

ANALISIS DATA MINING DAN WAREHOUSING

Rutman Lumbantoruan
Posma Sariguna Johnson Kennedy

Fakultas Ekonomi, Universitas Kristen Indonesia,
Jakarta, Indonesia

Abstrak

Data mining adalah salah satu langkah dalam proses *Knowledge Discovery in Database (KDD)*, suatu pengakuan proses pencarian. *Data mining* menggunakan algoritma tertentu untuk mencari pola dalam database. *Machine Learning* adalah area dalam kecerdasan buatan yang berhubungan dengan pengembangan teknik yang dapat diprogram dan belajar dari data masa lalu. Pengenalan pola, *data mining* dan *machine learning* sering digunakan untuk merujuk pada hal yang sama. Bidang ini mengambil probabilitas dan statistik, kadang-kadang optimasi. *Machine Learning* menjadi alat analisis dalam data mining.

1. Pendahuluan

Proses yang disebut Knowledge Discovery in Database (KDD). Data mining adalah salah satu dari langkah-langkah dalam proses KDD. Data mining menggunakan algoritma khusus untuk pola pencarian dalam database. Tulisan ini akan menjelaskan proses penambangan data Selanjutnya, ini akan digunakan untuk mengembangkan sistem pakar. Proses menggunakan RapidMiner 4.0 software dengan pohon keputusan. Hasil dari proses data mining menggunakan metode pohon keputusan menunjukkan bahwa data adalah konvergen dalam jenis tindakan. Peningkatan jumlah informasi atau data yang disimpan dalam format elektronik. Akumulasi data sangat besar diperkirakan bahwa jumlah informasi di dunia setiap 20 bulan dan ukuran serta jumlah database meningkat lebih cepat, data center, baik yang ada *di could computing*.

Selain database relasional, lingkungan data warehouse meliputi ekstraksi, transportasi, transformasi, sebuah pengolahan analisis online (OLAP) mesin, alat analisis klien, dan aplikasi lain yang mengelola proses pengumpulan data dan memberikan kepada pengguna bisnis. ETL Alat dimaksudkan untuk mengekstrak, mengubah dan memuat data ke Data Warehouse untuk keputusan-keputusan. Dilakukan secara manual dengan menggunakan kode SQL yang dibuat oleh programmer. Tugas ini membosankan dalam banyak kasus karena terlibat banyak sumber daya, coding kompleks dan jam kerja yang lebih. Di atas semua itu, menjaga kode ditempatkan tantangan besar di antara programmer. Kesulitan-kesulitan ini dieliminasi suatu alat karena mereka sangat kuat dan mereka menawarkan banyak keuntungan di semua tahapan proses mulai dari ekstraksi, pembersihan data, data yang profil, transformasi, debugging dan loading ke data warehouse jika dibandingkan dengan metode lama.

Ada sejumlah alat yang tersedia di pasar untuk melakukan proses data sesuai untuk bisnis / persyaratan teknis. Berikut adalah beberapa dari Alat Nama Nama perusahaan Informatica Informatica Perusahaan DT / Studio Embarcadero Teknologi ,Data Tahap IBM,Ab initio Ab initio Software Corporation ,Data Junction Pervasive Software,Oracle Warehouse Builder Oracle Corporation ,Microsoft SQL Server Integration Microsoft,Transform On Demand Solonde Solunde Solusi ETL transformasi Manajer. Sebuah cara yang umum untuk memperkenalkan data warehouse adalah untuk merujuk pada karakteristik data

Peningkatan jumlah informasi atau data yang disimpan dalam format elektronik. Akumulasi data sangat besar diperkirakan bahwa jumlah informasi di dunia setiap 20 bulan dan ukuran serta jumlah database meningkat lebih cepat. Data penjualan data ritel, kebijakan dan klaim asuransi, data yang riwayat medis dalam kesehatan, data universitas, data perawatan, data keuangan di bidang perbankan dan sekuritas, data pertanian, data kelautan, data pertambangan merupakan contoh dari jenis data yang volume datanya yang sudah maksimal sudah terkumpul di pusat data center baik yang ada *di cloud computing*.

Berdasarkan latar belakang masalah di atas, rumusan masalah yang di ajukan dalam tulisan ini adalah apakah dan informasi yang tersimpan dalam format elektronik dapat memberikan kualitas pengambilan keputusan, serta bagaimana efisiensi data yang ada di data center.

2. Tinjauan Pustaka

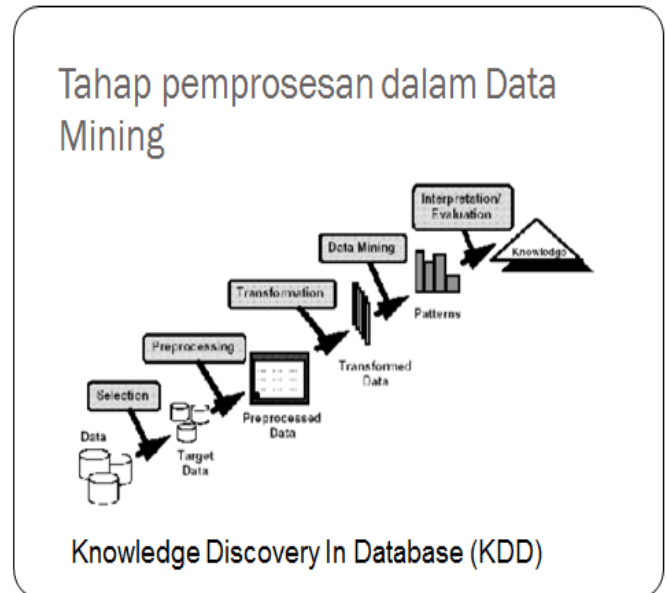
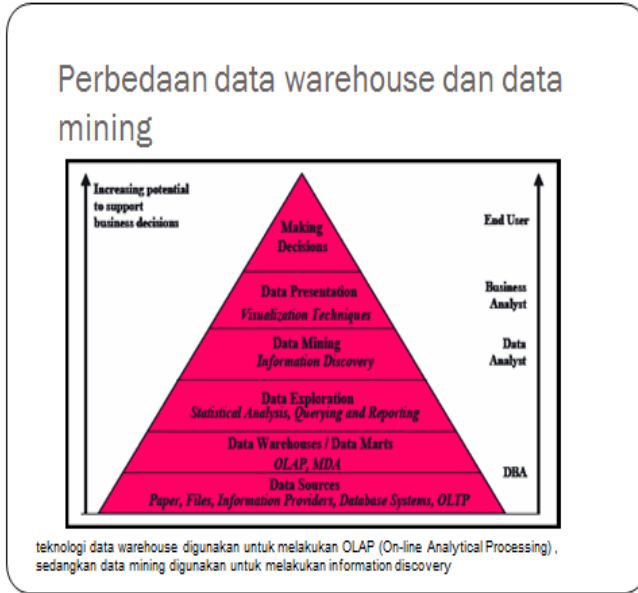
Karakteristik data mining sebagai berikut

Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan database. Beberapa metode yang sering disebut-sebut dalam literatur data mining antara lain clustering, lassification, association rules mining, neural network, genetic algorithm dan lain-lain (Pramudiono, 2007).

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 (Han, 2006). 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 (Witten, 2005).

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).



Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohon kredit, image atau signal atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya (Santoso, 2007). Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition jarang digunakan karena termasuk bagian dari data mining (Santoso, 2007).

Machine Learning adalah suatu area dalam artificial intelligence atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, data mining dan machine learning sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi. Machine learning menjadi alat analisis dalam data mining. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam gambar 2.4 (Santoso, 2007).

Penyimpanan data menjadi lebih mudah karena ketersediaan sejumlah besar daya komputasi dengan biaya rendah, biaya tenaga pemrosesan dan penyimpanan yang sudah lebih baik kualitas dan kuantitas membuat data yang murah. Ada juga pengenalan metode pembelajaran mesin baru untuk representasi pengetahuan berdasarkan logika pemrograman, selain analisis statistik tradisional data. Metode baru cenderung komputasi intensif maka permintaan untuk kekuatan pemrosesan lebih lanjut. Bit penyimpanan data / byte dihitung sebagai berikut:

- 1 byte = 8 bit
- 1 kilobyte (K / KB) = 2¹⁰ bytes = 1.024 bytes
- 1 megabyte(M/MB) = 2²⁰ bytes = 1.048.576 bytes
- 1 gigabyte(G/GB) = 2³⁰ bytes = 1.073.741.824 bytes
- 1 terabyte(T/TB) = 2⁴⁰ bytes = 1.099.511.627.776 bytes

1 petabyte(P/PB)=	2^{50}	bytes	=1.125.899.906.842.624	bytes	
1 exabyte(E/EB)	=	2^{60}	bytes	=1.152.921.504.606.846.976	byte
1 zettabyte(Z/ZB)	=	1 000 000 000 000 000 000 000			bytes
1 yottabyte(Y/YB)	=	1 000 000 000 000 000 000 000 000			bytes

Informasi adalah jantung dari operasi bisnis dan pengambil keputusan bisa memanfaatkan data yang disimpan untuk mendapatkan pemahaman yang berharga tentang bisnis. Database Sistem Manajemen memberikan akses ke data yang tersimpan tapi ini hanya sebagian kecil dari apa yang bisa diperoleh dari data. Sistem tradisional on-line proses transaksi, baik data yang masuk ke database dengan cepat, aman dan efisien tapi tidak dimanfaatkan dengan baik, tidak memberikan dan bermakna. Data harus memberikan pengetahuan lebih lanjut tentang bisnis dengan melampaui data secara eksplisit disimpan untuk menurunkan pengetahuan tentang bisnis. Data Mining, juga disebut sebagai data arkeologi, Data center, data yang permanen yang artinya sumber proses penggalian pengetahuan tersembunyi dan volume data mentah yang volume yang besar dan dapat dipergunakan untuk membuat keputusan bisnis penting. Di sinilah Data Mining atau Knowledge Discovery in Database (KDD) memiliki manfaat yang jelas bagi perusahaan apapun.

3. Model Penelitian

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 probabilistik (Santoso, 2007). 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 asosiatif 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 asosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence) (Pramudiono, 2007).

Ada beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu algoritma klasik yang sering dipakai yaitu algoritma apriori. Ide dasar dari algoritma ini adalah dengan mengembangkan frequent itemset. Dengan menggunakan satu item dan secara rekursif mengembangkan frequent itemset dengan dua item, tiga item dan seterusnya hingga frequent itemset dengan semua ukuran. Untuk mengembangkan frequent set dengan dua item, dapat menggunakan frequent set item. Alasannya adalah bila set satu item tidak melebihi support minimum, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi support minimum tersebut. Secara umum, mengembangkan set

dengan fc-item menggunakan frequent set dengan k – 1 item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi database.

Dalam asosiasi terdapat istilah antecedent dan consequent, antecedent untuk mewakili bagian “jika” dan consequent untuk mewakili bagian “maka”. Dalam analisis ini, antecedent dan consequent adalah sekelompok item yang tidak punya hubungan secara bersama (Santoso, 2007). Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturan-aturan yang cukup kuat tingkat ketergantungan antar item dalam antecedent dan consequent. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran support dan confidence. Support adalah rasio antara jumlah transaksi yang memuat antecedent dan consequent dengan jumlah transaksi. Confidence adalah rasio antara jumlah transaksi yang meliputi semua item dalam antecedent dan consequent dengan jumlah transaksi yang meliputi semua item dalam *antecedent*.

$$S = \frac{\Sigma(Ta+Tc)}{\Sigma(T)} \dots\dots\dots(2.1)$$

Keterangan :

- S* = *Support*
- $\Sigma(Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequencent*
- $\Sigma(T)$ = Jumlah transaksi

$$C = \frac{\Sigma(Ta+Tc)}{\Sigma(Ta)} \dots\dots\dots(2.2)$$

Keterangan :

- C* = *Confidence*
- $\Sigma(Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequencent*
- $\Sigma(Ta)$ = Jumlah transaksi yang mengandung *antecedent*

Langkah pertama algoritma apriori adalah, support dari setiap item dihitung dengan men-scan database. Setelah support dari setiap item didapat, item yang memiliki support lebih besar dari minimum support dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item. Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap

kandidat 2-itemset ini dihitung support-nya dengan men-scan database. Support artinya jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-itemset. Setelah support dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum support dapat

ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2.(Pramudiono, 2007).

Untuk selanjutnya iterasi iterasi ke-k dapat dibagi lagi menjadi beberapa bagian :

1. Pembentukan kandidat itemset. Kandidat k-itemset dibentuk dari kombinasi (k-1)-itemset yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma apriori adalah adanya pemangkasan kandidat k-itemset yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Penghitungan support dari tiap kandidat k-itemset. Support dari tiap kandidat k-itemset didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tersebut. Ini adalah juga ciri dari algoritma apriori yaitu diperlukan penghitungan dengan scan seluruh database sebanyak k-itemset terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k item atau k-itemset ditetapkan dari kandidat k-itemset yang support-nya lebih besar dari minimum support. Kemudian dihitung confidence masing-masing kombinasi item. Iterasi berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi. (Pramudiono, 2007)

Secara ringkas algoritma apriori sebagai berikut :

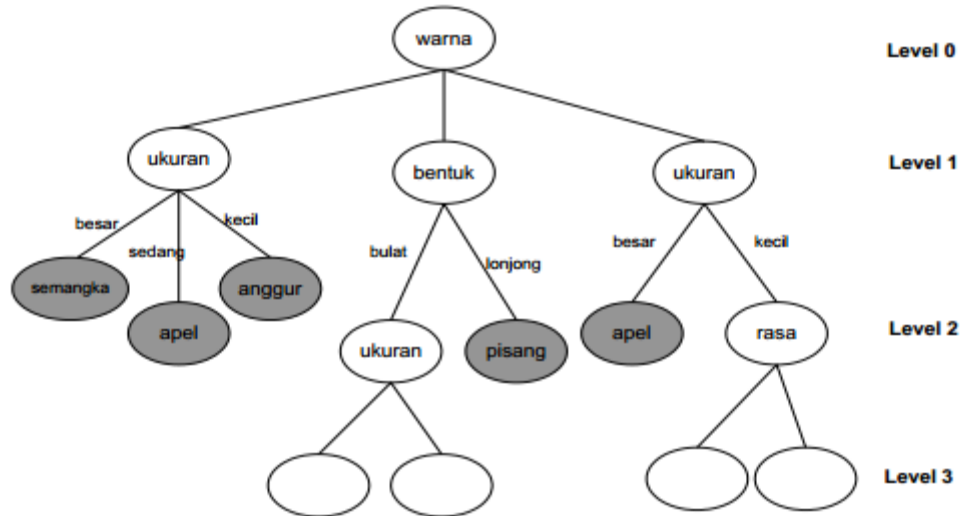
```
Create L1 = set of supported itemsets of cardinality one  
Set k to 2  
while (Lk-1 ≠ ∅) {  
  Create Ck from Lk-1  
  Prune all the itemsets in Ck that are not  
  supported, to create Lk  
Increase k by 1  
}  
The set of all supported itemsets is L1 ∪ L2 ∪ ... ∪ Lk
```

Selain algoritma apriori, terdapat juga algoritma lain seperti FP-Grwoth. Perbedaan algoritma apriori dengan FP-Growth pada banyaknya scan database. Algoritma apriori melakukan scan database setiap kali iterasi sedangkan algoritma FP-Growth hanya melakukan sekali di awal (Bramer, 2007).

Decision Tree

Dalam decision tree tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Seperti yang diilustrasikan pada gambar 2.6, 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 (Santoso, 2007).



Gambar 2. 6 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 (Santoso, 2007). 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 dan nilainya
2. Label/keluaran data biasanya bernilai diskrit
3. Data mempunyai missing value (nilai dari suatu atribut tidak diketahui)

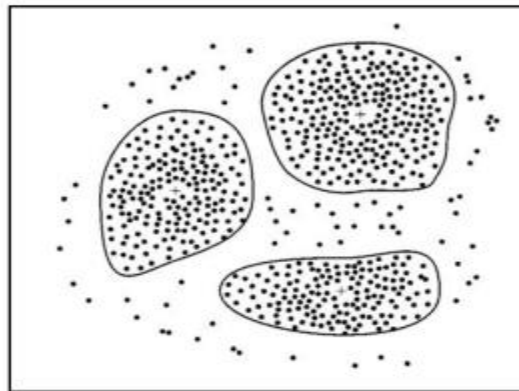
Dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Untuk membuat decision tree perlu memperhatikan hal-hal berikut ini :

1. Atribut mana yang akan dipilih untuk pemisahan obyek
2. Urutan atribut mana yang akan dipilih terlebih dahulu
3. Struktur tree
4. Kriteria pemberhentian
5. Pruning

(Santoso, 2007)

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 2.7 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 2.7. 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-cluster yang lain. Dalam metode ini tidak diketahui sebelumnya berapa jumlah cluster dan bagaimana pengelompokannya (Santoso, 2007).

Software Aplikasi

Software aplikasi terdiri atas program yang berdiri sendiri yang mampu mengatasi kebutuhan bisnis tertentu. Aplikasi memfasilitasi operasi bisnis atau pengambilan keputusan manajemen maupun teknik sebagai tambahan dalam aplikasi pemrosesan data konvensional. Software aplikasi digunakan untuk mengatur fungsi bisnis secara real time (Pressman, 2005).

Desain Model Aplikasi




Desain model dari aplikasi terdiri dari physical model dan logical model. Physical model dapat digambarkan dengan bagan alir sistem. Logical model dalam sistem informasi lebih menjelaskan kepada pengguna bagaimana nantinya fungsi-fungsi di sistem informasi secara logika akan bekerja.

Logical model dapat digambarkan dengan DFD (*Data Flow Diagram*) dan kamus data (*Data Dictionary*). Adapun penjelasan dari alat bantu dalam desain model adalah sebagai berikut :

1. Diagram Konteks (*Context Diagram*)

Diagram konteks adalah sebuah diagram sederhana yang menggambarkan hubungan antara proses dan entitas luarnya. Adapun simbol-simbol dalam diagram konteks seperti dijelaskan pada tabel 2.2.


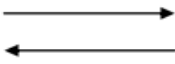


Tabel 2.2 Simbol-Simbol *Context Diagram*

Simbol	Keterangan
	Proses , menunjukkan suatu proses untuk menerima masukan dan menghasilkan keluaran.
	Entitas luar , merupakan sumber atau tujuan dari aliran data dari atau ke sistem. Entitas luar merupakan lingkungan luar sistem
	Arus data atau aliran data , yaitu komponen yang menggambarkan aliran data dari satu proses ke proses lainnya

2. DFD (*Data Flow Diagram*)

DFD merupakan suatu model logika yang menggambarkan asal data dan tujuan data yang keluar dari sistem, serta menggambarkan penyimpanan data dan proses yang mentransformasikan data. DFD menunjukkan hubungan antara data pada sistem dan proses pada sistem. Beberapa simbol yang digunakan dalam DFD diterangkan pada tabel 2.3

Tabel 2.3 Simbol-Simbol DFD

Simbol	Keterangan
	Proses yang berfungsi untuk menunjukkan transformasi dari masukan menjadi keluaran
	Arus data atau aliran data , yaitu komponen yang menggambarkan aliran data dari satu proses ke proses lainnya
	Tempat penyimpanan , yaitu komponen yang digunakan untuk menyimpan kumpulan data, penyimpanan data bisa berupa <i>file</i> , <i>database</i> , maupun bagian dari <i>record</i>
	Entitas luar , merupakan sumber atau tujuan dari aliran data dari atau ke sistem. Entitas luar merupakan lingkungan luar sistem.

Perancangan Perangkat Lunak

Proses perancangan sistem membagi persyaratan dalam sistem perangkat keras atau perangkat lunak. Kegiatan ini menentukan arsitektur sistem secara keseluruhan. Perancangan perangkat lunak melibatkan identifikasi dan deskripsi abstraksi sistem perangkat lunak yang mendasar dan hubungannya (Sommerville, 2003). Sebagaimana persyaratan, desain didokumentasikan dan menjadi bagian dari konfigurasi software (Pressman, 1997). Tahap desain meliputi perancangan data, perancangan fungsional, dan perancangan antarmuka.

1. Perancangan data. Perancangan data mentransformasikan model data yang dihasilkan oleh proses analisis menjadi struktur data yang dibutuhkan pada saat pembuatan program (coding). Selain itu juga akan dilakukan desain terhadap struktur database yang akan dipakai.
2. Perancangan fungsional. Perancangan fungsional mendeskripsikan kebutuhan fungsi-fungsi utama perangkat lunak.
3. Perancangan antarmuka. Perancangan antarmuka mendefinisikan bagaimana pengguna (user) dan perangkat lunak berkomunikasi dalam menjalankan fungsionalitas perangkat lunak.

Analisis dan Pengujian Unit

Pada tahap ini, perancangan perangkat lunak direalisasikan sebagai serangkaian program atau unit program. Kemudian pengujian unit melibatkan verifikasi bahwa setiap unit program telah memenuhi spesifikasinya (Sommerville, 2003). Program sebaiknya dirilis setelah dikembangkan, diuji untuk memperbaiki kesalahan yang ditemukan pada pengujian untuk menjamin kualitasnya (Padmini, 2005). Terdapat dua metode pengujian yaitu :

- 1) Metode white box yaitu pengujian yang berfokus pada logika internal software (source code program).

- 2) Metode black box yaitu mengarahkan pengujian untuk menemukan kesalahan-mesalahan dan memastikan bahwa input yang dibatasi akan memberikan hasil aktual yang sesuai dengan hasil yang dibutuhkan. Pada tahap pengujian, penulis melakukan metode black box yaitu menguji fungsionalitas dari perangkat lunak saja tanpa harus mengetahui struktur internal program (source code).

4. Kesimpulan

Kesimpulan yang dapat diambil dalam pengerjaan aplikasi, *Data Mining* ini dapat digunakan untuk menampilkan informasi hasil dari analisis data. Informasi yang ditampilkan berupa nilai *support* dan *confidence* hubungan antara database. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut. Hasil dari proses *data mining* ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih.

Untuk pengembangan Aplikasi *Data Mining* lebih lanjut, dapat menggunakan algoritma lain, misal algoritma *FP-Growth*. Perbedaannya adalah algoritma *apriori* harus melakukan *scan database* setiap kali iterasi, sedangkan algoritma *FP-Growth* hanya melakukan satu kali *scan database* diawal

DAFTAR PUSTAKA

- Siagian P. Sondang, Sistem Informasi Manajemen, 2014 – Bumi Aksara, Jakarta
- Jogiyanto Hartono, Sistem Teknologi Informasi Bisnis, 2013, Salemba 4 , Jakarta
- Bramer, Max, 2007, “*Principles of Data Mining*”, Springer, London.
- Chintakayala, Padmini. 2005. “*Beginners Guide for Software Testing : Symbiosys Technologies*”.
- Davies, and Paul Beynon, 2004, “*Database Systems Third Edition*”, Palgrave Macmillan, New York.
- Elmasri, Ramez and Shamkant B. Navathe, 2000, “*Fundamentals of Database Systems. Third Edition*”, Addison Wesley Publishing Company, New York.
- Han, J. and Kamber, M, 2006, “*Data Mining Concepts and Techniques Second Edition*”. Morgan Kauffman, San Francisco. Kadir, Abdul, 1999, “*Konsep dan Tuntunan Praktis*
- <http://www.ijdc.net/index.php/ijdc/article/viewFile/110/87>
- http://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBoQFjAA&url=http%3A%2F%2Flib.ugm.ac.id%2Fdigitasi%2Fupload%2F2227_Sri%2520Mulyana.pdf&ei=XnMeVIqgKsK2uATEjIDwCg&usg=AFQjCNE-PqDcWEHd5idTZ4lviiGUmbK4qw