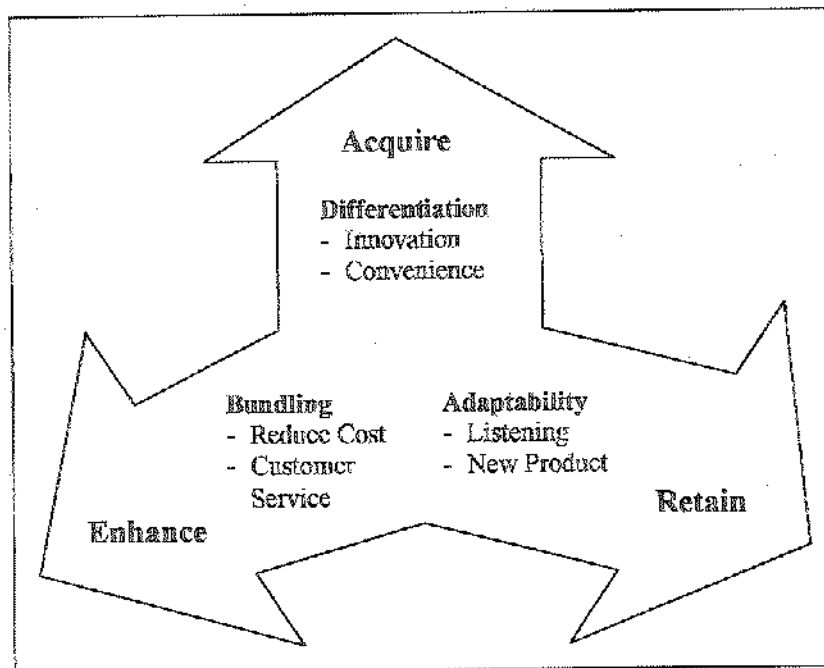
2.3.5 Aspek CRM dari Sudut Pandang Pelanggan

Berfokus kepada pelanggan berarti memiliki kekonsistenan, ketergantungan dan interaksi yang menyenangkan dengan pelanggan pada setiap pertemuan. Penerapan strategi bisnis CRM pada kenyataannya membutuhkan pengembangan beberapa aplikasi yang terintegrasi mencakup semua aspek dari kebutuhan *front office*, seperti kebutuhan untuk mengotomatisasi pelayanan pelanggan, penjualan dan pemasaran. Dengan menerapkan aplikasi CRM, perusahaan mengharapkan untuk membuat program penyimpanan informasi pelanggan yang lebih baik guna memaksimalkan pendapatan.

Ada tiga aspek dari CRM (Gambar 2.3) yaitu : *acquisition, enhancement* dan *retention*. Masing-masing memiliki dampak yang berbeda terhadap hubungan dengan pelanggan (Kalakota dan Robinson, 1999, p113).

1. Mendapatkan pelanggan baru (*acquisition*). Memperoleh pelanggan baru dengan menawarkan produk atau pelayanan yang mendorong peningkatan *performance*, kenyamanan dan inovasi.
2. Meningkatkan keuntungan dari pelanggan yang telah ada (*enhancement*). Meningkatkan nilai dari pelanggan yang telah ada dengan menjual lebih banyak produk lagi kepada mereka sehingga dapat meningkatkan pendapatan. Hal ini juga mendorong terjalinnya hubungan yang lebih erat dengan pelanggan. Aspek ini berkaitan dengan *cross-selling* dan *up-selling*.

3. Mempertahankan pelanggan yang menguntungkan (*retention*). Dengan berfokus pada pelayanan dalam segala situasi dan dengan menawarkan apa yang diinginkan pelanggan dapat menghindari kehilangan pelanggan.



Gambar 2.3 Aspek CRM

Sumber : Kalakota dan Robinson, 1999, p114

2.3.6 Hubungan CRM dengan *Data Mining*

Data mining dalam kaitannya dengan CRM (Berson, 2000) merupakan sistem yang mengatur hubungan antara perusahaan dengan pelanggannya. Sistem tersebut harus dapat melakukan segmentasi pasar termasuk di dalamnya pelanggan serta prospeknya. Aplikasi *data mining* membantu menemukan pola-pola yang berguna untuk prediksi terhadap pelanggan. Hal yang harus diperhatikan adalah *data mining* dan CRM harus terintegrasi dengan baik.

Data mining membantu dalam menemukan target pemasaran, penjualan dan pelayanan yang akurat terhadap pelanggan. Umumnya *data mining* mencari jawaban dari pertanyaan-pertanyaan yang seperti berikut, yaitu :

1. Pelanggan mana yang memiliki prospek yang baik untuk penawaran produk baru?
(Aspek *acquisition* dari CRM)
2. Produk mana yang terbaik untuk dapat ditawarkan kepada pelanggan yang telah ada?
(Aspek *enhancement* dari CRM)
3. Pelanggan mana yang cenderung akan beralih dari perusahaan? (Aspek *retention* dari CRM)

Jawaban terhadap pertanyaan-pertanyaan serupa di atas dapat meningkatkan hubungan dengan pelanggan dan meningkatkan *Return On Investment (ROI)* bagi perusahaan.

Tujuan dari *data mining* adalah membuat model-model bagi pengambil keputusan yang memperkirakan respon yang akan terjadi dengan berdasarkan analisis terhadap kejadian-kejadian yang telah lalu.

Aplikasi *data mining* dilihat dari sudut pandang CRM terdiri atas :

1. *Customer retention*. Program *customer retention* berawal dari pembuatan model terhadap pelanggan yang pernah beralih dari perusahaan dengan mengidentifikasi pola-pola penyebabnya. Kemudian model tersebut diaplikasikan pada pelanggan saat ini yang mempunyai kesamaan faktor dengan pola penyebab tersebut sehingga dapat dilakukan tindakan pencegahan.
2. *Sales dan customer service*. Jika informasi pelanggan tersedia, maka program yang dibuat berdasarkan *rules* dapat merekomendasikan produk secara otomatis. Program ini seperti *market-basket analysis* yaitu analisis dari *database* transaksi untuk

menemukan sekumpulan produk yang sering muncul secara bersamaan dalam satu transaksi pembelian. Analisis ini telah dimanfaatkan dalam menentukan rasio *cross-selling*, perbaikan penempatan produk di rak-rak barang dan lantai, serta perbaikan *layout* dari katalog produk pada *website*.

3. *Marketing*. Dengan memiliki informasi yang akurat tentang pelanggan, maka dapat dilakukan kampanye mempertahankan pelanggan, analisis *lifetime value* dari pelanggan, dan membuat target promosi yang tepat sehingga dapat mengurangi biaya pemasaran.
4. *Risk assessment* dan *fraud detection*. Dengan memiliki data pelanggan maka perusahaan dapat mengurangi resiko yang berkaitan dengan pola pembayaran (*payment pattern*) serta dapat mendeteksi tindakan-tindakan ilegal yang mungkin dilakukan oleh pelanggan. Umumnya program ini digunakan oleh perusahaan asuransi dan bank dalam memberikan pinjaman.

2.4 Cross-selling sebagai Salah Satu Strategi Dalam CRM

2.4.1 Definisi Cross-selling

Cross-selling adalah proses penawaran produk-produk dan pelayanan-pelayanan baru kepada pelanggan yang telah ada (Berson, 2000, p264). Salah satu bentuk dari *cross-selling* adalah *up-selling* yaitu proses penawaran produk kepada pelanggan yang telah ada berkaitan dengan transaksi pelanggan tersebut yang telah lalu.

Menurut Kalakota dan Robinson (1999) *cross-selling* merupakan suatu bentuk strategi yang menguntungkan bagi departemen pemasaran pada perusahaan. Dengan mengimplementasikan sebuah strategi *cross-selling* lengkap dengan aplikasi yang

diperlukan untuk mengetahui hubungan pelanggan dengan perusahaan dan mengidentifikasi tipe pelanggan dapat menambah penjualan perusahaan.

Menurut Kalakota dan Robinson (1999) strategi *up-selling* tergantung dengan terintegrasinya kuantitas persediaan barang dan dukungan pelayanan pelanggan. *Up-selling* juga memberikan pengaruh bagi departemen pemasaran dalam usaha untuk meningkatkan penjualan produk dengan penambahan kuantitas produk yang dilihat dari data-data pelanggan mengenai status pelanggan.

Cross-selling merupakan salah satu penerapan dari aspek *enhancement* dalam CRM karena tujuannya adalah untuk meningkatkan nilai dari pelanggan yang telah ada dengan menjual lebih banyak produk kepada pelanggan tersebut. *Cross-selling* akan menghasilkan *win-win solution* dimana perusahaan dan pelanggan sama-sama memperoleh keuntungan. Perusahaan memperoleh keuntungan dari penjualan yang meningkat sehingga meningkatkan pendapatan. Sedangkan pelanggan memperoleh keuntungan karena memperoleh barang atau pelayanan sesuai dengan kebutuhan pelanggan tersebut.

2.4.2 Hubungan Cross-selling dengan Data Mining

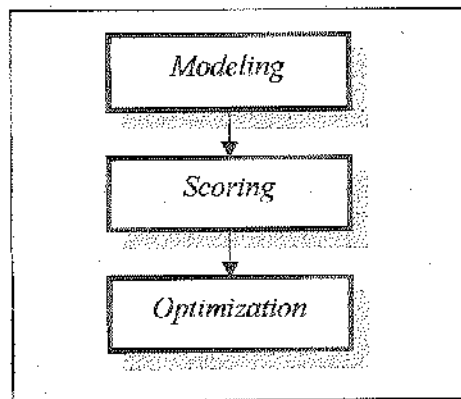
Penggunaan *data mining* untuk *cross-selling* diawali dengan data mengenai perilaku pembelian pelanggan (*customer buying behavior*) yang dilakukan oleh pelanggan yang telah ada untuk menghasilkan model penawaran *cross-selling* untuk setiap produk atau pelayanan. Kemudian model tersebut diterapkan pada data baru pelanggan yang akan diprediksi. Kunci pokok dalam *cross-selling* adalah pengoptimalan penawaran produk kepada semua pelanggan sehingga penawaran yang diterima oleh

pelanggan memberikan keuntungan maksimal bagi pihak perusahaan dan pihak pelanggan.

2.4.3 Tahapan Perancangan *Data Mining* untuk *Cross-selling*

Tahapan-tahapan dalam perancangan *data mining* untuk *cross-selling* dibagi menjadi tiga tahap (Berson, 2000, pp268-269). Tahap-tahap yang dilakukan adalah :

1. *Modeling*. Merupakan suatu proses dimana algoritma data mining menganalisis data, membuat persamaan matematika (model) yang dapat digunakan untuk memprediksi perilaku pembelian pelanggan (*customer buying behavior*). Untuk *cross-selling* diperlukan sebuah model untuk setiap penawaran.
2. *Scoring*. Setelah model dibuat, setiap model dapat diterapkan pada data pelanggan yang ingin diprediksi. Nilai prediksi (*score*) yang dihasilkan adalah *output* dari model tersebut yang menyatakan probabilitas pendugaan dari seorang pelanggan akan membeli produk tersebut di masa yang akan datang. *Scoring* diakhiri dengan sebuah matriks nilai dengan satu baris untuk setiap pelanggan dan satu kolom untuk setiap *score* dari penawaran *cross-selling*.
3. *Optimization*. Optimisasi terhadap matriks hasil *scoring* dilakukan dengan cara memilih yang terbaik dari segala kemungkinan penawaran yang ada untuk setiap pelanggan.



Gambar 2.4 Tahapan Perancangan *Data Mining* Untuk *Cross-selling*
 Sumber : Berson, 2000, p268

2.5 Database dan Structured Query Language (SQL)

2.5.1 Database

Dalam pengertian umum *database* diartikan sebagai gabungan dari elemen-elemen data yang berhubungan dan terorganisir. *Database* terbagi dalam beberapa kategori umum, yaitu :

1. *Paper-based*. Merupakan *database* yang paling sederhana yang disimpan dalam bentuk kumpulan kertas dokumen yang terorganisasi.
2. *Legacy Mainframe*. Biasa dikenal dengan *Virtual Storage Access Method (VSAM)*. *Legacy Mainframe* menggunakan kemampuan *mainframe* untuk melakukan proses penyimpanan dan pengaksesan data.
3. *Dbase*. Mengandung *Index Sequential Access Method (ISAM)* yang merupakan metode pengaksesan data secara berurutan yang memiliki *index*. Pada umumnya menggunakan file yang terpisah untuk setiap tabelnya. Contoh dari *database* yang menggunakan sistem ini adalah *Dbase*, *Foxpro*, *Microsoft Access*, *Paradox*.

4. *Relational Database Management System (RDBMS)*. Merupakan sistem *database* untuk jumlah *user* yang besar dengan integritas data yang lebih baik.. Struktur perintahnya disebut dengan *Structured Query Language (SQL)*.
5. *Object-oriented Database*. Menggunakan sistem objek dalam penyimpanan data. Data disimpan bukan dalam bentuk tabel melainkan dalam bentuk objek-objek yang terpisah.

2.5.2 *Structured Query Language (SQL)*

SQL merupakan bahasa standart untuk pengolahan *database*, sering disebut dengan *sequel* saja. SQL mulai dikembangkan pada akhir tahun 70-an di Laboratorium IBM, San Jose, California (Kurniawan, 2001, p85).

SQL adalah bahasa generasi ke-4 yang relatif lebih mudah untuk dipelajari dan dipahami karena perintah-perintahnya dinyatakan dalam bahasa yang sederhana dan mempunyai struktur linier dalam baris yang berurutan (Gaos, 2001, p57).

SQL terbagi menjadi dua komponen yaitu : *Data Definition Language (DDL)* dan *Data Manipulation Language (DML)*. DDL adalah perintah-perintah untuk mendefinisikan data. DDL mencakup perintah-perintah *create*, *alter* dan *drop*. Sedangkan DML mencakup perintah-perintah untuk memanipulasi data yang telah didefinisikan sebelumnya. DML mencakup perintah-perintah *select*, *insert*, *update* dan *delete*.

2.6 On Line Transaction Processing (OLTP)

OLTP adalah suatu proses yang menyediakan mekanisme transaksi pada suatu *database*. Pada proses ini, *design* pada *database* harus bersifat cepat untuk *transaction* (*insert, update, delete*). Hal ini bisa didapat dengan cara menormalisasi tabel-tabel pada *database*.

Keuntungan dari normalisasi adalah mengurangi redudansi data yang terdapat pada tabel-tabel. Hal ini memungkinkan proses transaksi yang cepat, karena data yang perlu di-*update* pada tabel menjadi lebih sedikit.