Bab 10

Data Mining

POKOK BAHASAN:

- ✓ Model Data Mining
- ✓ Tahapan dalam Data Mining
- ✓ Fungsionalitas dalam Data Mining
- ✓ Teknik-teknik Data Mining

TUJUAN BELAJAR:

Setelah mempelajari materi dalam bab ini, mahasiswa diharapkan mampu:

- ✓ Memahami pemodelan Data Mining
- ✓ Memahami setiap tahapan dalam Data Mining
- ✓ Memahami fungsionalitas dalam Data Mining
- ✓ Memahami beberapa teknik yang digunakan dalam Data Mining

10.1. PENDAHULUAN

Seiring dengan perkembangan teknologi, semakin berkembang pula kemampuan kita dalam menggumpulkan dan meng olah data. Penggunaan sistem komputerisasi dalam berbagai bidang baik itu dalam transaksi-transaksi bisnis, maupun untuk kalangan pemerintah dan sosial, telah menghasilkan data yang berukuran sangat besar. Data-data yang terkumpul ini merupakan suatu tambang emas yang dapat digunakan sebagai informasi dalam dunia bisnis.

Aplikasi basis data telah banyak diterapkan dalam berbagai antara lain bidang manajemen, manajemen data untuk industri, ilmu pegetahuan, administrasi pemerintah dan bidang-bidang lainnya. Akibatnya data yang dihasilkan oleh bidang-bidang tersebut sangatlah besar dan berkembang dengan cepat. Hal ini menyebabkan timbulnya kebutuhan terhadap teknik-teknik yang dapat melakukan pengolahan data sehingga dari data-data yang ada dapat diperoleh informasi penting yang dapat digunakan untuk perkembangan masing-masing bidang tersebut.

Istilah data mining sudah berkembang jauh dalam mengadaptasi setiap bentuk analisa data. Pada dasarnya data mining berhubungan dengan analisa data dan penggunaan teknik-teknik perangkat lunak untuk mencari pola dan keteraturan dalam himpunan data yang sifatnya tersembunyi.

Data mining diartikan sebagai suatu proses ekstraksi informasi berguna dan potensial dari sekumpulan data yang terdapat secara implisit dalam suatu basis data. Banyak istilah lain dari data mining yang dikenal luas seperti knowledge mining from databases, knowledge extraction, data archeology, data dredging, data analysis dan lain sebagainya [AGR-93].

Dengan diperolehnya informasi-informasi yang berguna dari data-data yang ada, hubungan antara item dalam transaksi, maupun informasi informasi-yang potensial, selanjutnya dapat diekstrak dan dianalisa dan diteliti lebih lanjut dari berbagai sudut pandang.

Informasi yang ditemukan ini selanjutnya dapat diaplikasi kan untuk aplikasi manajemen, melakukan query processing, peng ambilan keputusan dan lain sebagainya. Dengan semakin ber kembang nya kebutuhan akan informasi-informasi, semakin banyak pula bidang-bidang yang rnenerapkan konsep data mining.

10.2. MODEL DATA MINING

Dalam perkembangan teknologi data mining, terdapat model atau mode yang digunakan untuk melakukan proses penggalian informasi terhadap data-data yang ada. Menurut IBM model data mining dapat dibagi menjadi 2 bagian yaitu: *verification model* dan *discovery model.*

10.2.1. *VERIFICATION MODEL*

Model ini menggunakan perkiraan (hypothesis) dari pengguna, dan melakukan test terhadap perkiraan yang diambil sebelumnya dengan menggunakan data-data yang ada. Penekanan terhadap model ini adalah terletak pada user yang bertanggung jawab terhadap penyusunan perkiraan (hypothesis) dan permasalahan pada data untuk meniadakan atau menegaskan hasil perkiraan (hypothesis) yang diambil.

Sebagai contoh misalnya dalam bidang pemasaran, sebelum sebuah perusahaan mengeluarkan suatu produk baru kepasaran, perusahaan tersebut harus memiliki informasi tentang kecenderungan pelanggan untuk membeli produk yang akan di keluarkan. Perkiraan (hypothesis) dapat disusun untuk mengidentifikasikan pelanggan yang potensial dan karakteristik dari pelanggan yang ada. Data-data tentang pembelian pelanggah sebelumnya dan data tentang keadaan pelanggan, dapat digunakan untuk melakukan perbandingan antara pembelian dan karakteristik pelanggan untuk menetapkan dan menguji target yang telah diperkirakan sebelumnya. Dari keseluruhan operasi yang ada selanjutnya dapat dilakukan penyaringan dengan cermat sehingga jumlah perkiraan (hypothesys) yang sebelumnya banyak akan menjadi semakin berkurang sesuai dengan keadaan yang sebenarnya.Permasalahan utama dengan model ini adalah tidak ada informasi bare yang dapat dibuat, melainkan hanya pembuktian atau melemahkan perkiraan (hypothesys) dengan data-data yang ada sebelumnya. Datadata yang ada pada model ini hanya digunakan untuk membuktikan mendukung perkiraan (hypothesis) yang telah diambil sebelumnya. Jadi model ini sepenuhnya tergantung pads kemampuan user untuk melakukan analisa terhadap permasalahan yang ingin digali dan diperoleh informasinya.

10.2.2. *DISCOVERY MODEL*

Model ini berbeda dengan verification model, dimana pada model ini system secara langsung menemukan informasi-informasi penting yang tersembunyi dalam suatu data yang besar. Data-data yang ada kemudian dipilah-pilah-untuk-menemukan suatu pola, trend yang ada, dan keadaaan umum pada saat itu tanpa adanya campur tangan dan tuntunan dari pengguna. Hasil temuan ini menyatakan fakta-fakta yang ada dalam datadata yang ditemukan dalam waktu yang sesingkat rnungkin.Sebagai contoh, misalkan sebuah bank ingin menemuan kelompok-kelompok pelanggan yang dapat dijadikan target suatu produk yang akan di keluaran.

Pada data-data yang ada selanjutnya diadakan proses pencarian tanpa adanya proses perkiraan (hypothesis) sebelumnya. Sampai akhirnya semua pelanggan dikelompokan berdasarkan karakteristik yang sama**.**

10.3. KEBUTUHAN DAN TANTANGAN DALAM DATA MINING

Untuk memperoleh efektifitas dalam data mining, seseorang harus melakukan evaluasi kebutuhan dan memperhitungkan tantangan-tantangan apa saja yang mungkin dihadapinya dalam me ngembangkan suatu teknik data mining. Hal-hal yang harus diper hatikan tersebut antara lain adalah sebagai berikut

10.3.1. PENANGANAN BERBAGAI TIPE DATA

Karena ada bermacam data dan basis data yang digunakan dalam berbagai aplikasi, seseorang mungkin saja berpikir bahwa suatu sistem knowledge discovery harus bisa melakukan proses data mining yang efektif terhadap berbagai jenis data. Selanjutnya, banyak aplikasi basis data memuat tipe data yang kompleks seperti data terstruktur, objek data kompleks, data multimedia, data spasial dan data sementara, data transaksi dan lain sebagainya.

Oleh karena adanya beragam tipe data, tujuan yang berbeda dari data mining, maka adalah tidak realistis untuk mengharapkan bahwa suatu sistem data mining mampu menangani semua jenis data. Sistem data mining harus dikonstruksikan secara

khusus untuk tipe-tipe data khusus seperti dalam basis data relasional, basis data transaksi, basis data spasial, basis data multimedia dan lain sebagainya.

10.3.2. EFISIENSI DARI ALGORITMA DATA MINING

Untuk secara efektif melakukan ekstraksi informasi dari sejumlah besar data, algoritma yang digunakan untuk mewujudkannya haruslah efisien untuk basis data yang besar. Yaitu, waktu eksekusi dari algoritma tersebut haruslah sesuai dan realistis untuk data dengan ukuran besar.

10.3.3. KEGUNAAN, KEPASTIAN DAN KEAKURATAN HASIL

Informasi yang diperoleh harus secara akurat menggambarkan isi basis data dan berguna untuk aplikasi terkait. Kekurangsempurnaan yang ada haruslah dapat diekspresikan dengan suatu ukuran yang pasti dalam bentuk aturan-aturan kuantitif dan perkiraan-perkiraan yang masuk akal. Noise dan data-data yang tidak diperlukan harus ditangani dengan rapi dalam sistem data mining. Hal ini juga akan memotivasi suatu studi sistematik untuk mengukur kualitas dari informasi yang dihasilkan, termasuk seberapa menariknya dan tingkat kepercayaannya yang dapat diukur secara statistik, analitis dan menggunakan model simulasi.

10.3.4. EKSPRESI TERHADAP BERBAGAI JENIS HASIL

Berbagai macam jenis informasi dapat diperoleh dari sejumlah besar data. Seseorang mungkin ingin menguji informasi yang diperoleh dan sudut pandang yang berbeda dan menampilkannya dalam bentuk yang berbeda. Ini menuntut kita untuk mengekspresikan permintaan datamining dan informasi yang diperoleh dalam sebuah bahasa tingkat tinggi atau graphical user interface yang baik, sehingga program dapat digunakan oleh para pemakai biasa yang bukan ahli, dan hasil yang diperoleh dapat dimengerti serta langsung digunakan oleh pemakainya. Oleh karenanya, sistem harus bisa mengadopsi teknik-teknik penyajian informasi yang baik.

10.3.5. MEMPEROLEH INFORMASI DARI SUMBER-SUMBER DATA YANG BERBEDA

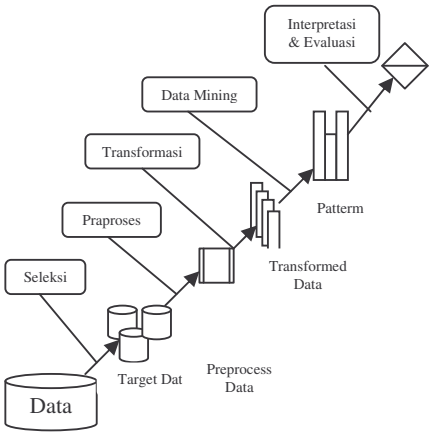
Dengan adanya LAN (Local Area Network) dan WAN (Wide Area Network) yang tersebar secara luas dewasa ini, termasuk Internet, maka terhubunglah - berbagai sumber data yang terdistribusi secara luas dan membentuk suatu basis data heterogen. Untuk memperoleh informasi dari berbagai sumber dan dalam berbagai format dengan berbagai semantik data menimbulkan tantangan baru dalam data mining. Di lain pihak, datamining bisa membantu mengungkapkan informasi-informasi yang ada dalam suatu basis data heterogen, dimana hal tersebut sulit untuk diwujudkan dengan sebuah sistem query sederhana. Lebih lanjut, ukuran data yang besar, distribusi yang luas dad data dan kompleksitas dari proses komputasi beberapa metode data mining, semakin memotivasi pengembangan algoritma untuk paralel data mining dan data mining untuk basis data terdistribusi.

10.3.6. PROTEKSI DAN KEAMANAN DATA

Ketika data dapat diperlihatkan dari berbagai sudut pandang dan dalam level abstrak yang berbeda, hal ini akan mengancam tujuan dari proteksi dan keamanan data, dan pelanggaran terhadap sifat kerahasiaan informasi. Sangatlah penting untuk mempelajari apakah penemuan informasi yang berguna itu akan mengakibatkan pelanggaran kerahasiaan dan ukuran keamanan yang diperiukan untuk menghalangi akses terhadap data-data yang sifatnya sensitif.

10.4. TAHAPAN DALAM DATA MINING

Data-data yang ada, tidak dapat langsung diolah dengan menggunakan sistem data mining. Data-data tersebut harus dipersiapkan terlebih dahulu agar hasil yang diperoleh dapat lebih maksimal, dan waktu komputasinya lebih minimal. Proses persiapan data ini sendiri dapat mencapai 60 % dari keseluruhan proses dalam data mining. Adapun tahapan-tahapan yang harus dilalui dalam proses data mining antara lain:



Gambar 11.1. Tahapan Data Mining

- Basis Data Relasional

Dewasa ini, hampir semua Data bisnis disimpan dalam basis data relasional. Sebuah model basis data relasional dibangun dari serangkaian tabel, setiap tabel disimpan sebagai sebuah file. Sebuah tabel relasional terdiri dari baris dan kolom. Kebanyakan model basis data relasional saat ini dibangun diatas lingkungan OLTP. OLTP (Online Transaction Processing) adalah tipe akses yang digunakan oleh bisnis yang membutuhkan transaksi konkuren dalam jumlah besar. Bentuk data yang tersimpan dalam basis data relasional inilah yang dapat diolah oleh sistem data mining.

- Ekstraksi Data

Data-data yang dikumpulkan dalam proses transaksi seringkali ditempatkan pada lokasi yang berbeda-beds. Maka dari itu dibutuhkan kemampuan dari sistem utuk dapat mengumpulkan data dengan cepat. Jika data tersebut disimpan dalam kantor regional, seringkali data tersebut di upload ke sebuah server yang lebih terpusat. Ini bisa dilakukan secara harian, mingguan, atau bulanan tergantung jumlah .data, keamanan dan biaya. Data dapat diringkas dulu sebelum dikirimkan ke tempat penyimpanan pusat. Sebagai contoh, sebuah toko perangkat keras mungkin mengirim data yang menunjukan bahwa 10 rol kabel telah terjual pada hari ini oleh karyawan nomer 10 dibanding pengiriman data detail transaksi.

- Transformasi Data

Transformasi data melakukan peringkasan data dengan mengasumsikan bahwa data telah tersimpan dalam tempat penyimpanan tunggal. Pada langkah terakhir, data telah di ekstrak dari banyak basis data ke dalam basis data tunggal. Tipe peringkasan yang dikerjakan dalam langkah ini mirip dengan peringkasan yang dikerjakan selama tahap ekstraksi. Beberapa perusahaan memilih untuk menngkas data dalam sebuah tempat penyimpanan tunggal. Fungsi fungsi Agregate yang sering digunakan antara lain: summarizations, averages, minimum, maximum, dan count.

- Pembersihan Data

Data-data yang telah terkumpul selanjutnya akan mengalami proses pembersihan. Proses pembersihan data dilakukan untuk membuang record yang keliru, menstandarkan attribut-attribut, merasionalisasi struktur data, dan mengendalikan data yang hilang. Data yang tidak konsisten dan banyak kekeliruan membuat hasil data mining tidak akurat. Adalah sangat penting untuk membuat data konsisten dan seiagam. Pembersihan data juga dapat membantu perusahaan untuk mengkonsolidasikan record. ini sangat berguna ketika sebuah perusahaan mempunyai banyak record untuk seorang pelanggan.Setiap record atau file pelanggan mempunyai nomor pelanggan yang sama, tetapi informasi dalam tiap filenya berbeda.

- Bentuk Standar

Selanjutnya setelah data mengalami proses pembersihan maka data ditranfer kedalam bentuk standar. Bentuk standar adalah adalah bentuk data yang akan diakses oleh algoritma data mining. Bentuk standar ini biasanya dalam bentuk spreadsheet like. Bentuk spreadsheet bekerja dengan baik karena baris merepresentasikan kasus dan kolom merepresentasikan feature.

- Reduksi Data dan Feature

Setelah data berada dalam bentuk standar spreadsheet perlu dipertimbangkan untuk mereduksi jumlah feature. Ada beberapa alasan untuk mengurangi jumlah feature dalam spreadsheet kita. Sebuah bank mungkin mempunyai ratusan feature ketika hendak memprediksi resiko kredit. Hal ini berarti perusahaan mempunyai data dalam jumlah

yang sangat besar. Bekerja dengan data sebanyak ini membuat algoritma prediksi menurun kinerjanya.

- Menjalankan Algoritma

Setelah semua proses diatas dikerjakan, maka algoritma data mining sudah siap untuk dijalankan.

10.5. FUNGSIONALITAS DALAM DATA MINING

Kebutuhan akan Data mining semakin dirasakan dalam berbagai bidang. Data mining bersifat dependen terhadap aplikasi terkait, ini berarti untuk aplikasi basis data yang berbeda, maka teknik data mining yang digunakannya mungkin juga akan berbeda. Hal ini dikarenakan terdapat kelebihan dan kekurangan dari masing-masing metode pencarian informasi, sehingga kita harus menyesuaikan antara keperluan dan kebutuhan akan informasi dengan penerapan teknik pencarian yang akan digunakan. Untuk memberikan gambaran yang lebih jelas tentang macam-macam informasi yang dapat ditemukan dalam sekumpulan data, berikut akan diberikan sedikit bahasan rinci mengenai hal tersebut.

10.5.1. MINING ASSOCIATION RULE

Mining association rules atau pencarian aturan-aturan hubungan antar item dari suatu basis data transaksi atau basis data relasional, telah menjadi perhatian utama dalam masyarakat basis data. Tugas utamanya adalah untuk menemukan suatu himpunan hubungan antar item dalam bentuk A1A...AAm => B1A...ABn dimana A, (for i E {1,...,m}) dan B; (for j C {1,...,n}) adalah himpunan atribut nilai, dari sekumpulan data yang relevan dalam suatu basis data. Sebagai contoh, dari suatu himpunan data transaksi, seseorang mungkin menemukan suatu hubungan berikut, yaitu jika seorang pelanggan membeli selai, ia biasanya juga membeli roti dalam satu transaksi yang sama. Oleh karena proses untuk menemukan hubungan antar item ini mungkin memerlukan pembacaan data transaksi secara berulang-ulang dalam sejumlah besar data-data transaksi untuk menemukan pola-pola hubungan yang berbeda-beda,

maka waktu dan biaya komputasi tentunya juga akan sangat besar, sehingga untuk menemukan hubungan tersebut diperlukan suatu algoritma yang efisien dan metode-metode tertentu.

10.5.2. GENERALISASI, PENCATATAN DAN KARAKTERISASI DATA MULTI LEVEL

Salah satu aplikasi data mining dan analisa data yang paling sering digunakan dalam hubungannya dengan basis data sistem produksi adalah generalisasi dan pencatatan data, yang juga dikenal dengan beberapa nama lain seperti on-line analytical processing (OLAP), basis data multi dimensi, data cubes, abstraksi data, dan lain sebagainya. Generalisasi dan pencatatan data ini menampilkan karakteristik umum terhadap sekumpulan data yang dispesifikasikan oleh pemakai dalam basis data.

Data dan obyek dalam basis data seringkali memuat informasi yang mendetail pada level primitif. Sebagai contoh, item relasi dalam suatu basis data sales mungkin saja mengandung atribut level primitif tentang informasi item seperti nomor item, nama item, tanggal pembuatan, harga dan lain sebagainya. Seringkali kita menginginkan untuk mencatat sejumlah besar himpunan data dan menampilkannya dalam level tingkat tinggi. Misalnya seseorang mungkin ingin mencatat sejumlah besar himpunan item yang terhubung ke beberapa sales untuk memberikan

10.5.3. KLASIFIKASI DATA

Aplikasi lain yang penting dari data mining adalah kemampuannya untuk melakukan proses klasifikasi pada suatu data dalam jumlah besar. Hal ini sering disebut mining classification rules. Sebagai contoh, sebuah dealer mobil ingin mengkiasifikasikan pelanggannya menurut kecenderungan mereka untuk menyukai mobil jenis tertentu, sehingga para sales yang bekerja disitu akan mengetahui siapa yang harus didekati, kemana katalog mobil jenis baru harus dikirim, sehingga hal ini akan sangat membantu dalam hal promosi.

Klasifikasi data adalah suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalarn sebuah basis data, dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. 'Untuk membentuk sebuah model klasifikasi, suatu sampel basis data 'E' diperlakukan sebagai training set, dimana setiap tupel terdiri dari himpunan yang sama yang memuat atribut yang beragam seperti tupel-tupel yang terdapat dalam suatu basis data yang besar 'W'. Setiap tupel diidentifkasikan dengan sebuah label atau identitas kelas. Tujuan dari klasifikasi ini adalah pertama-tama untuk menganalisa training data dan membentuk sebuah deskripsi yang akurat atau sebuah model untuk setiap kelas berdasarkan feature-feature yang tersedia di dalam data itu.

Deskripsi dari masing-masing kelas itu nantinya akan digunakan untuk mengklasifikasikan data yang hendak di test dalam basis data 'W', atau untuk membangun suatu deskripsi yang lebih balk untuk setiap kelas dalam basis data. Contoh untuk model ini adalah prediksi terhadap resiko pemberian kredit. Data terdiri dari orang orang yang telah menerima kredit. Sebagian kreditur menjalankan kewajiban dengan balk, dan sebagian lagi tidak. Data mining, harus mampu mendefinisikan atribut atribut apa yang paling berpengaruh dalam hal ini.

10.5.4. ANALISA CLUSTER

Pada dasamya clustering terhadap data adalah suatu proses untuk mengelompokkan sekumpulan data tanpa suatu atribut kelas yang telah didefinisikart sebelumnya, berdasarkan pada prinsip konseptual clustering yaitu memaksimalkan dan juga meminimalkan kemiripan intra kelas. Misalnya, sekumpulan obyek-obyek komoditi pertama-tama dapat di clustering menjadi sebuah himpunan kelas-kelas dan lalu menjadi sebuah himpunan aturan-aturan yang dapat diturunkan berdasarkan suatu klasifikasi tertentu.

Proses untuk mengelompokkan secara fisik atau abstrak obyek-obyek ke dalam bentuk kelas-kelas atau obyek-obyek yang serupa,disebut dengan clustering atau unsupervised classification. Melakukan analisa dengan clustering, akan sangat membantu untuk membentuk partisi-parti si yang berguna terhadap sejumlah besar

himpunan obyek dengan didasarkan pada prinsip "divide and conquer" yang mendekomposisikan suatu sistem skala besar, menjadi komponen-komponen yang lebih kecil, untuk menyederhanakan proses desain dan implementasi. Perbedaan utama antara Clustering Analysis dan klasifikasi adalah bahwa Clustering Analysis digunakan untuk memprediksi kelas dalam format bilangan real dad pada format katagorikal atau Boolean.

10.5.5. PENCARIAN POLA, SEKUENSIAL

Fungsi pola sekuensial menganalisa sekumpulan rekord pada suatu periode waktu, misalnya untuk menganalisa trend. Anggaplah kita memiliki suatu basis data yang ukurannya besar, yaitu basis data transaksi dimana setiap transaksi terdiri dari nomor pelanggan, waktu transaksi dan item-item yang ditransaksikan. Suatu pola dapat ditampilkan dalam contoh sebagai berikut, pelanggan biasanya membeli gula Iangsung melakukan transaksi membeli kopi. Dari semua transaksi membeli gula ternyata hampir seluruhnya terdapat transaksi membeli kopi. Maka dari pola-pola yang ada ini dapat dijadikan masukan bahwa telah terjadi suatu kecendrungan (trend) dari pelanggan dimana setiap pelanggan melakukan transaksi membeli gula maka akan diikuti oleh transaksi membeli kopi. Untuk itu pihak manajemen dapat menempatkan letak item kopi dekat dengan item gula. Sehingga memudahkan pelanggan untuk melakukan transaksi selanjutnya.

10.6. TEKNIK-TEKNIK DATA MINING

Perkembangan bidang data mining yang semakin pesat, menimbulkan banyak tantangan baru, aplikasi-aplikasi dari metode dan teknik, statistik serta sistem basis data yang ada tidak dapat secara langsung menyelesaikan masalah-masalah yang ada dalam data mining.

Oleh karena itu maka perlu dilakukan studi-studi terkait untuk menemukan metode data mining baru atau suatu teknik terintegrasi untuk sebuah sistem data mining yang efektif dan efisien. Dalam konteks ini, data mining itu sendiri telah menjadi suatu bidang baru yang independen. Telah banyak kemajuan dalam hal riset dan pengembangan dari data mining, juga banyak teknik data mining dan sistem baru yang

akhir-akhir ini dikembangkan. Klasifikasi skema yang berbeda dapat digunakan untuk mengkategorikan metode dan sistem data mining dengan didasarkan pada jenis basis data yang akan dipelajari, dan teknik apa yang akan digunakan.

- Jenis Basis Data yang akan dijadikan obyek.

Suatu sistem data mining dapat diklasifikasikan menurut jenis basis data dimana proses data mining tersebut dilakukan. Sebagai contoh, sebuah sistern adalah relationar data miner jika sistem tersebut menemukan informasi dad basis data relasional, atau suatu object oriented data miner bila informasi diperoleh dari basis data yang berorientasi pada obyek. Secara umum, data miner dapat digolongkan menurut jenis basis data apa yang diolahnya seperti misalnya basis data relasional, basis data transaksi, basis data yang berorientasi obyek, basis data deduktif, basis data spasial, basis data multimedia, basis-data-heterogen, dan lain sebagainya.

- Jenis informasi yang hendak dicari

Beberapa jenis informasi dapat dihasilkan dad proses data mining ini, termasuk *association rules, characteristic rules, classification rules, discriminant rules, clustering, sequential pattern, dan deviation analysis* [AGR-93]. Lebih lanjut, ada kiasifikasi lainnya menurut level abstraksi dari informasi yang diperoleh, antara lain generalized knowledge, primitive level knowledge dan multiple level knowledge. Suatu sistem data mining yang fleksibel dapat menggali informasi pada berbagai level abstraksi.

- Teknik yang hendak digunakan.

Cara kiasifikasi yang lainnya adalah berdasarkan teknik yang digunakan. Misalnya, dikategorikan berdasarkan metode kendalinya seperti autonomous knowledge miner, data driven miner, query driven miner dan interactive data miner. Dapat juga dikategorikan berdasarkan pendekatan yang dipakai dalam melakukan data mining, yaitu generalization based mining, statistics and mathematical based mining, integrated approach mining dan lain sebagainya.Diantara berbagai macam klasifikasi yang ada,

hasil penelitian menunjukkan ada satu skema utama yang menjadi patokan, yaitu jenis informasi yang dibutuhkan. Mengapa demikian, adalah karena dengan kiasifikasi ini akan memberikan gambaran yang jelas mengenai teknik dan kebutuhan datamining yang beragam. Metode-metode pencarianinformasi yang ada berdasarkan jenis informasinya seperti association rules, characterization rules, classification rules, sequence patterns, clustering dan ;ain-Iainnya telah diteliti secara mendalam. Untuk proses pencarian suatu informasi tertentu, berbagai pendekatan seperti pendekatan secara statistik, pendekatan berorientasi pada basis data yang besar dan sebagainya akan dibandingkan dengan penekanan utama pada basis data, dimana efektifitas dan efisiensi merupakan salah satu tujuan utamanya.

10.6.1. MARKET BASKET ANALYSIS

Fungsi Association Rules seringkali disebut dengan *"market basket analysis"*, yang digunakan untuk menemukan relasi atau korelasi diantara himpunan item2. Market Basket Analysis adalah Analisis dari kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjaannya.

Fungsi ini paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis. Tipe association rule bisa dinyatakan sebagai misal : "70% dari orangorang yang membeli mie, juice dan saus akan membeli juga roti tawar". Aturan asosiasi mengcapture item atau kejadian dalam data berukuran besar yang berisi data transaksi. Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "basket data." Aturan asosiasi yang didefinisikan pada basket data, digunakan untuk keperluan promosi, desain katalog, segmentasi customer dan target pemasaran. Secara tradisional, aturan asosiasi digunakan untuk menemukan trend bisnis dengan menganalisa transaksi customer. Dan dapat digunakan secara efektif pada bidang Web Mining yang diilustrasikan sebagai berikut : pada Web access log, kita menemukan bahwa aturan asosiasi : "A and B implies C," memiliki nilai confidence 80%, dimana A, B, dan C adalah halaman Web yang bisa diakses. Jika seorang user mengunjungi halaman A dan B, maka terdapat 80% kemungkinan dia akan

mengunjungi halaman C juga pada session yang sama, sehingga halaman C perlu diberi direct link dari A atau B. Informasi ini dapat digunakan untuk membuat link secara dinamik ke halaman C dari halaman A atau B sehingga user dapat melakukan direct link ke halaman C. Informasi semacam ini digunakan untuk melakukan link ke halaman produk yang berbeda secara dinamik berdasarkan interaksi customer.

Apa Itu Kaidah Asosiasi?

- Kaidah asosiasi penambangan
  - Pertama kali diusulkan oleh Agrawal, Imielinski dan Swami [AIS93]
- Diberikan:
  - Suatu database transaksi
  - Setiap transaksi adalah suatu himpunan item-item
- Cari seluruh kaidah asosiasi yang memenuhi kendala minimum support dan minimum confidence yang diberikan user.
- Contoh:

  30% dari transaksi yang memuat bir juga memuat popok 5% dari transaksi memuat item-item berikut:
  - 30% : confidence dari kaidah ini
  - 5% : support dari kaidah ini
- Kita berminat untuk mencari seluruh kaidah ketimbang memeriksa apakah suatu kaidah berlaku.

Definisi Umum

- **Itemset:** himpunan dari item-item yang muncul bersama-sama
- **Kaidah asosiasi**: peluang bahwa item-item tertentu hadir bersama-sama.

  oX $\rightarrow$ Y dimana X n Y = 0
- **Support, supp(X)** dari suatu itemset X adalah rasio dari jumlah transaksi dimana suatu itemset muncul dengan total jumlah transaksi.
- **Konfidence (keyakinan)** dari kaidah X . Y, ditulis **conf(X . Y)** adalah
  - **conf(X $\rightarrow$ Y)=supp(X $\cup$ Y) / supp(X)**

  - Konfindence bisa juga didefinisikan dalam terminologi peluang bersyarat **conf(X → Y)=P(Y|X)=P(X ∩ Y)/P(X)**

- Database transaksi menyimpan data transaksi. Data transaksi bisa juga disimpan dalam suatu bentuk lain dari suatu database mxn.

Ukuran Support

- Misalkan **I={I1, I2, …,Im}** merupakan suatu himpunan dari literal, yang disebut item-item.
- Misalkan **D={T1, T2, …, Tn}** merupakan suatu himpunan dari n transaksi, dimana untuk setiap transaksi **T∈ D, T ⊆ I**.
- Suatu himpunan item **X ⊆ I** disebut itemset.
- Suatu transaksi T memuat suatu itemset X jika X ⊆ T.
- Setiap itemset X diasosiasikan dengan suatu himpunan transaksi **TX ={T∈ D | T ⊇ X}** yang merupakan himpunan transaksi yang memuat itemset X.
- Support **supp(X)** dari itemset X sama dengan **|TX|/|D|.**
- Didalam setiap item adalah nilainilai yang menyatakan besaran item terjual.

TID	Item A	Item B	Item C	Item D
T1	1	0	1	14
T2	0	0	6	0
T3	1	0	2	4
T4	0	0	4	0
T5	0	0	3	1
T6	0	0	1	13
T7	0	0	8	0
T8	4	0	0	7
T9	0	1	1	10
T10	0	0	0	18

Gambar 11.2. Bentuk Transaksi Database

- Item A muncul dalam 3 transaksi (|**TA**|) yakni di transaksi T1, T3,dan T8.
- Ada sebanyak 10 transaksi (|**D**|)

- Supp(A)= **|TA|/|D| = 3/10 = 0.3**
- Kombinasi CD muncul didalam 5 transaksi ((|**TCD**|) yakni di transasi T1, T3, T5, T6, dan T9.
- Supp(CD)= **|TCD|/|D|= 5/10 = 0.5**
- **Frequent itemset** didefinisikan sebagai itemset dimana support-nya lebih besar atau sama dengan minsupport yang merupakan ambang yang diberikan oleh user.
- Jika minsupport diberikan oleh user sebagai ambang adalah 0.2, maka frequent itemset adalah semua itemset yang supportnya besar sama dengan 0.2, yakni A, C, D, AC, AD, CD, ACD
- Dari frequent itemset bisa dibangun kaidah asosiasi sbb:

  A → C **C → A** **A → D**

  D → A **C → D** **D → C,**

  A,C → D **A,D → C** **C,D →A**

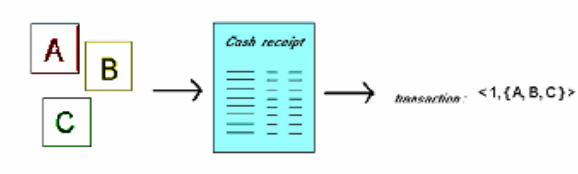
Itemset	s
A	0.30
B	0.10
C	0.80
D	0.70
AB	0.00
AC	0.20
AD	0.30
BC	0.10
BD	0.10
CD	0.50
ABC	0.00
ABD	0.00
ACD	0.20
BCD	0.10
ABCD	0.00

Gambar 11.3. Hasil nilai support untuk setiap items

Analisis dari kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjaannya.

Gambar 11.4. Keranjang Belanja

- Diberikan :
  - Suatu database transaksi customer (misal, keranjang belanja), dimana setiap transaksi dalah suatu himpunan item-item (misal produk)
- Cari:
  - Grup item-item yang sering dibeli secara bersama-sama



Gambar 11.5. Bentuk Transaksi Keranjang Belanja

- Mengekstraksi informasi perilaku pembelian
  - "**IF** membeli bir dan sosis, **THEN** juga membeli mostar dengan peluang tinggi"
- Informasi yang bisa ditindak-lanjuti:
  - Bisa menyarankan Tata letak toko yang baru dan campuran produk
  - Bisa menyarankan Produk apa untuk diletakkan dalam promosi ?
- Menganalisis tabel transaksi

Person	Basket
A	Chips, Salsa, coke, crakers, cookies, beer
B	Lettuce, Spinach, Oranges, Cellery, Apples, Grapes
C	Chips, Salsa, Frozen Pizza, Frozen cake
D	Lettuce, Spinach, Milk, Butter

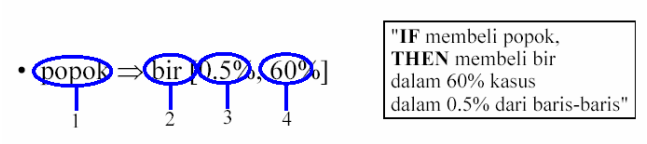
Gambar 11.6. Bentuk Analisa Keranjang Belanja

- Bisakah kita membuat hipotesa?
  - Chips => Salsa Lettuce => Spinach

Dasar Kaidah Asosiasi:

- Kaidah asosiasi penambangan:
  - Mencari pola yang sering muncul, asosiasi, korelasi, atau struktur sebab musabab diantara himpunan item-item atau objek-objek dalam database transaksi, database relasional, dan penyimpanan informasi lainnya
- Kepemahaman:
  - Sederhana untuk dipahami
- Kegunaan:
  - menyediakan informasi yang bias ditindaklanjuti
- Efisiensi:
  - ada algoritma pencarian yang efisient
- Aplikasi:
  - Analisis data keranjang pasar, pemasaran silang, rancangan katalog, analisis lossleader, clustering, klasifikasi, dsb.
- Format penyajian kaidah asosiasi yang biasa:
  - popok . bir [0.5%, 60%]
  - beli:popok . beli:bir [0.5%, 60%]
  - "**IF** membeli popok, **THEN** membeli bir dalam 60% kasus. Popok dan bir dibeli bersama-sama dalam 0.5% dari baris-baris dalam database."

- Penyajian lainnya (digunakan dalam buku Han):
  - Beli (x, “popok”) ⇒ beli (x, “bir”) [0.5%, 60%]
  - Major (x, "CS") ^ mengambil (x, "DB") ⇒ grade(x,"A") [1%, 75%]



1. **Antecedent**, left-hand side (LHS), body
2. **Consequent**, right-hand side (RHS), head
3. **Support**, frekuensi ("dalam berapa besar bagian dari data benda-benda dalam LHS dan RHS terjadi bersama-sama")
4. **Confidence**, kekuatan ("jika LHS terjadi, bagaimana kirakira RHS terjadi")

- **Support:** menunjukkan frekuensi dari kaidah didalam transaksi.

$$\text{support}(A \Rightarrow B\ [\ s, c\ ]) = p(A \cup B) = \underline{\text{support}(\{A,B\})}$$

- **Confidence:** menunjukkan persentasi dari transaksi yang memuat A yang juga memuat B.

$$\text{confidence}(A \Rightarrow B\ [\ s, c\ ]) = p(B|A) = p(A \cup B) / p(A) = \underline{\text{support}(\{A,B\}) / \text{support}(\{A\})}$$

- **Minimum support σ :**
  - High $\Rightarrow$ sedikit itemset yang sering
    $\Rightarrow$ sedikit kaidah yang sah yang sangat sering terjadi
  - Low $\Rightarrow$ banyak kaidah yang sah yang jarang terjadi
- **Minimum confidence γ :**
  - High $\Rightarrow$ sedikit kaidah, tetapi selurhnya "hampir secara logika true"
  - Low $\Rightarrow$ banyak kaidah, banyak diantaranya sangat "takpasti"
- **Nilai-nilai biasanya:** σ = 2 s/d 10 %, γ = 70 s/d 90 %
- **Transaksi:**
  - Relational format Format Kompak

    < Tid, item > < Tid, itemset >

    < 1, item1 > < 1, {item1,item2}>

    < 1, item2 > < 2, {item3}>

    < 2, item3 >
- **Item vs itemsets :** elemen tunggal vs. himpunan item

- **Support dari suatu itemset I:** jumlah transaksi yang memuat I
- **Minimum support σ:** ambang untuk support
- **Frequent itemset :** dengan support = σ

10.6.2. ALGORITMA APRIORI

Persoalan association rule mining terdiri dari dua sub persoalan :

1. Menemukan semua kombinasi dari item, dise but dengan frequent itemsets,yang memiliki support yang lebih besar daripada minimum support.
2. Gunakan frequent itemsets untuk men-generate aturan yang dikehendaki.Semisal, ABCD dan AB adalah frequent, maka didapatkan aturan AB -> CD jika rasio dari support(ABCD) terhadap support(AB) sedikitnya sama dengan minimum confidence. Aturan ini memiliki minimum support karena ABCD adalah frequent.

Algoritma Apriori yang bertujuan untuk menemukan frequent itemsets dijalankan pada sekumpulan data. Pada iterasi ke -*k*, akan ditemukan semua itemsets yang memiliki *k* items, disebut dengan *k* -itemsets. Tiap iterasi berisi dua tahap. Misal *Oracle Data Mining F*k merepresentasikan himpunan dari frequent *k* -itemsets, dan *C*k adalah himpunan candidate *k*-itemsets (yang potensial untuk menjadi frequent itemsets). Tahap pertama adalah men-generate kandidat, dimana himpunan dari semua frequent (*k*- 1) itemsets, *F*k-1, ditemukan dalam iterasi ke-(*k*-1), digunakan untuk men-generate candidate itemsets *C*k. Prosedur generate candidate memastikan bahwa *C*k adalah superset dari himpunan semua frequent *k*-itemsets. Struktur data hash-tree digunakan untuk menyimpan *C*k. Kemudian data di-scan dalam tahap penghitungan support. Untuk setiap transaksi, candidates dalam *C*k diisikan ke dalam transaksi, ditentukan dengan menggunakan struktur data hash-tree hashtree dan nilai penghitungan support dinaikkan. Pada akhir dari tahap kedua, nilai *C*k diuji untuk menentukan yang mana dari candidates yang merupakan frequent. Kondisi penghitung (terminate condition) dari algoritma ini dicapai pada saat *F*k atau *C*k+1 kosong.

Inti dari algoritma apriori :

- Gunakan frequent (k – 1)-itemsets untuk membangun kandidat frequent k-itemsets.
- Gunakan scan database dan pencocokan pola untuk mengumpulkan hitungan untuk kandidat itemsets

Penyumbatan dari apriori : generasi kandidat

- Himpunan kandidat yang besar sekali:
  - 10^4 frequent 1-itemset akan membangun 10^7 kandidat 2-itemsets.
  - Untuk menemukan suatu pola yang sering dari ukuran 100, misal, $\{a_1, a_2, \ldots, a_{100}\}$, seseorang perlu membangun $2^{100} \approx 10^{30}$ kandidat.
- Scan database berkali-kali:
  - Perlu (n +1) scans, n adalah panjang dari pola terpanjang

Dalam praktek:

- Untuk pendekatan apriori dasar, jumlah atribut dalam baris biasanya lebih kritis ketimbang jumlah baris transaksi
- Contoh:
  - 50 atribut masing-masing memiliki 1-3 nilai, 100.000 baris (tidak sangat buruk)
  - 50 atribut masing-masing memiliki 10-100 nilai, 100.000 baris (cukup buruk)

- Perhatian:
  - Satu atribut bisa memiliki beberapa nilai berbeda
  - Algoritma kaidah asosiasi biasanya memperlakukan setiap pasangan atribut-nilai sebagai satu atribut (2 atribut dengan masingmasing 5 nilai => "10 atribut")

Ada beberapa cara untuk mengatasi problem dalam algoritma apriori ini berikut, Perbaikan Kinerja Apriori :

1. **Hitungan itemset berbasis hash**:

   Suatu *k*-itemset yang hitungan ember hash terkaitnya dibawah ambang tidak bisa frequent.

2. **Reduksi transaksi**:

   Suatu transaksi yang tidak memuat frequent *k* itemset apapun adalah sia-sia dalam scan berikutnya.

3. **Partisi**:

   Itemset apapun yang potensial frequent dalam DB haruslah frequent dalam paling tidak satu dari partisi dari DB

4. **Sampling**:

   Penambangan atas suatu subset dari data yang diberikan, menurunkan ambang support suatu metoda untuk menentukan kelengkapan.

- **Diberikan:** (1) database transaksi, (2) setiap adalah suatu daftar dari item-item yang dibeli (dibeli seorang customer pada suatu kunjungan)

Transaction ID	Items Bought
100	A,B,C
200	A,C
400	A,D
500	B,E,F

Frequent Itemset	Support
{A}	3 or 75%
{B} dan {C}	2 or 50%
{D}, {E} dan {F}	1 or 25%
{A,C}	2 or 50%
Pasangan item lainnya	max 25%

- **cari: seluruh** kaidah dengan minimum support dan confidence
- If min. support 50% dan min. confidence 50%, then A => ***C*** [50%, 66.6%], ***C*** => ***A*** [50%, 100%].
- Langkah-langkah untuk mencari nilai minimum support dam confidence dengan algoritma apriori

  STEP 1: cari *frequent itemsets*: himpunan item-item yang memiliki minimum support.

  - Disebut trik Apriori: suatu subset tak hampa dari suatu frequent itemset haruslah juga suatu frequent itemset:
    - Artinya, jika $\{AB\}$ adalah suatu frequent itemset, kedua $\{A\}$ dan $\{B\}$ harus juga frequent itemsets.

- Secara iteratif cari frequent itemsets dengan ukuran dari 1 hingga *k (k-*itemset*)*

STEP 2: gunakan frequent itemsets untuk membangun kaidah asosiasi..

- Jika **{bir,popok, kacang}** frequent, maka **{bir, popok}** juga frequent.
- Setiap transaksi yang memiliki {beer, popok, kacang} juga memuat {bir, popok}.
- Jika {A,B} memiliki support paling tidak a , maka A dan B keduanya memiliki support paling tidak a.
- Jika A atau B memiliki support kecil dari a maka {A, B} memiliki support lebih kecil dari a.

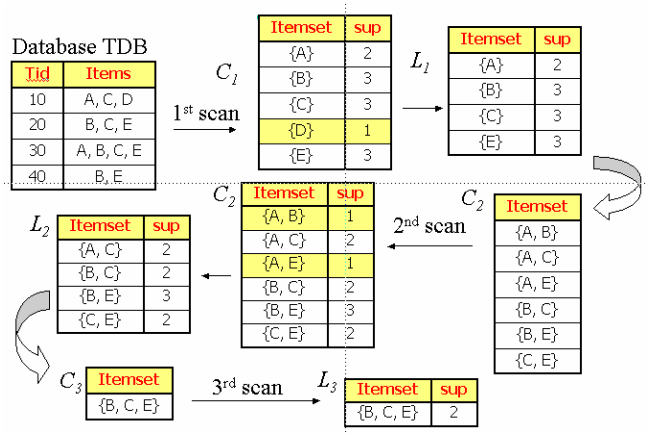
Step Gabungan: *Ck* dibangun dgn menggabungkan *Lk-1*dengan dirinya

Step Pemangkasan: setiap (*k-1*)-itemset yg bukan frequent tidak boleh menjadi suatu subset dari suatu frequent *k*-itemset.

Pseudo-code: *Ck*: Kandidate itemset dari ukuran k; *Lk* : Frequent itemset dari ukuran k.

L1 = {frequent items};
for (*k* = 1; *Lk* !=0; *k*++) **do begin**
    Ck+1 = {kandidat dibangun dari *Lk* };
for each transaksi *t* dalam database do naikkan hitungan dari seluruh kandidat dalam *Ck+1* yang dimuat dalam *t*
    Lk+1 = {kandidat dalam *Ck+1* dengan min_support}
end
return .*k Lk*;

Bentuk Ilustrasi Algoritma Apriori



Gambar 11.7. Ilustrasi Algoritma Apriori

Contoh apriori :

TID	CID	Item	Price	Date
101	201	Computer	1500	1/4/99
101	201	MS Office	300	1/4/99
101	201	MCSE Book	100	1/4/99
102	201	Hard disk	500	1/8/99
102	201	MCSE Book	100	1/8/99
103	202	Computer	1500	1/21/99
103	202	Hard disk	500	1/21/99
103	202	MCSE Book	100	1/21/99

- Dalam contoh ini untuk kaidah asosiasi
  {Computer} → {Hard disk}
  - Jumlah seluruh transaksi adalah 3 (oleh customer 201 dua kali yakni pada 1/4/99 dan 1/8/99, customer 202 sekali yakni pada 1/21/99. Catatan perhatikan Customer dan tanggal transaksi)

- Jumlah transaksi Computer dan Hard Disk adalah 1 (oleh customer 202 pada 1/21/99)
- Jumlah transaksi hanya Computer adalah 2 (pada 1/4/99 oleh 201 dan pada 1/21/99 oleh 202)

  > Support(Computer Hard disk) = 1/3=33.3%

  > Conf(Computer Hard disk) = 1/2=50%

- Bagaimana dengan {Computer} → {MCSE book}
  - Jumlah seluruh transaksi adalah 3 (oleh customer 201 dua kali, customer 202 sekali. Catatan perhatikan Customer dan tanggal transaksi)
  - Jumlah transaksi Computer dan MCSE book adalah 2 (oleh customer 201 dan 202)
  - Jumlah transaksi hanya Computer adalah 2

    > Support(Computer {MCSE book)= 2/3 = 66.6%

    > Conf(Computer {MCSE book)= 2/2 = 100%

- Berapa support dari 2-itemset {Computer , Hard disk} ?
  - Jumlah transaksi 2-itemset {Computer, Hard disk}adalah 1.
  - Jumlah transaksi keseluruhan adalah 3.

    > Support dari 2-itemset {Computer, Hard disk} adalah 1/3=33.3%

- Berapa support dari 1-itemset {Computer}?
  - Jumlah transaksi 1-itemset {Computer}adalah 2.
  - Jumlah transaksi keseluruhan adalah 3.

    > Support dari 1-itemset {Computer} adalah 2/3=66.6%

- **2 Step dalam kaidah asosiasi penambangan:**
  - Cari seluruh itemsets yang supportnya diatas minimum support yang diberikan oleh user. Kita sebut itemsets ini itemsets besar.
  - Untuk setiap itemset besar L, carilah seluruh kaidah asosiasi dalam bentuk a (L-a) dimana a dan (L-a) adalah himpunan bagian L yang tak hampa.
- **Step 2 adalah jelas yang dikaitkan dengan step 1:**
  - Ruang pencarian eksponensial
  - Ukuran dari transaksi database

Supp(Computer)=2/3=66.7%, supp(MS Office)=1/3=33.3%

Supp(MCSE Book)=3/3=100%, supp(Hard Disk)=2/3=66.7%

Supp(Computer,MSOffice)=1/3=33.3%

Supp(Computer,MCSE Book)=2/3=66.7%

Supp(Computer,Hard Disk)=1/3=33.3%

Supp(MCSE Book, MS Office)=1/3=33.3%

Supp(MCSE Book, Hard Disk)=2/3=66.7%

Supp(MSOffice,Hard Disk)=0/3=0%

Supp(Computer, MCSE Book,MSOffice)=1/3=33.3%

Supp(Computer, MCSE Book, Hard Disk)=1/3=33.3%

Supp(MCSE Book, MSOffice,Hard Disk)=0/3=0%

Supp(Computer,MCSE Book, MSOffice,HardDisk)=0/3=0%

Asosiasi dengan minsupport 60% adalah:

Computer → MCSE Book, MCSE Book → Computer

MCSE Book → Hard Disk, Hard Disk → MCSE Book

Conf(Computer → MCSE Book)=2/2=100%

Conf(MCSE Book → Computer)=2/3=66.7%

Conf(MCSE Book → Hard Disk)=2/3=66.7%

Conf(Hard Disk → MCSE Book)=2/2=100%

Jadi, asosiasi yang memenuhi minsupport 60% dan minconfidence 80% adalah:

Hard Disk → MCSE Book dan

Computer → MCSE Book

RINGKASAN:

- Pada dasarnya data mining berhubungan dengan analisa data dan penggunaan teknik-teknik perangkat lunak untuk mencari pola dan keteraturan dalam himpunan data yang sifatnya tersembunyi.
- Data mining diartikan sebagai suatu proses ekstraksi informasi berguna dan potensial dari sekumpulan data yang terdapat secara implisit dalam suatu basis data
- Tantangan-tantangan dalam Data Mining meliputi : penanganan berbagai tipe data, efisiensi dari algoritma data mining, kegunaan, kepastian dan keakuratan hasil, ekspresi terhadap berbagai jenis hasil dan data yang diambil dari berbagai sumber yang berbeda.
- Tahapan dalam Data Mining meliputi : proses seleksi, pembersihan data, tranformasi, implementasi teknik data mining dan interprestasi hasil
- Fungsionalitas dalam Data Mining meliputi mining association rule, karakterisasi data multilevel, klasifikasi data, analisa cluster, dan pencarian pola sekuensial
- Teknik-teknik dalam Data Mining yang bisa diterapkan antara lain : *market basket analysis* dan Algoritma Apriori.

LATIHAN SOAL :

1. Apa perbedaan antara klasifikasi dan clustering ?
2. Apa peranan visualisasi informasi dalam data mining ?
3. Definisikan support dan confidence untuk aturan asosiasi
4. Jelaskan mengapa aturan asosiasi tidak dapat digunakan secara langsung untuk prediksi, tanpa analisis yang lebih lanjut atau domain pengetahuan !
5. Perhatikan table Purchase berikut ini :

Transid	Custid	Date	Item	Qty
111	201	5/1/2002	Ink	1
111	201	5/1/2002	Milk	2
111	201	5/1/2002	Juice	1
112	105	6/3/2002	Pen	1
112	105	6/3/2002	Ink	1
112	105	6/3/2002	Water	1
113	106	5/10/2002	Pen	1
113	106	5/10/2002	Water	2
113	106	5/10/2002	Milk	1
114	201	6/1/2002	Pen	2
114	201	6/1/2002	Ink	2
114	201	6/1/2002	Juice	4
114	201	6/1/2002	Water	1
114	201	6/1/2002	Milk	1

Simulasikan algoritma untuk menemukan frequent itemset pada table degan minimum support = 90 persen, lalu cari aturan asosiasi dengan minimum confidence = 90 persen.