

BAB II

TINJAUAN PUSTAKA

II.1. Sistem Informasi

II.1.1. Sistem

Sistem adalah sekumpulan komponen yang saling berhubungan, bekerja sama untuk mencapai suatu tujuan dengan menerima input serta menghasilkan output dalam proses transformasi yang teratur .

Sistem adalah sebuah tatanan yang terdiri atas sejumlah komponen fungsional (dengan tugas/fungsi khusus) yang saling berhubungan dan secara bersama-sama bertujuan untuk memenuhi suatu proses/pekerjaan tertentu. Sebagai contoh, sistem kendaraan terdiri dari: komponen starter, komponen pengapian, komponen penggerak, komponen pengerem, komponen kelistrikan- speedometer, lampu dan lain-lain. Komponen-komponen tersebut diatas memiliki tujuan yang sama yaitu untuk membuat kendaraan tersebut bisa dikendarai dengan nyaman dan aman. Contoh lain yaitu sistem perguruan tinggi, yang terdiri dari dosen, mahasiswa, kurikulum, dan lain-lain. Sistem ini bertujuan untuk menghasilkan mahasiswa-mahasiswa yang memiliki kemampuan di bidang ilmunya. (Jogiyanto ; 2005: 2)

II.1.2. Konsep Dasar Sistem

Terdapat dua kelompok pendekatan di dalam mendefinisikan sistem, yaitu menekankan pada prosedurnya dan yang menekankan pada komponen atau

elemennya. Pendekatan sistem yang lebih menekankan pada prosedur mendefinisikan sistem sebagai berikut.

Suatu sistem adalah suatu jaringan kerja dari prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau menyelesaikan suatu sasaran yang tertentu.

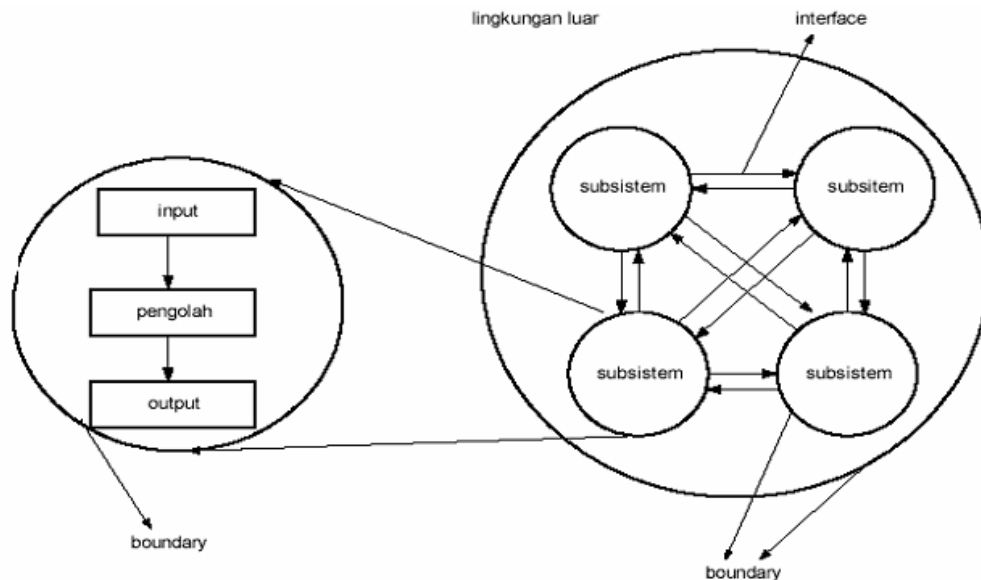
Pendekatan sistem yang merupakan jaringan kerja dari prosedur lebih menekankan urutan operasi di dalam sistem. Prosedur didefinisikan oleh Richard F. Neuschel sebagai berikut ini.

Suatu prosedur adalah suatu urutan operasi klerikal (tulis menulis), biasanya melibatkan beberapa orang di dalam satu atau lebih department, yang diterapkan untuk menjamin penanganan yang seragam dari transaksi-transaksi bisnis yang terjadi.

Pendekatan sistem yang menekankan pada komponen akan lebih mudah di dalam mempelajari suatu sistem untuk tujuan analisis dan perancangan suatu sistem. Untuk menganalisis dan merencanakan suatu sistem, analisis dan perancangan sistem harus dimengerti terlebih dahulu mengenai komponen-komponen atau elemen-elemen atau subsistem-subsistem dari sistem tersebut. (Jogiyanto ; 2005: 2)

II.1.3. Karakteristik Sistem

Karakteristik sistem dapatlah digambarkan sebagai berikut :



Gambar II.1. Karakteristik Sistem
(Sumber : Jogiyanto; 2005:6)

1. Komponen Sistem (*Components*)

Suatu sistem terdiri dari sejumlah komponen yang saling berinteraksi, yang artinya saling bekerja sama membentuk satu kesatuan. Komponen-komponen sistem atau elemen-elemen sistem dapat berupa suatu subsistem atau bagian-bagian dari sistem. Setiap sistem tidak peduli betapapun kecilnya, selalu mengandung komponen-komponen atau subsistem-subsistem. Setiap subsistem mempunyai sifat-sifat dari sistem untuk menjalankan suatu fungsi tertentu dan mempengaruhi proses sistem secara keseluruhan. Jadi, dapat dibayangkan jika dalam suatu sistem ada subsistem yang tidak berjalan/berfungsi sebagaimana mestinya. Tentunya sistem tersebut tidak akan berjalan mulus atau mungkin juga sistem

tersebut rusak sehingga dengan sendirinya tujuan sistem tersebut tidak tercapai.

2. Batas Sistem (*Boundary*)

Batas sistem (boundary) merupakan daerah yang membatasi antara suatu sistem dengan sistem yang lainnya atau dengan lingkungan luarnya. Batas sistem ini memungkinkan suatu sistem dipandang sebagai satu kesatuan. Batas suatu sistem menunjukkan ruang lingkup (scope) dari sistem tersebut.

3. Lingkungan Luar Sistem (*Environments*)

Lingkungan luar dari suatu sistem adalah apapun diluar batas dari sistem yang mempengaruhi operasi sistem. Lingkungan luar sistem dapat bersifat menguntungkan dan dapat juga bersifat merugikan sistem tersebut. Lingkungan luar yang menguntungkan merupakan energi dari sistem dan dengan demikian harus tetap dijaga dan dipelihara. Sedang lingkungan luar yang merugikan harus ditahan dan dikendalikan, kalau tidak maka akan mengganggu kelangsungan hidup dari sistem.

4. Penghubung (*Interface*)

Penghubung merupakan media penghubung antara satu subsistem dengan subsistem lainnya. Melalui penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke yang lainnya. Keluaran (output) dari satu subsistem akan menjadi masukan (input) untuk subsistem lainnya dengan melalui penghubung. Dengan penghubung satu

subsistem dapat berintegrasi dengan subsistem yang lainnya membentuk satu kesatuan.

5. Masukan (*Input*)

Masukan adalah energi yang dimasukkan ke dalam sistem. Masukan dapat berupa masukan perawatan (*maintenance input*) dan masukan sinyal (*signal input*). Maintenance input adalah energi yang dimasukkan supaya sistem tersebut dapat beroperasi. Signal input adalah energi yang diproses untuk didapatkan keluaran. Sebagai contoh didalam sistem komputer, program adalah *maintenance input* yang digunakan untuk mengoperasikan komputernya dan data adalah signal input untuk diolah menjadi informasi.

6. Keluaran (*Output*)

Keluaran adalah hasil dari energi yang diolah dan diklasifikasikan menjadi keluaran yang berguna dan sisa pembuangan. Keluaran dapat merupakan masukan untuk subsistem yang lain atau kepada super sistem. Misalnya untuk sistem komputer, panas yang dihasilkan adalah keluaran yang tidak berguna dan merupakan hasil sisa pembuangan, sedang informasi adalah keluaran yang dibutuhkan.

7. Pengolah (*Process*)

Suatu sistem dapat mempunyai suatu bagian pengolah yang akan merubah masukan menjadi keluaran. Suatu sistem produksi akan mengolah masukan berupa bahan baku dan bahan-bahan yang lain menjadi keluaran berupa barang jadi. Sistem akuntansi akan mengolah data-data transaksi

menjadi laporan-laporan keuangan dan laporan-laporan lain yang dibutuhkan oleh manajemen.

8. Sasaran (*Objectives*) atau Tujuan (*Goal*)

Suatu sistem pasti mempunyai tujuan atau sasaran. Kalau suatu sistem tidak mempunyai sasaran, maka operasi sistem tidak akan ada gunanya. Sasaran dari sistem sangat menentukan sekali masukan yang dibutuhkan sistem dan keluaran yang akan dihasilkan sistem. Suatu sistem dikatakan berhasil bila mengenai sasaran atau tujuannya. Perbedaan suatu sasaran (*objectives*) dan suatu tujuan (*goal*) adalah, *goal* biasanya dihubungkan dengan ruang lingkup yang lebih luas dan sasaran dalam ruang lingkup yang lebih sempit. Bila merupakan suatu sistem utama, seperti misalnya sistem bisnis perusahaan, maka istilah *goal* lebih tepat diterapkan. Untuk sistem akuntansi atau sistem-sistem lainnya yang merupakan bagian atau subsistem dari sistem bisnis, maka istilah *objectives* yang lebih tepat. Jadi tergantung dari ruang lingkup mana memandang sistem tersebut. Seringkali tujuan (*goal*) dan sasaran (*objectives*) digunakan bergantian dan tidak dibedakan. (Jogiyanto ; 2005: 4)

II.1.4. Informasi

Informasi adalah data yang dapat dianalogikan dengan data – data , yang belum di kelolah dan harus diolah untuk menjadi informasi yang akurat. Agar informasi yang penulis sajikan lebih bermanfaat maka terlebih dahulu dibuat aliran informasi yang lebih jelas dan lengkap. Berkaitannya dengan penyedia

informasi bagi manajemen dalam mengambil suatu keputusan, yang diperoleh harus berkualitas, maka kualitas dari informasi tergantung :

1. Akurat

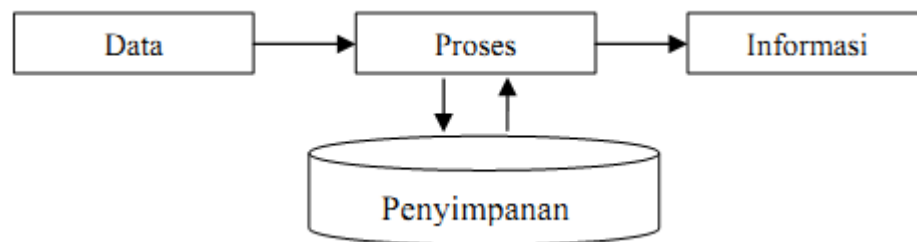
Akurat berarti bahwa informasi harus bebas dari kesalahan - kesalahan dan tidak biasa (menyesatkan) dan jelas mencerminkan maksudnya. Informasi harus akurat karena dari sumber informasi sampai ke penerimaan informasi kemungkinan banyak terjadi gangguan yang dapat merubah informasi atau merusak informasi tersebut.

2. Relevansi

Berarti informasi tersebut mempunyai manfaat untuk pemakaiannya. Relevansi informasi untuk tiap-tiap orang yang satu dengan yang lainnya berbeda.

3. Tepat waktu

Tepat waktu berarti bahwa informasi yang datang pada penerimaan tidak boleh terlambat. Informasi yang sudah usang tidak akan mempunyai nilai lagi. Karena informasi merupakan landasan dalam pengambilan keputusan. Dewasa ini mahalnya nilai informasi disebabkan harus cepatnya informasi itu di dapat, sehingga diperlukan teknologi-teknologi yang muktahir untuk mendapatkan, mengolah dan mengirimkannya. (Jogiyanto ; 2005: 10) Hal ini dapat dilihat seperti gambar dibawah ini :



Gambar II.2. Proses Data Menjadi Informasi
(Sumber : Kusrini ; 2008:15)

Sedangkan Sistem Informasi adalah kumpulan suatu sistem jaringan yang menghubungkan beberapa jaringan, yang mana telah dirangkai dengan sempurna dengan suatu struktur yang terbentuk untuk melaksanakan suatu kegiatan atau fungsi utama suatu perusahaan (Vervolt Keith ; 2003: 105).

Sistem informasi adalah sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, membantu dan mendukung kegiatan operasi, bersifat manajerial dari suatu organisasi dan membantu mempermudah penyediaan laporan yang diperlukan. Sistem informasi dapat juga dianggap sebagai suatu departemen dan diperlukan pemisahan fungsi, dimana fungsi – fungsi tersebut dapat terdiri dari sistem analis, programmer, operator, yang dibawah oleh manajer sistem informasi (Zaki Baridwan ; 2009: 19).

Berdasarkan definisi di atas dapat disimpulkan bahwa sistem informasi adalah sekumpulan prosedur organisasi yang pada saat dilaksanakan akan memberikan informasi bagi pengambilan keputusan atau untuk pengendalian

berinteraksi untuk menghasilkan informasi yang menunjang pengambilan keputusan operasi, manajerial, strategik.

II.2. Sistem Bisnis Cerdas

II.2.1. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar. (Turban et al ; 2005). Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose ; 2005).

Selain definisi di atas beberapa definisi juga diberikan seperti, “*data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.” (Pramudiono; 2006). “*Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.”(Pramudiono; 2006).

“*Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.” (Larose; 2005). “*Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.” (Larose; 2005).`

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain : (Larose; 2005)

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang baik.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Berdasarkan definisi-definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Hubungan yang dicari dalam *data mining* dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya dalam dimensi produk, dapat di lihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungan juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek. (Ponniah; 2001).

Sementara itu, penemuan pola merupakan keluaran lain dari *data mining*. Misalkan sebuah perusahaan yang akan meningkatkan fasilitas kartu kredit dari pelanggan, maka perusahaan akan mencari pola dari pelanggan-pelanggan yang ada untuk mengetahui pelanggan yang potensial dan pelanggan yang tidak potensial.

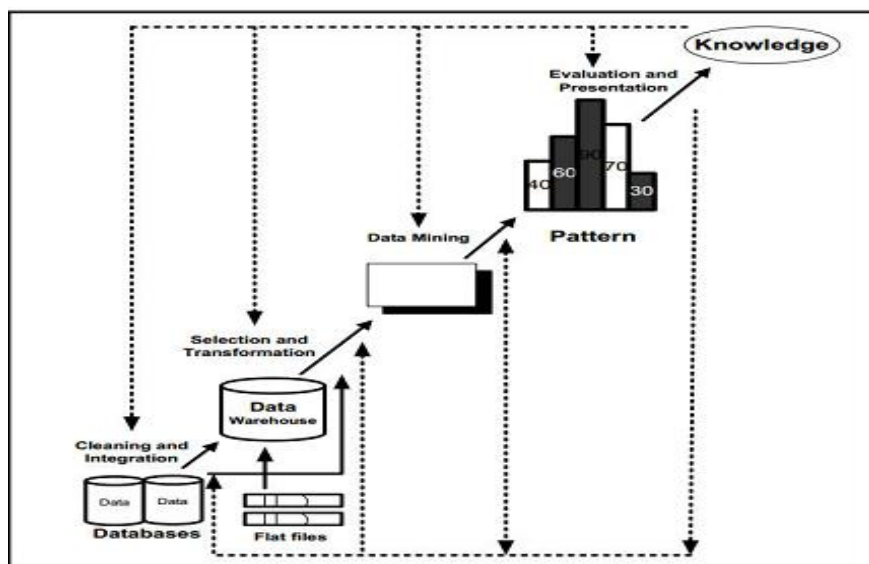
Beberapa definisi awal dari *data mining* meyertakan fokus pada proses otomatisasi. Berry dan Linoff, (2004) dalam buku *Data Mining Technique for Marketing, Sales, and Customer Support* mendefinisikan *data mining* sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semi otomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti (Larose; 2005).

Tiga tahun kemudian, dalam buku *Mastering Data Mining* mereka memberikan definisi ulang terhadap pengertian *data mining* dan memberikan pernyataan bahwa “jika ada yang kami sesalkan adalah frasa secara otomatis maupun semi otomatis, karena kami merasa hal tersebut memberikan focus berlebih pada teknik otomatis dan kurang pada eksplorasi dan analisis”. Hal tersebut memberikan pemahaman yang salah bahwa *data mining* merupakan produk yang dapat dibeli dibandingkan keilmuan yang harus dikuasai (Larose; 2005).

Pernyataan tersebut menegaskan bahwa dalam *data mining* otomatisasi tidak menggantikan campur tangan manusia. Manusia harus ikut aktif dalam setiap fase dalam proses *data mining*. Kehebatan kemampuan algoritma *data mining* yang terdapat dalam perangkat lunak analisis yang terdapat saat ini memungkinkan terjadinya kesalahan penggunaan yang berakibat fatal. Pengguna mungkin menerapkan analisis yang tidak tepat terhadap kumpulan data dengan menggunakan pendekatan yang berbeda. Oleh karenanya, dibutuhkan pemahaman tentang statistik dan struktur model matematika yang mendasari kerja perangkat lunak (Larose; 2005).

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstrasi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep –konsep yang akan dipelajari.

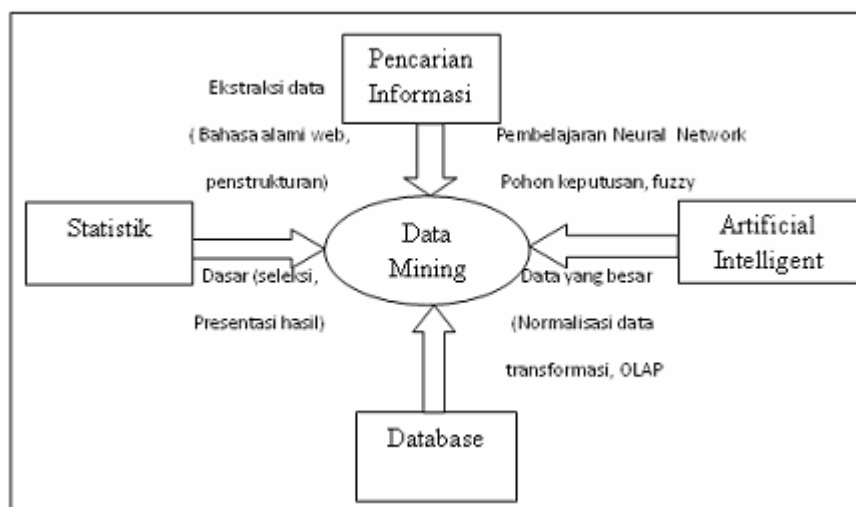
Knowledge Discovery in database (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini *data mining* merupakan satu langkah dari proses KDD.



Gambar II.3. Proses Data Mining
(Sumber : *Beta Noranita dan Nurdin Bahtiar, 2010*)

Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya. Alasan menggunakan data mining dari sudut pandang komersial karena (1) meledaknya volume data yang dihimpun dan disimpan dalam data warehouse seperti data *web*, *e-commerce*, penjualan di *department store*, transaksi bank/*credit card*, (2) proses komputasi yang dapat diupayakan, (3) kuatnya tekanan kompetitif untuk dapat menyediakan yang lebih baik,

layanan-layanan custom-isasi dan informasi sedang menjadi produk yang berarti. Operasi data mining menurut sifatnya dibedakan menjadi dua, yaitu bersifat (1) prediksi (*Prediction driven*) untuk menjawab pertanyaan apa dan sesuatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, querying dan pelaporan (misal: *spreadsheet dan pivot table*), analisis multidimensi (*dimensional summary*); OLAP (Online Analytic processing) serta analisis statistic. (2) penemuan (*discovery driven*) bersifat transparan dan untuk menjawab pertanyaan “mengapa?”. Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi *database*, analisis keterkaitan (*link analysis*) dan deteksi deviasi. (Fajar Astuti Hermawati ; 2009 :3).



Gambar II.4. Bidang Ilmu Data Mining
Sumber: (Kusrini, Emha Taufiq Luthfi, 2009)

Selain definisi di atas beberapa definisi juga diberikan seperti, *Data Mining*, sering juga disebut *knowledge discovery in database*, 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* sekarang jarang digunakan karena ia termasuk bagian dari data mining. (Budi Santoso;2007:10)

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. (Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2016; Heroe Santoso, I Putu Hariyadi, Prayitno)

Data mining, sering juga disebut sebagai knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. (Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2016; Heroe Santoso, I Putu Hariyadi, Prayitno)

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. (Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2016; Heroe Santoso, I Putu Hariyadi, Prayitno)

II.2.2. Teknik Data Mining

Beberapa teknik dan sifat data mining menurut (Fajar Astuti Hermawati ; 2009 :14). adalah sebagai berikut :

1. Klasifikasi

Klasifikasi adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori (atau klas) yang telah di definisikan sebelumnya. Disebut juga dengan '*Supervised learning*'. Berikut beberapa aplikasi dari klasifikasi :

a). Penjualan langsung (*direct marketing*)

Tujuan : mengurangi costsurat meyurat dengan menentukan (*targeting*) satu set konsumen yang mempunyai kesamaan dalam membeli produk telepon selular baru.

Pendekatan :

- Gunakan data penjualan untuk suatu produk telepon selular
- Kita mengetahui pelanggan yang memutuskan untuk membeli dan yang memutuskan untuk tidak membeli. Keputusan $\{buy, don't buy\}$ ini menentukan *class attribute*.
- Himpunan bermacam demografi, gaya hidup dan *company-interaction* sehubungan dengan informasi mengenai pelanggan tertentu. Misalkan: Tipe bisnis, dimana mereka tinggal, beberapa banyak mereka membayar, dll.
- Gunakan informasi tersebut sebagai atribut *input* untuk mempelajari suatu model klasifikasi.

b). Fraud Detection

Tujuan : memprediksi kasus-kasus transaksi curang dengan menggunakan kartu kredit.

Pendekatan :

- Gunakan transaksi kartu kredit dan informasi pemegang kartu kredit sebagai atributnya. Misalkan : kapan seorang pelanggan membeli, apa yang di beli, apa selalu membayar tepat waktu,dsb.
- Beri label-label transaksi sebelumnya sebagai transaksi '*fraud*' atau '*fair*' dan bentuk ini menjadi *class attribute*.
- Pelajari satu model untuk *class* transaksi tersebut.
- Gunakan model ini untuk mendeteksi kecurangan dengan mengobservasi transaksi kartu kredit tiap *account*.

c). Customer Attrition/Churn

Tujuan : untuk memprediksi pelanggan mana yang akan berpindah ke competitor kita.

Pendekatan :

- Gunakan *record* transaksi dengan pelanggan yang lalu maupun yang sekarang untuk mendapat atribut, seperti: seberapa sering pelanggan menghubungi, dimana dia menghubungi, pada hari apa dia paling

- sering menghubungi, status keuangannya, status perkawinannya, dsb.
- Beri label pelanggan sebagai 'setia' atau 'tidak setia'.
- Temukan suatu model untuk 'loyalty'.

2. Regresi

Memprediksi nilai dari suatu variable kontinyu yang diberikan berdasarkan nilai dari variable yang lain, dengan mengamsusikan sebuah model ketergantungan linier atau nonlinier. Teknik ini banyak dipelajari dalam statistika, bidang jaringan saraf tiruan (*neural network*). Contoh aplikasinya :

- Memprediksi jumlah penjualan produk baru berdasarkan pada belanja promosi/iklan
- Memprediksi kecepatan angin sebagai suatu fungsi suhu, kelembapan, tekanan udara, dsb.
- *Time Series prediction* dari indeks *stock market*.

3. Klustering

Mempartisi data-set menjadi beberapa sub-set atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set properti yang dishare bersama, dengan tingkat similitas yang tinggi dalam satu kelompok dan tingkat similitas antar kelompok yang rendah. Disebut juga dengan '*unsupervised learning*'.

Jika diberikan sejumlah titik data yang masing-masing mempunyai sejumlah atribut dan dengan menggunakan satu ukuran similaritas, dapat ditemukan klaster-klaster sedemikian sehingga:

- Titik-titik data dalam satu klaster mempunyai similaritas yang lebih besar
- Titik-titik data dalam klaster yang berbeda mempunyai similaritas yang kecil.

Ukuran similaritas yang digunakan

- *Euciledean Distance* jika atributnya kontinyu
- Permasalahan lain – ukuran tertentu

Aplikasi dari klasterisasi diantaranya adalah :

a). Market Segmentation:

Tujuan : membagi pasar kedalam sub-set pelanggan yang berbeda, dimana suatu sub-set mungkin dapat dipilih sebagai target pasar yang dicapai dengan satu kombinasi pemasaran yang berbeda.

Pendekatan :

- Kumpulkan atribut dari pelanggan yang berbeda berdasarkan pada informasi tempat tinggal dan gaya hidup.
- Tentukan klaster dari pelanggan-pelanggan yang sama.

- Hitung kualitas klaster dengan mengobservasi pola daya beli pelanggan pada klaster yang sama versus dari klaster yang berbeda.

b). Document Clustering:

Tujuan : untuk mendapatkan kelompok dokumen yang mempunyai kesamaan berdasarkan pernyataan atau kata-kata penting yang muncul dalam dokumen tersebut.

Pendekatan :

Untuk mengenali kata-kata yang sering muncul dalam tiap dokumen. Dari suatu pengukuran similaritas yang didasarkan pada frekuensi *term* yang berbeda. Gunakan pengukuran ini untuk membentuk klaster-klaster.

Pencapaian : *Information Retrieval* dapat dimanfaatkan untuk menghubungkan suatu dokumen baru atau mencari *term* ke dokumen –dokumen yang di klaster.

4. Kaidah Asosiasi

Mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut. Contoh : 90% orang yang berbelanja di suatu supermarket yang membeli roti juga membeli selai, dan 60% dari semua orang yang berbelanja membeli keduanya.

Jika diberikan sekumpulan *record* yang masing-masing terdiri dari sejumlah item dari kumpulan yang diberikan akan menghasilkan

aturan ketergantungan (*dependency rules*) yang akan memprediksi kejadian dari satu item berdasarkan kejadian item lainnya.

Contoh aplikasi kaidah asosiasi adalah sebagai berikut:

a. Marketing and sales promotion

Misalkan diketahui aturan ketergantungan dimana

{Bagels,...} → {Potato Chips}

Potato Chips sebagai consequent => dapat digunakan untuk menentukan apa yang dapat dilakukan untuk meningkatkan penjualan.

Bagels in the antecedent => dapat digunakan untuk melihat produk mana yang akan terkena dampak jika toko tersebut tidak lagi menjual *bagels*.

Bagels in antecedent and Potato chips in consequent => dapat digunakan untuk melihat produk apa yang harus dijual dengan *bagels* untuk mempromosikan penjualan *Potato Chips!*.

b. Supermarket shelf management

Tujuan : untuk mengenali item-item yang dibeli bersama-sama oleh cukup banyak pelanggan.

Pendekatan : memproses data *point-of-sale* yang dikumpulkan dengan pemindai *barcode* untuk menemukan ketergantungan antar-*item*.

Aturan klasik jika seorang pelanggan membeli diaper dan susu maka dia juga akan membeli *beer*. Sehingga jangan kaget jika anda akan menemukan enam pak *beer* yang ditunpuk dekat *diapers*.

c. Inventory Management

Tujuan : seorang pelanggan perusahaan-perbaikan-peralatan mengharapkan keaslian dari perbaikan produk konsumen dan menjaga pelayanan dengan menggunakan suku cadang yang baik untuk mengurangi jumlah kunjungan ke rumah pelanggan.

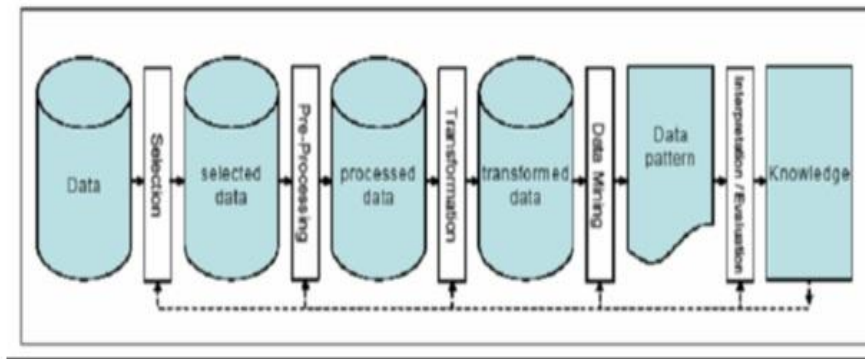
Pendekatan: memproses data peralatan dan suku cadang yang dibutuhkan pada perbaikan sebelumnya di tempat pelanggan yang berbeda dan menemukan pola-pola kejadian yang berulang.

5. Pencarian Pola Sekuensial (*Sequence Mining*)

Mencari sejumlah *event* yang secara umum terjadi bersama sama. contoh, dalam satu set urutan DNA, ACGTC diikuti oleh GTCA setelah suatu celah selebar 9 dengan probabilitas sebesar 30%.

Jika diberikan sekumpulan *obyek*, dengan masing-masing obyek dihubungkan dengan waktu kejadiannya maka dapatkan pola yang memprediksi ketergantungan sekuensial (*sequential dependencies*) yang kuat diantara kejadian-kejadian yang berbeda.

Pola-pola sekuensial pertama, pada dasarnya dibentuk dengan cara mencari semua kemungkinan pola yang ada. Nilai-nilai kejadian dalam pola diatur berdasarkan urutan waktu kejadian.



Gambar II.5. Proses dari Data Mining
(Sumber: *Statistical Package for the Social Science ; 2004*)

II.3. Algoritma Apriori

Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk pada golongan ini adalah metode Generalized Rule Induction dan Algoritma Hash Based. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis. Analisis asosiasi atau association rule adalah teknik data mining untuk menemukan aturan asosiatif antara suatu item. Algoritma Apriori yang bertujuan untuk menemukan frequent itemsets dijalankan pada sekumpulan data. Analisis Apriori didefinisikan suatu proses untuk menemukan semua aturan apriori yang memenuhi syarat minimum untuk support dan syarat minimum confidence. Support adalah nilai penunjang, atau persentase kombinasi sebuah item dalam

database. Rumus support sebagai berikut: $\text{Support (A)} = (\text{jumlah transaksi mengandung A} / \text{Total transaksi}) \times 100\%$. Sedangkan confidence adalah nilai kepastian yaitu kuatnya hubungan antar item dalam sebuah apriori. Confidence bisa dicari setelah pola frekuensi munculnya sebuah item ditemukan. Rumus untuk menghitung confidence adalah sebagai berikut: Contoh misal ditemukan aturan $A \Rightarrow B$ maka: $\text{Confidence } P(B|A) = (\text{Total transaksi mengandung A dan B} / \text{transaksi mengandung A}) \times 100\%$. (Kusrini;2009)

Algoritma apriori adalah salah satu algoritma yang melakukan pencarian frequent itemset dengan menggunakan teknik association rule. Untuk mencari association rule dari suatu kumpulan data, tahap pertama yang harus dilakukan adalah mencari frequent itemset terlebih dahulu. Frequent itemset adalah sekumpulan item yang sering muncul secara bersamaan. Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu support dan confidence. Support adalah nilai penunjang atau persentase kombinasi sebuah item dalam database, sedangkan confidence adalah nilai kepastian yaitu kuatnya hubungan antar item dalam aturan asosiasi. Proses utama yang dilakukan dalam algoritma apriori untuk mendapat frequent itemset yaitu :

b. Join (penggabungan).

Proses ini dilakukan dengan cara pengkombinasian item dengan yang item lainnya hingga tidak bisa terbentuk kombinasi lagi.

c. Prune (pemangkasan).

Proses pemangkasan yaitu hasil dari item yang telah dikombinasikan kemudian dipangkas dengan menggunakan minimum support yang telah ditentukan.

Konsep Aturan Asosiasi Rule.

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiasi antara kombinasi item. Contoh dari aturan asosiasi dari analisa pembelian di suatu pasar swalayan adalah dapat diketahui berapa besar kemungkinan seseorang 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. Metodologi dasar analisis asosiasi

a. Analisis Pola Frekuensi Tinggi

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

$$\text{Support A} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

Sedangkan nilai support dari 2 item diperoleh dari rumus 2 berikut.

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi Mengandung A dan B}}{\sum \text{Transaksi}}$$

b. Pembentukan Aturan Asosiasi

Pembentukan Aturan Asosiasi Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasif “ jika A maka B “. Nilai confidence dari aturan “ jika A maka B “ diperoleh dari rumus berikut :

$$\text{Confidence } P(B | A) = \frac{\sum \text{Transaksi Mengandung A dan B}}{\sum \text{Transaksi Mengandung A}}$$

Walaupun akhir-akhir ini dikembangkan banyak algoritma yang lebih efisien dari Apriori seperti FP-growth, LCM dsb, tetapi Apriori tetap menjadi algoritma yang paling banyak diimplementasikan dalam produk komersial untuk data mining karena apriori dianggap yang paling mapan.

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 ini, 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. Support disini artinya jumlah transaksi dalam database yang

mengandung kedua item dalam kandidat 2-itemset. Setelah support dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum support dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2. Untuk selanjutnya pada iterasi ke-k dapat dibagi lagi menjadi beberapa bagian :

- a). 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.
 - b). 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 algoritme Apriori dimana diperlukan penghitungan dengan scan seluruh database sebanyak k-item set terpanjang.
 - c). 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.
 - d). Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian 1.
- (Fajar Astuti Hermawati ; 2009)

II.4. Sekilas Tentang SQL Server 2008

II.4.1. Pengenalan Database, DBMS dan RDBMS

Basis data (atau *database*) adalah kumpulan data yang saling berelasi. Data sendiri merupakan fakta mengenai obyek, orang, dan lain-lain. Data dinyatakan dengan nilai (angka, deretan, karakter, atau simbol) (Kusrini ; 2007 : 2)

Basis data bertujuan untuk mengatur data sehingga diperoleh kemudahan, ketepatan, dan kecepatan dalam pengambilan kembali.

Banyak manfaat yang dapat kita peroleh dengan menggunakan basis data. Manfaat/kelebihan basis data diantaranya adalah :

1. Kecepatan dan kemudahan (*Speed*)

Dengan menggunakan basis data pengambilan informasi dapat dilakukan dengan cepat dan mudah. Basis data memiliki kemampuan untuk dalam mengelompokkan, mengurutkan bahkan perhitungan dengan matematika. Dengan perancangan yang benar, maka penyajian informasi akan dapat dilakukan dengan cepat dan mudah.

2. Kebersamaan pemakai (*sharability*)

Sebuah basis data dapat digunakan oleh banyak user dan banyak aplikasi. Untuk data-data yang diperlukan oleh banyak bagian/orang, tidak perlu dilakukan pencatatan di masing-masing bagian, tetapi cukup dengan satu basis data untuk dipakai bersama.

3. Pemusatan kontrol data

Karena cukup dengan satu basis data untuk banyak keperluan, pengontrolan terhadap data juga cukup dilakukan satu tempat saja. Jika adalah perubahan alamat mahasiswa misalnya, maka tidak perlu kita mengupdate semua data di masing-masing bagian tetapi cukup hanya di satu basis data.

4. Efisiensi ruang penyimpanan (*space*)

Dengan pemakaian bersama. Kita tidak perlu menyediakan tempat penyimpanan di berbagai tempat, tetapi cukup satu saja sehingga ini akan menghemat ruang penyimpanan yang dimiliki oleh sebuah organisasi. Dengan teknik perancangan basis data yang benar, kita akan dapat menyederhanakan penyimpanan sehingga tidak semua data harus disimpan.

5. Keakuratan (*Accuracy*)

Penerapan secara ketat aturan tipe data, domain data, keunikan data, hubungan antar data, dan lain-lain, dapat menekan ketidakakuratan dalam pemasukan/penyimpanan data.

6. Ketersediaan (*Availability*)

Dengan basis data kita dapat mem-backup data, memilah-milah data mana yang masi diperlukan dan data mana yang perlu kita simpan ke tempat lain. Hal ini mengingat pertumbuhan transaksi suatu organisasi dari waktu ke waktu membutuhkan media penyimpanan yang semakin besar.

7. Kelengkapan (*Completeness*)

8. Keamanan (*Security*)

Kebanyakan DBMS dilengkapi dengan fasilitas manajemen pengguna. Pengguna diberikan hak akses yang berbeda-beda sesuai dengan kepentingan dan posisinya. Basis data bisa diberikan password untuk membatasi orang yang mengaksesnya.

9. Kemudahan dalam pembuatan program aplikasi baru

Penggunaan basis data merupakan bagian dari perkembangan teknologi. Dengan adanya basis data pembuatan aplikasi bisa memanfaatkan kemampuan dari DBMS, sehingga pembuat aplikasi tidak perlu mengurus penyimpanan data, tetapi cukup mengatur interface untuk pengguna.

10. Pemakaian secara langsung

Basis data memiliki fasilitas untuk melihat datanya secara langsung dengan tool yang di sediakan oleh DBMS. Untuk melihat data, langsung ke tabel ataupun dengan menggunakan query. Biasanya yang menggunakan fasilitas ini adalah user yang sudah ahli, atau database administrator.

11. Kebebasan Data (*Data Independence*)

Jika sebuah program telah selesai dibuat, dan ternyata ada perubahan isi/struktur data. Maka dengan basis data, perubahan ini hanya perlu dilakukan pada level DBMS tanpa harus membongkar kembali program aplikasinya.

12. User view

Basis data menyediakan pandangan yang berbeda-beda untuk tiap-tiap pengguna. Misalnya kita memiliki data-data dari perusahaan yang bergerak di bidang retail. Data yang berupa data barang, penjualan, dan pembelian. Ada beberapa jenis pengguna yang memerlukan informasi terkait dengan data perusahaan tersebut. Mereka adalah pelanggan, kasir, bagian gudang, bagian akuntansi dan juga manajer. Tidak semua data dapat diakses pengguna.

Beberapa operasi dasar basis data, yaitu:

1. Pembuatan basis data
2. Penghapusan basis data
3. Pembuatan file/tabel
4. Penghapusan file/tabel
5. Pengubahan tabel
6. Penambahan/pengisian
7. Pengambilan data
8. Penghapusan data

DBMS merupakan suatu sistem perangkat lunak yang memungkinkan user (pengguna) untuk membuat, memelihara, mengontrol, dan mengakses database secara praktis dan efisien. Dengan DBMS, user akan lebih mudah mengontrol dan memanipulasi data yang ada.

Sedangkan RDBMS atau *Relationship Database Management System* merupakan salah satu jenis DBMS yang mendukung adanya *relationship* atau hubungan antar tabel. Di samping RDBMS, terdapat jenis DBMS lain, misalnya *Hierarchy DBMS*, *Object Oriented DBMS*, dan sebagainya. (Kusrini ; 2007 : 5)

1. Beberapa Istilah Database.

- a. *Table*

Sebuah tabel merupakan kumpulan data (nilai) yang diorganisasikan ke dalam baris (*record*) dan kolom (*field*). Masing-masing kolom memiliki nama yang spesifik dan unik.

- b. *Field*

Field merupakan kolom dari sebuah table. *Field* memiliki ukuran type data tertentu yang menentukan bagaimana data nantinya tersimpan.

- c. *Record*

Record merupakan sebuah kumpulan nilai yang saling terkait.

- d. *Key*

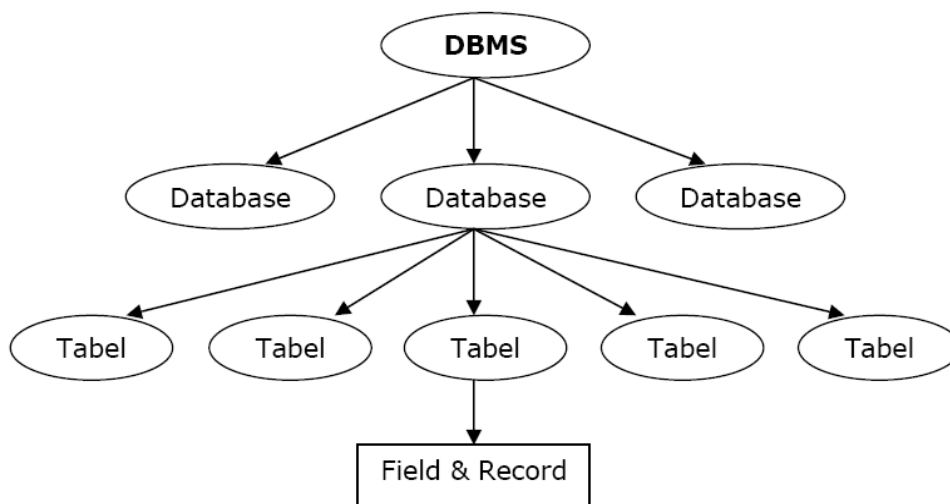
Key merupakan suatu field yang dapat dijadikan kunci dalam operasi tabel. Dalam konsep database, *key* memiliki banyak jenis diantaranya *Primary Key*, *Foreign Key*, *Composite Key*, dan lain-lain.

e. *SQL*

SQL atau *Structured Query Language* merupakan suatu bahasa (*language*) yang digunakan untuk mengakses database. *SQL* sering disebut juga sebagai *query*. (Achmad Solihin ; 2010 : 7)

2. Hierki Database

Dalam konsep *database*, urutan atau hierarki database sangatlah penting. Urutan atau hierarki *database* digambarkan seperti pada gambar II.7. dibawah ini



Gambar II.6. Hirarki Database
(Sumber : Achmad Solihin ; 2010:7)

II.4.2. SQL Server 2008

SQL Server 2008 R2 (dahulu nama kode SQL Server “Kilimanjaro”) diumumkan di TechEd 2009, dan dirilis untuk manufaktur pada 21 April 2010. SQL Server 2008 R2 menambahkan fitur tertentu untuk SQL Server

2008 termasuk data master manajemen sistem dicap sebagai Master Data Services , manajemen pusat data entitas master dan hirarki. Juga *Multi Server Management*, terpusat konsol untuk mengelola beberapa SQL Server 2008 *instance* dan jasa, termasuk database relasional, *Reporting Services*, *Analysis Services & Integration Services*. Adapun tampilan utama SQL Server 2008 R2 :



Gambar II.7. SQL Server 2008 R2
(Sumber : Achmad Solihin ; 2010:7)

SQL Server 2008 R2 meliputi sejumlah layanan baru, termasuk *PowerPivot* untuk *Excel* dan *SharePoint*, *Master Data Services*, *StreamInsight*, *Report Builder 3.0*, *Reporting Services Add-in* untuk *SharePoint*, sebuah lapis fungsi Data dalam Visual Studio yang memungkinkan kemas database berjenjang sebagai bagian dari aplikasi, dan *Utility SQL Server* yang bernama UC (*Utility Control Point*), bagian dari AMSM (Aplikasi dan *Multi-Server Management*) yang digunakan untuk mengelola beberapa SQL Server. *Microsoft SQL Server* tersedia dalam beberapa edisi, dengan set fitur yang berbeda dan menargetkan pengguna yang berbeda.

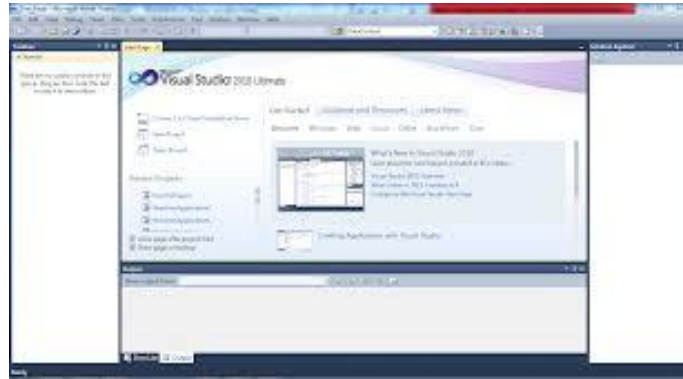
II.5. *Microsoft Visual Studio 2010*

Microsoft Visual Studio merupakan sebuah perangkat lunak lengkap (suite) yang dapat digunakan untuk melakukan pengembangan aplikasi, baik itu aplikasi bisnis, aplikasi personal, ataupun komponen aplikasinya, dalam bentuk aplikasi console, aplikasi Windows, ataupun aplikasi Web. Visual Studio mencakup kompiler, SDK, Integrated Development Environment (IDE), dan dokumentasi (umumnya berupa MSDN Library). Kompiler yang dimasukkan ke dalam paket Visual Studio antara lain Visual C++, Visual C#, Visual Basic, Visual Basic .NET, Visual InterDev, Visual J++, Visual J#, Visual FoxPro, dan Visual SourceSafe.

Microsoft Visual Studio dapat digunakan untuk mengembangkan aplikasi dalam native code (dalam bentuk bahasa mesin yang berjalan di atas Windows) ataupun managed code (dalam bentuk Microsoft Intermediate Language di atas .NET Framework). Selain itu, Visual Studio juga dapat digunakan untuk mengembangkan aplikasi Silverlight, aplikasi Windows Mobile (yang berjalan di atas .NET Compact Framework). (Wahana Komputer;2012).

Gambar berikut menampilkan tampilan utama Microsoft Visual Studio

2010 :



Gambar II.8. Tampilan Microsoft Visual Studio 2010