

**DATA MINING:  
SOLUSI PENGEMBANGAN PENGETAHUAN BERDASARKAN BASIS  
DATA**

**Edhy Tri Cahyono**  
**Program Studi D3 Manajemen Informatika**  
**STMIK Jenderal Achmad Yani Yogyakarta**  
[etc0202@yahoo.com](mailto:etc0202@yahoo.com)

**ABSTRAK**

*Data adalah catatan atas fakta-fakta, kejadian-kejadian atau peristiwa-peristiwa yang telah terjadi. Dalam suatu organisasi atau institusi, apapun bentuk dan jenisnya, selama organisasi itu masih hidup dan berkembang, maka data pasti akan senantiasa mengalami pertumbuhan. Pertumbuhan ini akan senantiasa terjadi sepanjang organisasi masih hidup dan belum mati. Akibatnya seiring dengan perjalanan waktu, data di suatu organisasi akan terus dan terus bertambah besar volumenya. Data dalam volume yang besar ini jika tidak dimanfaatkan dengan pengelolaan yang baik justru akan menjadi beban bagi organisasi. Bisa jadi biaya yang ditimbulkan dalam pemeliharaan data tidak sebanding dengan manfaat yang bisa diberikan. Padahal sesungguhnya data yang menggambarkan peristiwa atau kejadian dari sekian waktu yang sudah berjalan bisa digunakan untuk dasar perumusan atas sesuatu hal sehingga menjadi suatu pengetahuan bagi organisasi tersebut. Ibarat pengalaman kehidupan bagi seseorang, semakin lama semakin banyak yang dia alami, akan membuat seseorang menjadi lebih cerdas atau lebih berpengetahuan manakala dia mau belajar dari pengalaman hidupnya tersebut. Data mining adalah suatu metode pengambilan pengetahuan yang dibangun dengan menggunakan data-data historis yang sudah terkumpul sekian lama. Tulisan berikut akan membahas aspek-aspek berkaitan dengan data mining, baik menyangkut pengertian, proses, teknik-teknik serta pada bidang apa saja dapat diterapkan.*

**Kata Kunci:** data tumbuh, pengembangan pengetahuan.

**PENDAHULUAN**

Data adalah catatan atas fakta-fakta, kejadian-kejadian atau peristiwa-peristiwa yang telah terjadi. Dalam suatu organisasi atau institusi, apapun bentuk dan jenisnya, selama organisasi itu masih hidup dan berkembang, maka data pasti akan senantiasa mengalami pertumbuhan. Pertumbuhan ini akan senantiasa terjadi sepanjang organisasi masih hidup dan belum mati. Akibatnya seiring dengan perjalanan waktu, data di suatu organisasi akan terus dan terus bertambah besar volumenya. Data yang besar ini jika tidak dimanfaatkan dengan pengelolaan yang baik justru akan menjadi beban bagi organisasi. Bisa jadi biaya yang ditimbulkan dalam pemeliharaan data tidak sebanding dengan manfaat yang bisa diberikan. Padahal sesungguhnya suatu data yang mencatat peristiwa atau kejadian dari sekian waktu yang sudah berjalan bisa disimpulkan atau

dirumuskan sehingga menjadi suatu pengetahuan bagi organisasi tersebut. Ibarat pengalaman kehidupan bagi seseorang, semakin lama semakin banyak yang dia alami, akan membuat seseorang menjadi lebih cerdas atau lebih berpengetahuan manakala dia mau belajar dari pengalaman hidupnya tersebut. Data mining adalah suatu metode pengambilan pengetahuan yang dibangun dengan menggunakan data-data historis yang sudah terkumpul sekian lama.

Kehadiran data mining ini tak lepas dari masalah yang dihadapi oleh organisasi akan meledaknya jumlah pertumbuhan data yang dikelolanya. Pertumbuhan data organisasi selain karena aktifitas organisasi itu sendiri, juga disebabkan oleh dua hal yaitu pemahaman bahwa segala keputusan organisasi mestinya didasari oleh data yang berkualitas dan lengkap dan yang kedua kenyataan bahwa penggunaan teknologi informasi telah memungkinkan pengumpulan data dalam jumlah yang amat besar. Sebagai contoh suatu toko swalayan atau pusat perbelanjaan yang setiap hari melakukan jutaan transaksi, baik transaksi penjualan barang, pembelian barang, pembayaran hutang ataupun transaksi-transaksi lainnya, bisa diperkirakan berapa jumlah data yang dikumpulkannya dalam waktu satu minggu, satu bulan, satu semester dan satu tahun, tentu data akan tumbuh amat besar dengan cepatnya. Atau suatu bank yang setiap hari melayani ribuan transaksi, maka dalam waktu periode tertentu data yang dikelolanya juga akan semakin besar. Secara teknis data yang besar tersebut bisa ditangani karena adanya teknologi informasi, tetapi pertanyaannya adalah apakah data yang besar tersebut telah memberi manfaat yang sebanding dengan biaya untuk mengelola data tersebut? Hal inilah yang menjadi persoalan bagi banyak organisasi hari-hari ini. Banyak organisasi yang memiliki data besar tetapi sedikit manfaatnya, bisa dikatakan data yang banyak ibarat kuburan data saja, kondisi inilah yang disebut *rich of data but poor of information*.

## **PENGETAHUAN DATA, INFORMASI, PENGETAHUAN DAN DATA MINING**

Data adalah sekumpulan catatan atas fakta-fakta dari suatu kejadian atau peristiwa yang terjadi dalam suatu waktu dan tempat tertentu. Segala fakta yang dicatat dan dikumpulkan akan menjadi data. Data yang diolah dengan cara tertentu sehingga memberi makna yang lebih bagi penerimanya disebut informasi. Dengan informasi yang diolah berdasarkan fakta fakta yang disimpan dalam basis data maka penerima informasi akan bisa mengambil keputusan dengan

tepat yang didasarkan atas fakta. Informasi-informasi yang sejenis dan berlangsung berulang kali akan bisa disimpulkan polanya, sehingga pengenalan pola-pola tertentu tersebut bisa menjadi suatu pengetahuan akan hal-hal tertentu. Proses menemukan pola-pola tertentu inilah akan menjadi pengetahuan bagi organisasi yang memiliki data tersebut, sehingga dengan pengumpulan pengetahuan yang komprehensif tentu akan sangat membantu organisasi dalam pengelolaan dan mencapai tujuannya secara efektif dan efisien. Proses untuk membangun pengetahuan berdasarkan database itulah yang disebut Data Mining. Dalam istilah yang lain Data Mining bisa dimaknai dengan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual atau bisa juga dimaknai dengan analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Kata mining sendiri sebenarnya punya makna usaha untuk mendapatkan sedikit barang berharga dari sejumlah material dasar.

## **METODOLOGI DATA MINING**

Bagaimana sesungguhnya data mining bekerja untuk membangun pengetahuan? Teknik yang dipakai untuk menemukan atau menggali hal-hal penting yang sebelumnya belum diketahui adalah dengan pemodelan. Yang dimaksud pemodelan adalah kegiatan untuk membangun sebuah model pada situasi yang telah diketahui jawabannya dan kemudian menerapkannya pada situasi lain yang akan dicari jawabannya.

Data mining melakukan pencocokan model ke atau menentukan pola dari data yang diobservasi. Ada dua pendekatan matematis yang digunakan dalam pencocokan model yaitu statistik yang memberikan efek non-deterministik dan logik yang murni deterministik. Pada kenyataannya pendekatan statistik lebih banyak dipakai mengingat ketidakpastian yang ada dalam proses pembangkitan data di dunia nyata.

Pada prakteknya aplikasi data mining kebanyakan bertujuan untuk membuat prediksi dan deskripsi. Beberapa variabel atau field-field basis data digunakan untuk memprediksi nilai-nilai variabel masa mendatang yang diperlukan tetapi saat ini belum diketahui. Deskripsi berfokus pada penemuan

pola-pola tersembunyi dari data yang ditelaah. Prediksi dan deskripsi pada datamining dilakukan melalui berbagai cara yaitu:

- Klasifikasi yaitu fungsi pembelajaran yang memetakan (mengklasifikasikan) sebuah unsur (item) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan.
- Regresi yaitu fungsi pembelajaran yang memetakan sebuah unsur data ke sebuah variabel prediksi bernilai nyata.
- Pengelompokan (*clustering*) yaitu proses deskripsi yang banyak digunakan dalam mengidentifikasi sebuah himpunan terbatas pada kategori atau cluster untuk mendeskripsikan data yang ditelaah. Kategori ini dapat bersifat eksklusif dan ekshaustif mutual atau mengandung representasi yang lebih kaya seperti kategori yang hirarkis atau saling menumpu (*overlapping*).
- Peringkasan yaitu metodologi untuk menemukan deskripsi ringkas dari sebuah himpunan data, misal mentabulasikan mean dan deviasi standar untuk semua field tabel.
- Pemodelan Kebergantungan yaitu penemuan sebuah model yang mendeskripsikan kebergantungan yang signifikan antara variabel-variabel. Model ini ada dua tingkat yakni tingkat struktural yang menspesifikasikan variabel yang secara lokal bergantung satu sama lain dan tingkat kuantitatif yang menspesifikasikan tingkat kebergantungan dengan menggunakan skala numerik.
- Pendeteksian perubahan dan deviasi yang berfokus pada penemuan perubahan yang paling signifikan di dalam data dari nilai-nilai yang telah diukur sebelumnya.

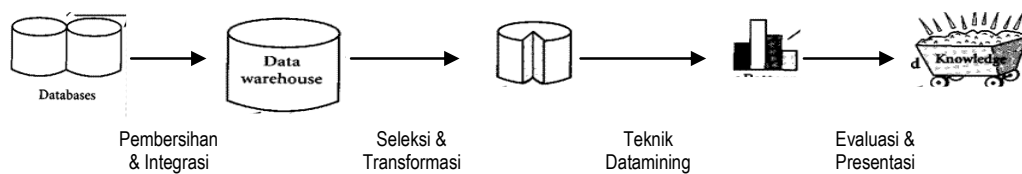
## PROSES DATA MINING

Untuk melakukan proses data mining dari suatu database, tidaklah bisa langsung dilakukan dengan seketika. Ada beberapa langkah dan tahapan yang harus dilakukan sebelum proses data mining itu sendiri dijalankan. Disamping itu, pada umumnya proses data mining berjalan secara interaktif, karena kadang kala hasil data mining pada awalnya tidak sesuai dengan harapan analisnya sehingga perlu dilakukan desain ulang prosesnya.

Hal lain yang perlu diperhatikan juga bagi seorang analis data mining adalah perlunya pemahanan akan data dan proses diperolehnya data tersebut.

Juga perlu memahami mengapa menerapkan data mining dan target yang ingin dicapai. Sehingga secara garis besar sudah ada hipotesa mengenai aksi-aksi yang dapat diterapkan dari hasilnya nanti. Pemahaman tersebut akan sangat membantu dalam mendesain proses dan pemilihan teknik yang akan dipakai dalam data mining..

Ada sejumlah tahapan yang perlu dilakukan dalam proses data mining yakni Pembersihan dan integrasi data, Seleksi dan transformasi data, Aplikasi teknik data mining, Evaluasi dan presentasi atas hasil temuan pola untuk menghasilkan aksi.



Gambar 1 Tahap-tahap data mining

#### 1. Tahap Pembersihan dan Integrasi

Tahap pertama yakni pembersihan dan integrasi data. Pembersihan data adalah proses yang dimaksudkan untuk membuang data yang tidak valid, sehingga data sumber yang akan dipakai untuk proses data mining selanjutnya sudah terjamin validitasnya. Hal ini perlu dilakukan mengingat pada umumnya data yang diperoleh apakah dari sumber internal maupun eksternal selalu saja mengandung unsur kesalahan. Juga kadang database sumber memiliki atribut-atribut yang kurang relevan dengan kepentingan analisis dalam data mining. Data-data yang diragukan validitasnya dan atribut-atribut yang kurang relevan itulah yang perlu dibersihkan atau dibuang sehingga data yang akan dijadikan sumber betul-betul hanya data yang valid dan relevan. Dengan proses ini juga bisa menjadikan ukuran database yang lebih kecil dan ringan sehingga bisa mengefisienkan proses pengelolaannya.

Sedangkan proses integrasi data adalah proses penggabungan data dari beberapa sumber. Bisa jadi sumber data untuk data mining berasal dari berbagai database, sehingga perlu diintegrasikan melalui beberapa penyesuaian.. Penyesuaian bisa berkaitan dengan perbedaan atribut antar database ataupun perbedaaan kategori dalam suatu atribut yang sama, atau bisa juga sama atribut tapi beda domain. Bisa juga

dalam proses integrasi ini sekaligus dilakukan pembersihan atas data-data yang tidak perlu untuk kepentingan data mining. Misalnya integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Proses integrasi dari berbagai database akan diwujudkan dalam suatu data warehouse yang dibangun dengan struktur tertentu yang dipandang tepat untuk kepentingan data mining.

Data ware house berbeda dengan database yang biasa dipakai untuk OLTP (*online transaction processing*). Database di suatu organisasi, pada umumnya untuk menyimpan data transaksi hariannya seperti mencatat penjualan dan pembelian, pembayaran gaji pegawai, administrasi surat menyurat, administrasi pengiriman barang atau pembayaran transaksi yang lainnya. Sedangkan data warehouse dipakai untuk menunjang proses analisis untuk pengambilan keputusan yang lazim juga disebut dengan OLAP (*online analytical processing*). Berbeda dengan OLTP yang hanya memakai operasi query yang sederhana dan berulang-ulang, query untuk OLAP biasanya lebih rumit, bersifat adhoc, dan tidak melibatkan operasi data update. OLAP juga tidak memakai data operasi sehari-hari begitu saja, tetapi memakai data yang sudah terangkum dengan model data yang disebut data cube. Data cube adalah presentasi data multidimensi seperti jenis barang, waktu, lokasi dan sebagainya. Dimensi pada data cube dapat dibuat bertingkat, contohnya dimensi lokasi dapat dibagi menjadi kota, propinsi dan negara. Sedangkan dimensi waktu mencakup jam, hari, minggu, bulan, tahun dan sebagainya. Dengan ini pemakai dapat dengan mudah mendapat rangkuman informasi dari tingkatan dimensi yang lebih luas atau umum seperti negara atau tahun dengan operasi yang disebut *roll-up*. Sebaliknya dengan operasi *drill-down*, pemakai dapat menggali informasi dari tingkatan dimensi yang lebih detil seperti data harian atau data di lokasi yang spesifik. Data cube yang tersedia pada data warehouse memungkinkan pemakai untuk menganalisa data operasi sehari-hari dengan berbagai sudut pandang, dan sangat berguna untuk mengevaluasi suatu asumsi bisnis.

## 2. Tahap Seleksi dan Transformasi

Selanjutnya setelah memiliki dataware house, tahap kedua adalah seleksi dan transformasi data, yaitu proses untuk memilih dan mengubah data menjadi bentuk yang sesuai untuk kepentingan data mining. Hal ini berkaitan dengan diperlukannya kesesuaian antara format data dengan teknik data mining yang akan dipakai. Ketidaksesuaian akan menghasilkan proses yang kualitasnya tidak bisa diandalkan, sehingga proses seleksi dan transformasi data ini akan sangat menentukan kualitas hasil dari data mining nantinya.

### 3. Tahap Penerapan Aplikasi Teknik Data mining.

Tahap ketiga dari proses data mining adalah penerapan aplikasi teknik data mining. Dari sekian banyak teknik data mining, beberapa diantaranya yang populer dipakai antara lain *association rule mining*, *classification* dan *clustering*. *Association rule mining* adalah teknik mining untuk menemukan aturan *assosiatif* antara suatu kombinasi item. Sebagai contoh, aturan *assosiatif* bisa diterapkan untuk analisa pembelian di suatu toko retail atau pasar swalayan untuk mengetahui berapa besar kemungkinan seorang pelanggan membeli popok bayi bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran berkaitan dengan penjualan untuk kombinasi barang tertentu. Penting tidaknya suatu aturan *assosiatif* dapat diketahui dengan dua parameter yaitu *support* dan *confidence*. *Support* adalah persentase kombinasi item tersebut dalam database dan *confidence* adalah hubungan antar item dalam aturan *assosiatif*. Algoritma yang paling populer dikenal sebagai Apriori dengan paradigma *generate and test*, yaitu pembuatan kandidat kombinasi item yang mungkin berdasar aturan tertentu lalu diuji apakah kombinasi item tersebut memenuhi syarat *support* minimum. Kombinasi item yang memenuhi syarat tersebut disebut *frequent itemset*, yang nantinya dipakai untuk membuat aturan-aturan yang memenuhi syarat *confidence* minimum. Algoritma baru yang lebih efisien bernama FP-Tree.

Sedangkan *Classification* adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”,

berupa *decision tree*, formula matematis atau *neural network*. *Decision tree* adalah salah satu metode *classification* yang paling populer karena mudah untuk diinterpretasi oleh manusia. Algoritma *decision tree* yang paling terkenal adalah C4.5, tetapi akhir-akhir ini telah dikembangkan algoritma yang mampu menangani data skala besar yang tidak dapat ditampung di *main memory* seperti RainForest. Metode-metode *classification* yang lain adalah *Bayesian*, *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, dan *k-nearest neighbor*. Proses *classification* biasanya dibagi menjadi dua fase yaitu *learning* dan *test*. Pada fase *learning*, sebagian data yang telah diketahui kelas datanya diumpungkan untuk membentuk model perkiraan. Kemudian pada fase *test*, model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

Berbeda dengan *association rule mining* dan *classification*, dimana kelas data telah ditentukan sebelumnya, *clustering* melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan *clustering* dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu *clustering* sering digolongkan sebagai metode *unsupervised learning*. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau cluster. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi. Banyak algoritma *clustering* memerlukan fungsi jarak untuk mengukur kemiripan antar data, diperlukan juga metode untuk normalisasi bermacam atribut yang dimiliki data. Kategori algoritma *clustering* yang banyak dikenal adalah metode partisi dimana pemakai harus menentukan jumlah  $k$  partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi. Metode lain yang telah lama dikenal adalah metode hierarki yang terbagi dua cara, yakni *bottom-up* yang menggabungkan cluster kecil menjadi cluster lebih besar dan *top-down* yang memecah cluster besar menjadi cluster yang lebih kecil. Kelemahan metode ini adalah bila salah satu penggabungan atau pemecahan dilakukan pada tempat yang salah, tidak akan didapatkan cluster yang optimal. Pendekatan yang banyak diambil adalah



menggabungkan metode *hierarki* dengan metode *clustering* lainnya. Akhir-akhir ini dikembangkan juga metode berdasar kepadatan data, yaitu jumlah data yang ada di sekitar suatu data yang sudah teridentifikasi dalam suatu cluster. Bila jumlah data dalam jangkauan tertentu lebih besar dari nilai ambang batas, data-data tersebut dimasukkan dalam cluster. Kelebihan metode ini adalah bentuk cluster yang lebih fleksibel. Algoritma yang terkenal adalah DBSCAN.

Dari penerapan teknik data mining terhadap data warehouse yang sudah diseleksi, maka akan diperoleh suatu pola atau model tertentu terhadap hal-hal tertentu. Pola-pola inilah yang perlu ditindaklanjuti di tahap berikutnya.

#### 4. Tahap Evaluasi dan Presentasi

Tahap terakhir dari proses data mining adalah evaluasi dan presentasi atas pola yang ditemukan dari proses penerapan teknik data mining di tahap sebelumnya. Evaluasi ini dimaksudkan untuk menemukan hal-hal yang menarik atau dianggap bernilai. Dari proses di tahap penerapan aplikasi teknik data mining biasanya akan dihasilkan pola-pola yang khas atau model prediksi. Pola yang khas ataupun model prediksi inilah yang akan dievaluasi lebih lanjut untuk dinilai apakah hipotesa yang ada memang tercapai atautkah tidak. Jika tercapai berarti berhasil, tetapi jika hasil yang diperoleh tidak sesuai hipotesa maka ada beberapa alternatif yang dapat dilakukan misalnya menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

Berdasarkan pola temuan tersebut maka pekerjaan terpenting adalah bagaimana memformulasikan langkah apa yang akan diambil sebagai pilihan atas penemuan pengetahuan tersebut. Tentu ini sangat menentukan masa depan organisasi sehingga dalam proses perumusan kebijakan ini bisa melibatkan pihak-pihak yang tidak terlibat dalam proses data mining tetapi mempunyai peran penting dalam penentuan kebijakan organisasi. Untuk membantu pemahaman semua pihak yang dihadirkan, maka penggambaran bagaimana pengetahuan ditemukan secara komprehensif perlu dijelaskan, dan akan lebih bagus lagi jika bisa disertai dengan visualisasi-visualisasi tertentu sehingga akan semakin

memudahkan penerimaan akan pengetahuan tersebut. Dan yang yang terpenting adalah pengambilan kebijakan atau keputusan langkah lanjutannya.

## **PENERAPAN DATA MINING**

Banyak aspek yang bisa mendapatkan manfaat dari penerapan data mining. Bagi sebuah perusahaan bisnis, sentuhan data mining bisa dipakai untuk membantu para manajer baik di area pemasaran, keuangan, sumber daya manusia maupun produksi atau operasi. Manajer pemasaran misalnya, dengan data mining bisa mengembangkan databasenya menjadi data ware house yang bisa dipakai untuk mendapatkan pengetahuan tentang karakteristik penjualan dan pembelian yang terjadi. Dari aspek pembeli saja, dengan melakukan analisis tertentu bisa diketahui karakteristiknya berkaitan dengan usia, jenis kelamin, status maupun perilaku pembelian mereka. Dengan mengetahui karakteristik pembeli dan perilaku pembeliannya maka bisa diperoleh pengetahuan yang akan bisa membantu dalam program-program pemasarannya. Program pemasaran yang tepat tentu akan bisa meningkatkan omset penjualannya. Juga dengan mengetahui pola perilaku pembelian para customernya, bisa dipakai untuk memutuskan penataan barang-barang di etalase penjualan, misalnya apakah susu anak dan popok bayi perlu didekatkan ataukah tidak. Apakah mie instan perlu didekatkan dengan rokok ataukah yang lain. Bahkan pengetahuan tentang pola pembelian bisa juga dipakai untuk menentukan program-program diskon. Mengenali perilaku pembeli, juga bisa dipakai untuk menilai loyalitas customernya, sehingga bisa dirumuskan aksi pemeliharaan dan pelayanan kepada customer yang tepat. Selanjutnya penemuan pola penjualan dan pembelian atau belanja perusahaan, bisa dipakai manajer keuangan untuk menyiapkan cash flow secara tepat sehingga naik turunnya kesediaan cash bisa diantisipasi tanpa harus mengganggu kinerja perusahaan.

Di bidang asuransi, Australian Health Insurance Commision bisa menggunakan data mining untuk mengidentifikasi layanan kesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi. Dan dari hasil analisis tersebut bisa dilakukan langkah-langkah penentuan kebijakan sehingga bisa berhasil menghemat satu juta dollar per tahunnya. Dengan cara sama tentunya bisa juga dilakukan data mining untuk transaksi-transaksi

asuransi yang lain sehingga bisa ditemukan pola yang bisa membantu penghematan anggaran.

Di bidang e-commerce, misalnya seseorang yang membeli buku lewat amazon.com bisa dibantu untuk memutuskan buku yang tepat yang sesuai minatnya. Dengan pengetahuan yang dibangun berdasarkan data mining, sistem bisa menyarankan kepada pembeli tersebut buku lain apa saja yang seharusnya dia beli yang relevan dengan pilihan buku yang akan dibelinya tersebut.

Di industri perbankan, dengan data mining bisa dipakai untuk mengidentifikasi apakah sudah terjadi penipuan terhadap penggunaan kartu kredit atau penentuan keputusan terhadap pemberian kredit bagi seorang calon nasabah. Pola-pola transaksi nasabah bisa diidentifikasi dari berbagai database perbankan yang ada sehingga histori seseorang berkaitan dengan transaksi perbankannya bisa dijadikan dasar tindakan persetujuan pencairan suatu kredit.

Di bidang kesehatan, data-data rekam medik yang terjadi bertahun-tahun dan terus berkembang di berbagai rumah sakit tentu bisa sangat bermanfaat jika bisa dijadikan dasar pengembangan pengetahuan untuk membantu kebijakan-kebijakan dan tindakan di bidang medis. Sebagai contoh misal suatu obat yang banyak dipakai ternyata kemudian ditemukan adanya hal yang menyebabkan tidak boleh dikonsumsi lagi. Seandainya saat itu ada banyak pasien rumah sakit yang bisa jadi dirawat inap atau mungkin rawat jalan banyak yang memakai tentu akan sangat sulit untuk langsung bisa menghentikan dengan seketika. Tetapi dengan data mining tentu hal tersebut bukan sesuatu yang sulit karena dengan cepat akan diketahuai siapa saja pasien yang menggunakan obat tersebut dan segera bisa dihubungi untuk menghentikan penggunaan obatnya serta diberi obat penggantinya.

## **KESIMPULAN**

Sesungguhnya memiliki data yang besar adalah ibarat memiliki suatu sumber daya alam yang tersimpan di dalam bumi. Sumber daya alam atau material tersebut tidak punya makna sampai ada proses penambangan material dilakukan dan dilanjutkan dengan pengolahan menjadi barang-barang yang bermanfaat bagi kehidupan. Demikian juga dengan memiliki data besar, data tersebut tidak akan bermakna manakala dibiarkan tersimpan rapi di gudang penyimpanan data dan tidak ada proses untuk membuka dan mengolahnya menjadi sesuatu yang berguna bagi pemiliknya. Melakukan data mining berarti

melakukan upaya pemanfaatan data yang senantiasa tumbuh membesar volumenya dengan teknik-teknik tertentu sehingga bisa diperoleh pola atau model yang bisa memberi pengetahuan yang berguna bagi pemilik data atau organisasi, sehingga organisasi akan semakin 'berpengetahuan' atau 'cerdas' dan dengannya tujuan organisasi bisa dicapai secara efektif dan efisien.

#### DAFTAR PUSTAKA

- Santosa, Budi, 2007, *Data Mining: Teknik dan Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.
- Turban, Efrain, 2005, *Decision Support System and Intelligent System-7th Ed. (Sistem Pendukung Keputusan dan Sistem Cerdas, Jilid 1, diterjemahkan oleh Dwi Prabandani)*, Andi, Yogyakarta.
- Kantardzic, Mehmed, 2003, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons.
- Sucahyo, Y.G. 2003, *Penerapan Data Mining: Permasalahan Apa Saja yang Bisa Diselesaikannya?*, <http://www.ilmukomputer.com>, 2009.
- Pramudiono, Iko, 2003, *Pengantar Data Mining: Manambang Permata Pengetahuan di Gunung Data*, <http://www.ilmukomputer.com>, 2009.
- Moertini, Veronika S., 2002, *Data Mining sebagai Solusi Bisnis*, INTEGRAL Vol 7 No 1.