

## RANCANG BANGUN PENGEMBANGAN POLA PENGETAHUAN MELALUI SISTEM PEMBELAJARAN ONLINE

**Hidayatulah Himawan**

Jurusan Teknik Informatika UPN "Veteran" Yogyakarta  
Jl. Babarsari no 2 Tambakbayan 55281 Yogyakarta Telp (0274)-485323  
e-mail : if.iwan@gmail.com

### **Abstrak**

*Sistem pembelajaran yang efektif dan efisien tanpa mengurangi hasil dan nilai kompetensi yang ingin dihasilkan, menjadi output (keluaran) yang diinginkan oleh tiap komponen yang terlibat di dalam dunia pendidikan. Salah satu sistem pembelajaran yang berkembang saat ini adalah pembelajaran jarak jauh atau yang biasa disebut dengan e-learning.*

*Sistem pembelajaran jarak jauh atau e-learning telah banyak dikembangkan oleh kalangan akademik. Namun apa yang diinginkan agar sistem pendidikan bisa berjalan lebih efektif dan efisien tanpa mengurangi kualitas output yang dihasilkan masih sangat jauh dari apa yang diharapkan. Hal ini disebabkan terlalu besarnya sistem database yang ada tanpa adanya proses yang dapat membuat para pengguna lebih mudah menggunakannya.*

*Data mining bisa menjadi salah satu solusi agar sistem pencarian yang berjalan pada database bisa dilakukan dengan menggunakan pola-pola pengetahuan yang lebih baik. Pembagian Proses pencarian data kuliah akan mejadi lebih mudah jika dilihat dari berbagai kategori yang ada. Mulai dari sistem yang bersifat objek oriented ataupun yang non-object oriented. Baik secara struktural ataupun non-struktural.*

*Selain itu tingkat kebiasaan kita sebagai pengguna yang ingin mendapatkan suatu sistem yang "user friendly" masih sangat diharapkan. Untuk itulah, penelitian ini akan mengembangkan pengolahan data mining pada pengembangan pola pengetahuan melalui sistem pembelajaran online.*

*Metodologi yang digunakan dalam penelitian ini adalah model air terjun (water fall) yang dipadukan dengan System Development Life Cycle (SDLC). Yaitu dengan melakukan analisis sistem, desain sistem, implementasi sistem dan operasi serta pemeliharaan sistem. Dimana keempat sistem tersebut dapat dilakukan secara berulang. Sedangkan bahasa pemrograman yang digunakan menggunakan PHP dan MySql sebagai basis datanya. Sedangkan desainnya menggunakan macromedia Dreamweaver.*

**Kata kunci** : e-learning, data mining, SDLC

### **1. PENDAHULUAN**

Sistem pembelajaran memiliki banyak variasi pada penerapannya. Secara umum pembelajaran memiliki proses yang efektif dan efisien dalam proses yang dijalankan. Keinginan agar proses pembelajaran menjadi lebih praktis tanpa mengurangi fungsi pembelajaran adalah keinginan yang hendak dicapai oleh tiap komponen pembelajaran. Keinginan agar tiap sistem pembelajaran membawa kepada peningkatan mutu pendidikan akan tercapai dengan sendirinya. Hal ini bukanlah hal yang mudah untuk dicapai. Karena membutuhkan komitmen dari semua pihak yang terlibat. Sehingga pengembangan sistem pendukung pembelajaran sangat diperlukan pada saat ini. Termasuk teknologi dan sistem yang akan digunakan.

Salah satu sistem pembelajaran yang dikembangkan pada saat ini adalah sistem pembelajaran *online*. Pembelajaran online memiliki banyak definisi. Namun secara aplikatif sistem pembelajaran online adalah suatu model pembelajaran yang menggunakan media internet secara online melalui web, sehingga proses pembelajaran dilakukan secara tidak langsung atau tatap muka antara pendidik dan siswa. Sistem pembelajaran online sudah banyak dipergunakan oleh lembaga pendidikan pada saat ini. Namun penggunaan dan efektifitas dalam aplikasinya masih sangat jauh dari yang diharapkan.

Sistem pembelajaran online membutuhkan teknologi pendukung yang mampu memproses berbagai aplikasi permintaan yang diberikan oleh pengguna. Salah satu teknologi yang dapat dikembangkan adalah penggunaan basis data pada sistem pembelajaran online. Basis data menjadi salah satu komponen teknologi yang harus dapat mengolah berbagai data dan aktifitas permintaan informasi yang diberikan oleh pengguna. Teknologi pendukung ini diberikan sebagai bentuk pelayanan Baik dari sistem aplikasi hingga kepada pengenalan sistem pencarian yang ada. Termasuk didalamnya pengenalan pola pengetahuan atau biasa dikenal dengan teknologi data mining atau data warehouse pada pembelajaran online.

Teknologi dan kemampuan pemrosesan pada basis data yang digunakan pada saat ini memungkinkan untuk menyimpan data dengan komponen dan pengelolaan data yang bersifat kompleks. Namun dengan kondisi basis data yang kurang mendukung pada pemrosesan yang terjadi, maka mulai terjadi berbagai permasalahan. Seperti komponen data yang tidak sesuai dengan data yang ada, atau pengolahan data yang berbeda dari hasil keluaran yang diinginkan oleh individu atau perusahaan. Sehingga data sangat perlu untuk disimpan dan dikelola dengan baik. Namun, proses penemuan pengetahuan (knowledge) dari data yang disimpan adalah lebih penting

dari proses yang ada. Karena sistem pencarian yang cepat, tepat dan akurat juga diperlukan. Agar data yang diperlukan oleh individu atau perusahaan akan dapat diperoleh dengan tepat.

Sudah diketahui bahwa informasi adalah suatu hal yang sangat penting dalam mendukung dan menunjang setiap kegiatan. Baik itu yang bersifat bisnis ataupun kegiatan lainnya. Terlebih ketika informasi itu dibutuhkan dalam suatu kegiatan pendidikan yang ada pada sistem pembelajaran. Sistem pemrosesan data yang bersifat tradisional sangat baik dalam menyimpan data secara cepat ke dalam basis data, namun dalam hal pemrosesan data dengan memberikan hasil berupa analisa terhadap data yang ada masih kurang baik, padahal ketika analisa tersebut diberikan akan sangat mendukung tiap kegiatan yang dijalankan. Sehingga pengolahan data mining pada pengembangan pola pengetahuan melalui sistem pembelajaran online akan dapat dilaksanakan sesuai dengan kebutuhan pengembangan pengetahuan yang ada.

## 2. KEASLIAN PENELITIAN

Berdasarkan penelitian yang pernah dilakukan oleh Wicaksono, V (2007) tentang faktor-faktor yang mendukung efektifitas penggunaan e-learning terlihat bahwa variabel bebas secara umum dapat memiliki pengaruh terhadap penggunaan *e-learning* dengan memperhatikan aspek kecepatan, sarana pengajaran, kemudahan serta aspek kognitifnya. Sehingga para pengguna khususnya para mahasiswa bisa berinteraksi secara terbuka dengan staff pengajar atau pembimbing mereka. Sistem pembelajaran yang dihasilkan memiliki kompleksitas dalam hal pemasukan data sistem pembelajaran pada tiap mata kuliahnya, serta sistem pembaharuan data pembelajaran pada mata kuliah yang diberikan. Yang berakibat pada tingkat kesulitan yang dihasilkan menjadi lebih tinggi. Sehingga efektifitas sistem pembelajaran yang seharusnya bisa berjalan dengan baik, kurang mendapatkan hasil yang optimal.

Penelitian lain yang bisa kita bandingkan adalah penelitian yang dilakukan oleh Herlina T., R. (2007), dimana pengolahan dan penerapan data *warehouse* dan data *mining* untuk sistem pendukung keputusan dalam kegiatan akademik di Multimedia Training Center berjalan dengan optimal. Penelitian dilakukan dengan memprediksi apakah mahasiswa akan dapat lulus dengan tepat waktu beserta analisa terhadap faktor-faktor yang mempengaruhinya. Prediksi yang dilakukan berdasarkan pada data nilai perkuliahan yang dijalani oleh mahasiswa, informasi terhadap kendala yang diberikan mahasiswa pada sistem pembelajaran online, serta interaksi antar mahasiswa dan dosen pembimbing. Data yang diperoleh diolah oleh sistem sehingga menghasilkan analisa terhadap kemampuan mahasiswa dalam menjalani sistem pembelajaran. Pengolahan yang dilakukan hanya sebatas pada nilai yang di berikan oleh dosen pembimbing dengan tingkat kedekatan terhadap mahasiswa yang ada. Sehingga objektivitas penilaian yang diberikan masih belum berdasarkan pada hasil analisa sistem yang ada.

Selain itu penelitian yang dilakukan oleh Sari, R (2008) tentang analisa kemampuan sumber daya manusia terhadap penggunaan sistem pembelajaran online memiliki aspek yang tinggi terhadap aktifitas pembelajaran yang ada. Pengguna dapat berinteraksi secara langsung dengan pembimbing sehingga komunikasi yang dilakukan dapat berlangsung secara online. Fasilitas komunikasi telah tersedia melalui menu "komunikasi" pada sistem yang dihasilkan. Pengguna dapat melakukan pertukaran data melalui sistem yang ada, sehingga pengolahan data menjadi komponen penting dari sistem yang dihasilkan. Pengolahan dilakukan secara satu per satu terhadap data yang dikirimkan. Sehingga informasi yang diberikan menjadi lebih lama. Dan efektifitas terhadap sistem yang ada menjadi berkurang.

Penelitian yang dilakukan pada saat ini berbeda dengan penelitian yang telah dilakukan. Sistem pengolahan data yang diberikan lebih kepada pengolahan sistem pencarian terhadap data yang diperlukan oleh para pengguna. Sistem pengolahan ini memiliki sistem pengembangan pola pengetahuan terhadap data dan informasi. Sehingga sistem pencarian data dan informasi serta komunikasi antar pengguna menjadi lebih efektif.

## 3. DASAR TEORI

Sistem merupakan jaringan dari elemen-elemen yang saling berhubungan, membentuk suatu kesatuan untuk melaksanakan suatu tujuan pokok dari sistem tersebut (Kadir, 2003). Informasi adalah data yang telah diolah menjadi sebuah bentuk yang berarti bagi penerimanya dan bermanfaat dalam mengambil keputusan saat ini dan masa mendatang (Davis, 1999). Menurut Wahyono, informasi merupakan hasil dari pengolahan data menjadi bentuk yang lebih berguna bagi yang menerimanya yang menggambarkan suatu kejadian-kejadian nyata dan dapat digunakan sebagai alat bantu untuk pengambilan suatu keputusan (Wahyono, 2003). Sistem informasi adalah kombinasi antara prosedur kerja, informasi, orang, dan teknologi informasi yang diorganisasikan untuk mencapai tujuan (Kadir, 2003). Kualitas informasi sangat dipengaruhi oleh tiga hal pokok yaitu :

1. Relevansi (*relevancy*)  
Informasi akan relevan jika memberikan manfaat bagi pemakainya.
2. *Time lines* (tepat waktu)  
berarti informasi yang datang pada penerima tidak boleh terlambat.
3. Akurasi (*accuracy*)

Informasi dikatakan akurat jika informasi tersebut tidak menyesatkan, bebas dari kesalahan-kesalahan dan harus jelas mencerminkan maksudnya.

Beberapa hal yang dapat berpengaruh terhadap keakuratan sebuah informasi antara lain :

- Kelengkapan Informasi (*completeness*) berarti informasi yang dihasilkan terdiri dari satu kesatuan informasi yang menyeluruh dan mencakup berbagai hal yang terkait di dalamnya.
- Kebenaran Informasi (*correctness*) berarti informasi yang dihasilkan benar sesuai dengan perhitungan yang ada di dalamnya.
- Keamanan Informasi berarti informasi akan dapat diakses oleh pihak-pihak yang berkepentingan dengan informasi tersebut.

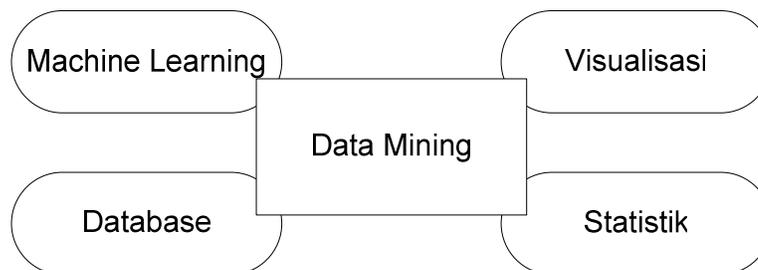
Sistem informasi merupakan suatu sistem yang dibuat manusia yang terdiri dari komponen-komponen dalam organisasi untuk mencapai suatu tujuan yaitu menyajikan informasi. Sistem informasi di dalam suatu organisasi mempertemukan kebutuhan pengolahan transaksi, mendukung operasi, bersifat manajerial dan strategi dari suatu organisasi dan menyediakan pihak luar tertentu dengan laporan-laporan yang diperlukan (Wahyono, 2003).

### **Pola, Data Mining dan Machine learning**

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan objek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, objek-objek ini bias berupa pasien, mahasiswa, pemohon kredit, image atau signal atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya. Biasanya subjek ini disebut dengan pengenalan pola atau *pattern recognition*.

Pada proses pengenalan pola, ada suatu sistem yang disebut dengan data mining. Data mining sering juga disebut *knowledge discovery in database* (KDD), yaitu sebuah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data besar. Keluaran dari data mining ini bias dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah *pattern recognition* sekarang jarang digunakan karena ia termasuk dari bagian data mining.

Didalam suatu proses pencarian data mining, kecerdasan buatan sangat berhubungan dengan pengembangan teknik-teknik yang akan dijalankan. Ini disebut sebagai suatu poses pada *machine learning* , yaitu suatu area dalam artificial intelligence yang mengembangkan teknik yang bias diprogramkan dan belajar dari data dan informasi masa lalu sebagai dasar pengembangan pola sistem yang baru. Bidang ini bersinggungan ilmu probabilitas dan statistic serta optimasi. *Machine learning* menjadi alat analisa dalam data mining. Hubungan antar komponen dapat dilihat pada gambar 2.1



**Gambar 1.** Gambar Gabungan Komponen Data Mining

Pada gambar 1 terlihat bahwa ada 4 komponen yang melingkupi tiap kegiatan yaitu *machine learning*, visualisasi, statistik dan *database*. Semua komponen tersebut diolah didalam suatu prose yang dinamakan data mining.

Ada perbedaan yang bisa dijelaskan dari proses yang digambarkan. Yaitu komponen statistik lebih berdasarkan pada teori dan lebih fokus kepada hipotesis yang digunakan. Sedangkan *machine learning* lebih bersifat heuristic dan berfokus pada perbaikan performansi dari suatu teknik *learning* atau pembelajaran, yang juga meliputi *real-time learning robotic*. Sedangkan data mining merupakan gabungan teori dan heuristic, serta terfokus pada seluruh proses penemuan pengetahuan (*knowledge*) atau pola termasuk *cleaning*, *learning* dan visualisasi dari hasilnya.

### **Variabel, Fitur dan Attribute**

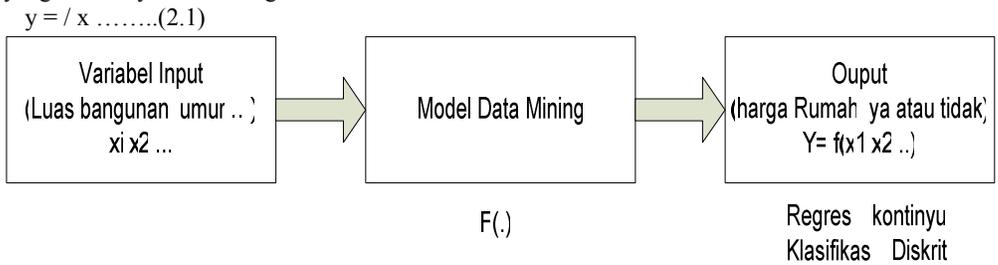
Suatu observasi, *example*, *pattern* (pola) atau objek biasanya ditandai oleh beberapa attribute. Misalnya objek orang ditandai dengan attribute tinggi badan, jenis kelamin, warna kulit, bentuk muka, atau lain sebagainya. Attribute ini sering disebut dengan variable. Juga ada yang menyebutnya sebagai suatu fitur. Jadi ketiga hal tersebut adalah sama. Dari sederet variabel akan dikelompokkan menjadi *input* (variabel *independent*,

*predictor*) dan *output* (variabel *dependent*, response atau label). Format data dinyatakan dalam bentuk matriks dimana baris menyatakan objek atau observasi dan kolom menyatakan attribute atau fitur atau variabel. Dalam proses pengurangan dimensi data orang menyebut sebagai *feature extraction* dan *feature selection*. Kedua istilah ini mengurangi dimensi variabel atau kolom dari data.

**Supervised dan Unsupervised Learning**

Metode pembelajaran (*learning*) secara umum dibagi menjadi kedalam dua pendekatan. Yaitu *supervised* dan *unsupervised*. Pada pendekatan *unsupervised learning*, metode dapat diterapkan tanpa adanya latihan (*training*) dan tanpa guru (*teacher*). Guru disini adalah label dari data. Misalkan kita punya sekelompok pengamatan atau data tanpa label (*output*) tertentu, maka dalam unsupervised learning kita mengelompokkan data tersebut ke dalam beberapa kelas yang kita kehendaki. Ini dilakukan karena data yang ada memang tidak mempunyai label. Label menandai kemana data dikelompokkan. Untuk melakukan tugas (*task*), dapat diterapkan metode *unsupervised learning*. Metode yang termasuk didalam kelompok ini adalah metode *klustering* dan *self organizing map* (SOM).

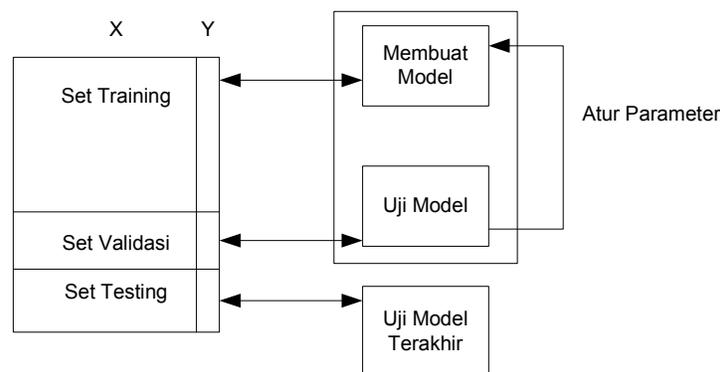
Metode pembelajaran lainnya adalah supervised learning. Yaitu metode pembelajaran dengan menggunakan pelatih dan latihan. Banyak teknik dan metode *pattern recognition* dalam kategori ini, termasuk didalamnya adalah regresi, analisis diskriminan (LDA), artificial neural network (ANN), dan support vector machine (SVM). Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, kita menggunakan contoh data yang mempunyai output atau label selama prose latihan. Di sini kita menemukan fungsi yang bias dinyatakan sebagai :



**Gambar 2.** Gambar Proses Hubungan antar Variabel

Pada gambar 2 terlihat hubungan antar variabel input, x, output dan y pada supervised learning. Proses ini akan berjalan terus hingga didapatkan suatu bentuk hasil akhir dari model yang diinginkan. Antara variabel input yang ada, akan diolah menjadi suatu output dengan menggunakan model yang telah ditetapkan oleh sistem. Sehingga keluaran dari proses data akan dijalankan pada sistem.

Data untuk melatih kita sebut *set training*. Fungsi atau model yang ditemukan akan diuji dengan set data validasi. Ada yang membagi data hanya ke dalam dua (2) kelompok, yaitu *set training* dan *set testing*. Ada juga yang membagi menjadi tiga (3) bagian, yaitu *set training*, validasi dan *testing*. Fungsi set validasi dan testing sebenarnya sama, yakni untuk menguji apakah fungsi atau model yang digunakan sudah cukup bagus performansinya ketika diterapkan untuk memprediksi data pada masa yang akan datang. Jika performansi pada saat validasi atau testing belum memenuhi harapan kita bias mengatur nilai parameter model untuk mendapatkan model dengan performansi yang lebih baik. Setelah fungsi tersebut ditemukan dan performansinya cukup bagus, kemudian dapat dilakukan penugasan untuk mengelompokkan objek tertentu yang belum diketahui labelnya atau membuat prediksi. Proses ini dapat dilihat pada gambar 3.



**Gambar 3.** Pembagian Data menjadi Set Training, Validasi dan Testing

Pada gambar 3 data untuk training dan testing terdiri dari vector atau matrik input dan output (label). Matrik atau vector input biasa diberi symbol X dan vector output diberi symbol Y. dalam unsupervised learning tidak terdapat data output atau Y.

### Jenis Nilai Variabel

Berdasarkan nilai yang ada, maka variabel dapat dikelompokkan menjadi beberapa bagian. Yaitu :

1. Nominal  
Variabel yang nilainya berupa symbol, nilainya sendiri hanya berfungsi sebagai label atau member nama, tidak ada hubungan antar nilai nominal, tidak bisa diurutkan atau diukur jaraknya dan hanya uji persamaan yang bias dilakukan.  
Sebagai contoh dalam variabel cuaca, ada *hujan* dan *tidak hujan*. Ini adalah nominal, tidak bias diurutkan atau dihitung jarak antar hujan dan tidak hujan.
2. Ordinal  
Variabel yang nilainya berupa simbol tetapi bias diurutkan, tidak bias diukur jaraknya, tidak bias dijumlahkan. Kadang perbedaannya dengan variabel nominal kurang tegas atau hamper sama.  
Sebagai contoh variabel temperatur mempunyai nilai panas, sedang, dingin. Nilai ini bias diurutkan dari panas-sedang-dingin.
3. Interval  
Variabel yang nilainya bias diurutkan, dan diukur dengan tetap dan unit yang sama. Misalkan variabel temperature diukur dalam derajat Fahrenheit atau Celsius. Perbedaan nilai interval bisa dihitung. Nilai nol (0) tidak didefinisikan secara mutlak.
4. Rasio  
Variabel yang mempunyai nilai nol (0) yang mutlak. Sebagai contoh variabel jarak yang diukur dalam centimeter. Jarak antara suatu objek dengan dirinya sendiri adalah nol (0). Nilai variabel rasio diperlakukan sebagai bilangan riil. Semua operasi aritmatika seperti penjumlahan, pengurangan, pembagian dan sebagainya. Bias dilakukan terhadap nilai rasio.

Dalam prakteknya kebanyakan pengukuran dinyatakan sebagai kontinyu (ordinal) atau diskrit (kategoris, nominal).

### Transformasi Data

Sebelum menggunakan data dengan metode atau teknik yang akan dipakai, harus dilakukan proses *preprocessing* terhadap data. Hal ini dilakukan untuk mendapatkan hasil analisis yang lebih akurat dalam pemakaian teknik-teknik machine learning maupun data mining. Dalam beberapa hal, *pre-processing* data juga bermanfaat untuk pengurangan waktu komputasi terutama untuk *large scale problem*. *Preprocessing* bisa membuat nilai data menjadi lebih kecil tanpa merubah informasi yang dikandungnya. Ada beberapa cara transformasi data yang dilakukan sebelum kita menerapkan suatu metode. Contoh model data pada gambar 2.4.

$$X = \begin{matrix} & X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{31} & X_{32} & \dots & X_{3n} \\ & & & & \dots \\ & & & & X_{mn} \end{matrix}$$

Gambar 4. Model Data Matriks

Dimana n adalah jumlah variabel atau attribute dan m adalah banyaknya observasi, maka beberapa cara untuk mentransformasi data adalah sebagai berikut :

1. *Centering*

Dengan metode *centering* proses terjadi dengan mengurangi setiap data dengan rata-rata dari setiap attribut yang ada. Misalkan proses yang ingin dilakukan adalah mentransformasikan data dalam suatu kolom dengan cara *centering*, maka bias digunakan rumus sebagai berikut :

$$\hat{A} = A - \bar{A} \dots \dots \dots (2.2)$$

Dimana :

$\hat{A}$  adalah vektor hasil setelah *centering*

A adalah vektor kolom

$\bar{A}$  adalah rata-rata dari kolom yang bersangkutan

Kita harus melakukan proses ini untuk semua kolom dari  $i = 1$  sampai  $i = n$ . dari data yang sudah di *centering* bisa kita lakukan beberapa operasi untuk mendapatkan besaran baru. Adapun operasi tersebut adalah :

a. *Matriks Scatter*

*Scatter* adalah jarak antar variabel A1 dengan A2 atau A1 dengan A3, dan seterusnya. *Scatter* antara setiap variabel dengan variabel yang lain bias dihitung dengan rumus :

$$Scatter = \hat{A}'\hat{A} \dots\dots\dots (2.3)$$

b. *Matriks Kovarian*

Seperti *scatter* tapi kita bagi setiap entri dalam matriks dengan jumlah data  $m$ .

$$Kovarian = \frac{\hat{A}'\hat{A}}{m-1} \dots\dots\dots (2.4)$$

2. *Normalisasi*

Setelah *centering* bisa juga dilanjutkan dengan proses berikutnya yaitu membagi setiap data yang sudah di *centering* dengan standar deviasi dari atribut yang bersangkutan.

$$\hat{A} = \frac{A - \bar{A}}{\sigma_X} \dots\dots\dots (2.5)$$

3. *Scaling*

*Scaling* adalah prosedur merubah data sehingga berada dalam skala tertentu. Skala ini bisa antara (0,1), (-1,1) atau skala lain yang dikehendaki. Jika nilai maksimum tiap kolom adalah  $A_{max}$  dan nilai minimumnya adalah  $A_{min}$ , untuk mengubah data ke skala baru, untuk setiap data bisa dilakukan operasi

$$\hat{A} = \frac{A - A_{min}}{A_{max} - A_{min}} * (Batas\ atas - Batas\ bawah) + Batas\ bawah \dots\dots\dots(2.6)$$

Jika ingin data berada dalam skala (-1,1), maka bias menggunakan rumus berikut :

$$\hat{A} = \frac{A - A_{min}}{A_{max} - A_{min}} * (1 - (-1)) + (-1) \dots\dots\dots(2.7)$$

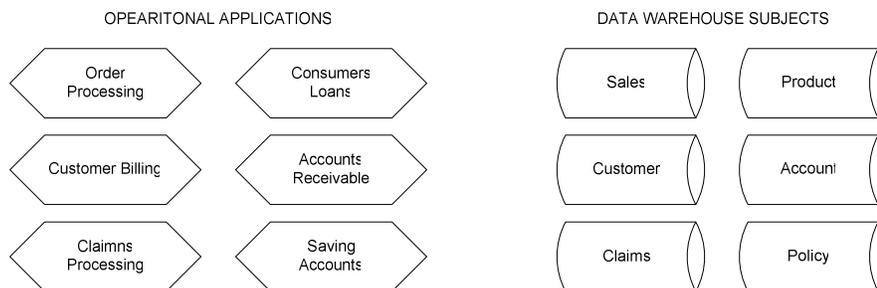
**Data Warehouse**

Data *warehouse* merupakan kumpulan data yang berorientasi subjek, terintegrasi, tidak dapat di *update*, memiliki dimensi waktu, yang digunakan untuk mendukung proses manajemen pengambilan keputusan dan kecerdasan bisnis (Mallach, 2000).

Berikut ini merupakan karakteristik utama dari data *warehouse* :

1. Berorientasi subjek

Data *warehouse* adalah tempat penyimpanan berdasarkan subjek bukan berdasarkan aplikasi. Subjek merupakan bagian dari suatu perusahaan. Contoh subjek pada perusahaan manufaktur adalah penjualan, konsumen, inventori dan lain sebagainya. Gambar di bawah ini menunjukkan perbedaan mengenai data *warehouse* dan *database* operasional



**Gambar 5.** Perbedaan Data Warehouse dan Database Operasional  
 (Sumber : Poniah, 2001, h.21.)

Pada gambar 5 terlihat ada perbedaan yang jelas antara operational applications dan data warehouse subject. Data yang dimiliki memiliki perbedaan identitas dalam aplikasi yang digunakan.

Untuk lebih jelasnya mengenai perbedaan antara *database* operasional dan *database warehouse* bisa dilihat pada table 2.1 di bawah ini.

**Tabel 1.** Perbedaan *Database* Operasional dan Data *Warehouse*

JENIS	DATABASE OPERATIONAL	DATABASE WAREHOUSE
Isi data	Bernilai sekarang atau <i>up-to-date</i>	Arsip, <i>history</i> , rangkuman
Struktur Data	Di optimasi untuk transaksi, normalisasi	Di optimasi untuk <i>query</i> yang kompleks, <i>unnormalisasi</i>
Frekuensi data	Tinggi	Sedang-Rendah
Tipe Akses	<i>Read, update, delete</i>	<i>Read</i>
Penggunaan users	<i>Update</i> secara terus menerus	<i>Update</i> secara periodik
	Banyak	Lebih sedikit

2. Terintegrasi

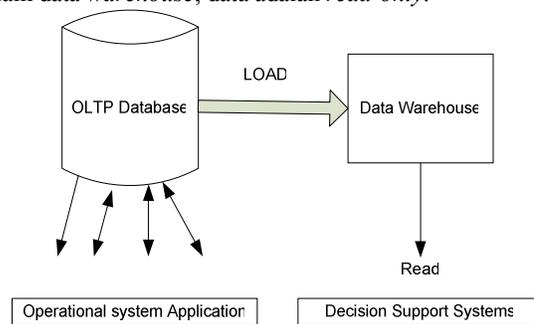
Data pada sumber berbeda dapat di-*encode* dengan cara yang berbeda. Sebagai contoh, data jenis kelamin dapat di-*encode* sebagai 0 dan 1 di satu tempat dan "m" dan "f" di tempat lain. Di dalam data *warehouse*, *encode* tersebut dibersihkan atau dibuat ke dalam satu format sehingga mereka distandarisasi dan konsisten. Data yang terintegrasi mengatasi inkonsistensi dan menyediakan istilah yang seragam di organisasi keseluruhan, juga format waktu dan data yang bervariasi.

3. *Time Variant*

Waktu adalah dimensi penting yang harus didukung oleh semua data *warehouse*. Data untuk analisis dari berbagai sumber berisi berbagai *point* waktu (misal hari, minggu, bulan)

4. *Nonvolatile*

Data dalam *database* operasional akan secara berkala atau periodik dipindahkan ke dalam data *warehouse* sesuai dengan jadwal yang susah ditentukan. Misal perhari, perminggu, perbulan dan lain sebagainya. Sekali masuk ke dalam data *warehouse*, data adalah *read-only*.



**Gambar 6.** Data *Warehouse* adalah *Nonvolatile*  
 (Sumber : Poniah, 2001, h.24)

Pada gambar 2.6 di bawah ini bisa dilihat bahwa *database* OLTP (*Online Transaction Processing*) bisa dibaca, di-*update* dan dihapus. Tetapi pada *database warehouse* hanya bisa dibaca.

5. Ringkas

Jika diperlukan, data operasional dikumpulkan ke dalam ringkasan-ringkasan

6. Granularity

Pada sistem operasional data dibuat secara *real-time* sehingga untuk mendapatkan informasi langsung dilakukan proses *query*. Pada data *warehouse* menganalisis harus memperhatikan *level of detail* misalkan perhari, ringkasan perbulan, atau lain sebagainya.

7. Tidak ternormalisasi

Data di dalam sebuah data *warehouse* biasanya tidak ternormalisasi dan sangat *redundant*.

**Data Mining**

Menurut Turban, dkk (2003, h.15) *Data Mining* (DM) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Definisi informasi menurut Turban, dkk (2003, h.15) adalah informasi merupakan kumpulan fakta (data) yang terorganisasi dengan cara tertentu sehingga bermakna bagi penerimanya. Dengan kata lain informasi berasal dari data yang telah diproses, dimana proses tersebut dikenal dengan istilah data *processing*. Sebagai contoh, jika nama mahasiswa dimasukkan ke dalam data rata-rata nilai, maka akan menjadi sebuah informasi bahwa mahasiswa dengan nama, misal x, yang terdapat di dalam kumpulan data nama mahasiswa tadi, memiliki rata-rata nilai, misal 3,5. Sedangkan pengetahuan terdiri dari informasi yang telah diorganisasi dan diproses untuk memberikan pemahaman, pengalaman, pembelajaran yang terakumulasi, atau keahlian yang diterapkan ke masalah atau proses bisnis.

Berikut ini karakteristik utama dan sasaran data *mining* :

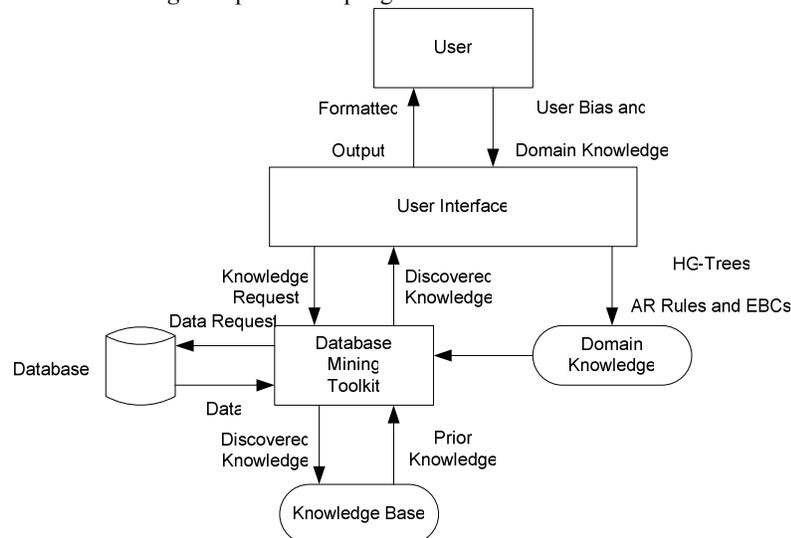
1. Data sering dikubur pada *database* yang sangat besar, yang kadang-kadang berisi data dari beberapa tahun
2. Lingkungan data *mining* biasanya adalah arsitektur klien/server atau arsitektur berbasis web
3. Peranti-peranti baru yang *sophisticated*, meliputi piranti visualisasi canggih, membantu memindah informasi atau mengubur informasi dalam file-file perusahaan atau arsip catatan publik.
4. Pemilik seringkali adalah pengguna akhir diberdayakan oleh data riil dan piranti query lainnya untuk mengajukan pertanyaan khusus dan mendapatkan jawaban secara cepat dengan sedikit atau tanpa keterampilan pemrograman
5. Pemaksaan sering melibatkan penemuan hasil yang tidak diharapkan dan mengharuskan pengguna akhir berpikir kreatif
6. Piranti data *mining* sudah digabung dengan *spreadsheet* dan piranti pengembang perangkat lunak lainnya. Jadi data yang sudah di data *mining* dapat dianalisis dan diproses dengan cepat dan mudah.
7. Karena ada sejumlah besar data dan usaha pencarian *massif*, maka pemrosesan paralel untuk data *mining* kadang-kadang perlu digunakan.

Piranti data *mining* menemukan pola-pola di dalam data dan bahkan menyimpulkan aturan dari data tersebut. Ada tiga tipe metode menurut Nemati dan Barko (Turban, dkk; 2005; h.342) yang digunakan untuk mengidentifikasi pola-pola di dalam data :

1. Model Sederhana (*Query* berbasis SQL, OLAP, Pertimbangan Manusia)
2. Model Sedang (Regresi, Pohon Keputusan, *Clustering*)
3. Model Kompleks (Jaringan Saraf, Induksi aturan lain)

### Model Data Mining

Kesadaran bahwa di dalam data tersebut terdapat pengetahuan yang tersembunyi yang dapat dimanfaatkan membawa manusia ke usaha penemuan pengetahuan di dalam *database* atau data *mining*. Gambar 7 menunjukkan model data *mining* atau penemuan pengetahuan di dalam database.



Gambar 7. Model Data Mining atau Penemuan Pengetahuan di dalam Database

### Teknik Data Mining

Teknik dan piranti data mining dapat digolongkan berdasarkan struktur data dan algoritma yang digunakan. Teknik utama yang digunakan adalah :

1. Metode Statistik

Meliputi *Regresi Linier* dan *Non-Linear*, Poin Estimasi, *Dalil Bayes* (Distribusi Probabilitas), Korelasi, dan *Analisa Cluster*

2. Pohon Keputusan

Pohon keputusan digunakan dalam klasifikasi dan metode *clustering*. Pohon keputusan membagi masalah ke dalam beberapa subset yang terpisah dengan bertolak dari generalisasi ke informasi yang semakin spesifik

3. Pemikiran Berbasis Kasus

Dengan menggunakan kasus histori, pendekatan pemikiran berbasis kasus dapat digunakan untuk mengenali berbagai pola

4. Komputasi Saraf

Jaringan saraf menggunakan banyak node terkoneksi (yang beroperasi dalam suatu cara yang serupa dengan bagaimana saraf otak manusia berfungsi). Pendekatan ini menguji sejumlah besar data histori untuk berbagai pola

5. Agen Cerdas

Salah satu dari pendekatan paling menjanjikan untuk mendapatkan kembali informasi dari *database*, terutama informasi eksternal, adalah penggunaan agen cerdas. Aplikasi data *mining* berbasis web umumnya dimungkinkan oleh agen perangkat lunak cerdas.

6. Algoritma genetika

Dengan menetapkan jumlah tertentu dari hasil akhir yang mungkin dicapai, algoritma genetika berusaha menentukan solusi baru dan lebih baik. Algoritma genetika digunakan untuk aturan *clustering* dan asosiasi

7. Piranti-piranti lainnya

Beberapa piranti lain dapat digunakan, meliputi induksi aturan dan visualisasi data

#### 4. KESIMPULAN

Kesimpulan yang bisa diambil :

1. Proses data mining di dalam suatu pengolahan data dapat dilakukan dengan basis data yang besar agar hasil akhir yang dicapai dapat dilihat pola yang terbentuk.
2. Pengambilan keputusan pada suatu sistem dapat diambil berdasarkan analisa dari suatu pola yang terbentuk, baik keputusan sebagai pendukung ataupun keputusan utama dalam suatu sistem.
3. Pola pengetahuan yang ada bisa didapatkan dengan mengamati hasil akhir yang didapatkan sesuai dengan data-data yang ada.

#### 5. DAFTAR PUSTAKA

- Armadyah A., 2008, Perancangan dan Pembuatan Data Warehouse pada Perpustakaan STMIK Amikom Yogyakarta, MTI Universitas Gadjah Mada, Yogyakarta.
- Herlina T., 2007, Pengolahan dan Penerapan Data Mining untuk Sistem Pendukung Keputusan dalam Kegiatan Akademik di MMTK Yogyakarta, Teknik Elektro UGM, Yogyakarta.
- Kadir, A., 2003, Pengenalan Sistem Informasi, Andi Offset, Yogyakarta.
- Nugroho, B., 2008, Aplikasi E-Learning dengan PHP & Editor Dreamweaver, Universitas Atma Jaya Yogyakarta, Yogyakarta.
- Poniah, P., 2001, Data Warehouse Fundamentals : A Comprehensive Guide for IT Professional. New York : John Wiley & Sons
- Purbo, Onno W., Wahyudi, Aang Arif, 2001, Mengenal E-Commerce, Elex Media Komputindo, Jakarta
- Sanjaya, R., Leong, M., 2008, Mudah Membangun WEB E-Learning, Universitas Atma Jaya Yogyakarta, Yogyakarta
- Santosa, B., 2007, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis, Graha Ilmu, Edisi Pertama, Yogyakarta.
- Sari, R., 2008, Analisa Kemampuan SDM dalam Penggunaan E-Learning, Teknik Elektro UGM, Yogyakarta.
- Turban, 2005, *Decision Support Systems and Intelligent Systems* (Sistem Pendukung Keputusan dan Sistem Cerdas) jilid 1., Andi Offset : Yogyakarta
- Wicaksono, V., 2007, Faktor Pendukung Efektifitas Penggunaan E-Learning, Teknik Elektro UGM, Yogyakarta.