

BAB II

LANDASAN TEORI

Penyusunan skripsi ini memerlukan beberapa teori untuk digunakan sebagai acuan / landasan teori. Penulisan dalam laporan skripsi ini menggunakan teori yang diperoleh dari buku yang dapat dipakai sebagai acuan. Sebelum melangkah lebih lanjut terlebih dahulu akan dibahas pengertian landasan teori yang melatar belakangnya.

2.1. PENGERTIAN ALGORITMA DATA MINING

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa 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.¹

Data mining adalah proses menerapkan metode ini untuk data dengan maksud untuk mengungkap pola-pola tersembunyi. Dengan arti lain Data mining adalah proses untuk penggalian pola-pola dari data. Data mining menjadi alat yang semakin penting untuk mengubah data tersebut menjadi informasi.²

¹ Julius Hermawan, *Membangun Decision Support System*, Pengertian Data Mining, Penerbit Andi, Yogyakarta, 2005, h. 33-34

² Kusri & Emha Taufiq Luthfi, *Algoritma Data Mining*, Data Mining, Penerbit Andi, Yogyakarta, 2009, h. 3-5

Hal ini sering digunakan dalam berbagai praktek profil, seperti pemasaran, pengawasan, penipuan deteksi dan penemuan ilmiah. Telah digunakan selama bertahun-tahun oleh bisnis, ilmuwan dan pemerintah untuk menyaring volume data seperti catatan perjalanan penumpang penerbangan, data sensus dan supermarket scanner data untuk menghasilkan laporan riset pasar.

Alasan utama untuk menggunakan data mining adalah kelebihanannya dalam membantu proses analisis koleksi pengamatan perilaku. Data tersebut rentan terhadap collinearity karena diketahui keterkaitan. Fakta yang tak terelakkan data mining adalah bahwa subset/set data yang dianalisis mungkin tidak mewakili seluruh domain, dan karenanya tidak boleh berisi contoh-contoh hubungan kritis tertentu dan perilaku yang ada di bagian lain dari domain . Untuk mengatasi masalah semacam ini, analisis dapat ditambah menggunakan berbasis percobaan dan pendekatan lain, seperti Choice Modelling untuk data yang dihasilkan manusia. Dalam situasi ini, yang melekat dapat berupa korelasi dikontrol untuk, atau dihapus sama sekali, selama konstruksi desain eksperimental

Beberapa teknik yang sering disebut-sebut dalam literatur Data Mining dalam penerapannya antara lain: clustering, classification, association rule mining, neural network, genetic algorithm dan lain-lain. Yang membedakan persepsi terhadap Data Mining adalah perkembangan teknik-teknik Data Mining untuk aplikasi pada database skala besar.

Sebelum populernya Data Mining, teknik-teknik tersebut hanya dapat dipakai untuk data skala kecil saja.

2.2. LANGKAH – LANGKAH DATA MINING

Untuk memanfaatkan data mining dan untuk mempelajari pola dari data yang dimiliki oleh suatu perusahaan dan mampu memberikan prediksi berdasarkan pola data yang ada, maka diperlukan langkah-langkah penentuan definisi data mining dari awal sampai dengan data mining siap untuk memberikan prediksi.

Langkah-langkah tersebut adalah :

1. Definisi permasalahan bisnis yang ingin diketahui.
Langkah pertama dalam pembuatan data mining adalah definisi permasalahan bisnis yang ingin dijawab, misalnya ingin mengetahui apakah seorang customer berpotensi memiliki kredit macet, atau mengidentifikasi seorang customer apakah akan pindah ke kompetitor bisnis kita, dan lain sebagainya. Setelah menemukan pertanyaan bisnis yang perlu dijawab oleh data mining, selanjutnya tentukan tipe tugas dasar untuk menjawab pertanyaan bisnis tersebut. Tugas dasar yang menjadi dasar algoritma data mining adalah klasifikasi, regresi, segmentasi, asosiasi dan sequence analysis. Dengan mengetahui tugas dasar, anda memiliki pedoman kira-kira algoritma mana yang bisa digunakan untuk menjawab pertanyaan bisnis.
2. Mempersiapkan data yang menjadi sumber untuk data mining untuk dipelajari polanya. Setelah menentukan definisi masalah, langkah

berikutnya adalah mencari data yang mendukung definisi masalah anda. Sumber data dapat berasal dari OLTP maupun dari cube OLAP. Bila sumber data berasal dari OLTP, sebaiknya pastikan datanya telah konsisten. Bila belum konsisten, SSIS bisa digunakan untuk melakukan cleansing data sampai data tersebut konsisten.

3. Menentukan porsi data yang digunakan men-training data mining berdasarkan algoritma data mining yang telah dibuat. Setelah persiapan data selesai dilakukan, langkah berikutnya adalah memberikan sebagian data kedalam algoritma data mining. Algoritma data mining perlu mempelajari pola data yang diberikan guna menarik informasi dan trend dari data tersebut. Istilah itu dikenal dengan sebutan training data mining.
4. Validasi apakah data mining memberikan prediksi yang akurat. Setelah training data selesai dilakukan, data mining tersebut perlu di “uji” atau di”validasi” keakuratannya terhadap data testing. Biasanya tidak hanya 1 algoritma data mining yang diimplementasikan ke dalam suatu data mining. Berarti data yang dipersiapkan pada persiapan data mengandung data yang digunakan untuk training dan data yang digunakan untuk testing. Pemilihan data training dan data testing dapat dilakukan secara manual ataupun secara acak oleh MySql.

2.3. PROSES DATA MINING

1. Pembersihan data untuk membuang data yang tidak konsisten dan noise.
2. Integrasi data merupakan penggabungan data dari beberapa sumber.
3. Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk di-mining.
4. Aplikasi teknik Data Mining.
5. Evaluasi pola yang ditemukan untuk menemukan yang menarik/bernilai.
6. Presentasi pengetahuan menggunakan dengan teknik visualisasi.

2.4. PENGERTIAN NEAREST NEIGHBOR

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.³

Sehingga menurut Kusrini & Emha Taufiq Lutfhi, Nearest Neighbor (NN) merupakan suatu metode untuk mengklasifikasikan suatu data baru berdasarkan similaritas dengan label data. Didalam metode ini, similaritas biasanya memakai matrik jarak dan satuan jarak umumnya menggunakan Euclidian.

Jenis NN :

- 1-NN : Pengklasifikasian dilakukan terhadap 1 labeled data terdekat.

³ Kusrini & Emha Taufiq Luthfi, *Algoritma Data Mining*, Algoritma Nearest Neighbor, Penerbit Andi, Yogyakarta, 2009, h. 93-95

- K-NN : Pengklasifikasian dilakukan terhadap K labeled data terdekat, dengan $K > 1$

Algoritma 1-NN :

- Hitung jarak antara data baru ke setiap label data
- Tentukan 1 label data yang mempunyai jarak yang minimal
- Klasifikasikan data baru ke dalam label data tersebut

Algoritma K-NN :

- Tentukan K
- Hitung jarak antara data baru ke setiap label data
- Tentukan K label data yang mempunyai jarak yang minimal
- Klasifikasikan data baru ke dalam label data yang mayoritas

Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Berikut beberapa kelebihan dan kelemahan metode Nearest Neighbor dibandingkan dengan Metode AHP, diambil dari halaman web <http://www.digilib.uns.ac.id/upload/dokumen/45072004200901002.rtf>.

Berikut Kelebihan Dan Kelemahan Metode *Nearest Neighbor* dengan Metode *Analytic Hierarchy Process* :

Kelebihan Nearest Neighbor	Kelebihan AHP
1. Dapat menghasilkan data yang kuat atau jelas (khususnya jika menggunakan turunan perkalian kuadrat pada besaran jarak).	1. Dapat menangani saling ketergantungan elemen-elemen dalam suatu sistem dan tak memaksakan pemikiran linier.
2. Efektif jika digunakan untuk data	2. Memberi suatu skala untuk

<p>yang besar.</p> <p>3. Mampu mencari nilai kedekatan hingga 1 jika data record sudah dalam skala besar.</p>	<p>mengukur hal-hal dan wujud suatu model untuk menetapkan prioritas.</p> <p>3. Melacak konsistensi logis dari pertimbangan-pertimbangan yang digunakan dalam menentukan prioritas.</p>
Kelemahan Nearest Neighbor	Kelemahan AHP
<p>1. Membutuhkan nilai K sebagai parameter.</p> <p>2. Jarak dari data percobaan tidak dapat jelas dengan tipe jarak yang digunakan dan dengan atribut yang digunakan untuk menghasilkan hasil yang terbaik, maka harus menggunakan semua atribut atau hanya 1 atribut yang telah pasti.</p> <p>3. Perhitungan harga sangat tinggi karena percobaan ini membutuhkan perhitungan jarak dari beberapa query untuk semua data percobaan.</p>	<p>1. Tidak dapat diterapkan pada suatu perbedaan sudut pandang yang sangat tajam/ ekstrim di kalangan responden.</p> <p>2. Metode ini mensyaratkan ketergantungan pada sekelompok ahli sesuai dengan jenis spesialis terkait dalam pengambilan keputusan.</p> <p>3. Responden yang dilibatkan harus memiliki pengetahuan dan pengalaman yang cukup tentang permasalahan serta metode <i>AHP</i>.</p>

2.5 DESAIN INPUT

Menurut Pree, W (1991) Desain Input membahas tentang input data yang tidak terlepas dari alat-alat input (input device) yang akan digunakan, misalnya keyboard, card reader dan lain-lain. Alat input di golongkan menjadi dua golongan, yaitu alat input langsung (online input device) dan alat input tidak langsung (offline input device). Alat input langsung merupakan alat input yang langsung dihubungkan dengan CPU, misalnya keyboard, mouse, touch screen. Alat input tidak langsung adalah alat input yang tidak langsung dihubungkan dengan CPU, misalnya KTC (key-to-card), KTT (key-to-tape), dan KTD (key-to-disk).⁴

Input yang menggunakan alat input tidak langsung mempunyai tiga tahapan utama tersebut. Input yang menggunakan alat input langsung hanya terdiri dari dua tahapan utama saja (data capture dan data entry).

2.6. DESAIN OUTPUT

Menurut Rumbaugh, J. (1991) Output adalah produk dari sistem informasi yang dapat berupa hasil media kertas, hasil media lunak dan hasil dari suatu proses yang akan digunakan oleh proses lain yang tersimpan pada suatu media seperti tipe, disk atau kartu. Sedangkan output yang dimaksud dalam desain ini adalah output yang berupa tampilan di media kertas atau di layer video.⁵

⁴ Pree, W., Pomberger, G., Sikora, H. (1991) : *Construction Techniques of Graphic, Direct-Manipulation User Interfaces*. In : Post, F.H., Barth, W. (ed.) : Eurographics '91, Elsevier Science Publishers, Amsterdam.

⁵ James Rumbaugh, Michael Blaha, William Premerlani, Frederick Eddy, and William Lorenson, *Object-Oriented Modeling and Design*, Prentice Hall, 1991.

Output dapat diklasifikasikan ke dalam beberapa tipe, yaitu output intern (internal output) dan output ekstern (external output).

a. Output Internal

Adalah output yang dimaksudkan untuk mendukung kegiatan manajemen. Output ini akan tetap tinggal di perusahaan dan akan disimpan sebagai arsip atau dimusnahkan bila sudah tidak digunakan lagi. Output jenis ini dapat berupa laporan-laporan terinci, laporan-laporan ringkasan dan laporan-laporan lainnya.

b. Output Ekstern

Adalah output yang akan didistribusikan kepada pihak luar yang membutuhkannya. Contoh output ekstern adalah faktur, cek, tanda terima pembayaran dan lain-lain.

2.7 DESAIN DATABASE

Menurut Fatansyah (1999) database dapat didefinisikan dalam sejumlah sudut pandang, yaitu :

- a. Himpunan kelompok data (arsip) yang saling berhubungan yang diorganisasikan sedemikian rupa agar kelak dapat dimanfaatkan kembali dengan cepat dan mudah.
- b. Kumpulan data yang saling berhubungan yang disimpan secara bersama sedemikian rupa dan tanpa pengulangan yang tidak perlu, untuk memenuhi berbagai kebutuhan.
- c. Kumpulan file / tabel / arsip yang saling berhubungan yang disimpan dalam media elektronik.

Database adalah suatu kumpulan data terhubung yang disimpan secara bersama-sama pada suatu media tanpa menatap satu sama lain atau tidak perlu suatu kerangkapan data dengan cara-cara tertentu, sehingga mudah untuk digunakan atau ditampilkan kembali, dapat digunakan untuk satu atau lebih program aplikasi secara optimal, data disimpan tanpa mengalami ketergantungan pada program yang akan menggunkannya, data disimpan sedemikian rupa sehingga penambahan juga pengurangan dan modifikasi data dapat dilakukan dengan mudah dan terkontrol.⁶

2.8 PENGERTIAN VISUAL BASIC

Visual Basic 6.0 merupakan salah satu aplikasi pemrograman visual menggunakan pendekatan *Graphical User Interface* (GUI) yang lebih nyaman dan lebih mudah digunakan oleh programmer untuk membuat aplikasi, baik aplikasi kecil untuk keperluan sendiri ataupun *sistem enterprise* yang besar serta aplikasi terdistribusi yang dijalankan melalui internet.⁷

Visual Basic 6.0 memiliki antar muka yang dapat diatur sesuai selera penggunaannya, diantaranya :

1. *Title Bar*

Menunjukkan nama project yang sedang dibuat.

2. *Main Bar*

⁶ Fatansyah, *Basis Data*, Basis Data, Bandung : Informatika Bandung, 1999. h. 27-28

⁷ Arief Ramadhan, *SQL Server 2000 dan VB 6.0*, Visual Basic 6.0, PT Elex Media Komputindo, Jakarta, 2005 h. 51-55

Merupakan toolbar utama yang berisi icon-icon yang dapat diklik oleh pengguna untuk melakukan suatu perintah khusus secara cepat.

3. *Form*

Merupakan tempat untuk merancang tampilan program aplikasi yang dibuat.

4. *Form Layout Window*

Merupakan jendela yang menunjukkan posisi relatif form terhadap layar monitor saat form tersebut ditampilkan dilayar windows.

5. *Code Window*

Merupakan jendela yang digunakan untuk melihat atau mengetikan kode program. Setiap form memiliki sebuah kode window sendiri.

6. *Toolbox*

Terdiri atas beberapa class objek yang dapat digunakan untuk objek ke dalam proses pembuatan aplikasi.

7. *Project Explorer Window*

Merupakan jendela untuk menampilkan project-project, form-form, atau modul-modul yang terlibat dalam proses pembuatan program aplikasi.

8. *Propertise Window*

Terdiri atas beberapa class objek yang dapat digunakan untuk objek ke dalam proses pembuatan aplikasi.

Kelebihan Menggunakan Visual Basic 6.0 dengan Pascal.

Kelebihan Microsoft Visual Basic 6.0 adalah :

1. Kemampuan untuk mengkompilasi program dalam bentuk *Native Code* yaitu optimasi pada saat processor mengkompilasi dan menjalankan program tersebut.
2. Kecepatannya dalam mengakses program dimana hal ini hanya dapat ditemui pada aplikasi-aplikasi yang dikompilasi dengan bahasa pemrograman C++.
3. Ada beberapa dan hampir semua sistem database kompatibel ama vb6 seperti:
 - Microsoft Access
 - ODBC
 - SQL server 7
 - Paradox
 - MySQL
 - Excel
 - Oracle

Kekurangan Pascal adalah :

1. Masih menggunakan Sistem DOS.
2. Dalam pascal sebuah program yang memiliki baris *source code* lebih dari 3000 baris, pascal tidak mampu meng-compilanya. Karena dalam pascal baris *source code* hanya dibatasi sekitar 2000 baris.
3. Jika ingin membuat sebuah tulisan yang tampil di layar namun mengandung karakter tanda petik satu (‘), pascal tidak bisa meng-compilanya. Hal ini dikarenakan pascal dalam mengeluarkan inputan string dari keyboard harus diapit oleh tanda petik satu tadi.

4. Tidak mendukung pemrograman berorientasi objek.
5. Pascal tidak fleksibel dan banyak kekurangan yang dibutuhkan untuk membuat aplikasi yang besar.

2.9 PENGERTIAN MySQL

MySQL adalah sebuah perangkat lunak sistem manajemen basis data SQL (bahasa Inggris: database management system) atau DBMS yang multithread, multi-user, dengan sekitar 6 juta instalasi di seluruh dunia. MySQL AB membuat MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi GNU General Public License (GPL), tetapi mereka juga menjual dibawah lisensi komersial untuk kasus-kasus dimana penggunaannya tidak cocok dengan penggunaan GPL.⁸

Tidak sama dengan proyek-proyek seperti Apache, dimana perangkat lunak dikembangkan oleh komunitas umum, dan hak cipta untuk kode sumber dimiliki oleh penulisnya masing-masing, MySQL dimiliki dan disponsori oleh sebuah perusahaan komersial Swedia MySQL AB, dimana memegang hak cipta hampir atas semua kodenya. Kedua orang Swedia dan satu orang Finlandia yang mendirikan MySQL AB adalah: David Axmark, Allan Larsson, dan Michael "Monty" Widenius.

⁸ <http://id.wikipedia.org/wiki/MySQL> (10:53, 26 Maret 2011)

Kelebihan MySQL

MySQL memiliki beberapa keistimewaan, antara lain :

1. **Portabilitas.** MySQL dapat berjalan stabil pada berbagai sistem operasi seperti Windows, Linux, FreeBSD, Mac Os X Server, Solaris, Amiga, dan masih banyak lagi.
2. **Perangkat lunak sumber terbuka.** MySQL didistribusikan sebagai perangkat lunak sumber terbuka, dibawah lisensi GPL sehingga dapat digunakan secara gratis.
3. **Multi-user.** MySQL dapat digunakan oleh beberapa pengguna dalam waktu yang bersamaan tanpa mengalami masalah atau konflik.
4. **'Performance tuning'**, MySQL memiliki kecepatan yang menakjubkan dalam menangani query sederhana, dengan kata lain dapat memproses lebih banyak SQL per satuan waktu.
5. **Ragam tipe data.** MySQL memiliki ragam tipe data yang sangat kaya, seperti signed / unsigned integer, float, double, char, text, date, timestamp, dan lain-lain.
6. **Perintah dan Fungsi.** MySQL memiliki operator dan fungsi secara penuh yang mendukung perintah Select dan Where dalam perintah (*query*).
7. **Keamanan.** MySQL memiliki beberapa lapisan keamanan seperti level subnetmask, nama host, dan izin akses *user* dengan sistem perizinan yang mendetail serta sandi terenkripsi.

8. **Skalabilitas dan Pembatasan.** MySQL mampu menangani basis data dalam skala besar, dengan jumlah rekaman (records) lebih dari 50 juta dan 60 ribu tabel serta 5 milyar baris. Selain itu batas indeks yang dapat ditampung mencapai 32 indeks pada tiap tabelnya.
9. **Konektivitas.** MySQL dapat melakukan koneksi dengan klien menggunakan protokol TCP/IP, Unix socket (UNIX), atau Named Pipes (NT).
10. **Lokalisasi.** MySQL dapat mendeteksi pesan kesalahan pada klien dengan menggunakan lebih dari dua puluh bahasa. Meski pun demikian, bahasa Indonesia belum termasuk di dalamnya.
11. **Antar Muka.** MySQL memiliki antar muka (interface) terhadap berbagai aplikasi dan bahasa pemrograman dengan menggunakan fungsi API (Application Programming Interface).
12. **Klien dan Peralatan.** MySQL dilengkapi dengan berbagai peralatan (tool) yang dapat digunakan untuk administrasi basis data, dan pada setiap peralatan yang ada disertakan petunjuk online.
13. **Struktur tabel.** MySQL memiliki struktur tabel yang lebih fleksibel dalam menangani ALTER TABLE, dibandingkan basis data lainnya semacam PostgreSQL ataupun Oracle.