

Data Mining untuk Dunia Bisnis

Heribertus Yulianton

Fakultas Teknologi Informasi, Universitas Stikubank Semarang

email : heri@unisbank.ac.id

Abstrak : Pada artikel ini akan dijelaskan prinsip dasar penggunaan data mining dalam lingkungan Business Intelligence (BI) untuk mendapatkan pengetahuan tentang bisnis.

Kata kunci : business intelligence, data mining

PENDAHULUAN

Bisnis tidak berjalan pada ruang tertutup. Oleh karena itu keberhasilan dalam menjalankan bisnis tidak bergantung pada bagaimana cara menjalankan bisnis, akan tetapi pada bagaimana menjalankannya jika dibandingkan dengan lainnya. Kunci untuk membuat perbedaan terletak pada pemakaian data yang tersimpan pada sistem yang digunakan untuk bisnis sehari-hari.

Hasil utama dari pengolahan data tersebut yaitu kemampuan untuk memprediksi fakta lingkungan bisnis. Hal ini membuat kita mampu melakukan tindakan proaktif dibandingkan hanya reaktif. Data memberikan pengetahuan tak ternilai tentang masa depan. Data membantu kita mengoperasikan bisnis secara optimal, yaitu dengan cara mengembangkan *market share*, meningkatkan *customer share*, dan membangun loyalitas pelanggan lewat fokus layanan yang jelas.

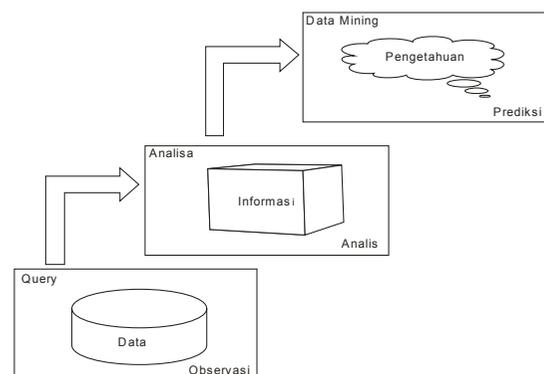
Permasalahannya adalah bagaimana cara mendapatkan pengetahuan ini dari data yang dimiliki. Kuncinya yaitu data mining, yang menjadi teknik paling populer untuk membangun sistem pendukung keputusan cerdas, mengaplikasikan alat dan metode yang digunakan di matematika statistik dan mesin pembelajaran.

Evolusi Pengumpulan Pengetahuan

Menyimpan banyak data tentang proses bisnis hari demi hari. Data ini sumber informasi yang berharga tentang bisnis, cara pengoperasian bisnis, dan pelanggan bisnis. Menggunakan informasi yang terkandung dalam

semua data untuk memperoleh pengetahuan yang dibutuhkan untuk menjalankan bisnis merupakan jalan paling penting untuk meningkatkan persaingan.

Gambar 1 menunjukkan evolusi yang telah berjalan pada beberapa tahun belakangan. Dimulai dari dapat melakukan query terhadap data operasional, menghasilkan report atau chart. Langkah logis berikutnya adalah menganalisa data hasil dengan statistik tradisional atau alat OLAP, mencari *trend* atau mencoba memverifikasi hipotesis. Dapat mencoba memodelkan relasi dalam data kita untuk mencari perilaku bisnis kita dalam situasi tertentu. Setelah itu kita dapat menggunakan suatu model untuk memberikan tanda jika suatu situasi tertentu membutuhkan perhatian kita. Dalam langkah terakhir, kita dapat berbicara Business Intelligence (BI), yaitu pengetahuan tentang bisnis kita digunakan untuk menjalankan keputusan yang kita buat.



Gambar 1. Evolusi dari Query ke Data Mining

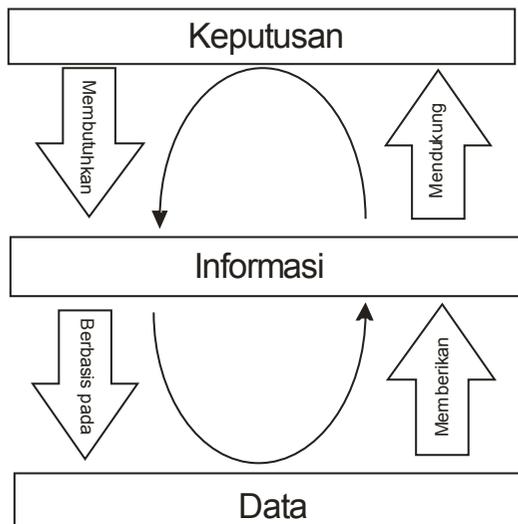
Langkah pertama dalam Gambar 1 berbasis pada pertanyaan, atau pengetahuan, yang kita masukkan sendiri dan kita validasi

terhadap data yang tersedia. Langkah ketiga dalam Gambar 1 yaitu data mining, dengan berbasis pada data asli akan menghasilkan pengetahuan. Pengetahuan baru ini digunakan untuk memodelkan bisnis kita tanpa bergantung pada asumsi yang tidak berasal dari data perusahaan kita.

Data mining bukan sekedar publisitas berlebihan, tetapi langkah logis dalam pengumpulan pengetahuan yang mungkin merupakan model perusahaan paling strategis. Melakukan langkah ini tidak berarti kita harus melakukan langkah sebelumnya, tetapi memperhatikan langkah-langkah tersebut akan membantu dalam persiapan. Data mining tidak akan mengubah sistem yang sudah ada menjadi usang, tetapi justru akan memperbesar sistem dengan memperkaya data yang digunakan.

Dari Data ke Keputusan

Business Intelligence (BI) adalah semua penggalian modal bisnis untuk mendapatkan keuntungan dari data yang tersedia, baik yang tersebar pada sistem yang berbeda, maupun yang terintegrasi dalam tempat penyimpanan yang terpusat. BI memberikan jalan untuk memperoleh pengetahuan yang dibutuhkan untuk membuat keputusan yang baik tentang bisnis kita, seperti ditunjukkan dalam Gambar 2.



Gambar 2. Keputusan, Informasi, dan Data

Kita dapat menginterpretasikan Gambar 2 sebagai sebuah siklus. Membuat keputusan membutuhkan informasi yang berbasis pada data, sedangkan data menyediakan informasi

untuk mendukung keputusan, demikian dalam hal ini data sendiri tidak memberikan pertimbangan atau interpretasi dan tindakan dasar. Konteks dan pemakaian data mengubahnya menjadi informasi. Menghubungkan potongan informasi memberikan petunjuk ke pengetahuan yang dapat mendukung keputusan.

Dalam situasi ideal, hasil keputusan tersebut diumpan balik ke lingkungan BI, melengkapi siklus yang ditunjukkan dalam Gambar 2. Hal ini memungkinkan keputusan dibuat berdasarkan pengetahuan nyata bukan berdasarkan pada perasaan sehingga menghasilkan organisasi yang mempunyai kemampuan belajar. Lingkungan yang optimal memungkinkan persilangan antara data warehousing dan data mining. Data warehousing memungkinkan akses ke data terintegrasi yang tak ternilai yang dapat ditambang. Data mining memberikan hasil yang dapat diintegrasikan kembali ke dalam data warehouse, dan menjadi bagian integral dari pengetahuan organisasi kita. Dari pengetahuan ini wilayah baru menarik untuk ditambang dapat ditemukan.

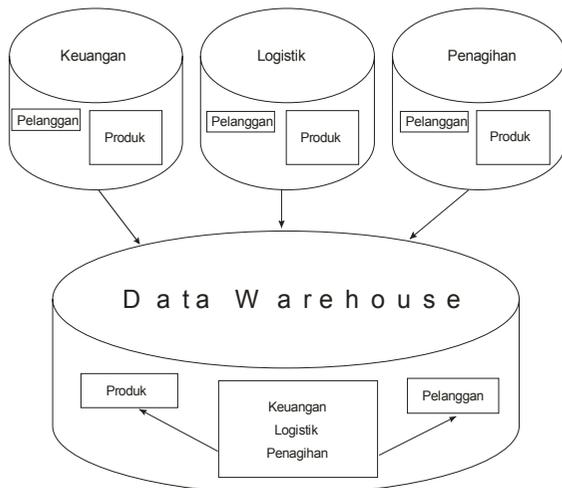
Banyak pengembangan saat ini diarahkan untuk mengintegrasikan data mining ke dalam lingkungan BI. Bahkan mungkin berakhir pada mesin basisdata, memperluas kemampuan query dengan suatu seperti “select 1000 kandidat paling mungkin untuk membeli produk X dari pelanggan kita”. Alat data mining aktual akan menjadi bagian integral dari lingkungan BI dan dapat digabungkan dengan aplikasi yang secara langsung mendukung alur bisnis kita.

Konsep Data Warehousing

Akan banyak terjadi manipulasi data pada lingkungan BI apalagi setelah menambahkan data mining. Saat itulah kita membutuhkan data warehouse.

Data warehouse sebenarnya tidak terlalu dibutuhkan, namun akan banyak membantu jika ada. Sebagian besar persiapan untuk data mining, telah dilakukan ketika lingkungan BI lengkap telah berada di tempat.

Ada beberapa alasan untuk membangun data warehouse. Dalam gambar 3, kita akan melihat bagaimana mengintegrasikan data dari beberapa sistem sumber ke dalam pandangan tunggal dari data kita.



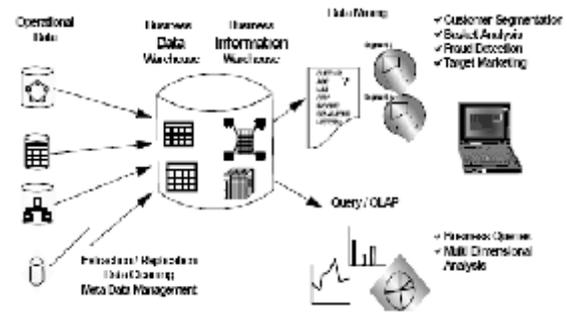
Gambar 3. Fokus Aplikasi dan Fokus Subjek

Pada gambar 3 menunjukkan bahwa fokus bergeser dari aplikasi operasional ke subjek yang penting untuk bisnis, seperti produk dan pelanggan.

Disamping integrasi data, kita membutuhkan beberapa sistem terpisah sebagai cara menggunakan data untuk beberapa alasan: query pada sistem operasional yang akan dijalankan terhadap model data yang tidak dirancang untuk hal ini, berjalan pada sistem yang mempunyai tujuan lain; query akan berebut sumberdaya jika transaksi berjalan pada sistem operasional, yang mengakibatkan penundaan yang tidak dapat diterima dalam pemrosesan waktu nyata; data yang selalu berubah mempersulit analisa; informasi harus berkorelasi antar aplikasi untuk menjelaskan semua relasi; data operasional lebih disesuaikan untuk kecepatan transaksi bukan untuk pemahaman manusia.

Akses khusus dari banyak user dapat juga menimbulkan kekhawatiran mengenai keamanan. Pengiriman informasi bisnis dari semua data ini dipagari oleh kapasitas bagian IT dalam organisasi kita.

Jawaban dari banyak masalah ini, walaupun tidak sederhana, yaitu dengan membangun lingkungan BI dengan penyimpanan terpusat untuk semua data perusahaan, berisi sejarah operasional sistem terintegrasi diperkaya dengan sumber lain yang mungkin diminati. Gambar 4 memberikan tinjauan komponen dari lingkungan BI secara umum.



Gambar 4. Lingkungan Business Intelligence

OLAP dan Data Mining

Alat Online Analytical Processing (OLAP) memberikan lingkungan interaktif sehingga user dapat menganalisa data bisnis pada “kecepatan berpikir” daripada harus menunggu hasil query selama sehari. OLAP hidup dengan fakta bahwa tiap query akan segera menghasilkan pertanyaan baru, yang harus diproses sebelum kita lupa apa yang kita cari.

Dengan OLAP kita hanya dapat menemukan informasi yang kita cari pada tempat pertama. Kita sebut ini analisa *verification-driven*. Sistem data mining akan keluar dan mencari semua informasi, tanpa interferensi manusia atau input, sehingga analisisnya *discovery-driven*. Sistem data mining menggunakan beberapa teknik untuk menentukan kunci relasi dan kecenderungan dalam data. Alat data mining dapat melihat pada banyak relasi pada waktu yang sama, mementingkan hal yang dominan atau luar biasa. Dengan cara ini kita dapat memperoleh pengetahuan bisnis baru dari data yang sudah ada.

Data mining dapat secara efektif menghadapi ketidakkonsistenan pada data. Walaupun sumber kita bersih, terintegrasi, dan tervalidasi, namun mungkin berisi data dunia nyata yang keliru. *Noise* ini dapat disebabkan oleh kesalahan input oleh user atau kesalahan pengisian formulir oleh pelanggan. Jika tidak terlalu sering muncul, alat data mining dapat mengabaikan *noise* dan masih mencari pola keseluruhan yang muncul pada data.

Di sisi lain, data dapat berisi pola yang hanya benar untuk subset kecil data. Efek ini secara statistik tidak penting, tetapi bisa jadi subset pelanggan ini yang paling penting karena

perilaku mereka. Algoritma data mining dapat mencari suatu lokalitas sedemikian hingga tidak akan tersesat dalam generalisasi.

Efek *noise* dan lokalitas adalah saling berlawanan. Bergantung pada pendekatan, alat mining dapat melihat *noise* sebagai efek penting dan mengabaikan lokalitas sebagai *noise*. Tidak ada aturan keras dan cepat mengenai hal ini, tetapi masalah ini dapat diselesaikan dengan aplikasi bisnis model data mining kita.

Data mining akan melihat data dari sudut yang berbeda pada waktu yang sama. Hal ini mencegah pengabaian atribut yang terlihat tidak relevan. Data mining akan mencari interdependensi antara atribut yang memungkinkan ekstraksi semua informasi yang relevan dari data, walaupun tersembunyi dalam kombinasi beberapa atribut.

Output data mining dapat memberikan fleksibilitas. Sebagai contoh, jika kita mempunyai anggaran untuk mengirim informasi ke 1000 orang tentang produk baru, query atau analisa OLAP langsung pada data tidak akan pernah bisa memilih secara tepat orang sejumlah tersebut dari database. Dengan tambahan atribut pada data yang dapat digunakan pada query atau analisa OLAP, data mining memungkinkan kita mencari 1000 orang yang paling mungkin untuk menjawab. Contoh ini juga menunjukkan bahwa data mining tidak menggantikan OLAP, tetapi memperluasnya.

Definisi Data Mining

Definisi data mining secara formal adalah proses mengekstrak informasi yang valid, bermanfaat, tak dikenal, dan dapat dipahami dari data dan menggunakannya untuk membuat keputusan bisnis.

Data mining bukan alat dalam kotak yang dapat secara sederhana dibeli dan dijalankan terhadap lingkungan BI kita, dan yang akan secara otomatis menghasilkan pengertian bisnis yang menarik.

Informasi yang diekstrak harus benar dan secara statistik berarti untuk mendukung keputusan yang cukup beralasan. Validitas berarti kebenaran dan juga kelengkapan. Kita tidak Cuma membutuhkan pelanggan yang benar dari database tetapi semua pelanggan. Hal ini

membutuhkan data asli dan proses data mining yang valid.

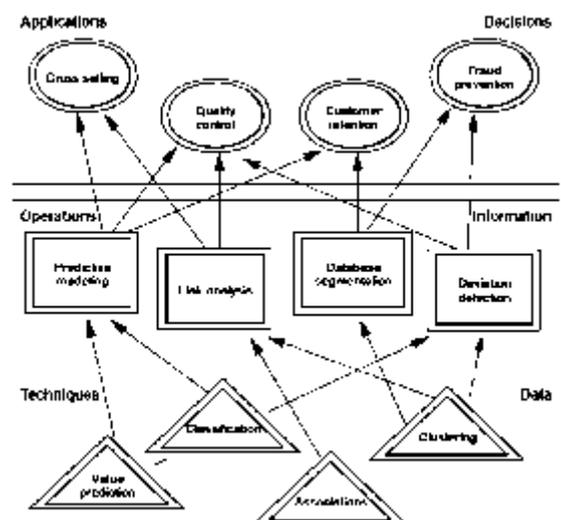
Proses data mining memberikan hasil yang benar dan berarti, namun pengetahuan ini harus bermanfaat untuk bisnis. Sebagai contoh, jika hasil menyebutkan bahwa kita harus memvariasi tindakan pemasaran ke dalam banyak channel, kita mungkin tidak dapat bertindak atas pengetahuan ini. Hasil juga harus memungkinkan kita bertindak sebelum kompetitor kita bertindak.

Data mining dimaksudkan untuk menghasilkan informasi baru. Jika proses hanya mengeluarkan hasil sederhana, dorongan untuk menggunakan data mining akan hilang. Sifat ini membedakan antara verifikasi dan penemuan.

Hasil proses data mining harus mudah dijelaskan dalam istilah bisnis. Atau mungkin sebagai contoh hanya sekedar model statistik yang dapat digunakan untuk mengelompokkan pelanggan. Model sendiri paling tidak memberikan pengertian tentang cara pelanggan dikelompokkan dan faktor yang mempengaruhi pengelompokkan ini. Di beberapa negara terdapat hukum yang meminta perusahaan memberikan pengertian ini.

Aplikasi dan Operasi Data Mining

Gambar 5 menunjukkan beberapa contoh aplikasi, operasi, dan teknik yang digunakan dalam data mining, dengan relasinya.



Gambar 5. Aplikasi, Operasi, dan Teknik

Gambar tersebut tidak beranggapan lengkap. Tetapi memberikan ide tiga level yang kita hadapi. Aplikasi terlihat pada level bisnis, tempat keputusan dibuat. Operasi ditangani oleh pakar data mining pada level informasi. Kemudian satu atau beberapa alat data mining digunakan untuk memberikan teknik mengoperasikan data, sebagaimana ditunjukkan pada bagian bawah Gambar 5.

Aplikasi aktual data mining bergantung sebagian pada bisnis kita, dan sebagian pada imajinasi kita dan pakar mining. Tabel 1 memberikan pandangan aplikasi data mining yang telah digunakan. Kita distribusikan sejumlah contoh atas tiga kategori utama.

Tabel 1. Area Aplikasi Data Mining

Manajemen Pemasaran	Manajemen Resiko	Manajemen Proses
Target Pemasaran	Peramalan	Optimisasi Inventori
Manajemen Relasi	Retensi Pelanggan	Quality Control
Manajemen Channel	Analisa Erosi	Peramalan Kebutuhan
Optimisasi Berbasis Pasar	Underwriting	Business Scorecards
Penjualan Silang	Analisa Kompetitif	
Segmentasi Pasar	Healthcare Fraud	
Analisa Penggunaan Web		

Sebagaimana terlihat dalam Gambar 5, aplikasi didukung oleh operasi data mining. Kategori utama operasi ini dijelaskan dibawah ini beserta dengan contohnya.

Pemodelan prediktif: memrediksi nilai atribut dengan menggunakan contoh. Contoh aplikasi: menentukan kategori resiko pada pelanggan baru, atau memrediksi kemungkinan pelanggan menjawab surat.

Segmentasi basisdata: menggunakan atribut untuk mencari kelompok *record* yang tiap kelompok mempunyai kesamaan atribut,

dan perbedaan antar kelompok cukup jelas. Contoh aplikasi: mengelompokkan pelanggan berdasarkan sifatnya, atau sebagai langkah persiapan untuk pemodelan prediktif.

Analisa hubungan: mencari hubungan antar *records* dalam transaksi atau atas waktu. Contoh aplikasi: menganalisa produk mana yang dijual secara bersama untuk mengoptimalkan tata ruang toko atau inventori. Operasi tipe ini dapat digunakan untuk menganalisa pertanyaan atau rangkaian pengobatan medis.

Deteksi penyimpangan: mencari *records* atau rangkaian *records* dalam basisdata yang berisi nilai yang tidak diinginkan. Contoh aplikasi: mengidentifikasi pola perilaku curang atau mengontrol kualitas proses produksi.

Dorongan Bisnis

Perlu diperhatikan bahwa, jika lingkungan saat ini memenuhi kebutuhan bisnis kita, maka kita tidak butuh sesuatu yang lebih. Tetapi mungkin kita ingin tahu mengapa investasi kita tidak menghasilkan lebih dari yang telah dihasilkan. Lebih jauh lagi ingin mengetahui tentang pelanggan yang berpindah ke perusahaan lain, dan mengapa mereka melakukan itu. Hal itulah yang menjadi pintu masuknya data mining, lebih dikarenakan kebutuhan bisnis dibandingkan kemungkinan teknis.

Berikut ini adalah contoh dorongan bisnis terhadap data mining. Pemenuhan pasar: sulit untuk mencari pelanggan yang tepat untuk produk kita, atau produk yang tepat untuk pelanggan kita. Batas industri kabur: merger, pengambilalihan, dan diversifikasi membuat perubahan mendadak dalam pasar. Diferensiasi kurang jelas: tidak jelas pelanggan macam apa yang kita atau kompetitor kita layani atau akan layani. Channel alternatif yang berkembang pesat: internet mendapatkan banyak perhatian dari pelanggan tertentu. Pengurangan waktu ke pasar: memperpendek siklus waktu yang dibutuhkan untuk tetap mengikuti, atau bahkan memrediksi, perilaku pelanggan.

Tema yang sama dari dorongan ini adalah kekurangan informasi tentang perilaku pasar. Pada saat yang sama banyak data tersedia, tetapi tidak ada cara jelas untuk mengekstrak informasi yang dibutuhkan.

Kemudahan Teknologi

Ada tiga pengembangan teknologi yang penting yang memungkinkan aplikasi bisnis data mining. Pertama, penelitian pada aplikasi teknik mesin pembelajaran digunakan untuk menangani pertanyaan bisnis praktis, memungkinkan jawaban yang tidak hanya menarik perhatian secara saintifik, tetapi juga berelasi pada dunia nyata.

Kedua, pengembangan teknologi hardware dan software, membangun sistem yang dapat menyaring banyak data dalam waktu nyata. Hal ini berarti dapat membuat prediksi tentang perilaku sebelum terjadi secara aktual, memungkinkan membuat keputusan yang tepat dalam waktu yang tepat.

Ketiga, konektivitas terbuka antar basisdata dan alat, membuat kemudahan akses ke sembarang data dari sembarang paket software.

Dengan kata lain, pengembangan ini memungkinkan data mining karena kita mempunyai cukup kekuatan, kita tahu cara mengaplikasikannya, dan datanya tersedia. Juga berarti bahwa sekarang cukup mungkin untuk setiap orang dengan latarbelakang statistik, untuk mengaplikasikan data mining pada banyak data untuk menjawab pertanyaan bisnis nyata.

Faktor Organisasi

Budaya dalam organisasi harus mendukung aliran data dan informasi yang dibutuhkan, selain harus mengakomodasi aliran hasil dari upaya data mining. Berarti kita membutuhkan budaya komunikasi yang terbuka. Dalam budaya ini, orang bekerja sama secara aktif dalam pertukaran informasi. Hal ini secara khusus dibutuhkan dalam interaksi antara bagian bisnis dan bagian teknis. Orang harus mau menerima informasi baru, dan berdasarkan informasi tersebut, mengubah cara mereka bekerja.

Jika ada orang yang memproteksi datanya dan tidak mau membagi informasi baik yang dibutuhkan maupun yang dihasilkan, maka organisasi perlu mengambil upaya konsultasi baik internal maupun eksternal untuk mengubah hal ini. Hal ini bukan pekerjaan yang mudah, namun penting demi keberhasilan upaya data mining.

Orang yang terlibat dalam aktivitas data mining dapat dibagi menjadi tiga peran umum. Berikut ini adalah penjelasan dari tiap peran.

Pakar domain: orang yang tahu lingkungan, proses, pelanggan, dan kompetitor bisnis. Umumnya adalah orang dalam fungsi manajemen bisnis tinggi.

Administrator basisdata: orang yang tahu di mana dan bagaimana data perusahaan disimpan, cara mengaksesnya, dan bagaimana merelasikannya dengan penyimpanan data lain.

Ahli mining: orang dengan latar belakang analisa data yang paling tidak mempunyai pengetahuan statistik. Mereka dapat mengaplikasikan teknik data mining dan menginterpretasikan hasil secara teknis. Mereka harus dapat membangun relasi dengan pakar domain untuk petunjuk bisnis terhadap hasil mereka, dan dengan administrator basisdata untuk mengakses data yang dibutuhkan untuk aktivitas mereka.

Umumnya, dua peran pertama sudah ada pada perusahaan. Peran ketiga mungkin didapat dari luar ketika pertama kali perusahaan masuk ke proses data mining. Posisi ini ditempati orang dalam setelah pemindahan pengetahuan selesai.

Satu kesulitan utama dalam mencari orang yang tepat baik dari dalam maupun dari luar organisasi adalah variasi domain yang harus digabungkan dalam data mining. Dibutuhkan lingkungan bisnis, analisa, dan teknikal untuk menopang arus dua informasi dua arah secara kontinyu.

Arsitektur IT

Jika budaya organisasi harus mendukung arus informasi, maka arsitektur IT harus mendukung arus data. Kita butuh akses yang cepat, terukur, dan terbuka terhadap data yang tersedia dan fleksibilitas untuk mengekstrak dan memperbarui subset data dalam lingkungan yang akan digunakan untuk data mining. Contoh lingkungan yang baik adalah lingkungan yang mendukung BI dan mudah dalam membuat pasar data dari gudang pusat.

Disamping mengakses dan mengirim data, arsitektur IT juga harus bisa mempunyai kapasitas yang cukup untuk melakukan proses data mining, atau mudah untuk menambah kapasitas. Hal ini berarti menambah sistem

ekstra terdedikasi untuk mining, atau menjalankannya sebagai tambahan dari proses yang sudah ada dalam salah satu sistem.

Data tentu saja harus tersedia. Jumlah data mentah biasanya bukan masalah, tetapi jumlah data yang bersih, bermanfaat, relevan, dan terintegrasi mungkin kurang dari yang dipikirkan.

Tidak ada aturan tetap tentang jumlah data yang dibutuhkan untuk memulai mining. Sebagai aturan dasar, beberapa ribu *records*, dan sepuluh atau lebih atribut, adalah awal yang baik. Angka tersebut bergantung dari teknik data mining yang digunakan.

Alat yang digunakan untuk data mining harus dapat mendukung akses data, praproses, mining, visualisasi, penyimpanan, dan pemeliharaan hasil. Hal ini dapat didukung dari paket tunggal, atau mungkin membutuhkan beberapa alat. Integrasi antar alat adalah penting.

Kita juga harus memperhatikan skalabilitas alat yang akan digunakan. Kita akan selalu ingin menambah data ekstra, mengeksplorasi banyak sejarah, atau menerima hasil lebih cepat. Waktu pemrosesan seharusnya tidak bergerak lebih jauh dari linieritas jumlah data, baik jumlah atribut maupun jumlah *records*.

PENUTUP

Telah ditunjukkan pada artikel ini konsep business intelligence, data warehousing, Online Analytical Processing (OLAP), dan data mining. Pada artikel ini juga telah didiskusikan keuntungan data mining dan pekerjaan yang harus dilakukan ketika merencanakan untuk mengimplementasikan data mining.

DAFTAR PUSTAKA

1. Chang, E., Dillon, T., & Hussain, F. (2006). Trust and reputation for service-oriented environments: technologies for building business intelligence and consumer confidence.
2. Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The integration of business intelligence and knowledge

management, IBM Systems Journal, 41(4), 697—713.

3. Dhar, V., & Stein, R. (1997). Seven methods for transforming corporate data into business intelligence. Prentice Hall Upper Saddle River, NJ.
4. Kahaner, L. (1997). Competitive intelligence: how to gather, analyze, and use information to move your business to the top. Simon & Schuster.
5. Liautaud, B., & Hammond, M. (2000). E-business intelligence: turning information into knowledge into profit. McGraw-Hill Professional.
6. Luhn, H. P. (1958). A business intelligence system, IBM Journal of Research and Development, 2(4), 314—319.