BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Data Mining

Akibatnya, pengumpulan data adalah eksplorasi atau upaya untuk mengungkap pengetahuan yang sebelumnya tidak terlihat yang terkandung di dalam artefak, dan pada dasarnya adalah pemeriksaan apa yang telah disembunyikan di dalam kumpulan informasi yang besar (Davies, 2004). Data penambangan juga dapat digunakan sebagai langkah dalam proses penambahan nilai tambah secara manual pada kumpulan data sebagai sumber informasi yang belum dapat dipertanggungjawabkan (Pramudiano, 2007). KDD (Knowledge Discovery and Data Mining) juga dikenal sebagai data mining. Pengetahuan adalah metode untuk mengumpulkan dan menggunakan data historis untuk mendapatkan wawasan politik, atau ungkapan lain untuk menjelaskan dalam konteks kumpulan data yang tidak kecil (Santoso, 2007).

Penambangan data merupakan teknik untuk mendapatkan pola yang cukup berbeda dari sebelumnya dalam jumlah data yang tidak sedikit menurut Han (2006). Ini berfungsi sebagai lokasi untuk mengunggah data dan sebagai gudang untuk semua informasi lainnya. Informasi yang relevan dengan bidang lain, seperti sistem data, sistem manajemen data, sistem statistik, sistem mesin, sistem pemrosesan informasi, dan komputer sekolah menengah. Selain itu, data mining dibantu oleh jenis pengetahuan lain, seperti pengetahuan teori relativitas dan analisis data spasial dan temporal. Data mining didefinisikan sebagai proses menemukan pola dalam sejumlah besar data. Proses saat ini dapat dilakukan secara otomatis atau semi otomatis. Sudah ada manfaat, dan ada kemungkinan model itu sendiri akan membawa manfaat tersebut, dengan kemungkinan terbesar dalam bentuk manfaat ekonomi. Sejumlah besar data besar diperlukan (Witten, 2005). Ada ciri-ciri tertentu dari gudang data yang dapat diuraikan sebagai berikut:

1. Dengan tidak adanya bukti yang jelas, atau pengetahuan sebelumnya tentang keberadaannya, data menunjukkan bahwa seseorang mungkin memikirkan sesuatu yang belum pernah mereka lihat atau sadari telah mereka pikirkan.
2. Gudang data sering menggunakan data dalam jumlah besar. Salah, sejumlah besar data besar digunakan untuk membuat hasil lebih mudah ditafsirkan.
3. Untuk menyatakan dengan jelas ide-ide konkret, terutama yang berhubungan dengan strategi, data yang baik cukup membantu (Davies, 2004).

Dalam konteks definisi ini, dapat dilihat bahwa augmentasi data adalah suatu metode untuk memperoleh informasi yang bernilai dari kumpulan data yang besar, sehingga dapat menghasilkan suatu pola yang unik dari yang sebelumnya tidak diketahui keberadaannya. Dengan menggunakan metode ini, Anda dapat mengekstrak sejumlah kecil inti dari sejumlah besar inti yang lebih besar, yang kemudian akan digunakan untuk membuat bahan bangunan dasar. Karena itu, penambangan data memiliki keunggulan yang jelas di bidang-bidang seperti kecerdasan buatan (AI), pembelajaran mesin, analisis statis, dan fondasi data. Ada berbagai metode yang sering digunakan dalam penjelasan data, antara lain: pengelompokan, klasifikasi, asosiasi, syaraf tiruan, algoritma genetika, dan lain-lain (Pramudiono, 2007).

Pengenalan pola adalah studi tentang bagaimana mengklasifikasikan objek ke dalam beberapa tingkatan atau kategori dan mengidentifikasi pola dalam data. Tergantung pada aplikasinya, mungkin seorang siswa, seorang pemimpin, pemegang kartu kredit, gambar atau simbol, atau apa pun yang harus diklasifikasikan atau dikenali untuk fungsi pengembaliannya (Santoso, 2007). Penambangan data, juga dikenal sebagai KDD (Knowledge Discovery via Data Mining), adalah aktivitas memperoleh dan menggunakan data historis untuk mendapatkan wawasan tentang struktur kumpulan data. Hasil dari teknik data mining ini dapat dimanfaatkan dengan baik dalam memprediksi kejadian di masa depan. Akibatnya, isilah pengnalan pola sering digunakan sebagai kantong penyimpanan data (Santoso, 2007).

2.1.2 Pengertian Data Warehouse

Ada berbagai teori interpretasi data, namun semuanya memiliki estetika yang mendasar, menurut beberapa ahli. Richard D.H.W.H. Inmon, Gudang data adalah sekumupulan data yang mempunyai orientasi pada subjek, terintegrasi, bervariasi dalam waktu, dan meningkatkan proses pengambilan keputusan dalam manajemen. Menurut Vidette Poe, gudang data adalah dasar dari data analitik dan file hanya-baca yang berfungsi sebagai pengingat bagi sistem yang melacak niat pengguna. Menurut Paul Lane, "data gudang" mengacu pada data dunia nyata yang digunakan untuk tujuan seperti meminta informasi tentang dan menganalisis transaksi; sering kali mencakup informasi tentang riwayat transaksi serta data dari sumber lain. Data berasal dari staf bank analitis dan transaksional dan memiliki kemampuan untuk mengatur dan mengkonsolidasikan data dari berbagai sumber. Akibatnya, gudang data adalah metode desain data yang mendukung DSS (Sistem Pendukung Keputusan) dan EIS (Sistem Informasi Manajemen). Meskipun mudah untuk menganggap "data besar" sebagai "data dasar", desain "data besar" dan "data dasar" sangat berbeda secara konseptual. Penggunaan normalisasi dalam desain basis data jangka panjang bukanlah solusi yang tepat. Berdasarkan pengertian-pengertian di atas, ditetapkan bahwa gudang data merupakan dasar data yang dapat digunakan untuk penelitian dan analisis, tetapi tidak memenuhi kriteria data yang dapat digunakan untuk membantu mereka yang mencoba merumuskan.

2.1.3 Istilah-istilah Data Warehouse

Sebagai contoh masalah terkait data, pertimbangkan hal berikut :

1. Menurut O'Brien (2003, p21), sistem untuk melacak pikiran seseorang adalah apa adanya. 'Sistem pendukung keputusan' adalah peran baru untuk sistem informasi, yang berarti bahwa SPK adalah peran baru untuk sistem informasi yang menyediakan dukungan

interaktif dan sesuai permintaan pengguna akhir manajerial untuk proses pengambilan keputusan mereka.' Artinya, SPK merupakan peran baru untuk dukungan interaktif bagi pengguna akhir manajerial. Ini juga merupakan sistem yang menginformasikan pengguna tentang bagaimana sistem menganalisis situasi dan menghasilkan kesimpulan yang cocok untuk mereka.

2. Menurut Connoly-Begg (2002, p1067) “Data Mart is a subset of data warehouse that support the requirements of a specific department of business function,” yang berarti bahwa data mart adalah subset dari data warehouse yang mendukung penyebaran informasi dari suatu unit bisnis atau fungsi bisnis tertentu. Data mart adalah komponen kunci dari penyimpanan data dan dapat digunakan untuk pengarsipan dan analisis data di banyak unit bisnis, departemen, dan operasi.

   Beberapa perbedaan antara data mart dan gudang data dapat dijelaskan :

a. sebuah. Data Mart hanya melayani kebutuhan pengguna yang telah menjalin koneksi dengan satu unit bisnis atau fungsi.
b. Data Mart tidak memuat data yang relevan secara operasional secara detail, tidak seperti Gudang.
c. Karena datanya lebih ringkas, informasi di data mart lebih mudah dipahami daripada informasi di database warehouse.

   OLAP adalah kategori perangkat lunak yang memungkinkan analis dan manajer untuk mendapatkan wawasan tentang data melalui akses interaktif yang cepat, konsisten, ke berbagai kemungkinan pandangan informasi yang telah diubah dari data mentah untuk mencerminkan dimensi nyata dari perusahaan. sebagaimana dipahami oleh pengguna," yang berarti bahwa "OLAP adalah kategori teknologi perangkat lunak." OLAP juga merupakan teknik analisis database yang menggunakan tabel fakta dan dimensi untuk menganalisis berbagai kueri yang berasal dari kumpulan data besar.

3. Mallach (2000, p531) menuturkan bahwa, *Online Analytical Processing* (OLAP) "*OLAP is a category of sofware that enables analyst, managers, and executive to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user*", yang artinya OLAP adalah kategori teknologi software yang dapat memungkinkan penganalisa, manager, dan eksekutif untuk melihat data yang ada dengan akses yang cepat, konsisten dan interaktif sehingga dapat melihat informasi yang sudah di transformasi dari data mentah menjadi dimensi keadaan nyata yang dapat dimengerti dengan mudah oleh user. OLAP juga merupakan suatu pemrosesan database yang menggunakan tabel fakta dan dimensi untuk dapat menampilkan berbagai bentuk laporan, query dari data yang berukuran besar.

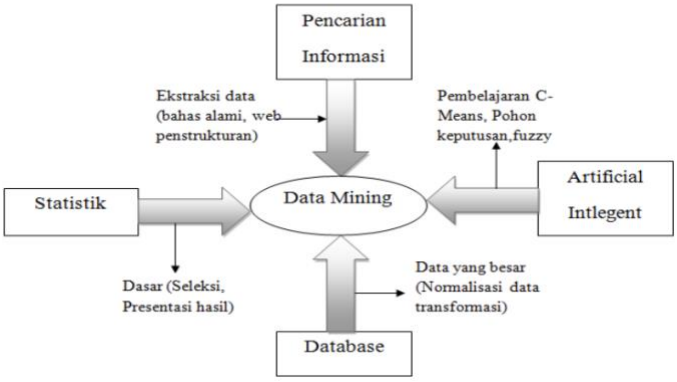
4. Menurut O'Brien (2003, p224), "OLTP adalah sistem pemrosesan transaksi waktu nyata," itulah yang dimaksud dengan TPS saat ini. "Transaction Processing System (TPS) adalah sistem informasi lintas fungsi yang memproses data harian dari transaksi bisnis", yang berarti TPS adalah sistem informasi lintas fungsi yang memproses data harian dari transaksi bisnis. OLTP dirancang untuk menyediakan akses simultan oleh beberapa pengguna ke kumpulan data bersama yang sama dan untuk menjalankan setiap proses yang diperlukan.

5. Selain itu, tabel fakta yang memuat kategori dan data rinci yang dapat digunakan sebagai insentif juga dapat mencantumkan waktu (dalam bentuk hari, minggu, dan tahun) sebagai dimensi tabel.

6. Tabel Fakta (Fact Table) merupakan tabel standar yang berisi data historis dan unique key karena key tersebut merupakan kombinasi dari ada kunci asing dan kunci utama yang muncul di setiap level. Unik dari tabel tersebut. Seperti dapat dilihat, Tabel Faktual memiliki berbagai jenis pengukuran, seperti yang berkaitan dengan dimensi

tabel dan lainnya yang tidak.

7. Data mining adalah proses untuk menganalisis data mentah yang belum dipahami, menurut Aldeman (2000, hlm. 145), sedangkan penambangan data adalah tentang menganalisis data juga menemukan pola yang tidak terlihat menggunakan metode yang ringkas dan sarana semi-otomatis." Demikian pula, "Data mining adalah tentang menemukan pola dalam data yang belum pernah ditemukan atau diprediksi sebelumnya," menurut Aldeman. Analisis data dan menemukan pola yang tidak terlihat dengan penggunaan maksud secara manual dan semi-otomatis adalah dua contohnya. Selain fakta bahwa ada baiknya menambahkan lebih banyak informasi ke data yang sudah ada, menambahkan nilai yang sudah ada berfungsi sebagai semacam alat pengumpulan pengetahuan. Pertimbangan pemrosesan data lainnya termasuk memahami makna yang mendasari data (KDD).

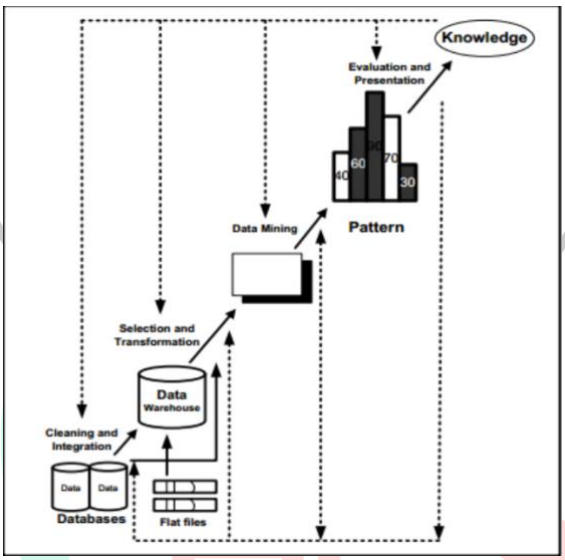


Gambar 2. 1 Gambar Bidang Ilmu Data Mining

2.1.4 Tahapan Data Mining

Sekalipun hanya ada beberapa langkah atau tahapan dalam prosesnya, data yang sedang dikirim dapat diubah menjadi banyak tahapan. Untuk

melaksanakan tugas ini, dilakukan baik di latar belakang atau melalui basis pengetahuan interaktif, seperti yang ditunjukkan pada **Gambar 2.2** di sebelah kanan halaman ini.



Gambar 2. 2 Gambar Tahap – tahap Data Mining (Han, 2006)

Dalam data mining, ada enam tahap, seperti yang ditunjukkan pada paragraf berikut :

1. Pembersihan data (Data Cleaning)

Pengurangan noise merupakan produk sampingan dari proses pengambilan data, termasuk data yang tidak konsisten atau tidak relevan. Biasanya data yang diperoleh dari database perusahaan dan laporan audit internal memiliki informasi yang tidak lengkap atau tidak benar, seperti data yang kadaluwarsa, data yang tidak lagi valid, atau hanya konsekuensi dari audit. Selain itu, beberapa atribut data tidak ada hubungannya dengan proses penambahan data. Selain itu, ini adalah ide yang baik untuk membuang data yang tidak relevan. Pembersihan data juga berdampak pada kinerja teknik data mining karena mengurangi volume dan kompleksitas data yang ditambang.

2. Integrasi data (Data Integration)

Selain mengintegrasikan data dari sumber yang berbeda, konsolidasi data sering disebut sebagai integrasi data. Tidak banyak data yang dibutuhkan untuk data mining, dan data tersebut berasal dari beberapa database atau file teks, bukan hanya satu. Nama atribut, jenis produk, dan nama pelanggan semuanya digunakan untuk mengidentifikasi entitas unik dalam data yang disertakan. Penggabungan data harus dilakukan dengan hati-hati, karena kesalahan apa pun dalam integrasi data dapat membahayakan hasil akhir dan menyebabkan masalah di masa mendatang. Sebagai contoh, jika integrasi data menggabungkan kategori produk yang berbeda untuk jenis produk yang berbeda, kemungkinan besar akan ditemukan korelasi antara produk yang sebenarnya tidak tersedia.

3. Seleksi Data (Data Selection)

Dalam database, data jarang digunakan untuk hal lain selain analisis yang digunakan database. Sebagai contoh, jika analisis keranjang belanja Anda menunjukkan bahwa seseorang siap membeli, Anda tidak perlu memberikan nama pembeli, cukup nama pembeli saja.

4. Transformasi data (Data Transformation)

Data ditransformasikan atau diubah ke dalam format yang sesuai untuk digunakan dalam pengumpulan data. Ada banyak cara untuk mengangkut data yang membutuhkan penggunaan format tertentu lebih spesifik. Dapat digunakan analisis asosiasi dan pengolahan data hanya dapat menerima informasi kategorikal. Akibatnya, data harus disimpan pada berbagai interval sebagai total berjalan. Transformasi data adalah istilah umum untuk proses ini.

5. Prosedur Penambangan

Proses mining adalah proses penting yang menggunakan metode untuk menemukan pengetahuan berharga dan tersembunyi dari data Anda.

6. Evaluasi pola (Pattern Evaluation)

Mengenali pola yang bermasalah perhatian Anda dari hasil yang Anda lihat. Pada fase ini, hasil data mining dianalisis dalam bentuk model statistik dan model prediksi untuk menentukan apakah hipotesis yang diberikan benar atau tidak. Beberapa pilihan tersedia untuk Anda jika hasil yang diperoleh tidak memenuhi harapan. Misalnya, hasil yang tidak terpengaruh oleh bug mungkin berguna.

2.1.5 Analisis Asosiasi

Asosiasi analisis, atau augmentasi data, adalah teknik untuk mengidentifikasi hubungan antara item dan trajectories asosiasi yang sesuai. Langkah pertama dalam melakukan analisis pelanggan di supermarket adalah menentukan kemungkinan bahwa setiap pelanggan akan membeli roti pada waktu yang sama dengan yang sebelumnya. Sebagai hasil dari informasi ini, pemilik toko dapat menggunakan alat kombinasi produk untuk menemukan produk baru dan meningkatkan kampanye pemasaran mereka. Saat digunakan untuk menganalisis supermarket isi keranjang belanja, tautan analisis menjadi semakin populer. Analisis tautan terkadang disebut sebagai analisis proses pembelian. Analisis linkage juga disebut sebagai data mining adalah teknik yang penting untuk banyak teknik lainnya. Teknik analisis yang paling banyak digunakan adalah yang dikenal sebagai ekstraksi frekuensi sampel, menarik perhatian banyak peneliti dalam pengembangan algoritma yang efisien. Aspek terpenting dari perilaku asosiatif dapat dipetik dari dua parameter: jarak (nilai dukungan), yang berfungsi sebagai proxy untuk hubungan antara berbagai faktor dalam database, dan kepercayaan (nilai pasti), yang mewakili tingkat

kepercayaan antara para pihak. Ada satu hal lagi yang harus disebutkan. Terminologi Asosiasi biasanya digunakan dalam format berikut: "smoke, coffee, match" (dukungan 40%, kepercayaan 50%). Artinya, terima kasih "Roti dan mentega termasuk dalam database 50.000 transaksi. Juga termasuk produk seperti sirup. Saat ini, 40% dari semua transaksi dalam database terdiri dari tiga item ini. Berikut ini juga dapat digunakan sebagai contoh: "Konsumen membeli rokok dan memiliki peluang lebih besar dari 50% untuk membeli korek api juga. Dalam hal ini, taruhannya tinggi karena ada lebih dari 40.000 transaksi sejak tahun lalu. Analisis asosiasis minimal adalah persyaratan. Proses menemukan semua aturan dan ketentuan yang memenuhi persyaratan minimum (dukungan minimal) dan persyaratan jaminan minimum didefinisikan di bawah (keyakinan minimum). Analisis tautan dibagi menjadi dua metode terpisah :

2.1.6 Analisa pola frekuensi tinggi

Dalam kalimat ini, kami mencari kecocokan elemen yang dapat menunjukkan prasyarat untuk tempat desimal data. Menunjukkan cara menghitung nilai support elemen tertentu, seperti pada **Gambar 2.3** dibawah ini :

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

Gambar 2. 3 Gambar Mencari Nilai Support A

Sebaliknya, nomor dukungan untuk item 2 dan 3 dapat diperoleh dengan mudah seperti yang ditunjukkan pada **Gambar 2.4** di bawah ini.

.

$$\text{Support } (A \cap B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}} \quad \ldots$$

Gambar 2. 4 Gambar Mencari Nilai Support A & B

2.1.7 Pembentukan aturan asosiatif

Pada titik ini, setelah semua sampel frekuensi tinggi ditemukan, saatnya untuk melihat aturan asosiatif yang memenuhi persyaratan minimal untuk kepercayaan dengan menghitung kepercayaan dari asosiasi A-B. Seperti ditunjukkan pada Gambar 2.5 di bawah ini, ketidakpercayaan dari aturan A-B dapat diperoleh sebagai berikut .

$$\text{Confidence} = P(B \mid A) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}} \quad ..$$

Gambar 2. 5 Gambar Mencari Confidence

Jika Anda ingin memahami proses penggunaan algoritma apriori, ini akan menunjukkan cara melakukannya. Menunjukkan hasil dukungan minimal 0,5 atau 2 seperti yang dapat dilihat dari **Tabel 2.1** dibawah ini.

Tabel 2. 1 Data Transaksi

Transaksi ID	**Item Set**
1	Item A, Item C, Item D
2	Item B, Item C, Item E
3	Item A, Item B, Item C, Item E
4	Item B, Item E

Dalam hal ini, kami mencari dukungan minimal 50% atau 0,50% (2 dari 4 transaksi) Untuk setiap kelompok, carilah nilai dukungan. Selain yang dijelaskan pada **Tabel 2.2** di bawah ini..

Tabel 2. 2 Nilai support 1 itemset

Itemset	**Support**
A	50%
B	75%
C	75%
D	25%
E	75%

Cara selanjutnya adalah mencari kandidat itemset untuk L 2 dan melampirkan item ke layer satu [A B, A C, A E, B C, B E, C E] (apriori-gen) butir D bukan termasuk dalam pencampuran dikarenakan kurangnya dukungan minimal. Kemudian, lihat total biaya perlindungan masing-masing komponen. Hasilnya dapat dilihat pada tabel **Tabel 2.3** dibawah ini.

Tabel 2. 3 Nilai support 2 itemset

Itemset	**Support**
A B	25%
A C	50%
A E	75%
B C	50%
B E	75%
C E	50%

Cara 4 : Menentukan itemset untuk menemukan dukungan. Hasilnya bisa dijelaskan pada **Tabel 2.4** dibawah ini.

Tabel 2. 4 Hasil itemset yang memenuhi support minimum

Itemset	**Support**
A C	50%
B C	50%
B E	75%
C E	50%

Cara kelima adalah lanjutan dari cara 2-4; langkah selanjutnya adalah menghubungkan item set pada layer 1 dan layer 2. Dengan mendapatkan hasil yang diketahui pada **Tabel 2.5** dibawah ini.

Tabel 2. 5 Hasil penggabungan 3 itemset

Itemset	**Gabungan 3 itemset**
A C + B C	ABC
A C + B E	ACB, ACE, ABE
A C + C E	ACE
B C + B E	BCE
B C + C E	BCE
B E + C E	BCE

Cara keenam adalah menghitung dukungan dari tiap-tiap calon set barang layer 3 seperti yang dijelaskan pada **Tabel.2.6** dibawah ini.

Tabel 2. 6 Nilai support 3 itemset

Itemset	**Support**
A B C	25%
A C E	25%
A B E	25%
B C E	50%

Cara ketujuh: L3 besar 3 item set "B C E" dan 3 item berikutnya dijatuhkan karena dukungan level tidak mencukupi.

Cara kedelapan : Untuk menemukan kriteria asosiatif, Anda juga memerlukan tingkat kepercayaan minimal 75%. Kriteria validitas asosiatif dapat dimodifikasi, seperti yang ditunjukkan pada **Tabel 2.7** di bawah ini.

Tabel 2. 7 Nilai confidence untuk setiap itemset

Aturan X - Y	**Support (X u Y)**	**Support X**	**Confidence**
B C → E	50%	50%	100%
B E → C	50%	75%	66.67%
C E → B	50%	50%	100%
A → C	50%	50%	100%
C → A	50%	50%	66.67%
B → C	50%	50%	66.67%
C → B	50%	50%	66.67%
B → E	75%	50%	100%
E → B	75%	50%	100%
C → E	50%	50%	66.67%
E → C	50%	50%	66.67%

2.1.8 Alogitma Apriori

Algoritma apriori merupakan algoritma yang terkenal untuk menemukan itemset yang sering muncul dengan memanfaatkan teknologi pemrosesan asosiatif. Algoritma Apriori menggunakan pengetahuan tentang sekumpulan item yang telah dipahami secara luas di masa lalu untuk menemukan informasi lebih lanjut. Algoritme Apriori mempertimbangkan ambang dukungan paling sedikit saat mengidentifikasi kandidat yang mungkin. Dua proses utama yang dilakukan oleh algoritma Apriori adalah sebagai berikut:

A. Bergabung

Dalam proses ini, setiap elemen dihubungkan ke item terkait hingga tidak ada lagi kemungkinan kombinasi. Algoritma Apriori ditransformasikan ke dalam beberapa proses yang dikenal dengan iterasi atau lintasan. Setiap iterasi menghasilkan sinyal frekuensi tinggi dengan panjang yang identik dengan sinyal frekuensi tinggi iterasi pertama dengan panjang tunggal. Pada iterasi awal, batasan setiap elemen dipenuhi dengan menggunakan database. Setelah menerima asuransi untuk setiap item, setiap item dengan asuransi yang lebih besar dari minimum yang disyaratkan diklasifikasikan sebagai item frekuensi tinggi dengan cakupan yang sama dengan atau biasanya melebihi cakupan untuk satu set item. Kitemset adalah istilah yang digunakan untuk menggambarkan himpunan berbasis K-element. Iterasi kedua menghasilkan dua itemset umum, masing-masing dengan dua input. Pertama, kedua elemen kandidat digabungkan menjadi satu himpunan. Setelah itu, setiap kandidat kedua dicari di database menggunakan penelusuran database.

B. Pemangkasan

Selama proses ini, hasil gabungan kemudian tidak disarankan menggunakan tingkat perlindungan dan kepercayaan pengguna yang minimal. Dukungan minimal yang dimaksud di sini adalah standar yang digunakan untuk membedakan antara calon yang berhasil dan yang tidak. Ambang minimal untuk dapat dipercaya adalah angka yang digunakan untuk menghitung skor minimum yang dapat diterima untuk mengaitkan dua item lengkap untuk setiap kandidat dan tiga set item terkait. Setelah mencapai persyaratan minimal untuk dua set item kandidat, dua set item kandidat yang memenuhi persyaratan ini dapat diidentifikasi sebagai dua set item secara teratur. Kedua item yang ditetapkan ini juga mewakili model frekuensi, dengan frekuensi sudut dua elemen. Untuk iterasi lebih lanjut, pie dapat diubah menjadi beberapa kantong (Sumber: Agrawal, R., Srikant, 2019). Membentukkandidat frequent set, himpunan kkandidat terbentuk dari kombinasi (k-1)-itemset yang diperoleh dari iterrasi lainnya. Salah satu fitur dari algoritma Apriori adalah kemampuan untuk

mengecualikan sekelompok kandidat dari k-item yang berisi k-1 yang hilang dari sampel dengan frekuensi tinggi relatif terhadap k-1 panjang.

1. Untuk mengurangi jumlah transaksi yang melibatkan setiap calon K-itemet, harus digunakan database. Ini mengurangi jumlah transaksi yang melibatkan setiap kandidat untuk K-itemet. Ini juga merupakan fitur dari algoritma Apriori yang diperlukan untuk mengatasi masalah melintasnya basis data dengan jumlah himpunan k-item terbanyak di sisi panjang..

2. Menentukan surat suara dengan nilai tertinggi, yang dapat berupa surat suara termasuk k-item atau k-itemet yang mendapat dukungan lebih dari minimum yang dipersyaratkan.

3. Jika frekuensi pola baru tidak dipahami, semua proses akan dihentikan. Jika tidak, maka K akan bergerak ke kiri dan menuju ke bagian 1. Meskipun algoritma Apriori lebih sederhana untuk dipahami dan diimplementasikan jika dibandingkan dengan algoritma lain yang secara alami terjadi selama proses asosiatif, algoritma Apriori juga memiliki kelemahan: dalam rangka untuk find frequent set, harus berulang kali menanyakan database untuk setiap kemungkinan kombinasi item. Ini mencegah perlunya beberapa jam untuk mengelola database. Untuk mendapatkan kombinasi faktor dari database, Anda juga harus membuat analisis rata-rata pergerakan 21 hari menggunakan algoritma Apriori dan kandidat FPGrowth Erwin. Dibandingkan dengan algoritma lain yang sering digunakan dalam proses asosiasi, algoritma Apriori mungkin lebih sederhana untuk dipahami dan digunakan, tetapi memiliki kelemahan: untuk menemukan himpunan umum, algoritma Apriori harus memperbarui basis datanya berkali-kali untuk setiap himpunan. kombinasi barang. Ini mencegah sejumlah besar waktu yang diperlukan untuk memelihara database. Selain itu, untuk mendapatkan kombinasi faktor database, kita harus melakukan analisis keranjang pasar ke-21 menggunakan algoritma

apriori dan kandidat FP Growth Erwin.

2.1.9 Definisi Data Mart

Data mart adalah susunan data yang digunakan dengan tujuan memenuhi keperluan tugas analitik individu atau kelompok, seperti melakukan analisis data untuk satu departemen atau melakukan operasi bisnis. Sebuah data center dapat melancarkan proses bisnis, seperti proses negosiasi, dimana hanya satu data center yang dapat melancarkan proses negosiasi (Srigunting, 2012).

2.1.10 Perbedaan Data Mart dengan Data Warehouse

Sesuai poniah (2019). Perbedaan yang paling menonjol adalah bahwa master data mart akan menggabungkan data mereka dengan sumber lain untuk memberikan hasil gudang data. Ini adalah kolaborasi antara pasar data dan gudang data. Adithama (2019) mencantumkan beberapa faktor lain yang mungkin membantu dengan data mart dan ekspansi data sebagai berikut:

1. Lingkup

Data lingkup gudang memiliki beberapa domain dan sering diimplementasikan dan dikelola oleh departemen kantor pusat seperti unit bisnis TI. Sering disebut sebagai perusahaan atau gudang data pusat. Sebaliknya, gudang data seringkali hanya berlaku untuk unit atau divisi yang relevan dari perusahaan tertentu; itu tidak menangkap semua informasi bisnis seperti gudang data.

2. Subjek

Single gudang data merupakan bentuk departemen data yang diperuntukan untuk lini bisnis (LOB).

3. Kirim Data

Data mart menarik data dari sistem yang lebih sedikit tetapi lebih besar, sedangkan data gudang menarik data dari sistem yang lebih sedikit tetapi lebih besar.

4. Ukuran Data

Menurut analisis ini, data mart kurang dapat terlihat perbedannya dari gudang data berdasarkan akurasinya, tetapi lebih pada penggunaan dan optimalisasinya. Definisi gudang yang paling menonjol adalah "Satu gudang lebih besar dari kerangka waktu cadangan."

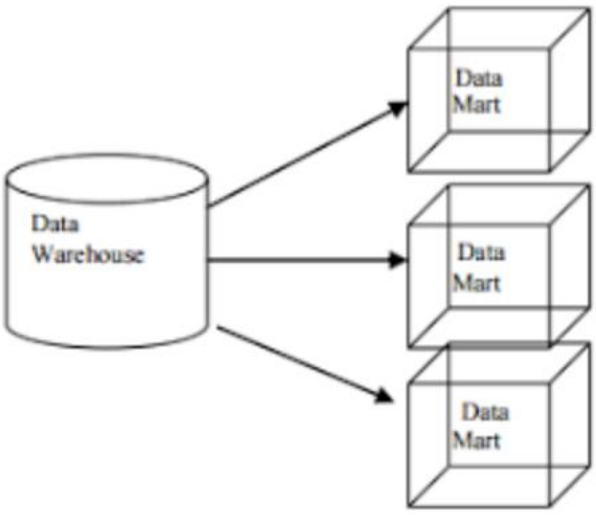
5. Waktu Implementasi

Data mart biasanya lebih mudah dibuat dan digunakan daripada gudang data karena lebih terorganisir secara berurutan. Data mart juga dapat digunakan sebagai dokumen "pembuktian konsensus" untuk pengembangan gudang data di seluruh perusahaan..

Ada dua pertimbangan utama dalam desain data-mart, dilansir dari Chhabra & Pahwa (2018), yaitu :

1. Mengandalkan data mart

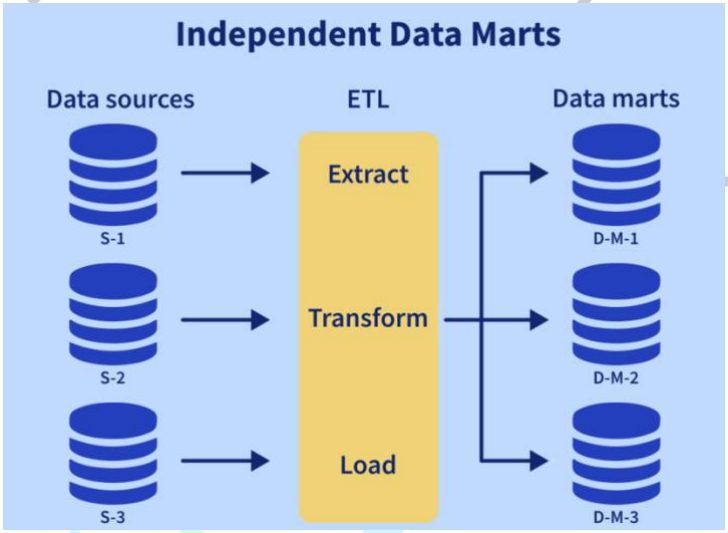
Ekspansi sebuah data dependen merupakan proses fisik atau logistik, yang merupakan bagian dari ekspansi data yang lebih besar. Dalam pendekatan ini, data mart disebut sebagai sumber data. Dalam pendekatan ini, item pertama adalah kumpulan data yang dibuat dari sejumlah kumpulan data serupa. Data warehouse yang bersangkutan terhubung ke data warehouse dan menerima data yang dibutuhkan dari data warehouse. Menurut metodologi ini, gudang data dibuat dari gudang data, yang berarti bahwa integrasi top-down tidak diperlukan. Akibatnya, ketergantungan penyimpanan data dapat dipenuhi,seperti yang dapat dijelaskan pada **Gambar 2.6** di bawah ini.



Gambar 2. 6 Gambar Dependt Data Mart

2. Data mart terpisah

Tahap kedua merupakan data independen. Dalam pendekatan ini, data independen pertama kali dibuat, kemudian data independen dibuat dari sejumlah ekspansi data independen. Karena fakta bahwa setiap gudang data beroperasi secara independen berdasarkan ketentuan perjanjian ini, integrasi gudang data diperlukan. Pendekatan ini juga dikenal sebagai pendekatan bottom-up dengan data warehouse terintegrasi untuk mengelola data warehouse; gudang data independen juga dimungkinkan dapat digambarkan seperti yang dapat dijelaskan pada **Gambar 2.7** di bawah ini.



Gambar 2. 7 Gambar Independent Data Marts

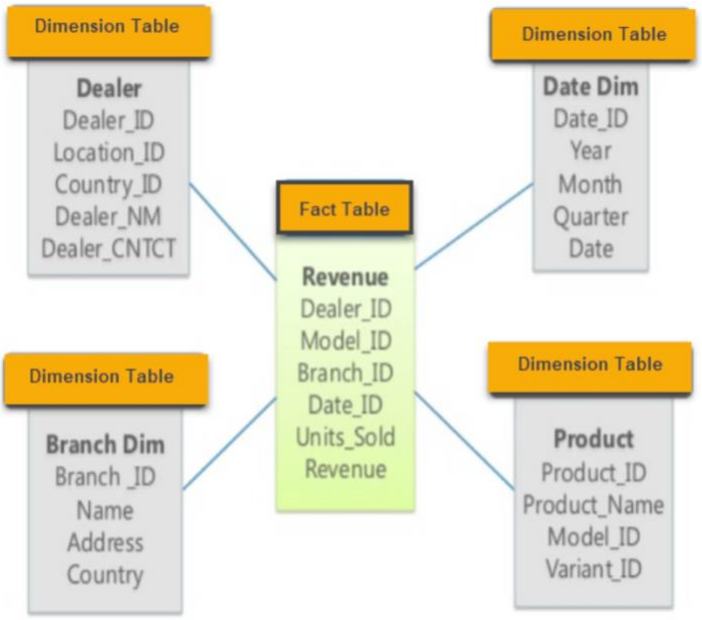
2.1.11 Extract-Transform-Load (ETL)

Dalam situasi ini, terapkan proses ETL untuk membuat gudang data. Data dari beberapa sumber dikumpulkan oleh penulis, yang diikuti dengan konsolidasi data tersebut menjadi satu sumber dan selanjutnya data tersebut dipindahkan ke penyimpanan lain. Selanjutnya, penulis akan menjelaskan bagaimana proses ETL berfungsi sebagai subset dari business intelligence. ETL adalah komponen intelijen bisnis yang memerlukan pengumpulan data dari banyak sumber, menerapkan kesalahan, mengubahnya menjadi format tunggal, dan menyimpannya di gudang data (Hocevar & Jaklic, 2010). Proses ETL membutuhkan tabel ETL dan proses ETL dimana tabel ETL terintegrasi dengan tabel OLTP dan data warehouse, sedangkan proses ETL mengubah

data dari database OLTP menjadi data warehouse berdasarkan tabel ETL (Warnars, 2009). Saraswati (2011) merinci proses ETL untuk setiap item berikut: A. Contohnya adalah proses mengekstrak data dari database; proses ini hanya mengekstrak data yang sudah usang daripada keseluruhan database yang aktif. B. Konversi database memerlukan perubahan struktur ke format standar untuk mengakui bahwa data yang diperoleh dari tipe pelanggan yang berbeda mungkin memiliki standar yang berbeda. Untuk memudahkan pelaporan, diperlukan standarisasi. Dibandingkan dengan load sendiri, transfer data merupakan proses yang sudah terjadi hingga tahap akhir (data warehouse). Nama harus konsisten selama konversi, dan skema pengkodean yang digunakan harus konsisten. Proses ETL untuk membangun data warehouse saat ini dimulai dengan pengumpulan data dari area penyimpanan data warehouse dan diakhiri dengan persiapan data melalui tipe data, konten, dan ukuran tipe lebar. Data yang telah dikonversi kemudian dikembangkan menjadi database.

2.1.12 Data Warehouse Modeling

Skema bintang adalah templat yang paling menonjol dan paling panjang di antara skema data-mart. Metode ini sering digunakan untuk mengembangkan atau membangun gudang data dan dimensional data mart. Tabel fakta ini memiliki satu atau lebih tabel dengan banyak dimensi. Skema bintang adalah gejala terpenting dari skema kepingan salju. Ini juga efektif dalam menghapus dasbor. Bintang disebut demikian karena representasi fisik dari suatu benda mencakup jenis bintang dengan tabel faktual di bagian atas dan tabel dimensi di bawahnya yang menampilkan titik-titik benda tersebut. Skema bintang memiliki satu tabel tengah, yang merupakan tabel fakta kabel yang menghubungkan ke beberapa dimensi tabel lainnya, seperti yang tercantum seperti yang ditunjukkan pada **Gambar 2.8** dibawah ini.



Gambar 2. 8 Gambar Star Schema Modeling

Salah satu karakteristik dari skema bintang adalah bahwa :

1. Setiap dimensi dalam skema bintang hanya diwakili oleh satu tabel dimensi.
2. Dimensi tabel harus memiliki sekumpulan atribut.
3. Bidang tabel menggunakan kunci asing untuk "bergabung" dengan tabel fakta.

2.1.13 Python Programming Languange

Python adalah bahasa untuk program komputer yang menyenangkan, logis, dan kreatif, yang dikembangkan oleh Python Software Foundation (2016). Python memiliki struktur data dinamis, pengetikan dinamis, dan pengkodean dinamis. Python memiliki sintaks yang mudah dipelajari yang mencegah kecurangan dan mengurangi biaya pemeliharaan perangkat lunak. Python menggunakan modul dan paket untuk mendorong penggunaan program dan kode berkelanjutan secara modular. Python dan kompiler standar sudah tersedia untuk platform apa pun dan dapat didistribusikan dengan cepat. Bahasa ini diciptakan pada tahun 1992 oleh Guido van Rossum dari Belgia.

Beberapa modul yang digunakan dalam machine learning antara lain sebagai berikut:

1. "Pandas adalah modul Python yang menyediakan struktur data yang cepat, fleksibel, dan ekspresif yang dirancang untuk membuat bekerja dengan data relasional atau berlabel menjadi lebih mudah dan lebih intuitif" (Pypi.org, 2018).
2. "Python Numerik," atau Numpy, adalah satu-satunya pustaka Python yang dirancang khusus untuk melakukan proses komputasi numerik. Sebaliknya, Array adalah kumpulan variabel dengan tipe data yang sama. Numpy menampilkan data dalam format array.

1.2 Tinjauan Studi

Di bawah ini dapat dikemukakan bahwa dalam Tinjauan Pustaka akan dicantumkan batas-batas pemeriksaan tomahawk yang akan digunakan dengan suatu tambah kop atau suatu tatanan yang akan dibangun menjelang tomahawk yang akan digunakan.

Nur Fitrianti Fahrudin (2019) dalam penelitian "Penerapan Algoritma Apriori untuk Analisis Keranjang Pasar". Analisis asosiatif menggunakan algoritme Apriori dapat membantu orang dengan kebutuhan khusus, seperti manajer toko, membuat rencana bisnis yang akan membantu mereka menjual barang di toko mereka. Setelah melakukan analisis asosiasi, maka perlu dilakukan lift ratio perhitungan untuk menentukan kualitas akhir dari asosiasi untuk menentukan asosiasi yang akan datang. Beberapa aturan memiliki threshold rate yang tinggi berdasarkan hasil pengujian yang menunjukkan penurunan lift rate, yaitu: mentega tidak sama dengan telur dengan menggunakan algoritma Apriori untuk menganalisa keranjang belanja. Dalam artikelnya yang berjudul "Analisis Keranjang Pasar Pada Mini Market Ayu dengan Apriori,"

Erin Elisa (2018). Penerapan algoritma apriori dalam teknik data mining sangat dan dapat mempercepat proses pembentukan tren yang efektif menggabungkan sampel produk dari penjualan barang kebutuhan pokok rumah

tangga ke mini market Ayu Tembesi di Batam dengan tingkat kepercayaan dan dukungan tertinggi adalah minyak dan susu dengan dukungan Dengan mengamati pembelian konsumen yang dilakukan berdasarkan kombinasi dua item tertentu, dimungkinkan untuk mengidentifikasi bias dalam pembelian. Pengetahuan baru dapat diperoleh berdasarkan hasil pengujian algoritma apriori, dan sistem yang baru dikembangkan dapat dilepaskan dari susunan barang untuk memfasilitasi pergerakan objek yang relevan. Dalam makalah 2018 berjudul "Implementasi Metode Association Rule Mining Dengan Algoritma Apriori Untuk Rekomendasi Promo Barang".

Andreas Aditya Christyan Putra, Hanny Haryanto, dan Erlin Dolphina. Dari pendataan mereka yang melibatkan 30 transaksi data dengan tingkat dukungan minimal 3 atau tingkat kepercayaan minimal 70%, hasilnya sebagai berikut. Hasil dari proses asosiatif cukup menjanjikan. Jika konsumen membeli Produk 3, konsumen juga akan membeli Produk 33 dengan diskon besar sebesar 80%. Jika seorang pelanggan membeli produk (9), mereka juga akan membeli produk (51) dengan faktor kepercayaan 75 persen. Jika pelanggan membeli produk, mereka juga akan membeli produk 9 dengan tingkat kepercayaan 9,75%. Menurut informasi yang diperoleh dari survei, konsumen secara konsisten membeli dua produk secara bersamaan, dengan tingkat diskonto rata-rata 80% antara satu produk dengan produk lainnya. Hasilnya, informasi yang diperoleh dapat digunakan untuk membuat produk promosi dengan menggunakan berbagai kombinasi produk. cocok untuk model transaksi.

"Aplikasi Data Mining Menggunakan Algoritma Apriori Untuk Analisis Penjualan Di XYZ Helm" adalah judul bagian Setyawan (2016) dalam disertasinya. Penelitian ini digunakan berpunca ganjaran penjualan alert yang didedikasikan untuk mengabdikan sejumlah 798 detil penjualan yang nanti terjamah mengabdikan qoth tercantum menetapkan dukungan etos dukungan dan keyakinan, bukan jumlah 126. Hasil penelitian menunjukkan bahwa empat dari setiap lima batang memiliki hari dengan setidaknya 40% dukungan dan kepercayaan 50%. Dimungkinkan untuk menggunakan barang yang tidak memenuhi persyaratan minimal, seperti lemburan berpunca pembelian. Karena tingginya tingkat tindakan bagian dalam proses

negosiasi, bar yang memegang etos support tinggi mungkin banyak ditemukan.

Yanto dan Khoiriyah (2015) menerbitkan makalah berjudul “Implementasi Data Mining Dengan Algoritma A Priori Untuk Pembelian Obat (Studi Kasus: Apotek Musi Rawas)”. hasilnya seperti gambar di bawah ini. hasil 2 asosiasi dengan support nilai minimum 50% dan nilai kepercayaan nilai minimal 77% untuk mendapatkan regimen jika membeli Amoxcilin, kemudian membeli Asamfenamat, dan jika membeli Cefadroxil, membeli Sanmol. Baik algoritme apriori manusia maupun algoritme berbasis sistem yang digunakan untuk permutasi menghasilkan hasil yang identik.

"Menentukan Pola Kecelakaan Lalu Lintas Menggunakan Metode Assosiation Rule Dengan Algoritma Apriori (Studi Kasus: Tingkat Kecelakaan Jalan) Tol Kabupaten Sleman" adalah judul penelitian oleh Hakim dan Fauzy dari tahun 2015. Penelitian ini menghasilkan 5 aturan asosiasi dengan nilai dukungan 20% dan nilai kepercayaan 90%, tetapi dilakukan 4 kali lipat dengan nilai dukungan yang sama untuk mendapatkan 1 aturan kesimpulan. Menurut SIM yang ada, jenis pekerjaan laki-laki dan swasta dibedakan dengan memiliki ambang batas cedar yang lebih tinggi daripada jenis ringan.

Dalam artikel 2016 mereka, "Implementasi Algoritma Apriori untuk Analisis Penjualan Berbasis Web," Nursikuwagus dan Hartono menyebutkan hal ini. Memahami produk khusus dapat membantu peneliti memperkuat undang-undang afiliasi. Aturan Asosiasi didasarkan pada penggunaan elemen himpunan dalam setiap transaksi. Karena itu, hasil yang diperoleh dapat digunakan untuk membantu pihak yang mengeluarkan keputusan tersebut.