BAB II

TINJAUAN PUSTAKA

2.1 Profil Perusahaan

2.1.1 Sejarah Perusahaan

PD. Wijaya Abadi adalah dealer resmi sepeda motor Honda, yang mendapat distribusi dari PT. Daya Adira Mustika, sebagai main dealer regional Jawa Barat. PD Wijaya Abadi didirikan pada tanggal 30 Juni 1996 dan beralamat di Jl. Kopo Sayati Km.7 No.84.

PD. Wijaya Abadi bergerak dalam penjualan sepeda motor merk Honda. Dalam melayani pembelian para konsumennya, PD. Wijaya Abadi memberikan layanan penjualan secara tunai maupun kredit, sehingga memudahkan konsumen dalam melakukan transaksi pembelian sepeda motor. Untuk penjualan secara kredit, pihak perusahaan melakukan kerja sama dengan beberapa *leasing company*. Di antaranya : PT. Adira Finance, PT. Federal International Finance, PT. Summit Oto Finance, dan PT. WOM Finance.

Seiring dengan semakin berkembangnya dunia bisnis di Jawa Barat, maka PD. Wijaya Abadi berusaha menjadi dealer resmi sepeda motor Honda yang dapat bersaing secara kompeten dengan pesaingnya. Hal ini ditunjukan dengan beberapa prestasi yang telah kami capai selama perusahaan didirikan. Prestasi yang terakhir kami capai adalah ranking 3 “The Best Growth” se-Jawa Barat tahun 2006, yang dinilai berdasarkan besarnya peningkatan jumlah penjualan dari tahun ke tahun oleh PT. Daya Adira Mustika.

Sesuai dengan visi perusahaan kami, maka PD. Wijaya Abadi berusaha menjadi mitra bisnis terpercaya bagi para pebisnis di Jawa Barat, dengan memberikan pelayanan terbaik kami, sehingga konsumen merasa puas. Karena kepuasan konsumen adalah harapan dan tujuan kami.

2.1.2 Visi dan Misi Perusahaan

Visi

Menjadi dealer terbaik.

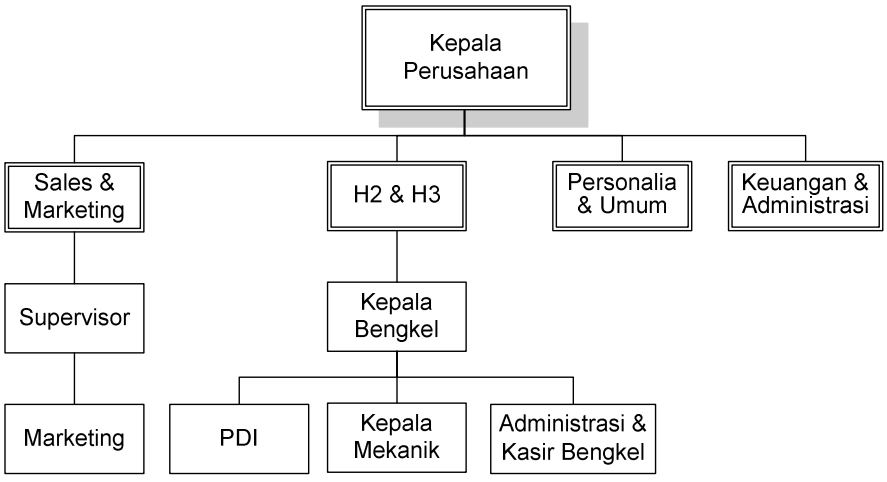
Misi

1. Membangun system perusahaan yang baik
2. Membangun sales team yang kuat
3. Membangun relasi dengan konsumen yang baik
4. Mencapai hasil penjualan yang baik

2.1.3 Struktur Organisasi dan Job Description

a. Struktur Organisasi PD. Wijaya Abadi Bandung

Struktur organisasi yang ada di PD. Wijaya Abadi Bandung dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Struktur Organisasi PD. Wijaya Abadi Bandung

b. Job Description

b.1 Kepala Perusahaan

Tugas, wewenang, dan tanggung jawab kepala cabang perusahaan :

1) Memberikan laporan kemajuan cabang kepada Direksi Pusat termasuk keuangannya.
2) Mengambil semua tindakan yang diperlukan agar cabang berjalan lancar.
3) Menjalankan Program Perusahaan untuk cabang itu/mengejar target.
4) Berhak atas promosi dan bonus jika cabang maju melebihi target Perusahaan.

b.2 Sales & Marketing

Tugas dan tanggung jawab Sales & Marketing :

1) Menentukan harga jual, produk yang akan dilaunching, jadwal kunjungan serta sistem promosi untuk memastikan tercapainya target penjualan.
2) Menganalisa dan mengembangkan strategi marketing untuk meningkatkan jumlah pelanggan dan layanan sesuai dengan target yang ditentukan.
3) Melakukan evaluasi kepuasan pelanggan dari hasil survey seluruh *sales team* untuk memastikan tercapainya target kepuasan pelanggan yang ditentukan.

4) Menerapkan budaya, sistem, dan peraturan intern perusahaan serta menerapkan manajemen biaya, untuk memastikan budaya perusahaan dan sistem serta peraturan dijalankan dengan optimal.

b.3 Personalia & Umum

Personalia mempunyai tugas sebagai berikut:

1) Mengendalikan dan menyelenggarakan kegiatan dibidang administrasi kepegawaian.
2) Melaksnakan Proses kegiatan Penggajian, kenaikan pangkat, kenaikan berkala, mutasi, kesejahteraan pegawai dan pembinaan pegawai.
3) Mengurus Proses Askes, Astek, Taspen dan proses pegawai yang telah mencapai usia pensiun dan penghargaan.
4) Memberikan saran-saran dan pertimbangan kepada Direksi tentang langkah-langkah atau tindakan yang perlu tentang kepegawaian.
5) Membuat laporan kegiatan Bagian Kepegawaian dan melaksanakan Tugas lain yang berhubungan dengan Tugasnya yang diberikan oleh atasan.

b.4 Keuangan & Administrasi

Tugas staff **administrasi** adalah membuat layanan administrasi dibawah pengawasan pimpinan/line managernya. Tugasnya (biasanya) meliputi admin, logistic, dan lainnya yang mendukung pelaksanaan administasi berjalan lancar. Tugas detailnyanya misalnya:

1) Menjaga dan meng-*update* informasi administasi mulai dari *office supply*, *stationaries*.

2) mempersiapkan *arrangement meeting detail*, absensi staff, serta
3) melakukan hal hal seperti surat menyurat dengan staff lainya,

Bagian **Keuangan** mempunyai tugas melaksanakan urusan keuangan di lingkungan Badan Perusahaan.

Bagian Keuangan menyelenggarakan fungsi:

1) penyusunan dokumen pelaksanaan anggaran Badan,
2) pelaksanaan urusan perbendaharaan Badan,
3) akuntansi pelaksanaan anggaran dan penyusunan laporan keuangan Badan.

b.5 Supervisor

Tugas utama supervisor adalah melakukan supervisi terhadap para staf pelaksanan rutinitas aktivitas bisnis perusahaan sehari-hari. Supervisor adalah level kepemimpinan yang tidak boleh membuat kebijakan yang bersifat strategis, tapi hanya menerjemahkan dan meneruskan kebijakan strategis atasannya kepada para bawahan untuk dikerjakan secara efektif dan produktif. Peran penting seorang supervisor adalah sebagai koordinator unit kerja.

b.6 Kepala Bengkel

Bertanggung jawab atas kelancaran operasional mekanis dan mesin-mesin dan peralatan produksi.

b.7 Kepala Mekanik

Bertanggung jawab atas perencanaan, pengkoordinasian, pengarahan, dan pengawasan atas pelaksanaan kegiatan maintenance dan repair mesin dan peralatan mekanik produksi.

2.2 Landasan Teori

2.2.1 Data

Menurut Abdul Kadir (2003: 29), data adalah deskripsi tentang benda, kejadian, aktivitas, dan transaksi, yang tidak mempunyai makna atau tidak berpengaruh secara langsung kepada pemakai. Data merupakan bentuk jamak dari *datum*, berasal dari bahasa Latin yang berbarti "sesuatau yang diberikan". Dalam penggunaan sehari-hari, data berarti suatu pernyataan yang diterima secara apa adanya. Pernyataan ini adalah suatu variabel yang bentuknya dapat berupa angka, kata-kata, atau citra.

Dalam keilmuan (ilmiah), fakta dikumpulkan untuk menjadi data. Data kemudian diolah sehingga dapat diutarakan secara jelas dan tepat sehingga dapat dimengerti oleh orang lain yang tak langsung mengalaminya sendiri, hal ini dinamakan deskripsi. Pemilahan banyak data sesuai dengan persamaan atau perbedaan yang dikandungnya dinamakan klasifikasi.

Secara konseptual, data adalah deskripsi tentang benda, kejadian, aktivitas, dan transaksi, yang tidak mempunyai makna atau tidak berpengaruh secara langsung kepada pemakai. Data dapat berupa nilai yang terformat, teks, citra, audio, dan video.

Data yang terformat adalah data dengan suatu format tertentu. Misalnya, data yang menyatakan tanggal atau jam, atau menyatakan nilai mata uang.

Teks adalah sederetan huruf, angka, dan simbol-simbol khusus (misalnya + dan $) yang kombinasinya tidak tergantung pada masing-masing item secara individual. Contoh teks adalah artikel majalah.

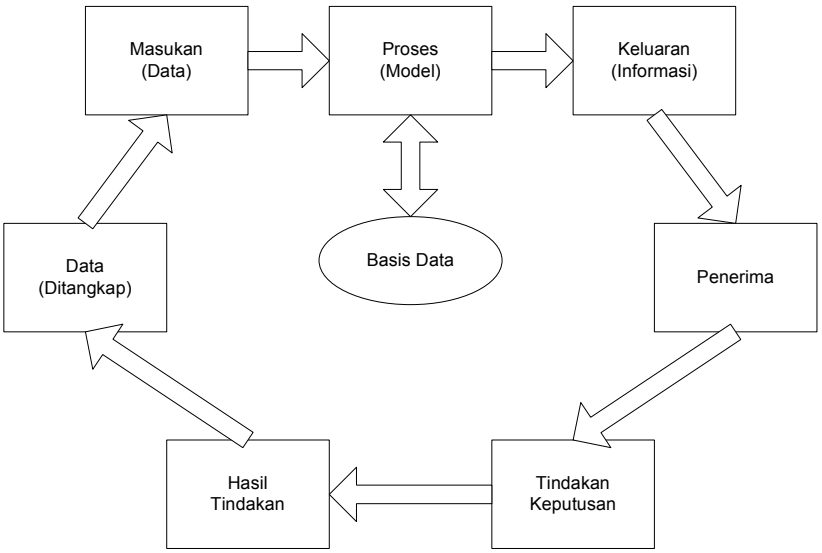
Citra (*image*) adalah data dalam bentuk gambar. Citra dapat berupa grafik, foto, hasil *rontgen*, dan tanda tangan, ataupun gambar yang lain.

Audio adalah data dalam bentuk suara. Instrumen music, suara orag atau suara binatang, gemericik air, detak jantung merupakan beberapa contoh data audio.

Video menyatakan data dalam bentuk sejumlah gambar yang bergerak dan bias saja dilengkapi dengan suara. Video dapat digunakan untuk mengabadikan suatau kejadian atau aktivitas.

2.2.2 Informasi

Informasi adalah data yang sudah diolah, dibentuk, atau dimanipulasi sesuai dengan keperluan tertentu atau hasil dari pengolahan data yang secara prinsip memiliki nilai atau *value* yang lebih dibandingkan data mentah. Informasi dapat juga dianggap seuatu data untuk diolah kembali dan menjadikan informasi sesuai dengan keperluan unit kerja tertentu.



Gambar 2. 2 Siklus Informasi

Menurut Burch dan Grudnitski (1989) Gambar 2.2 memperlihatkan siklus informasi, yaitu menggambarkan pengolahan data menjadi informasi dan pemakaian informasi untuk mengambil keputusan, hingga akhirnya dari tindakan hasil pengambilan keputusan tersebut dihasilkan data kembali.

Jadi, hal yang terpenting untuk membedakan informasi dengan data, informasi itu mempunyai kandungan "makna", sedangkan data tidak. Pengertian data di sini merupakan hal yang sangat penting, karena berdasarkan maknalah si penerima dapat memahami infomasi tersebut dan secara lebih jauh dapat menggunakannya untuk menarik suatu kesimpulan atrau bahkan mengambil keputusan.

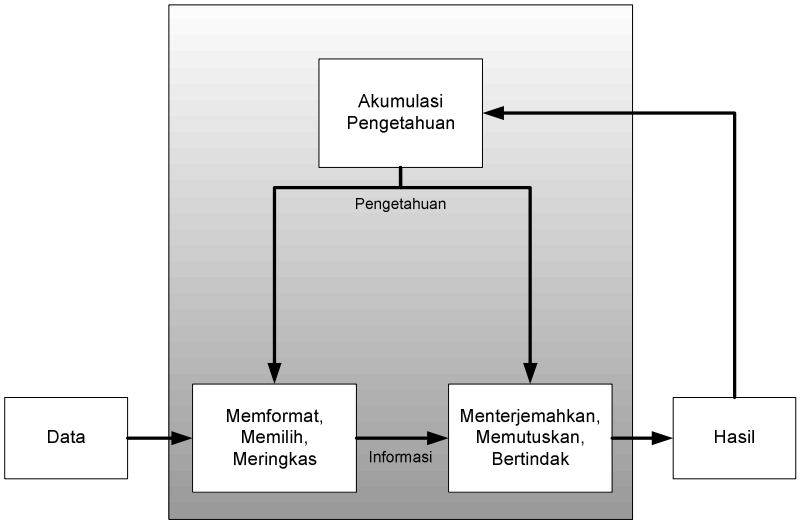
Informasi mempunyai tingkat kualitas yang ditentukan beberapa hal, antara lain:

1. *Akurat*, informasi harus bebas dari kesalahan-kesalahan dan harus jelas penyampaian maksudnya.

2. *Tepat pada waktunya*, informasi yang dating tidak boleh terlambat pada penerima.
3. *Relevan*, informasi harus mempunyai manfaat bagi pemakainya.
4. *Lengkap*, informasi berisi informasi yang dibutuhkan.
5. *Jelas*, isi informasi sesuai dengan keperluan pemakai.

2.2.3 Pengetahuan

Menurut Alter (1992), pengetahuan (*knowledge*) adalah kombinasi dari naluri, gagasan aturan, dan prosedur yang mengarahkan tindakan atau keputusan. Sebagai gambaran, informasi yang dipadukan dengan pengalaman masa lalu dan keahlian akan memberikan suatu pengetahuan yang tentu saja memiliki nilai yang tinggi. Sebuah gambaran tentang hubungan antara data, informasi, dan pengetahuan ditunjukkan pada gambar berikut.

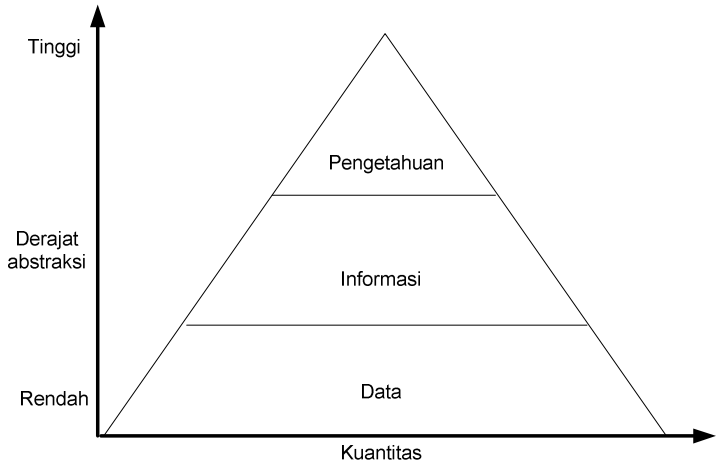


Gambar 2. 3 Hubungan data, informasi, dan pengetahuan

Gambar 2.3 memperlihatkan bahwa data diformat, dipilih, dan diringkas leh sisetm menjadi suatu informasi. Proses tersebut dilakukan berdasarkan suatu

pengetahuan tentang cara melakukannya. Selanjutnya, informasi yang dihasilkan dimasukkan ke tahap berikutnya dan diproses menjadi suatu hasil. Hasil ini diakumulasikan sebagai pengetahuan yang kemudian digunakan untuk melakukan pemrosesan data atau pemrosesan informasi.

Bersama-sama dengan data dan informasi, pengetahuan sering digambarkan dalam bentuk piramida, sperti gambar di bawah ini.



Gambar 2. 4 Gambaran data, informasi, dan pengetahuan menurut abstraksi

Gambar 2.4 menunjukkan bahwa dilihat dari derajat abstraksi, pengetahuanlah yang memiliki kuantitas paling abstrak, sedangkan jika dilihat dari kuantitas, pengetahuanlah yang memiliki kuantitas paling kecil.

2.2.4 Basis Data

Basis Data adalah kumpulan data yang terintegrasi atau saling berhubungan satu sama lain yang disimpan sedemikian rupa agar dapat dimanfaatkan kembali dengan cepat dan mudah.

Sistem basis data adalah sistem yang terdiri atas kumpulan file (tabel) yang saling berhubungan (dalam sebuah basis data disebuah sistem komputer) dan sekumpulan program (DBMS) yang memungkinkan beberapa pemakai dan/atau program lain untuk mengakses dan memanipulasi file-file (tabel-tabel) tersebut.

Dalam sebuah Basis Data, secara lengkap akan terdapat komponen-komponen utama yaitu:

1) Perangkat keras (*hardware*)
2) Sistem operasi (*operating system*),
3) Basis Data (*database*),
4) Sistem (aplikasi /perangkat lunak) pengelola basis data (DBMS),
5) Pemakai (*user*).

Ada beberapa jenis/tipe pemakai terhadap suatu sistem basis data yang dibedakan berdasarkan cara mereka berinteraksi terhadap sistem.

1) Pengguna Awam (*Naive User*) yaitu pemakai yang tidak berpengalaman, berinteraksi dengan sistem tanpa menulis program, tinggal menjalankan satu menu dan memilih proses yang telah ada atau telah dibuat sebelumnya oleh programmer.
2) User Mahir yaitu pemakai yang berinteraksi dengan sistem tanpa menulis modul program. Mereka menyatakan query (untuk akses data) dengan bahasa query yang telah disediakan oleh DBMS.
3) Programmer Aplikasi yaitu pemakai yang berinteraksi dengan basis data melalui *Data Manipulation Language* (DML), yang dibuat dengan bahasa C, Cobol dan lainnya.

4) User khusus yaitu pemakai yang menulis aplikasi basis data non konvensional, tetapi untuk keperluan-keperluan khusus, seperti untuk aplikasi AI, sistem pakar, pengolahan citra dan lain-lain.

2.2.4.1 Penerapan Basis Data

Bidang-bidang fungsional yang telah umum memanfaatkan basis data demi efisiensi, akurasi dan kecepatan operasi antara lain: bidang perbankan yang melakukan pengelolaan data nasabah/data tabungan/data pinjaman, pembuatan laporan-laporan akuntansi, pelayanan informasi pada nasabah/calon nasabah dan lain-lain. Kemudian bidang asuransi, bidang pendidikan/sekolah, telekomunikasi, rumah sakit dan lain-lain.

2.2.4.2 Kegunaan Basis Data

Penyusunan suatu basis data digunakan untuk mengatasi masalah masalah pada penyusunan data yaitu:

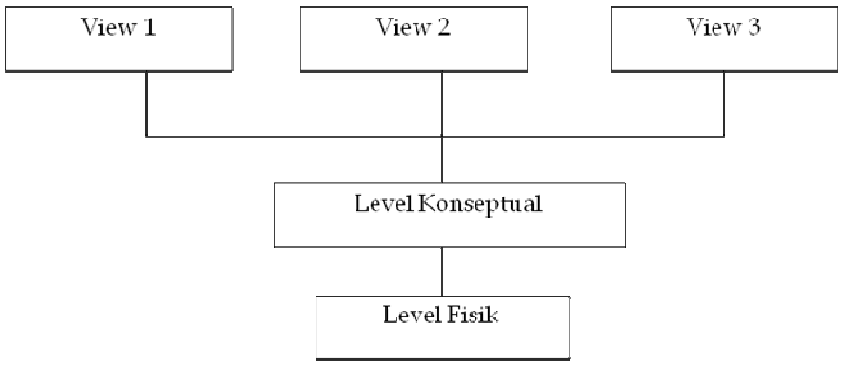
1) Redundansi dan Inkonsistensi data
2) Kesulitan dalam pengaksesan data
3) Isolasi data untuk standardisasi
4) Multiple User (banyak pemakai)
5) Masalah keamanan (*security*)
6) Masalah integritas (kesatuan)
7) Masalah data independence (kebebasan data)

2.2.4.3 Abstraksi Basis Data

Abstraksi data merupakan tingkatan/level dalam bagaimana pemakai melihat data dalam sebuah sistem basis data. Abstraksi data dalam sistem basis data dibagi menjadi tiga level yaitu:

1) Level Fisik (*Physical Level*), merupakan level terendah dalam abstraksi data, yang menunjukkan bagaimana (*how*) sesungguhnya suatu data disimpan.
2) Level Lojik/Konseptual (*Conceptual Level*),merupakan level lebih tinggi berikutnya dalam abstraksi data yang menggambarkan data apa (*what*) yang sebenarnya disimpan dalam basis data dan hubungan relasi yang terjadi antara data.
3) Level Pandangan Pemakai (*View Level*), merupakan level tertinggi dari abstraksi data yang hanya menunjukkan satu bagian dari keseluruhan basis data.

Hubungan antar level tersebut dapat digambarkan (Kristanto:1996) pada Gambar 2.5.



Gambar 2. 5 Abstraksi Data

2.2.4.4 Bahasa Basis Data

Sebuah bahasa basis data biasanya dapat dibagi ke dalam 2 bentuk yaitu:

1) *Data Definition Language* (DDL)

   Yaitu struktur/skema basis data yang menggambarkan/mewakili desain basis data secara keseluruhan dispesifikasikan dengan bahasa khusus.

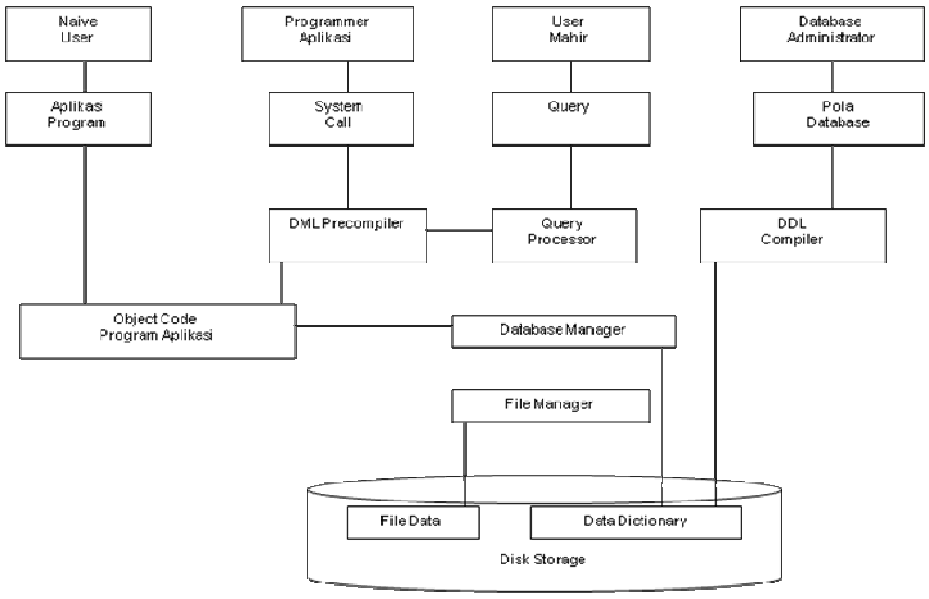
2) *Data Manipulation Language* (DML)

   Yaitu perintah-perintah yang digunakan untuk mengubah, memanipulasi, dan mengambil data pada basis data.

3) DCL (*Data Control Language*) yang berkaitan dengan pengaturan sekuritas terhadap basis data.

2.2.4.5 Struktur Sistem Secara Keseluruhan

Pada Gambar 2.6 memperlihatkan struktur sistem basis data secara keseluruhan.



Gambar 2. 6 Struktur sistem secara keseluruhan

2.2.5 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database.* Menurut Turban (2005), *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine lerarning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Perkembangan *data mining* yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi. Sebagai contoh, toko swalayan merekam setiap penjualan barang dengan memakai alat POS (*point of sales*). *Database* data penjualan tersebut bisa mencapai beberapa GB (Giga byte) setiap harinya untuk sebuah jaringan toko swalayan berskala nasional. Perkembangan internet juga punya andil cukup besar dalam akumulasi data. Tetapi pertumbuhan yang pesat dari akumulasi data itu telah menciptakan kondisi yang sering disebut sebagai ***rich of data but poor of information*** karena data yang terkumpul itu tidak dapat digunakan untuk aplikasi yang berguna. Tidak jarang kumpulan data itu dibiarkan begitu saja seakan-akan ***kuburan data (data tombs)***. Investasi yang besar di bidang IT untuk mengumpulkan data berskala besar ini perlu dijustifikasi dengan didapatnya nilai tambah dari kumpulan data ini.

Kebutuhan dari dunia bisnis yang ingin memperoleh nilai tambah dari data yang telah mereka kumpulkan telah mendorong penerapan teknik-teknik analisa data dari berbagai bidang seperti statistik, kecerdasan buatan dan sebagainya pada data berskala besar itu. Ternyata penerapan pada data berskala

besar memberikan tantangan-tantangan baru yang akhirnya memunculkan metodologi baru yang disebut data mining ini. Bermula dari penerapan di dunia bisnis, sekarang ini data mining juga diterapkan pada bidang-bidang lain yang memerlukan analisa data berskala besar seperti bioinformasi dan pertahanan negara.

Beberapa faktor yang mendukung perlunya dilakukan *data mining* adalah:

1. Data telah mencapai jumlah dan ukuran yang sangat besar
2. Telah dilakukan proses data warehousing
3. Kemampuan Komputasi yang semakin terjangkau
4. Persaingan bisnis yang semakin ketat

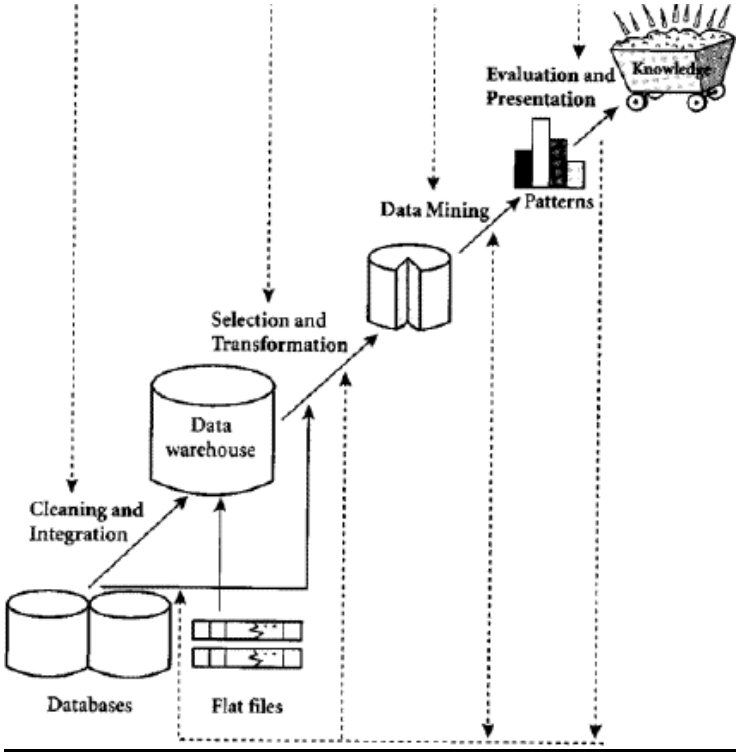
2.2.5.1 Model dalam Data Mining

Terdapat dua tipe atau mode operasi yang bias digunakan untuk mencari informasi yang dibutuhkan *user* lewat proses *data mining,* yaitu model verifikasi dan *knowledge discovery*.

Model verifikasi menggunakan pendekatan *top down* dengan mengambil hipotesa dari *user* dan memeriksa validitasnya dengan data sehingga bias dibuktikan kebenaran hipotesa.

Model *Knowledge Discovery* menggunakan pendekatan *bottom up* untuk mendapatkan informasi yang sebelumnya tidak diketahui. Model ini terbagi menjadi dua, yaitu *directed knowledge discovery* dan *undirected knowledge discovery.* Pada *directed knowledge discovery, data mining* akan mencoba mencari penjelasan nilai targaet *field* tertentu (seperti penghasilan, respons, usia,

dan lain-lain) terhadap *field-field* yang lain. Sedangkan pada *undirected knowledge discovery* tidak ada *field* karena komputer akan mencari pola yang ada pada data. Jadi, *undirected knowledge discovery* digunakan untuk mengenali hubungan / relasi yang ada pada data sedangkan *directed knowledge discovery* akan menjelaskan hubungan relasi tersebut.



Gambar 2. 7 Tahap-tahap *Data Mining*

Gambar 2.7 merupakan tahapan proses dalam *data mining* dan dapat dijelaskan sebagai berikut:

1. Data Selection

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (*Knowledge Discovery in*

Database) dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining,* disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Preprocessing/Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD.

Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, sperti kesalahan cetak (tipografi).

3. Transformasi

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining.* Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan cari dalam basis data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretasi / Evaluasi

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation.*

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2.5.2 Pengelompokan *Data Mining*

Menurut Larose (2005), *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan,yaitu:

1. Deskripsi

   Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cata untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

   Estimasi hamper sama dengan klasifikasi, kecuali variabel target estimasi lebih kea rah numeric daripada ke arah kategori.

3. Prediksi

   Prediksi hamper sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil aka nada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

   Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

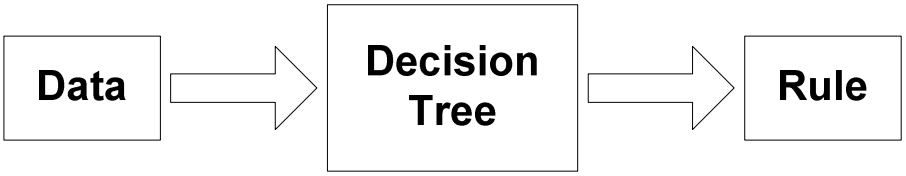
   Pengklusteran merupakan pengelompokan *record,* pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adlah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain.

6. Asosiasi

   Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut anlisis keranjang belanja.

2.2.5.3 *Decision Tree* (Pohon Keputusan)

Pohon keputusan merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan *data mining*. Pada dasarnya *Decision Tree* mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*).



Gambar 2. 8 Konsep Decision Tree

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Menurut Berry & Linoff (2004), sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang kebih kecil dengan menerapkan serangkaian aturan-aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain.

Sebuah pohon keputusan mungkin dibangun dengan saksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 (Larose, 2005).

Variabel tujuan biasanya dikelompokkan denga pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap *record* - terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu kelas.

Menurut Basuki & Syarif (2003), proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule,* dan menyederhanakan *rule.*

2.2.5.3.1 Konsep Data Dalam *Decision Tree*

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai criteria dalam pembentukan tree. Misalkan unutuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan temperature. Salah satu atribut merupakan atribut yng menyatakan data solusi per-item data yang disebut

dengan **target atribut**. Atribut memiliki nilai-nilai yang dinamakan dengan **instance**. Misalkan atribut cuaca mempunyai instance berupa cerah, berawan, dan hujan seperti tampak pada Gambar 2.9.

Nama	Cuaca	Angin	Temperatur	Main
Ali	cerah	keras	panas	tidak
Budi	cerah	lambat	panas	ya
Heri	berawan	keras	sedang	tidak
Irma	hujan	keras	dingin	tidak
Diman	cerah	lambat	dingin	ya

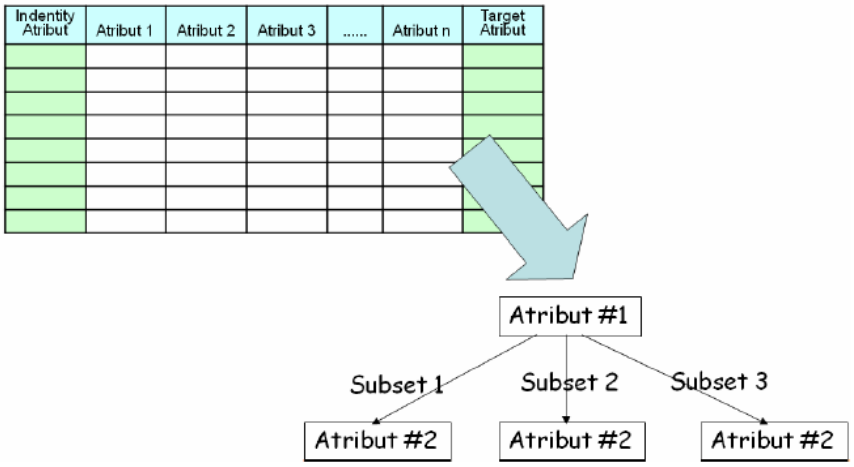
Sample

attribut

Target atribut

Gambar 2. 9 Konsep Data dalam Decision Tree

Untuk mengubah bentuk data (tabel) menjadi model tree, dalam kasus ini menggunakan algoritma ID3, setelah diperoleh model berupa tree, maka tree ini bisa dikonversikan ke dalam *rule* seperti tampak pada Gambar 2.10.



Gambar 2. 10 Dari Data Set (tabel) menjadi Tree

2.2.5.3.2 Atribut *Selection Measure*

Pengukuran informasi yang didapatkan (*information gain*) bertujuan untuk memilih suat atribut yang diuji (*test atribut*) di setiap node dalam struktur pohon. Atribut dengan nilai informasi tertinggi (*greatest entropy*) dipilih sebagai suatu atribut dalam suatu node yang sedang diproses.

Jika S merupakan suatu kumpulan dari s data sampel, atribut label kelas memiliki isi (*value*) m yang berbeda, C_i (for i=1..m). Jika s_i merupakan jumlah sampel S dalam kelas C_i dimana p_i merupakan probabilitas dari label sampel data dari C_i, maka:

$$I(s_1, s_2, .., s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i) \qquad (3.1)$$

Jika atribut A memiliki *v* isi yang berbeda $\{a_1, a_2, \ldots, a_v\}$, atribut A dapat digunakan untuk membagi S ke dalam subset *v*, $\{S_1, S_{2,\ldots,} S_v\}$, dimana S_j terdiri dari sampel dalam S yang memiliki isi a_j dari A. Jika A terpilih sebagai atribut yang diuji (atribut terbaik untuk dibagi atau dipartisi), kemudian subset ini alam berhubungan dengan cabang dari node yang berisi kumpulan S. Jika S_{ij} merupakan jumlah sampel dari kelas C_i di dalam suatu subset S_j, maka *Entropy*:

$$E(A) = \sum_{j=1}^{v} \frac{S_{1j} + \cdots + S_{mj}}{S} I(S_{1j}, .., S_{mj}) \qquad (3.2)$$

Dari persamaan (3.1) dan (3.2), maka :

$$Gain(A) = I(S_1, .., S_m) - E(A) \qquad (3.3)$$