

BAB III

LANDASAN TEORI

3.1. Sistem

Sistem adalah kumpulan obyek-obyek yang saling berinteraksi dan bekerja bersama-sama untuk mencapai tujuan tertentu dalam lingkungan yang kompleks (Andoko, 2013). Maksud dari adanya suatu sistem adalah untuk membantu pengguna dalam mencapai suatu tujuan tertentu. Suatu sistem mempunyai karakteristik atau sifat-sifat tertentu, yaitu mempunyai komponen-komponen (components), sistem data (boundary), lingkungan luar sistem (envirotments), penghubung (interface), masukan (input), keluaran (output), pengolahan (process) dan sasaran (objectives) atau tujuan (goal). Komponen-komponen sistem dapat berupa suatu subsistem atau bagian-bagian sistem. Batasan sistem merupakan daerah yang membatasi sistem dengan sistem lainnya atau dengan lingkungan luar sistem. Lingkungan luar dari sistem adalah apapun diluar batas sistem yang mempengaruhi operasi sistem (Widayanto & Wardati, 2013).

3.2. Pengelompokan Data

Data merupakan suatu nilai yang merepresentasikan deskripsi dari suatu objek atau kejadian. Pengelompokan data merupakan pengorganisasian sekumpulan data yang besar dengan cara membagi data tersebut kedalam beberapa kelompok (Ariawan et al., 2013). Kelompok-kelompok yang terbentuk akan mampu menjelaskan adanya persamaan maupun perbedaan dari keseluruhan data yang diteliti (Everitt et al., 2011). Seringkali

pengelompokan data digunakan untuk mempermudah dalam suatu proses analisis dan interpretasi dari suatu data yang besar, dengan membagi data tersebut kedalam kelompok-kelompok data. Pengelompokan data dapat mengungkapkan hubungan dan struktur di dalam data, yang sebelumnya tidak jelas menjadi lebih jelas (Suwarningsih, 2012).

3.3. Preprocessing Data

Data yang belum diproses disebut data mentah, dimana data harus disiapkan terlebih dahulu sebelum dapat dipakai dalam suatu proses. Data mentah atau data *real* cenderung mengandung kesalahan atau mengandung nilai-nilai yang menyimpang dari yang diharapkan (Kumar & Chadha, 2012). Data yang mengandung kesalahan dikarenakan data tersebut tidak lengkap ataupun tidak konsisten. Ketidaklengkapan data terjadi karena adanya atribut data yang tidak tersedia, hilangnya nilai untuk beberapa data (atribut) karena adanya penghapusan data yang dianggap tidak penting. Sedangkan data dianggap tidak konsisten karena pada saat pengumpulan data adanya instrumen yang rusak karena kesalahan manusia (*human error*) ataupun kesalahan komputer, adanya ketidaksamaan (tidak konsisten) dalam penamaan suatu data dengan data yang lain, yang merupakan suatu data yang sama (Han & Kamber, 2006).

Untuk mengurangi data yang tidak lengkap, mengandung noise (kesalahan) ataupun tidak konsisten dilakukannya suatu tahapan yang disebut *preprocessing data* yang merupakan bagian dari penambangan data (*data mining*). *Data mining* merupakan suatu proses pengekstrakan informasi penting pada data yang

berukuran besar (Maharani, 2009). Selain itu *data mining* juga dapat dikatakan sebagai sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis dari sekumpulan data yang berjumlah banyak. Data mining sering dianggap sebagai bagian dari *Knowledge Discovery in Database (KDD)*, yaitu sebuah proses yang mencari pengetahuan yang bermanfaat (Sunjana, 2010). Tujuan utama penerapan data mining adalah untuk memberikan prediksi dan uraian (deskripsi) (Aribowo, 2013). Proses awal sebelum dilakukannya data mining adalah *Preprocessing Data* (Ayub, 2007).

Preprocessing data merupakan langkah penting dalam proses penemuan pengetahuan, karena keputusan-keputusan yang berkualitas harus didasarkan pada data yang berkualitas (Kumar & Chadha, 2012). *Preprocessing data* sering kali digunakan untuk mengurangi kesalahan data dan sistematis bias dalam data mentah sebelum analisis apapun terjadi (Tong et al., 2011). Tugas utama dari *preprocessing data*, antara lain :

1. Pembersihan Data

Pembersihan data dilakukan dengan mengisi nilai yang hilang, mengidentifikasi atau menghapus data yang salah dan menyelesaikan ketidaksamaan atau inkonsistensi data.

2. Integrasi Data

Integrasi data adalah penggabungan data dari berbagai sumber penyimpanan data untuk menjadi suatu kesatuan data yang koheren.

3. Transformasi Data

Transformasi data dilakukan dengan proses normalisasi.

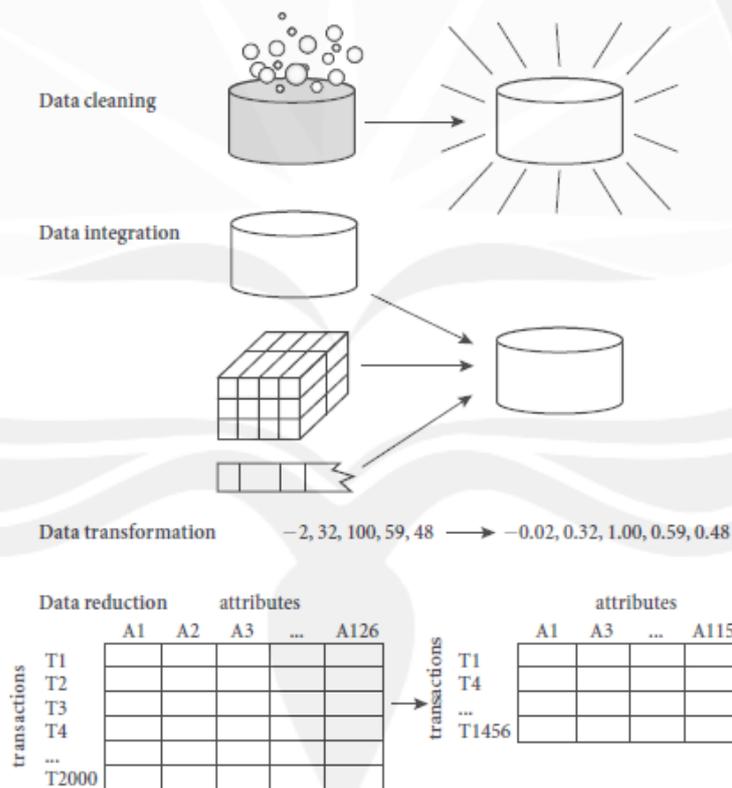
4. Reduksi Data

Reduksi data merupakan perolehan representasi penurunan volume tetapi menghasilkan hasil analisis yang sama atau mirip.

5. Diskritisasi Data

Diskritisasi data merupakan bagian dari reduksi data, tetapi dengan kepentingan tertentu, terutama untuk data numerik. (Han & Kamber, 2006)

Bentuk *preprocessing data* mengenai pembersihan data, integrasi data, transformasi data serta reduksi data dapat dilihat pada Gambar 3.1 dibawah ini.



Gambar 3.1. Bentuk Preprocessing Data
(Han & Kamber, 2006)

Pada studi kasus ini, menggunakan pembersihan data (*data cleaning*) sebagai tahapan awal, sebelum data

diolah untuk dikelompokkan, karena data yang ada tidak konsisten dan ada beberapa data yang tidak lengkap karena adanya nilai atribut yang hilang.

Pembersihan data (*data cleaning*) merupakan proses menghilangkan *noise* (kesalahan) dan data yang tidak relevan, yang pada umumnya data yang diperoleh baik dari *database* yang ada memiliki isian yang tidak sempurna seperti adanya data yang hilang, data yang tidak valid atau kesalahan dalam pengetikan (Asriningtyas & Mardhiyah, 2014). Dapat dikatakan tugas utama dari pembersihan data ialah mengisi data yang hilang, mengidentifikasi outlier dan kesalahan data, serta membenarkan data yang tidak konsisten. Pembersihan data, khususnya untuk data yang tidak konsisten dapat dilakukan dengan beberapa cara, yaitu :

1. Manual yaitu dengan mencari dan memperbaiki data yang tidak konsisten satu persatu.
2. Semi manual, yaitu dengan menggunakan bantuan sistem untuk mendeteksi data yang tidak konsisten, lalu memperbaikinya secara manual.
3. Otomatis, yaitu dengan menggunakan bantuan sistem sepenuhnya dalam mencari dan memperbaiki data yang tidak konsisten dengan menggunakan proses integrasi data (Han & Kamber, 2006).

3.4. Levenshtein Distance

Edit distance dalam komputer science dikenal dengan nama *Levenshtein Distance*. Nama tersebut diambil dari nama penemunya, yaitu Vladimir Levenshtein. *Levenshtein distance* ditemukan pada tahun 1965. Sejalan dengan waktu *Levenshtein Distance* lebih dikenal dengan *edit distance* (Mulyanto, 2010). Algoritma *Levenshtein*

Distance merupakan algoritma yang digunakan untuk mencari jumlah operasi *string* yang paling sedikit untuk mentransformasikan suatu *string* menjadi *string* yang lain. Algoritma ini digunakan dalam pencarian *string* dengan pendekatan perkiraan (*Approximate String Matching*) (Adiwidya, 2009).

Perhitungan *Levenshtein Distance* didapatkan dari matriks yang digunakan untuk menghitung jumlah perbedaan antara dua buah *string*. Perhitungan jarak antara dua *string* ini ditentukan dari jumlah minimum operasi perubahan untuk membuat *string* 'A' menjadi *string* 'B'. Jumlah operasi minimum tersebut disebut dengan *Levenshtein Distance* atau *edit distance*. Pada algoritma *Levenshtein Distance* ada tiga macam operasi utama, yaitu :

1. Operasi Pengubahan Karakter

Operasi pengubahan karakter berarti menukar karakter ke dalam suatu *string*.

2. Operasi Penambahan Karakter

Operasi penambahan karakter berarti menambahkan karakter ke dalam *string*.

3. Operasi Penghapusan Karakter

Operasi penghapusan karakter berarti menghapus atau menghilangkan karakter dari suatu *string*.

Algoritma ini berjalan mulai dari pojok kiri atas sebuah array dua dimensi yang telah diisi sejumlah karakter *string* awal dan *string* target dan diberikan nilai *cost*. Nilai *cost* dapat ditambah dengan nol ataupun satu, dengan ketentuan jika membandingkan huruf yang sama maka ditambah dengan nol dan jika

membandingkan huruf yang berbeda maka ditambah dengan satu.

$$m[s,t] = \min(m[s,t]+cost, m[s,t]+1, m(s,t)+1)$$

operasi edit, operasi tambah, operasi hapus

Nilai pada ujung kanan bawah menjadi nilai *edit distance* yang menggambarkan jumlah perbedaan dua *string* (Junedy, 2014).

Jarak (*distance*) pada metode ini adalah nilai minimum dari operasi pengubahan, penambahan dan penghapusan karakter yang dibutuhkan untuk merubah *string* asli (*s*) menjadi *string* pembanding (*t*), contoh :

1. Jika *s* adalah "aku" dan *t* adalah "aku", maka jarak antara kedua *string* adalah nol, dikarenakan tidak ada transformasi yang dibutuhkan. Maka dapat disimpulkan kedua *string* tersebut identik.
2. Jika *s* adalah "aku" dan *t* adalah "abu", maka jarak antara kedua *string* tersebut adalah satu, dikarenakan adanya pengubahan karakter dimana huruf "k" pada kata "aku" diubah menjadi huruf "b" pada kata "abu".

Proses algoritma *levenshtein distance* membandingkan kata "aku" sebagai *string* asli (*s*) dengan "abu" sebagai *string* pembanding (*t*), yaitu :

1. Membuat matrik

		a	b	u
	0	1	2	3
a	1			
k	2			
u	3			

2. Melakukan pencocokan dengan melakukan perbandingan dari setiap karakter s dengan karakter t, dengan ketentuan :

- a. Jika karakter pertama pada s dan karakter pertama pada t sama maka nilai pada cell (1,1) diisi dengan nol yang merupakan nilai matrik (x-1, y-1)

		a	b	u
	0	1	2	3
a	1	0		
k	2			
u	3			

- b. Jika karakter pertama pada s dan karakter kedua pada t berbeda, maka nilai cell didapat dari nilai terkecil dari :

- a) Nilai cell $(x-1, y-1)+1$, pada contoh nilaiya adalah dua.
- b) Nilai cell $(x-1, y)+1$, pada contoh nilaiya adalah tiga.
- c) Nilai cell $(x, y-1)+1$, pada contoh nilaiya adalah satu.

Maka jarak karakter pertama pada s dan karakter kedua pada t adalah satu.

		a	b	u
	0	1	2	3
a	1	0		
k	2	1		
u	3			

3. Proses terus berjalan sampai akhir karakter s dan karakter t.

		a	b	u
	0	1	2	3
a	1	0	1	2
k	2	1	1	2
u	3	2	2	1

Maka didapatkan jarak *string* asli (s) dengan *string* pembanding (t) adalah satu.

Semakin besar nilai jarak yang dihasilkan oleh operasi algoritma *levenshtein distance*, maka semakin besar perbedaan diantara kedua *string* tersebut (Junedy, 2014).

3.5. Bahasa Pemrograman C#

Sistem Pengelolaan Data Mahasiswa (SIPEMA) dibangun dengan menggunakan *Microsoft Visual Studio 2008* dengan menggunakan *C#* sebagai bahasa pemrogramannya. *C#* adalah sebuah bahasa pemrograman yang didesain untuk membangun aplikasi enterprise yang berjalan diatas *framework .NET* (Ferryzal, 2015).

3.6. Microsoft SQL Server 2005

Structured Query Language (SQL) adalah sebuah bahasa yang dipergunakan untuk mengakses data dalam

basis data relasional. Bahasa ini secara de facto merupakan bahasa standar yang digunakan dalam manajemen basis data relasional (Adelia & Setiawan, 2011).

Structured Query Language (SQL) terdiri dari dua bahasa yaitu *Data Definition Language (DDL)* dan *Data Manipulation Language (DML)*, yang mana secara umum implementasi setiap bahasa ini memiliki bentuk standar yang ditetapkan oleh *American National Standards Institute (ANSI)*. *SQL Server 2005* merupakan aplikasi *database* produk *Microsoft* yang memiliki fitur-fitur baru sehingga membuatnya menjadi suatu platform *database* yang sempurna, khususnya dalam mengolah data yang berskala besar dan aplikasi *e-commerce*. Fitur-fitur tersebut antara lain:

1. *Notification Services*

Fitur ini digunakan untuk mengirimkan dan menerima pesan atau pemberitahuan (*notification*) dengan tepat waktu dari *database* ke ribuan pengguna.

2. *Reporting Services*

Fitur ini digunakan untuk membuat laporan dari *SQL Server 2005*.

3. *Service Broker*

Fitur ini menyediakan infrastruktur yang digunakan untuk membangun keamanan pada aplikasi berkinerja tinggi (Kurnia et al., 2013).

Dapat dikatakan *SQL Server 2005* ialah perangkat lunak *Relational Database Management System (RDBMS)* yang andal. Sistem Pengelolaan Data Mahasiswa (*SIPEMA*) menggunakan *tools Microsoft SQL Server* sebagai *database-nya*, dengan menggunakan *Data Manipulation*

Language (DML) untuk memanipulasi data yang ada dalam suatu tabel. Perintah-perintah umum yang dilakukan pada *Data Manipulation Language* (DML) adalah :

1. SELECT untuk menampilkan data.
2. INSERT untuk menambahkan data baru.
3. UPDATE untuk mengubah data yang sudah ada.
4. DELETE untuk menghapus data (Adelia & Setiawan, 2011).

3.7. Promosi

Promosi merupakan fungsi bisnis yang bertugas untuk mengenali kebutuhan dan keinginan pelanggan, menentukan pasar sasaran mana yang dapat dilayani dengan sebaik-baiknya oleh perusahaan, serta merancang produk, jasa dan program yang tepat untuk melayani pasar tersebut (Novera, 2014). Ada lima jenis kegiatan promosi, antara lain :

1. Periklanan (advertising), yaitu bentuk promosi non personal dengan menggunakan berbagai media yang ditunjukkan untuk mempengaruhi pembelian konsumen.
2. Penjualan pribadi (personal selling), yaitu promosi secara personal dengan presentasi lisan dalam suatu percakapan dengan calon pembeli yang ditujukan untuk mempengaruhi pembelian konsumen.
3. Publisitas (publicity), yaitu suatu bentuk promosi non personal mengenai pelayanan atau kesatuan usaha tertentu dengan jalan mengulas informasi atau berita tentangnya.
4. Promosi penjualan (sales promotion), yaitu suatu bentuk promosi diluar ketiga bentuk di atas yang ditujukan untuk mempengaruhi pembelian konsumen.

5. Pemasaran langsung (direct marketing), yaitu suatu bentuk penjualan perorangan secara langsung ditujukan untuk mempengaruhi pembelian konsumen. (Kotler, 2005)

Tujuan utama dari promosi adalah modifikasi tingkah laku konsumen, menginformasikan, mempengaruhi dan membujuk serta mengingatkan konsumen sasaran tentang perusahaan dan produk atau jasa yang dijualnya (Swastha & Irawan, 2005).

3.8. Kantor Kerjasama dan Promosi

Kantor Kerjasama dan Promosi (KKP) adalah salah satu unit penunjang Universitas Atma Jaya Yogyakarta (UAJY), yang merupakan perangkat pelengkap UAJY yang mempunyai tugas memberikan dukungan pada pemimpin UAJY dibidang kerjasama dan promosi. Visi dan misi KKP antara lain :

1. Visi : menjadi unit pendukung untuk mengembangkan performa universitas dalam memperluas jangkauan promosi dan kerjasama institusi.
2. Misi : memberikan pelayanan yang terintegritas bagi terciptanya performa UAJY sebagai institusi pendidikan bermutu melalui peningkatan kinerja promosi dan kualitas kerjasama dengan lembaga relevan baik dalam maupun luar negeri.

Kantor Kerjasama dan Promosi mempunyai sasaran mutu, yaitu :

1. Memperluas jangkauan informasi untuk penjangking calon mahasiswa baru Program Strata Satu (S-1) ke minimal Seribu Sekolah Menengah Atas (SMA),

240 Gereja, dan 25 Lembaga terkait setiap setahun.

2. Meningkatkan jumlah SMA pendaftar program S-1 minimal empat puluh persen (40%) dari seluruh SMA yang mendapat informasi dan atau menjadi lokasi kegiatan promosi sebelum 2014.
3. Meningkatkan kinerja promosi unit-unit di lingkungan UAJY terutama di wilayah Jawa Tengah, Jawa Timur dan Jawa Barat sebelum 2014.
4. Membangun intensitas kerjasama melalui MoU baru dengan dua institusi baik institusi negeri maupun luar negeri dalam satu tahun.
5. Intensifikasi aktivitas kerjasama minimal satu kegiatan tiap MoU dalam satu tahun. (Yogyakarta, 2015).