

BAB II

LANDASAN TEORI

A. Tinjauan Pustaka

Tabel berikut ini akan menjelaskan beberapa hasil penelitian yang telah dilakukan dan memiliki kesamaan topik maupun metode yang dilakukan pada penelitian ini.

Tabel 2.1. Hasil Penelitian Sebelumnya

No.	Nama Peneliti, Tahun	Judul Penelitian	Hasil Penelitian
1.	Johan Oscar Ong, 2013	Implementasi Algoritma <i>K-Means Clustering</i> Untuk Menentukan Strategi Marketing President University	Dengan adanya pengelompokan-pengelompokan data seperti ini, diharapkan bagian marketing dapat melakukan pemasaran dengan strategi yang tepat untuk mendapatkan calon mahasiswa baru.
2.	Dwi Untari, 2014	Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif	Penerapan data mining dengan metode <i>Decision Tree C.4.5</i> untuk mencari karakteristik

		Menggunakan Metode <i>Decision Tree</i> C.4.5	mahasiswa berpotensi non-aktif.
3.	Ahmad Fikri Mauriza, 2014	Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Ums Menggunakan Metode <i>Naive Bayes</i>	Penggunaan metode <i>Naive Bayes</i> untuk menganalisis data dalam pengenalan pola dan memprediksi kelulusan mahasiswa.

Sumber Ong Oscar, J. 2013

B. Definisi *Clustering*

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*taining*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, 2007).

Metode *hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada

hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut.

Selanjutnya, berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010).

C. Definisi *K-Mean Clustering*

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:

1. Pilih jumlah *cluster k*.
2. Inisialisasi *k* pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random,
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

Sumber Santosa (2007)

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat cluster tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

D. Definisi PHP (*Professional Home Page*)

PHP adalah bahasa pemrograman *script* yang paling banyak dipakai saat ini. *PHP* banyak dipakai untuk memrogram situs *web* dinamis, walaupun tidak tertutup kemungkinan digunakan untuk pemakaian lain. Berikut ini dipaparkan sejarah *PHP* dari awal dikenalkan hingga saat ini :

1. *PHP /FI*

Pertama kali *PHP* dibuat dan diperkenalkan oleh Rasmus Lerdorf pada tahun 1995 menggunakan nama *PHP/FI*. Pada awalnya, *PHP/FI* merupakan bagian dari personal *Home Page Tools*.

Pada tahun 1997, dikeluarkan *PHP/FI* versi 2.0. Fungsi- fungsi pada *PHP/ FI* ditulis dengan menggunakan bahasa C. Karena telah memiliki fungsi khusus untuk mengakses *database*.

2. *PHP 3*

PHP 3 merupakan generasi baru hasil pengembangan *PHP/FI*. Banyak *developer* yang terlibat di dalamnya. Tak heran jika *PHP 3* dianggap sebagai tonggak awal bagi terciptanya *PHP* versi sekarang ini. Secara resmi peluncur *PHP 3.0* ialah Andi Gutmans dan Zeev Suraski pada tahun 1997. Setelah *PHP 3.0* dikeluarkan, mereka menyarankan untuk menghentikan proyek *PHP/FI* karena *PHP 3.0* masih lebih baik.

3. *PHP 4*

Pada pertengahan tahun 1999 Zend merilis *PHP* versi 4 diluncurkan untuk menangani kelemahan *PHP3*, yaitu penggunaan fungsi yang begitu *kompleks*. Kurangnya efisiensi waktu dan kinerja yang buruk diperbaiki dan ditulis ulang dari inti *PHP 3*. Dengan penambahan fitur baru, seperti *session*, *output buffering* dan penanganan *input*, menjadikan *PHP 4* aman dari berbagai jenis bahasa pemrograman berbasis *web*. Selain itu, inti perbedaan mereka terletak pada penggunaan *Zend Engine*. *Zend Engine* merupakan inti dari *PHP*. Sebagai bagian dari inti *PHP*, secara fungsional ia bertugas menangani *input*, menerjemahkan dan mengeksekusinya.

4. *PHP 5*

PHP versi 5 muncul untuk menangani kelemahan- kelemahan yang terdapat pada versi sebelumnya. *PHP* versi 5 dapat membuat *file swf* dan *applet java*. Secara resmi, *PHP* versi 5 diluncurkan pada Desember 2003. Fokus utamanya adalah mengoptimalkan penggunaan *PHP* untuk *OOP (Object Oriented Programming)* (Syafii dan Andi Pramono dkk, 2005).

E. Definisi *MySQL*

MySQL adalah sebuah program *databaseserver* yang mampu menerima dan mengirimkan datanya dengan sangat cepat, *multi user* serta menggunakan perintah *SQL (Structured Query Language)*. *MySQL* merupakan sebuah *databaseserver* yang *free*, artinya kita akan bebas menggunakan *database* ini untuk keperluan pribadi atau usaha tanpa harus membeli atau membayar lisensinya. (Nugroho B,2005)

Keunggulan *MySQL* :

1. Tidak ada *memory- leak* serta pemakaian memori yang sangat.
2. *MySQL* dapat menangani *databaseserelasional* dan dapat dipakai untuk *client/ server*.
3. *SoftwareMySQL* adalah *open source*, artinya kita dapat mengambil, memakai, dan mengubah *source*-nya dengan bebas tanpa biaya.
4. *MySQL* sangat cepat. Ia *multithreaded* yaitu setiap *query* diperlakukan sebagai *thread* tersendiri. *MySQL* memakai sistem alokasi memori *thread- based* yang sangat cepat.
5. *MySQL* dapat diakses oleh *client* menggunakan *protocol TCP/IP* pada semua *platform*. Pada windows, *client* dapat mengakses menggunakan *named- pipe*. Sementara itu UNIX (Linux) dapat memakai *domainsocket- file*.

F. Definisi *Datamining*

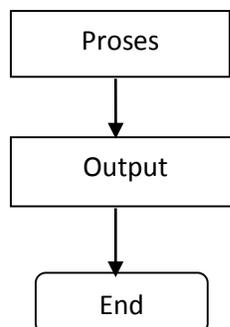
Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode *data mining* ini dapat digunakan untuk mengambil keputusan di masa depan. *Data mining* ini juga dikenal dengan istilah *pattern recognition*.

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis (*Santosa, 2007*).

G. Definisi *Flowchart*

Flowchart merupakan gambar atau bagan yang memperlihatkan urutan dan hubungan antar proses beserta instruksinya. Gambaran ini dinyatakan dengan simbol. Dengan demikian setiap simbol menggambarkan proses tertentu. Sedangkan hubungan antar proses digambarkan dengan garis penghubung.

Flowchart ini merupakan langkah awal pembuatan program. Dengan adanya *flowchart* urutan poses kegiatan menjadi lebih jelas. Jika ada penambahan proses maka dapat dilakukan lebih mudah. Setelah *flowchart* selesai disusun, selanjutnya pemrogram (*programmer*) menerjemahkannya ke bentuk program dengan bahasa pemrograman. Adapun alur *flowchart* dapat digambarkan pada gambar 2.1. berikut ini :



Sumber bayurha.blogspot.com/2014/03

Gambar 2. 1. Konsep *flowchart*

Terdapat 2 jenis *flowchart* yaitu sebagai berikut :

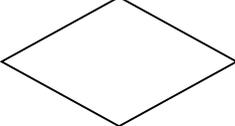
1. Sistem *Flowchart*

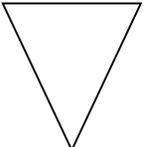
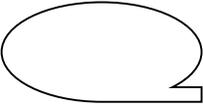
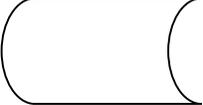
Sistemflowchart merupakan diagram alir yang menggambarkan suatu sistem peralatan komputer yang digunakan dalam proses pengolahan data serta hubungan antar peralatan tersebut. Sistem *flowchart* tidak digunakan untuk menggambarkan urutan langkah untuk memecahkan masalah , tetapi hanya untuk menggambarkan *prosedur* dalam sistem yang dibentuk.

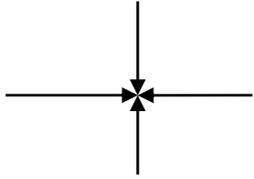
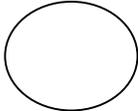
2. Program *Flowchart*

Merupakan bagan alir yang menggambarkan urutan logika dari suatu prosedur pemecahan masalah. Untuk menggambarakan flowchart program telah tersedia simbol-simbol standart. Berikut ini adalah gambar dari simbol-simbol standart yang digunakan pada flowchart program pada tabel 2.2. berikut ini :

Tabel 2.2. Simbol- simbol *flowchart*

No	Simbol	Fungsi
1		Simbol terminal, yaitu menyatakan permulaan atau akhir suatu program
2		Simbol proses, yaitu menyatakan suatu tindakan (proses) yang dilakukan oleh komputer
3		Simbol <i>output/ input</i> , untuk memasukkan data maupun menunjukkan hasil dari suatu proses
4		Simbol <i>decision</i> , yaitu menunjukkan suatu kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban : ya / tidak.
5		Simbol <i>predefined process</i> , yaitu menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberi harga awal

6		<p>Simbol <i>offline- storage</i>, menunjukkan bahwa data dalam simbol ini akan disimpan ke suatu media tertentu</p>
7		<p>Simbol <i>manual input</i>, memasukkan data secara manual dengan menggunakan <i>online keyboard</i></p>
8		<p>Simbol <i>manual</i>, yaitu menyatakan suatu tindakan(proses) yang tidak dilakukan oleh komputer</p>
9		<p>Simbol <i>punched card</i>, menyatakan <i>input</i> berasal dari kartu atau <i>output</i> ditulis kartu</p>
1 0		<p>Simbol <i>magnetic tape</i>, menyatakan <i>input</i> berasal dari kartu atau <i>output</i> disimpan ke <i>pita magnetis</i></p>
1 1		<p>Simbol <i>disk storage</i>, menyatakan <i>input</i> berasal dari disk atau <i>output</i> disimpan ke <i>disk</i></p>
1 2		<p>Dokumen merupakan simbol untuk data yang berbentuk informasi</p>

1 3		Simbol untuk <i>output</i> yang ditujukan suatu <i>device</i> , seperti <i>printer</i>
1 4		Arus atau <i>Flow-prosedur</i> yang dapat dilakukan dari atas kebawah, bawah keatas, dari kanan kekiri, dari kiri kekanan
1 5		Untuk menyimpan data
1 6		Untuk menyatakan sekumpulan langkah proses yang ditulis sebagai prosedur
1 7		<i>Connector</i> , suatu prosedur akan masuk dan keluar melalui simbol ini dalam lembar yang sama

Sumber bayurha.blogspot.com/2014/03

H. Definisi DFD (*Data Flow Diagram*)

DFD adalah suatu diagram yang menggunakan notasi-notasi untuk menggambarkan arus dari data sistem, yang penggunaannya sangat membantu untuk memahami sistem secara logika, tersruktur dan jelas. DFD merupakan alat bantu dalam menggambarkan atau menjelaskan.

Tujuan DFD adalah sebagai berikut:

1. Memberikan indikasi mengenai bagaimana data *ditransformasi* pada saat data bergerak melalui sistem.
2. Menggambarkan fungsi-fungsi dan sub fungsi yang *mentransformasi* aliran data.

I. Definisi ERD (*Entity Relationship Diagram*)

ERD merupakan suatu model untuk menjelaskan hubungan antar data dalam basisdata berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi.

Notasi- notasi simbolik yang digunakan dalam *Entity Relationship Diagram* adalah sebagai berikut:

1. Entitas

Entitas adalah segala sesuatu yang dapat digambarkan oleh data. Entitas juga dapat diartikan sebagai individu yang mewakili sesuatu yang nyata (eksistensinya) dan dapat dibedakan dari sesuatu yang lain.

2. Atribut

Atribut adalah pendeskripsian karakteristik dari entitas. Atribut digambarkan dalam bentuk lingkaran atau *elips*. Atribut yang menjadi kunci entitas atau key diberi garis bawah.

3. Relasi atau hubungan

Relasi menunjukkan adanya hubungan diantara sejumlah entitas yang berasal dari himpunan entitas yang berbeda.