

BAB II LANDASAN TEORI

2.1 Pengertian *Data mining*

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar. [3]

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* [6] *network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*.

“*data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.” [8]. “*Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.”[8]

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. [9]

2.2 Tahap-Tahap *Data mining*

Istilah *data mining* dan *knowledge discovery in databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut : [6]

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation/ Evaluation

Pola informasi yang dihasilkan dari proses *data mining*, perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.3 Teknik-teknik *Data mining*

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

a. *Association Rule Mining*

Association rules (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *market basket analysis*.

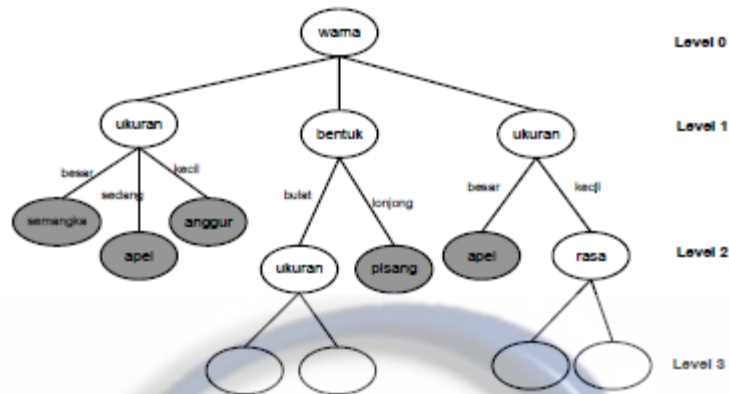
Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “if-then” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistic. [5]

Analisis asosiasi dikenal juga sebagai salah satu metode *data mining* yang menjadi dasar dari berbagai metode *data mining* lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut. dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan syarat minimum untuk confidence (*minimum confidence*). [2]

b. Classification

Dalam klasifikasi, terdapat target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. [1]

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, 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. [5]



Gambar 2.1 : Decision Tree [5]

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. [5]

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 :

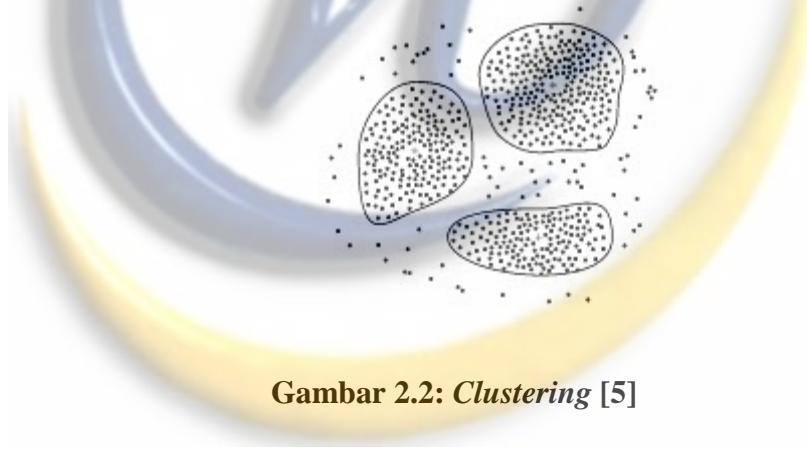
1. Atribut mana yang akan dipilih untuk pemisahan obyek
2. Urutan atribut mana yang akan dipilih terlebih dahulu
3. Struktur *tree*

4. Kriteria pemberhentian

5. Pruning [5]

c. *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 2.2: Clustering [5]

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 3. 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. [5]

Berikut ini adalah 9 algoritma penggalian data yang paling populer berdasarkan konferensi ICDM '06:

1. C4.5
2. k-Means
3. SVM
4. Apriori
5. EM
6. PageRank
7. AdaBoost
8. kNN
9. Naive Bayes

2.4 Implementasi (Penerapan *Data mining*)

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

1. Analisa pasar dan manajemen.

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

2. Analisa Perusahaan dan Manajemen resiko.

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

3. Telekomunikasi.

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

4. Keuangan.

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

5. Asuransi.

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

6. Olahraga.

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

7. Astronomi.

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

8. Internet Web surf-aid

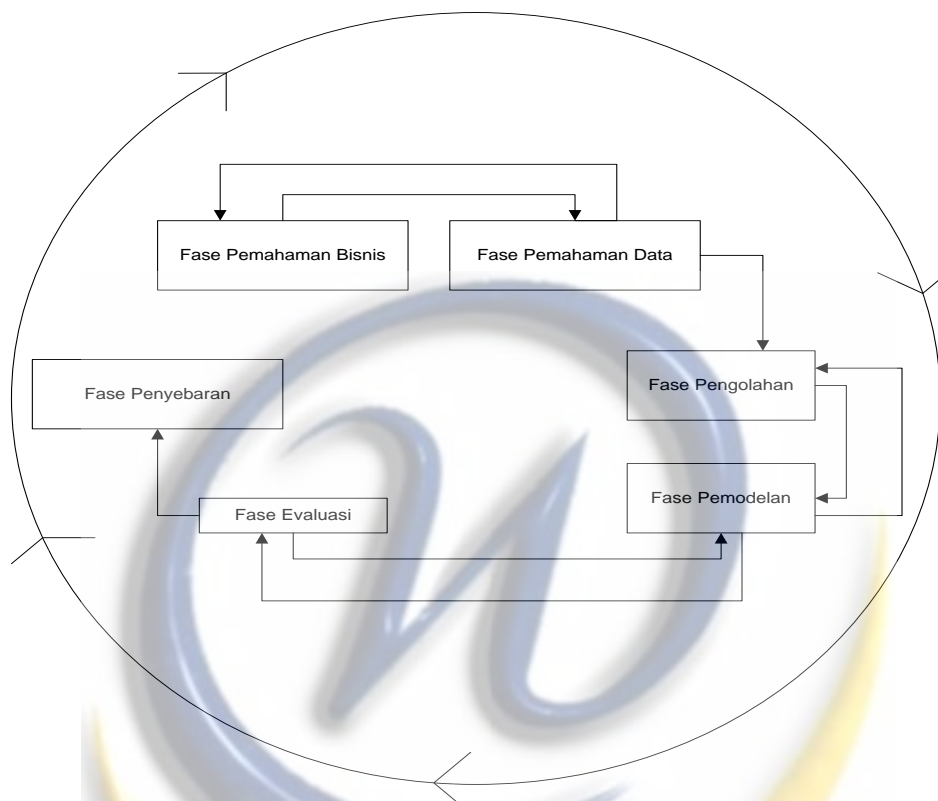
IBM Surf-Aid menggunakan algoritma *data mining* untuk mendata akses halaman Web khususnya yang berkaitan dengan pemasaran guna melihat perilaku dan minat customer serta melihat ke-efektif-an pemasaran melalui Web.

2.5 Metode Penelitian *Data mining*

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

Dalam CRISP-DM, sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam 6 fase (Gambar 2.4). Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antarfase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase *modeling*. Berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase

evaluation.



Gambar 2.3:Proses *Data mining* Menurut CRISP-DM [1]

Gambar 6 fase CRISP-DM

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.

2. Fase Pemahaman Data (*Data Understanding Phase*)
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.

3. Fase Pengolahan Data (*Data Preparation Phase*)
 - a. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
 - b. Pilih kasus dan variable yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variable jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.

4. Fase Pemodelan (*Modeling Fase*)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Kalibrasi aturan model untuk mengoptimalkan hasil.
 - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.

5. Fase Evaluasi (*Evaluation Phase*)

- a. Mengevaluasi satu atau lebih model yang digunakan dalam *fase* pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
- b. Menetapkan apakah terdapat model yang memenuhi tujuan pada *fase* awal.
- c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
- d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.

6. Fase Penyebaran (*Deployment Phase*)

- a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
- b. Contoh sederhana penyebaran: pembuatan laporan.

Contoh kompleks penyebaran: penerapan proses *data mining* secara parallel pada department lain. [4]






2.6 Metode Perancangan Sistem Berorientasi Aliran Data

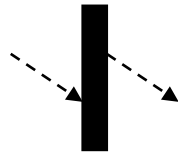
Pada saat informasi mengalir melalui perangkat lunak, data dimodifikasi oleh suatu deretan tranformasi. *Diagram aliran data* atau *Data Flow Diagram* (DFD) adalah suatu tekni grafis yang menggambarkan aliran informasi dan tranformasi yang diaplikasikan pada saat data bergerak dari input menjadi *output*. DFD dapat digunakan untuk menyajikan sebuah sistem atau perangkat lunak pada setiap tingkat abstraksi. [10]

Penulis menggunakan metode *waterfall* karena penulis dalam pengembangan sistem melakukan penganalisaan dari tahap implementasi sampai selesai kemudian melakukan pengujian serta mendengarkan keinginan pengguna. Jika ada permintaan

dari *user* maka penulis melakukan perbaikan atau penambahan pada tahapan tertentu saja (tidak perlu mengulang hingga ke tahapan pertama yaitu tahap analisis). Notasi DFD yang digunakan adalah :

Tabel 2.1 Simbol-simbol pada Diagram Alir Data (*Data Flow Diagram*)

No.	Simbol	Keterangan
1.		<i>External entity</i> (kata benda), sebuah elemen sistem (misalnya perangkat keras, seseorang, program yang lain) atau sistem yang lain yang menghasilkan informasi bagi transformasi oleh perangkat lunak atau menerima informasi yang dihasilkan oleh PL.
2.		Proses (kata kerja), menggambarkan kegiatan yang dilakukan oleh sistem, dapat berupa prosedur yang memanipulasi atau mengolah data.
3.		<i>Data Store</i> (kata benda), merupakan tempat penyimpanan data untuk direferensi atau diolah lagi lebih lanjut.
4.		<i>Data Flow</i> (kata benda), menggambarkan aliran data yang menunjukkan transportasi data atau informasi.
5.		<i>Control Flow</i> (kata benda), menggambarkan aliran kontrol data atau kejadian, dapat berupa nilai Boolean

No.	Simbol	Keterangan
6.		<i>Control Bar</i> , merupakan acuan untuk CSPEC (<i>control specification</i>) yang menjelaskan kebiasaan sistem dan mendefinisikan bagaimana proses diaktifkan sebagai konsekuensi dari suatu kejadian.

2.7 Kamus Data

Kamus data merupakan sebuah daftar yang terorganisasi dari elemen data yang berhubungan dengan system, dengan definisi yang tegas dan teliti sehingga pemakai dan analis akan memiliki pemahaman yang umum mengenai *input*, *output* komponen yang kalkulasi inter-mediate. [10]

Notasi yang digunakan untuk mengembangkan deskripsi isi adalah sebagai berikut:

Konstrik data	Notasi	Arti
Berurutan	=	Disusun atas
	+	Dan
Pilihan	[]	Baik ini atau
Pengulangan	{ }n	Pengulangan ke-n dari
	()	Data opsional
	* *	Komentar tidak dibatasi

Tabel 2.2 Notasi Deskripsi isi Kamus Data

2.8 Penggunaan Algoritma C4.5

Algoritma C 4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan. Algoritma C 4.5 merupakan

pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C 4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continue data*, dan *pruning*.

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target.

Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. [7]

Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi.

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan *probability* dari tiap-tiap *record* terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu kelas. Pohon keputusan juga dapat digunakan untuk

mengestimasi nilai dari variabel *continue* meskipun ada beberapa teknik yang lebih sesuai untuk kasus ini. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. [4]

Untuk menghitung *gain* digunakan rumus seperti tertera dalam persamaan 1 berikut: [1]

$$\text{Gain}(S,A) = \text{Entropy}(\text{Total}) - \sum_{i=1}^n \left(\left| \frac{S_i}{S} \right| \right) * \text{Entropy}(S_i)$$

Keterangan:

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- |Si| : jumlah kasus partisi ke-i
- |S| : jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut : [1]

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan:

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- Pi : proporsi dari si terhadap S

2.9 *My Structured Language (MySQL)*

2.9.1 Sejarah MySQL

MySQL dikembangkan oleh sebuah perusahaan Swedia bernama MySQL AB, yang pada saat itu bernama TcX Data Consult AB, sejak 1994-1995, meski cikal bakal kodenya bisa disebut sudah ada sejak 1979. Tujuan mula-mula TcX membuat MySQL pada waktu itu juga memang untuk mengembangkan aplikasi web untuk client – TcX adalah perusahaan pengembang software dan konsultan database.[11]

Pada saat itu Michael Widenius, atau “monty”, pengembang satu-satunya di TcX, memiliki aplikasi UNIREG dan rutin ISAM yang dibuat sendiri dan sedang mencari antarmuka SQL untuk ditempelkan di atasnya. Mula-mula TcX memakai mSQL, atau “mini SQL”. Barangkali mSQL adalah satu-satunya kode database open source yang tersedia dan cukup sederhana saat itu, meskipun sudah ada postgres. Namun ternyata, menurut Montry, mSQL tidaklah cukup cepat maupun fleksibel. Versi pertama mSQL bahkan tidak memiliki indeks. [11]

Setelah mencoba menghubungi David Hughes – pembuat mSQL ternyata mengetahui bahwa david tengah sibuk mengembangkan versi dua, maka keputusan yang diambil Montry yaitu membuat sendiri mes SQL yang antarmukanya mirip dengan mSQL tapi memiliki kemampuan yang lebih sesuai kebutuhan, maka lahirlah MySQL. [11]

Sebetulnya babak baru pengembangan MySQL yaitu saat MySQL dirilis sebagai produk open source sejati, dimulai saat perusahaan kecil MySQL AB ini menajalin kerja sama denga sebuah perusahaan database Amerika bernama Progress Software di tahun 2000. Progress membantu meyakinkan Montry dkk agar mengganti lisensi produk MySQL menjadi GPL. Selain itu, Progress juga memasuki bisnis open source dengan membentuk NuShpere, yang akan menjual MySQL khusus. [11]

Model bisnis mengemas software open source nampaknya baik di mata Progress, melihat kesuksesan distro-distro Linux dan perusahaan seperti Red Hat. Karena itu juga Progress berencana menyuntikan dana hingga \$2,5 juta dolar kepada

MySQL AB, dengan tujuan agar MySQL bisa menjadi lebih baik lagi. Satu lagi, NuSphere juga akan membuat table handler baru bernama Gemini, yang didasarkan pada software database Progress. Gemini akan menambahkan row-level locking dan transaksi MySQL. [11]

2.9.2 Apa itu MySQL

MySQL adalah sebuah server database SQL multiuser dan multithreaded yang dikembangkan oleh MySQL AB. MySQL adalah system manajemen basis data (*database management system*) adalah sekumpulan file yang saling berhubungan dan sekumpulan program yang memungkinkan pengguna mengakses dan memodifikasi file-file tersebut. Fungsi utama *system* basis data adalah menyediakan abstraksi data bagi pengguna tanpa perlu ditunjukkan bagaimana data tersebut disimpan dan dijaga. [11]

MySQL salah satu database relasional yang mendukung pemakaian SQL dan dirancang untuk pengguna aplikasi. Selain MySQL, beberapa produk sejenis yang sudah ada dipasaran, antara lain Oracle, Sybase, PostgreSQL, Informix, dan Microsoft SQL server.

Kesemua produk tersebut memungkinkan pengguna untuk mengolah data mereka didalam database tersentral pada computer pusat yang disebut dengan server. Sedangkan informasi yang dihasilkan dapat digunakan bersama-sama oleh beberapa user didalam *computer* localnya yang disebut *client*. [11]

Data yang diberikan server dapat disesuaikan dengan tingkat dan kepentingan pengguna. Sehingga data yang dikirim dapat berupa sebagian dari keseluruhan isi data, sesuai dengan permintaan user. Semua komunikasi yang terjadi mendukung perintah-perintah SQL. Saat ini makin banyak pembuatan aplikasi yang menggunakan arsitektur client/server, dalam pengaksesan database. Karena makin banyak pengguna aplikasi yang menuntut untuk dapat memperoleh informasi yang

actual dari bagian lain didalam suatu perusahaan, yang akan digunakan dalam proses pengambilan keputusan dibagian lainnya. [11]

Dengan menggunakan arsitektur ini dimungkinkan untuk mendapatkan integritas data yang cukup tinggi, karena semua user dapat bekerja dengan informasi yang sama. Pekerjaan dalam pemeliharaan data akan makin dimudahkan karena tugas seperti itu dapat dilakukan pada satu tempat saja, yaitu server. Pada awalnya MySQL dioperasikan hanya pada satu platform saja, namun sampai saat ini sudah berkembang jauh hingga dapat dioperasikan dalam berbagai platform, seperti windows 9X dan linux. Kehebatan database MySQL cukup diakui, terutama dalam system operasi linux. Disamping karena dukungannya kedalam berbagai platform, kecepatan aksesnya cukup bisa diandalkan. [11]

Kemudahannya dalam berbagai intergrasi aplikasi web (terutama PHP) cukup membantu dalam pengembangan system informasi online. Hebatnya lagi, MySQL dapat didownload secara bebas dan gratis pada pertengahan tahun 2001 MySQL, mendapatkan penghargaan dari linux Megazine Editor Choice Award, karena kemudahan instalasi, fleksibilitas dan ketersediannya dalam berbagai platform. Hal ini semakin mengukuhkan keberadaan MySQL dalam berbagai aplikasi berbasis database. Bukan hanya pada system operasi linux, tapi juga menyebar pada system operasi yang lain. [11]

2.10 Bahasa Pemrograman Visual Basic 6.0

Visual Basic merupakan bahasa pemrograman yang tercepat dan termudah untuk membuat suatu aplikasi dalam *Microsoft Windows*. Dengan menggunakan metode *Grapikal User Intervace*(GUI), *Visual Basic* memudahkan pemrogram untuk berinteraksi langsung dengan elemen-elemen untuk setiap bentuk pemrograman.

Visual Basic dibuat sebagai langkah pengembangan untuk menyesuaikan *Basic (Beginer All Puspose Symbolic Instruction Code)* yang berbasis DOS yang tidak mempunyai kemampuan menggunakan metode GUI dalam basis *Windows*.

Sebagai pemrograman berbasis *Windows*, *Visual Basic* mempunyai kemampuan untuk berintegrasi dengan seluruh aplikasi, seperti *Microsoft Word*, *Excel*, *Access* dan sebagainya. Kemampuannya yang hampir tidak terbatas, dapat digunakan untuk membuat aplikasi program mirip dengan *Microsoft Word*, *Excel*, ataupun *Game*, *Multimedia*, program perhitungan, dan sebagainya

Seiring dengan perkembangan, *Visual Basic* terus disempurnakan untuk kebutuhan *modernisasi* yang semakin meningkat, *Visual Basic 6* yang merupakan versi terakhir dari *Microsoft Visual Basic* diharapkan dapat menjawab semua tantangan akan kebutuhan computer.

Kelebihan dari *Visual Basic 6* [12]

1. Fasilitas *control*, *statemen* atau fungsi.
2. Kompiler yang sangat cepat.
3. Dapat mendukung database yang terintegrasi dengan variasi aplikasi yang sangat luas.
4. Serta kelebihan dari bahasa pemrograman lainnya yang relatif tidak tersruktur dalam bentuk pemrogramannya.

Struktur Aplikasi Dengan Bahasa Visual Basic aplikasi (Project) pada *Visual Basic 6* terdiri atas bagian-bagian:

1. *Form* adalah dimana anda mendesain program dengan meletakkan objek-objek yang merupakan rangkaian dari perintah-perintah yang akan dikerjakan oleh aplikasi tersebut
2. *Control* adalah yang mempunyai bentuk gambar grafis yang akan diletakkan diatas bidang kerja disebut form yang dapat berinteraksi dengan pemakai, seperti *textbox*, *labelbox*, *commandbutton*

Form dan *control* merupakan objek dalam pemrograman ini

3. *Properties* adalah variable yang melekat pada setiap objek. Contoh properties adalah nama, *caption*, ukuran,warna,posisi, dan isi. *Visual Basic* memberikan nilai baku dan nilai dapat diubah pada waktu program ini dijalankan.
4. *Methods* adalah prosedur yang sudah dibuat pada setiap objek yang sewaktu-waktu dapat digunakan sesuai dengan tujuan *methods* tersebut.
5. *Event Procedure* adalah kode yang berhubungan dengan setiap objek, yang menjelaskan tugasnya dengan *even* yang dimaksud. Kode ini akan beraksi apabila ada aksi dari *user* pada objek yang bersangkutan.

