GAES - PACE Book Publisher
https://digitalpress.gaes-edu.com/index.php/gaespace/index
ISBN: 978-623-97711-9-5
ISBN (e): xxx-xxx
Publication Year: (2022)

Data WareHouse dan Mining

Suhirman [1]
Universitas Teknologi Yogyakarta[1]
*E-mail: suhirman@staff.uty.ac.id

Defenisi Data Warehouse

Menurut Usama Fayyad (1996), Pengguna menerapkan keahliannya dalam hal masalah, dan komputer melakukan analisis data yang canggih untuk memilih data yang tepat dan menempatkannya dalam format yang sesuai untuk pengambilan keputusan. Menurut Klimavicius (2008), sistem data warehouse mempresentasikan sebuah sumber informasi untuk menganalisa pengembangan dan hasil dari sebuah perusahaan atau organisasi didalam lingkungan yang selalu berubah. Data di dalam data warehouse menggambarkan peristiwa dan status dari proses bisnis, produk dan jasa, tujuan dan unit-unit organisasi.

Data warehouse merupakan kumpulan dari data yang berorientasi subjek, terintegrasi, nonvolatile, dan mempunyai variansi waktu untuk mendukung pengambilan keputusan manajemen. Data warehouse (dalam bermacam bentuk) merepresentasikan sebuah basis data pusat bagi keseluruhan perusahaan untuk menyimpan dan mengakses data historis serta keberadaannya terpisah dari sistem operasional. Data warehouse merupakan suatu konsep dan kombinasi teknologi yang memfasilitasi organisiasi untuk mengelola dan memelihara data historis yang diperoleh dari system atau aplikasi operasional.

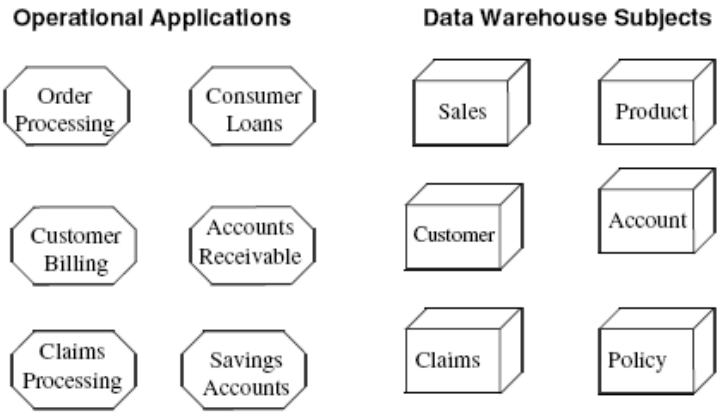
Dari beberapa definisi diatas dapat disimpulkan bahwa data warehouse merupakan basis data yang terpusat dan saling bereaksi untuk mengelola dan memelihara data historis yang berorientasi subjek, terintegrasi, nonvolatile dan mempunyai variansi waktu untuk mendukung pengambilan keputusan.

Karakteristik Data Werehouse

Data warehouse memiliki beberapa karakteristik, sebuah data warehouse memiliki karakteristik utama sebagai berikut:

Berorientasi Subjek

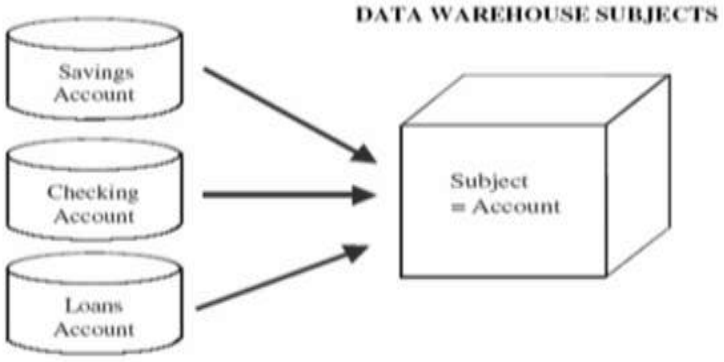
Karakteristik dari data warehouse yang pertama adalah berorientasi subjek, karakteristik ini pada data warehouse berarti bahwa data-data pada data warehouse diorganisir berdasarkan topik atau subjek bisnis. Sistem operasi klasik diorganisir pada seputar aplikasi yang dimiliki perusahan. Pada perusahaan asuransi misalnya, aplikasi-aplikasi yang dimiliki dan digunakan untuk memproses datadata seperti data mobil, data kehidupan pelanggan, data kesehatan pelanggan, serta data korban kecelakaan. Area-area subjek utama untuk perusahaan asuransi antara lain adalah pelanggan, polis, premium, dan data klaim. Untuk sebuah perusahaan manufaktur, area subjeknya antara lain adalah produk, SKU, penjualan, vendor, dan lain-lain. Setiap jenis perusahaan mempunyai sekumpulan subjek-subjek yang unik. Pada gambar 1 menggambarkan bahwa data warehouse berorientasi subjek.



Gambar 1. Orientasi subjek pada data warehouse

Terintegrasi

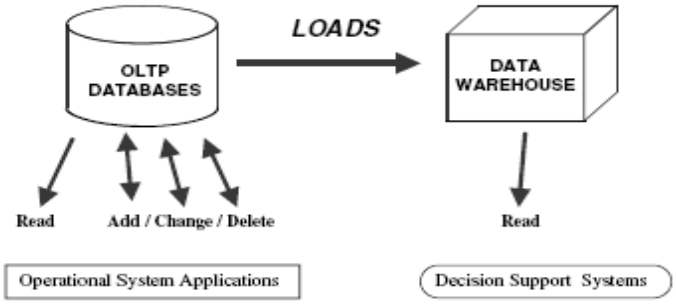
Karakteristik kedua dari data warehouse dan yang paling menonjol dari adalah integrasi. Integrasi merupakan aspek terpenting dari semua aspek yang dimiliki oleh data warehouse. Data-data dari berbagi sumber dimasukan ke dalam data warehouse. Selama proses pemuatan data ke dalam data warehouse, data dikonversi, direformasi, diurutkan kembali, diringkas, dan sebagainya. Hasilnya adalah saat data tersebut tersimpan pada data warehouse, data tersebut memiliki sebuah gambar fisik perusahaan yang tunggal. Gambar 2. ini merupakan ilustrasi integrasi yang terjadi ketika data dibawa dari lingkungan operasional yang berorientasi aplikasi ke data warehouse.



Gambar 2. Data werehouse terintegrasi

Tidak Berubah-ubah

Karakteristik ketiga yang dimiliki data warehouse adalah bahwa sebuah data warehouse bersifat nonvolatile (tidak berubah-ubah). Ilustrasi sifat nonvolatile dari data dan menunjukan bahwa data operasional mengakses dan memanipulasi satu record pada satu waktu terdapat pada gambar 3.



Gambar 3. Masalah nonvoatility

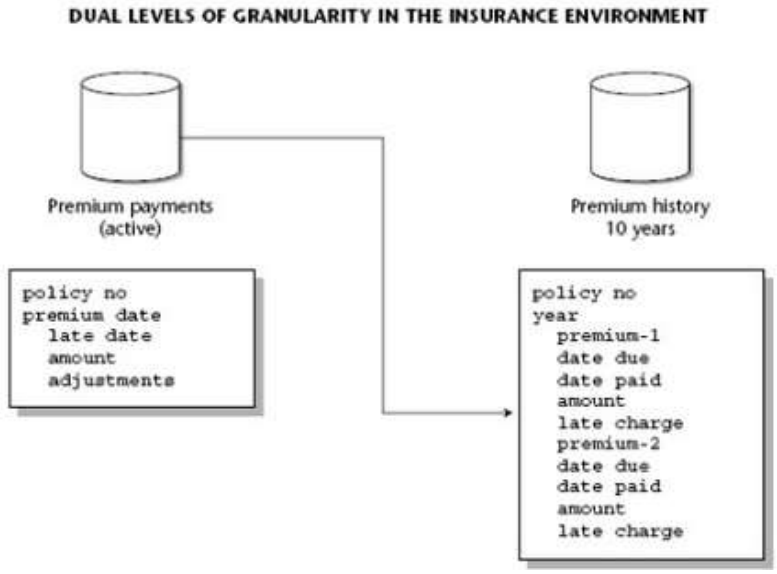
Pada lingkungan operasional, data selalu diperbaharui seperti yang biasa dilakukan, tetapi data pada data warehouse menunjukkan karakteristik yang sangat berbeda. Data dari data warehouse dimuat (biasanya, namun tidak selalu) dan diakses, namun data tersebut tidak diperbaharui atau diganti. Saat data pada data warehouse dimuat, data tersebut dimuat dalam sebuah snapshot dan mempunyai format statis. Ketika terjadi perubahan, sebuah snapshot baru ditambahkan. Dengan demikian, recordrecord historis dari data tetap tersimpan pada data warehouse.

Variansi Waktu

Karakteristik terakhir yang dimiliki oleh data warehouse adalah variansi waktu. Variansi waktu secara tidak langsung menyatakan bahwa setiap unit dari data dalam data warehouse akurat dalam kurunwaktu tertentu. Pada beberapa kasus, sebuah record mempunyai tanggal dan waktu transaksi. Tetapi pada setiap kasus, terdapat beberapabentuk penanda waktu untuk menunjukan rentang waktu dimana record tersebut akurat.

Granularity

Menurut Ponniah (2010) pada sistem operasional data dibuat secara real-time sehingga untuk mendapatkan informasi langsung dilakukan proses query. Pada data warehouse proses analisis harus memeperhatikan detail per level misalkan perhari, ringkasan perbulan, ringkasan pertiga-bulan. Granularitas menunjuk pada level perincian atau peringkasan yang ada pada unit-unit data dalam data warehouse. Semakin banyak detail yang ada, maka semakin rendah level granularitasnya. Semakin sedikit detail yang ada, maka semakin tinggi level granularitasnya. Semakin tinggi level granularitas maka query yang dapat ditangani oleh data warehouse semakin terbatas. Semakin rendah level granularitasnya maka query yang dapat ditangani oleh data warehouse semakin banyak dan jawaban query yang diperolehpun semakin detail. Pada gambar 4 akan mengilustrasi granularity sebuah data.

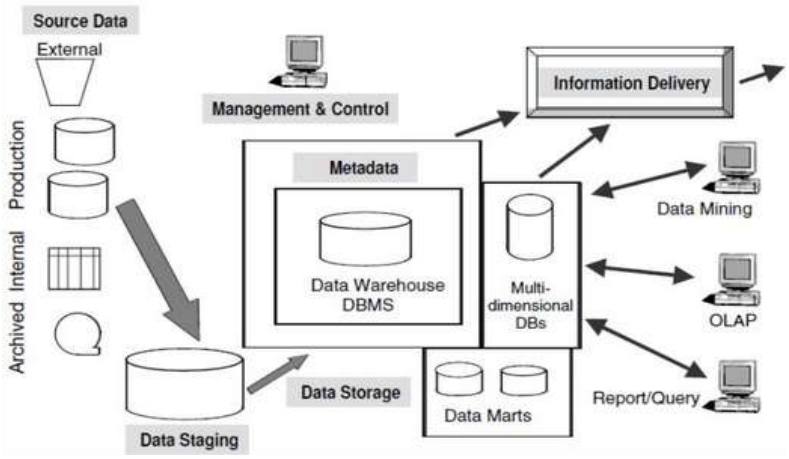


Gambar 4. Data granuality

Komponen Data Werehouse

Pada bagian ini akan dijelaskan secara singkat mengenai komponen-komponen data warehouse. Komponen data warehouse dapat digambarkan seperti yang terlihat pada gambar 5. Komponen source data terletak di sebelah kiri. Komponen data staging sebagai blok pembangunan berikutnya setelah source data. Pada bagian tengah, dapat dilihat komponen data storage yang mengelola data warehouse, komponen ini tidak hanya menyimpan dan mengelola data, tetapi juga menjaga bagian data yang disebut metadata repository. Komponen information delivery berada di sebelah kanan.

Komponen tersebut terdiri dari semua hal yang berkaitan dengan penyediaan informasi dari data warehouse bagi pengguna.



Gambar 5. Komponen data werehouse (ponniah, 2010)

Data Mart

Data mart merupakan sebuah struktur data yang didedikasikan untuk melayani kebutuhan analitis dari satu grup atau kelompok orang, misalnya seperti departemen akunting atau departemen keuangan. Data mart adalah suatu bagian pada data warehouse dan berada level departemen pada perusahaan atau organisasi yang mendukung pembuatan laporan dan analisa data pada suatu unit bagian atau operasi pada suatu perusahaan. Data mart menangani sebuah business process, misalkan penjualan, maka hanya proses penjualan saja yang ditangani pada data mart.

Perbedaan Data Werehouse dan Data Mart

Perbedaan yang mendasar secara keseluruhan data warehouse mengisi data kedalam dependent data mart dan kombinasi dari data mart menjadi sebuah data warehouse. Gambar 6 merupakan perbandingan antara data warehouse dan data mart.

	Data Warehouse	Data Marts
Scope	• Application Neutral • Centralized, Shared • Cross LOB/enterprise	• Specific Application Requirement • LOB, department • Business Process Oriented
Data Perspective	• Historical Detailed data • Some summary	• Detailed (some history) • Summarized
Subjects	• Multiple subject areas	• Single Partial subject • Multiple partial subjects • OLTP snapshots
Data Sources	• Many • Operational/ External Data	• Few • Operational, external data • OLTP snapshots
Implement Time Frame	• 9-15 months for first stage • Multiple stage implementation	• 4-12 months
Characteristics	• Flexible, extensible • Durable/Strategic • Data orientation	• Restrictive, non extensible • Short life/tactical • Project Orientation

Gambar 6. Data werehouse versus data mart

Beberapa aspek lainnya yang membedakan antara data mart dengan data warehouse menurut Aditama (2010) adalah sebagai berikut:

Lingkup

Sebuah data warehouse berhubungan dengan lebih dari satu area subjek dan biasanya diimplementasikan dan diatur oleh sebuah unit organisasional pusat seperti departemen IT perusahaan. Seringkali disebut dengan data warehouse pusat atau perusahaan. Sedangkan data mart biasanya hanya dibuat untuk departemen atau bagian dari perusahaan yang tertentu saja dan tidak mewakili seluruh informasi perusahaan seperti data warehouse.

Subjek

Sebuah data mart merupakan bentuk departemental dari data warehouse yang dirancang untuk sebuah garis bisnis tunggal (single line of business/LOB).

Sumber Data

Sebuah data warehouse umumnya mengambil data dari banyak sistem sumber, sedangkan data mart mengambil data dari sumber-sumber yang jumlahnya lebih sedikit.

Ukuran

Data mart tidak dibedakan dari data warehouse berdasarkan ukuran, tetapi dalam penggunaan dan manajemen. Satu definisi dari warehouse yang sangat besar adalah: "Suatu warehouse yaitu lebih besar daripada backup time window."

Waktu Iplementasi

Data mart biasanya lebih sederhana daripada data warehouse dan karena itu lebih mudah untuk dibuat dan dipelihara. Sebuah data mart juga dapat dibuat sebagai langkah "pembuktian konsep" terhadap pembangunan sebuah enterprisewide data warehouse.

Terdapat dua pendekatan utama untuk merancang data mart menurut Chhabra & Pahwa, (2014) yaitu: 1) Dependent data mart, Dependent data mart adalah sebuah perangkat fisik atau logis sebuah subset dari data warehouse yang lebih besar. Menurut pendekatan ini, data mart diperlakukan sebagai subset dari sebuah data warehouse. Pada pendekatan ini, yang pertama adalah sebuah data warehouse dibangun dari beberapa data mart yang berbeda. Data mart ini bergantung pada data warehouse dan mengekstrak data yang diperlukan dari data warehouse. Dalam pendekatan ini data mart dibangun dari sebuah data warehouse yang berarti tidak membutuhkan sebuah integrasi yang dikenal dengan pendekatan top-down; 2) Independent data mart, Pendekatan yang kedua adalah independent data mart. Pada pendekatan ini, pertama-tama independent data mart dibangun, kemudian data warehouse di bangun dari beberapa independent data mart. Dalam pendekatan ini semua data mart di desain secara independen sehingga integrasi antar data mart sangatlah diperlukan. Pendekatan ini juga disebut pendekatan bottom up dengan data mart yang terintegrasi untuk merancang sebuah data warehouse.

Pengertian Data Mining

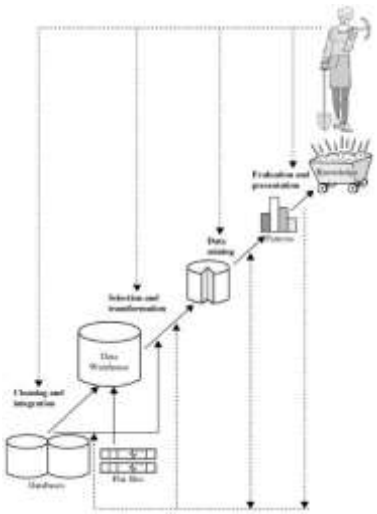
Menurut Han (2012), data mining adalah proses menemukan pola yang menarik, dan pengetahuan dari data yang berjumlah besar, yang terdapat dalam setiap informasi. Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatik yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstrasi dan menidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar. Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu – ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar.

Tahap-tahap Datamining

Istilah data mining dan knowledge discovery in databases (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses pentggalianinformasi tersembunyi dalam suatau basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat ditunjukkan pada gambar 7.



Gambar 7. Tahap-tahap knowledge discovery in database (Han, 2012)

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan knowledge base. Tahapan-tahapan tersebut, diantaranya:

Pembersihan Data

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang kita miliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Garbage in garbage out (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performasi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

Integrasi Data

Integrasi data dilakukan pada atribut-aribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan

menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

Transformasi Data

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klastering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

Aplikasi Teknik Data Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai. Kita akan membahas lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik data mining baru untuk penerapan di bidang bioinformatika seperti analisa hasil microarray untuk mengidentifikasi DNA dan fungsi-fungsinya.

Evaluasi Pola yang Ditemukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti: 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.

Presentasi Pola yang Ditemukan untuk Menghasilkan Aksi

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

Teknik-teknik Data Mining

Dengan definisi DM yang luas, ada banyak jenis teknik analisa yang dapat digolongkan dalam DM. Karena keterbatasan tempat, disini penulis akan memberikan sedikit gambaran tentang tiga teknik DM yang paling populer.

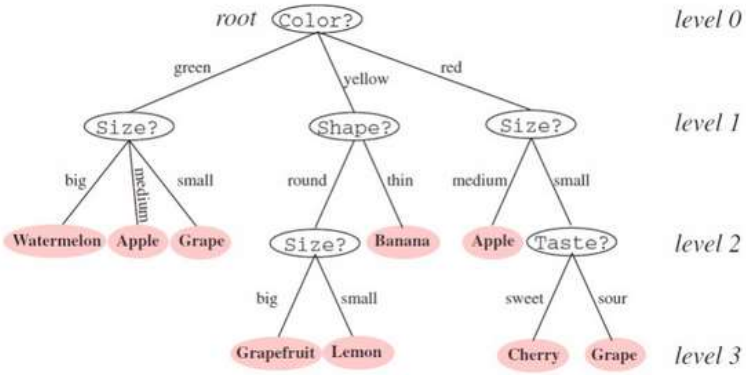
Association Rule Mining

Association rules (aturan asosiasi) atau affinity analysis (analisis afinitas) berkenaan dengan studi tentang "apa bersama apa". Sebagai contoh dapat berupa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang database transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan market basket analysis. Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan "if-then" atau "jika-maka". Aturan ini dihitung dari data yang sifatnya probabilistic.

Analisis asosiasi dikenal juga sebagai salah satu metode data mining yang menjadi dasar dari berbagai metode data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu prosentase kombinasi item tersebut. Dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence).

Classification

Dalam klasifikasi, terdapat target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Dalam decision tree tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Seperti yang diilustrasikan pada gambar 8, misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil. Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan.



Gambar 8. Decision tree

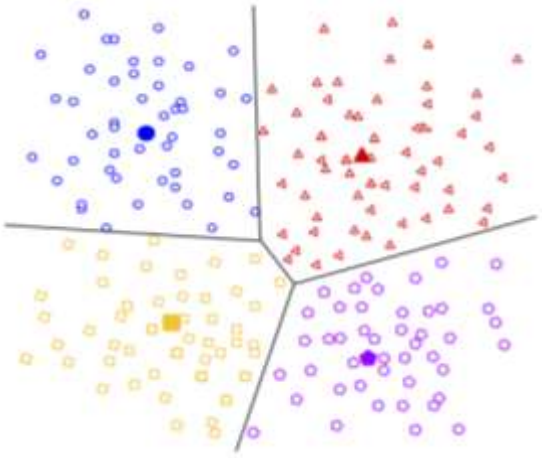
Ada beberapa macam algoritma decision tree diantaranya CART dan C4. 5. Beberapa isu utama dalam decision tree yang menjadi perhatian yaitu seberapa detail dalam mengembangkan decision tree, bagaimana mengatasi atribut yang bernilai continues, memilih ukuran yang cocok untuk penentuan atribut, menangani data training yang mempunyai data yang atributnya tidak mempunyai nilai, memperbaiki efisiensi perhitungan.

Decision tree sesuai digunakan untuk kasus-kasus yang keluarannya bernilai diskrit. Walaupun banyak variasi model decision tree dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri yang cocok untuk diterapkannya decision tree adalah sebagai berikut: 1) Data dinyatakan dengan pasangan atribut dannilainya; 2) Label/keluaran data biasanya bernilaidiskrit; 3) Data mempunyai missing value (nilai dari suatu atribut tidak diketahui)

Dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Untuk membuat decision tree perlu memperhatikan hal-hal berikut ini: 1) Atribut mana yang akan dipilih untuk pemisahanobyek; 2) Urutan atribut mana yang akan dipilih terlebihdahulu; 3) Strukturtree; 4) Kriteriapemberhentian; 5) Pruning.

Clustering

Clustering termasuk metode yang sudah cukup dikenal dan banyak dipakai dalam data mining. Sampai sekarang para ilmuwan dalam bidang data mining masih melakukan berbagai usaha untuk melakukan perbaikan model clustering karena metode yang dikembangkan sekarang masih bersifat heuristic. Usaha-usaha untuk menghitung jumlah cluster yang optimal dan pengklasteran yang paling baik masih terus dilakukan. Dengan demikian menggunakan metode yang sekarang, tidak bisa menjamin hasil pengklasteran sudah merupakan hasil yang optimal. Namun, hasil yang dicapai biasanya sudah cukup bagus dari segi praktis.



Gambar 9. Clustering

Tujuan utama dari metode clustering adalah pengelompokan sejumlah data/obyek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang semirip mungkin seperti diilustrasikan pada gambar 2.9. Dalam clustering metode ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin. Ini berarti obyek dalam satu cluster sangat mirip satu sama lain dan berbeda dengan obyek dalam cluster-clusteryang lain. Dalam metode ini tidak diketahui sebelumnya berapa jumlah cluster dan bagaimana pengelompokannya. Berikut ini adalah 9 algoritma penggalian data yang paling populer berdasarkan konferensi ICDM '06: 1) C4. 5; 2) k-Means; 3) SVM; 4) Apriori; 5) EM; 6) PageRank; 7) AdaBoost; 8) kNN; 9) NaiveBayes.

Implementasi (Penerapan Datamining)

Dalam bidang apa saja data mining dapat diterapkan? Berikut beberapa contoh bidang penerapan datamining:

Analisa Pasar dan Manajemen

Solusi yang dapat diselesaikan dengan data mining, diantaranya: Menembak target pasar, Melihat pola beli pemakai dari waktu ke waktu, Cross-Market analysis, Profil Customer, Identifikasi kebutuhan Customer, Menilai loyalitas Customer, InformasiSummary.

Analisa Perusahaan dan Manajemen Resiko

Solusi yang dapat diselesaikan dengan data mining, diantaranya: Perencanaan keuangan dan Evaluasi aset, Perencanaan sumber daya (Resource Planning), Persaingan (Competition).

Telekomunikasi

Sebuah perusahaan telekomunikasi menerapkan data mining untuk melihat dari jutaan transaksi yang masuk, transaksi mana sajakah yang masih harus ditangani secara manual.

Keuangan

Financial Crimes Enforcement Network di Amerika Serikat baru-baru ini menggunakan data mining untuk me-nambang trilyunan dari berbagai subyek seperti property, rekening bank dan transaksi keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigakan (seperti money laundry).

Asuransi

Australian Health Insurance Commision menggunakan data mining untuk mengidentifikasi layanan kesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi.

Olahraga

IBM Advanced Scout menggunakan data mining untuk menganalisis statistik permainan NBA (jumlah shots blocked, assists dan fouls) dalam rangka mencapai keunggulan bersaing (competitive advantage) untuk Tim New York Knicks dan Miami Heat.

Astronomi

Jet Propulsion Laboratory (JPL) di Pasadena, California dan Palomar Observatory berhasil menemukan 22 quasar dengan bantuan data mining. Hal ini merupakan salah satu kesuksesan penerapan data mining di bidang astronomi dan ilmu ruang angkasa.

Internet Websurf-aid

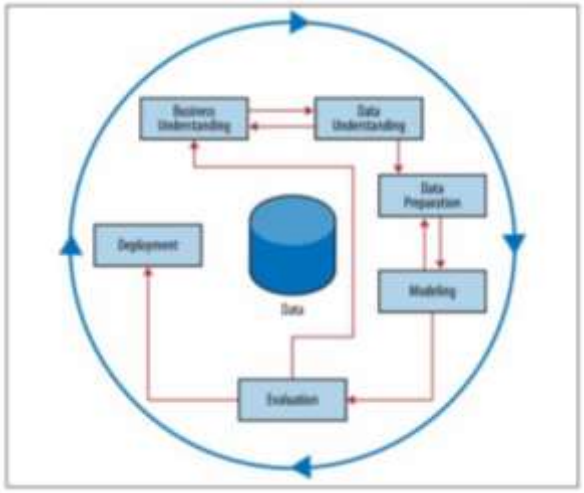
IBM Surf-Aid menggunakan algoritma data mining untuk mendata akses halaman Web khususnya yang berkaitan dengan pemasaran guna melihat prilaku dan minat customer serta melihat ke- efektif-an pemasaran melalui Web.

Metode Penelitian Data Mining

Metodologi penelitian data mining pada prinsipnya merupakan kegiatan pencarian pengetahuan atau lebih dikenal dengan Knowledge Discovery in Database. Dalam tahapan ini dibutuhkan alat bantu OLAP untuk melakukan explore terhadap database yang telah dibuat berupa data warehouse

Pembangunan aplikasi data mining untuk menampilkan data konsumen potensial telemarketing, merupakan kegiatan tahap kedua dari penelitian yang dilakukan. Metoda yang digunakan pada pembangunan aplikasi data mining ini adalah Cross-Industry Standard Process for Data mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industry seperti DaimlerChrysler, SPSS dan NCR, CRISP DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis ata unit penelitian.

Dalam CRISP-DM, sebuah proyek data mining memiliki siklus hidup yang terbagi dalam 6 fase (Gambar 2.10). Keseluruhan fase berururan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antarfase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase modeling. Berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase data preparation untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada faseevaluation.



Gambar 10. Proses Datamining (Larose, 2006)

Fase Pemahaman Bisnis (Business Understanding Phase)

1) Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secarakeseluruhan; 2) Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining; 3) Menyiapkan strategi awal untuk mencapai tujuan.

Pemahaman Data (Data Preparation Phase)

1) Mengumpulkandata; 2) Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuanawal; 3) Mengevaluasi kualitasdata; 4) Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola daripermasalahan.

Fase Pengolahan Data (Data Preparation Phase)

1) Siapkan dari data awal, kumpulan data yang akan digunakan untu keseluruahan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secaraintensif; 2) Pilih kasus dan variable yang ingin dianalisis dan yang sesuai analisis yang akandilakukan; 3) Lakukan perubahan pada beberapa variable jikadibutuhkan; 4) Siapkan data awal sehingga siap untuk perangkatpemodelan.

Fase Pemodelan (Modelinh Fase)

1) Pilih dan aplikasikan teknik pemodelan yang sesuai; 2) Kalibrasi aturan model untuk mengoptimalkan hasil; 3) Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan data mining yangsama; 4) Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yangsesuai dengan spesifikasi kebutuhan teknik data mining tertentu.

Fase Evaluasi (Evaluation Phase)

1) Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelanuntuk mendapatkan kualitas dan efektivitas sebelum disebarkan untukdigunakan; 2) Menetapkan apakah terdapat model yang memenuhi tujuan padafaseawal; 3) Menentukan apakah terdapat permasalah penting dari bisnis atau penelitian yang tidak tertangani denganbaik; 4) Mengambil keputusan berkaitan denga penggunaan hasil dari datamining.

Fase Penyebaran (Deployment Phase)

1) Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikanya proyek; 2) Contoh sederhana penyebaran: pembuatan laporan. Contoh kompleks penyebaran: penerapan proses data mining secara parallel pada department lain.

Daftar Pustaka

ADITAMA, T. Y, 2010, Manajemen Administrasi Rumah Sakit. Edisi II. Jakarta: Universitas Indonesia.

Chhabra R. and Pahwa P, 2014, Data Mart Designing and Integration Approaches. Internation Journal of Computer Science and Mobile Computing. IJCSMC, Vol 3, Issue 4, pp. 74-79. ISSN 2320-088X

Fayyad, U. , Shapiro, G. P. , and Smyth, P. ,1996, From Data Mining to Knowledge Discovery in Databases. American Association for Artificial IntelligenceMagazine, pp. 37-54.

Han, J. , Kamber, M. , and Pei, J. , 2012, Data Mining Concept and Techniques. 3nd, Morgan kaufmann

Klimavicius, M, 2008, Towards Development of Solution forBusiness Process-Oriented Data Analysis, World Academy of Science, Engineering and TechnologyInternational Journal of Computer and Information EngineeringVol:2, No:1.

Larose, Daniel T. . (2006). "Data Mining Methods and Models". Hoboken New Jersey: JohnWilley & Sons, Inc.

Larose, Daniel T. . (2006). "Discovering Knowledge in Data: An Introduction to Data Mining". USA: John Willey & Sons. Inc.

Ponniah, Paulraj, 2010, Data Warehousing Fundamentals for IT Professionals, Second Edition, John Wiley & Sons, Inc.