

BAB II

TINJAUAN PUSTAKA

2.1 Kelulusan Studi

Mahasiswa yang memenuhi persyaratan kelulusan ditetapkan dalam yudisium kelulusan Fakultas atau Program Studi dan ditetapkan dengan keputusan Rektor. Tanggal kelulusan ditetapkan berdasarkan tanggal yudisium kelulusan dan merupakan tanggal penetapan IPK akhir (transkrip nilai). Untuk menentukan kelulusan dan predikat, fakultas menggunakan dokumen atau format resmi hasil penilaian studi mahasiswa yang sudah dicek dan disahkan kebenarannya oleh pihak tertentu pada masing-masing jurusan.

Kelulusan dan predikat ditetapkan berdasarkan Indeks Prestasi Kumulatif (IPK) hasil dari semua ujian mata kuliah yang telah diwajibkan untuk diselesaikan oleh para mahasiswa, sesuai dengan ketentuan program dan ketentuan Sistem Kredit Semester (SKS) nya. Termasuk hasil Ujian Skripsi atau Tugas Akhir. Syarat kelulusan mahasiswa Universitas Muhammadiyah Jember pada Fakultas Teknik Informatika dapat dilihat dalam Buku Panduan Akademik Mahasiswa Tahun Angkatan 2011.

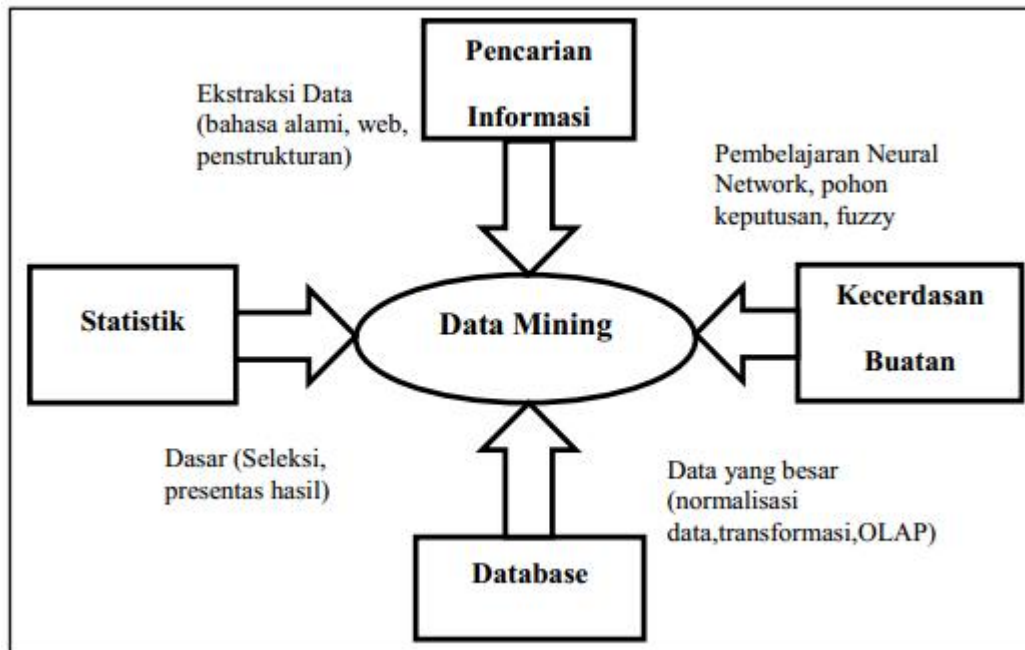
Setiap mahasiswa harus menyelesaikan beban studi yang telah ditetapkan. Beban studi mahasiswa dalam satu semester ditentukan atas dasar kemampuan akademik dan paket waktu yang tersedia bagi masing-masing mahasiswa. Beban studi program sarjana 140 SKS dan maksimal 154 SKS yang dijadwalkan 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan maksimal 14 semester. Sedangkan beban studi untuk Diploma III (tiga) minimal 110 SKS dan maksimal 120 SKS yang dijadwalkan 6 (enam) semester dan dapat ditempuh minimal 6 (enam) semester dan maksimal 10 (sepuluh) semester (Pedoman akademik Fakultas Teknik, 2011).

Berhasil tidaknya seorang mahasiswa belajar di Perguruan Tinggi dapat dilihat dari prestasi belajar yang diperolehnya. Tentu saja dalam mencapai gelar kesarjanaan membutuhkan waktu normal selama empat tahun, tetapi dalam prakteknya mahasiswa tidak selalu dapat menuntaskan studinya selama waktu normal yang telah ditentukan, dikarenakan beragamnya para mahasiswa yang memiliki bakat, minat, kompetensi dan motivasi yang berbeda-beda. Faktor-faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa dapat digolongkan dalam dua faktor: faktor internal dan eksternal. Faktor internal adalah faktor dari dalam diri mahasiswa yang mempengaruhi prestasi belajar, terdiri atas minat, intelegensi dan motivasi. Sedangkan faktor eksternal adalah faktor dari luar yang mempengaruhi prestasi belajar mahasiswa, terdiri dari peran dosen dalam mendidik mahasiswa, kondisi ekonomi orang tua mahasiswa dan metode pembelajaran yang digunakan (Lembang dan Fendjalang, 2015).

2.2 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban *et al*, 2005).

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar (Larose, 2005). Tan *et al*, (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*.



Gambar 2.1 Bidang ilmu data mining (Han *et al*, 2006)

Gambar 2.1 menunjukkan bahwa *data mining* memiliki akar panjang dari bidang ilmu seperti kecerdasan buatan (*Artificial Intelligent*), *machine learning*, statistik, *database*, dan juga informasi retrieval (Larose, 2006). 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. Salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*.

2.3 Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan. Perbedaan *Clustering* dengan grup, kalau grup berarti kelompok yang sama kondisinya kalau tidak ya pasti bukan kelompoknya. Tetapi kalau *cluster* tidak harus sama akan tetapi pengelompokannya berdasarkan kedekatan dari suatu karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak *euclidean*. Aplikasi *cluster* ini sangat banyak, karena hampir banyak dalam

mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja.

2.4 *K-Means*

K-Means merupakan algoritma untuk *cluster* objek berdasarkan atribut menjadi k partisi, dimana $k < n$. Secara Umum *K-Means Clustering* merupakan salah satu metode data *Clustering non-hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster* atau kelompok.

Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain.

Istilah-istilah dalam *K-Means* :

- a. **N data** : data set yang akan diolah sebanyak N data dimana N data tersebut terdiri dari atribut-atributnya
- b. **K centroid** : Inisialisasi dari pusat *cluster* data adalah sebanyak K dimana pusat-pusat awal tersebut digunakan sebagai banyaknya kelas yang akan tercipta. *Centroid* didapatkan secara random dari N data set yang ada.
- c. **Euclidian Distance**: merupakan jarak yang didapat dari perhitungan antara semua N data dengan K *centroid* dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut.

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} .$$

Keterangan :

x = Titik Data Pertama,

y = Titik Data Kedua,

n = Jumlah Karakteristik (Atribut) dalam Terminologi *Data Mining*,

$d(x,y)$ = *Euclidian Distance*, Jarak Antara Data Titik X Dan Titik Y menggunakan kalkulasi matematika

Kelemahan *K-Means* :

- a. Bila jumlah data tidak terlalu banyak, mudah untuk menentukan *cluster* awal.
- b. Jumlah *cluster*, sebanyak *K*, harus ditentukan sebelum dilakukan perhitungan.
- c. Tidak pernah mengetahui real *cluster* dengan menggunakan data yang sama, namun jika dimasukkan dengan cara yang berbeda mungkin dapat memproduksi *cluster* yang berbeda jika jumlah datanya sedikit.
- d. Tidak tahu kontribusi dari atribut dalam proses pengelompokan karena dianggap bahwa setiap atribut memiliki bobot yang sama.

Langkah-langkah dalam Algoritma *K-Means Clustering* :

1. Menentukan jumlah *cluster* .
2. Menentukan nilai *centroid*. Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut :

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} ,$$

3. Menghitung jarak antara titik *centroid* dengan titik tiap objek

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

4. Pengelompokan objek untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek.
5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

Contoh Kasus :

1. Dari banyak siswa diambil 12 siswa sebagai contoh untuk penerapan algoritma *K-Means* dalam penjurusan siswa sesuai dengan nilai rata-rata 8 mata pelajaran. Percobaan dilakukan dengan menggunakan parameter-parameter berikut :

- Jumlah *cluster* : 2
- Jumlah data :12
- Jumlah atribut : 8

Data (contoh kasus)

Tabel 2.1 Data Siswa

NO	NIS	Nama
1	1011.001	Anis Fadilah
2	1011.002	Didik Mulyana
3	1011.003	Eka Dwi Utami
4	1011.004	Fandik Lasmana
5	1011.005	Putri Khadijah
6	1011.006	Lina Nur Aini
7	1011.007	Mahmudin Ali
8	1011.008	M. Roemly
9	1011.009	Misbahus Surur
10	1011.010	Putra Adi Wijaya
11	1011.011	Syarif Kurniawan
12	1011.012	Ananda Anugrah

Tabel 2.2 Data Nilai Siswa

Mata Pelajaran								
NIS	Matematika	Fisika	Kimia	Biologi	Sejarah	Akutansi	Sosiologi	Geografi
1011.001	79	75	75	85	76	78	76	80
1011.002	84	76	79	77	76	77	75	81
1011.003	77	84	78	85	92	89	77	82
1011.004	78	86	84	77	78	77	75	75
1011.005	82	82	81	91	90	82	79	91
1011.006	75	75	70	89	75	75	79	75
1011.007	77	75	75	89	80	80	75	80
1011.008	77	76	70	77	77	80	75	80
1011.009	79	76	75	84	77	81	76	77
1011.010	80	75	75	75	75	78	77	79
1011.011	76	71	75	75	77	81	79	84
1011.012	80	77	75	78	78	77	77	80

Iterasi 1 (Contoh Kasus)

1. Penentuan pusat awal *cluster*

- a. Untuk penentuan awal di asumsikan:
- b. Diambil data ke- 2 sebagai pusat **Cluster Ke-1**: (84, 76, 79, 77, 76, 77, 75, 81)
- c. Diambil data ke- 5 sebagai pusat **Cluster Ke-2**: (82, 82, 81, 91, 90, 82, 79, 91)

2. Perhitungan jarak pusat *cluster*

- a. Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian distance*, kemudian akan didapatkan matrik jarak sebagai berikut :
- b. Rumus *Euclidian distance* : $d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- c. x = pusat *cluster*
- d. y = data

$$C1 = \sqrt{\frac{(84 - 79)^2 + (76 - 75)^2 + (79 - 75)^2 + (77 - 85)^2 + (76 - 76)^2}{(77 - 78)^2 + (75 - 76)^2 + (81 - 80)^2}} = 10.44$$

$$C2 = \sqrt{\frac{(82 - 79)^2 + (82 - 75)^2 + (81 - 75)^2 + (91 - 85)^2 + (90 - 76)^2}{(82 - 78)^2 + (79 - 76)^2 + (91 - 80)^2}} = 21,7$$

$$D^1 =$$

Tabel 2.3 Jarak terhadap Cluster

1	2	3	4	5	6	7	8	9	10	11	12	Pusat Cluster
10.4	0	24.1	14.1	24.02	15.6	16.4	11.8	11.1	6.8	12.2	6.5	C1
21.7	24.02	14.5	26.03	0	28.8	22.1	26.2	22.4	27.09	25.8	22.9	C2

- Setiap kolom pada matrik menunjukkan nilai jarak data terhadap pusat *cluster*. Baris pertama pada matrik menunjukkan nilai jarak data terhadap titik pusat *cluster* pertama, baris kedua pada matrik menunjukkan nilai jarak data terhadap titik pusat *cluster* kedua dan seterusnya.
- Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat.

Berikut ini akan ditampilkan data matriks pengelompokkan *group*, nilai 1 berarti data tersebut berada dalam *group*.

$$D^1 =$$

Tabel 2.4 *Group* Jarak terhadap Cluster

1	2	3	4	5	6	7	8	9	10	11	12	Pusat Cluster
1	1	0	1	0	1	1	1	1	1	1	1	C1
0	0	1	0	1	0	0	0	0	0	0	0	C2

Karena C1 memiliki 10 anggota maka perhitungan *cluster* baru menjadi sebagai berikut :

$$C1 = \left(\frac{79 + 84 + 78 + 75 + 77 + 77 + 79 + 80 + 76 + 80}{10}, \frac{75 + 76 + 86 + 75 + 75 + 76 + 76 + 75 + 71 + 77}{10}, \right. \\ \left. \frac{75 + 79 + 84 + 70 + 75 + 70 + 75 + 75 + 75 + 75}{10}, \frac{85 + 77 + 77 + 82 + 89 + 77 + 84 + 75 + 75 + 76}{10}, \right. \\ \left. \frac{76 + 76 + 78 + 75 + 80 + 77 + 77 + 75 + 77 + 76}{10}, \frac{78 + 77 + 77 + 75 + 80 + 80 + 81 + 78 + 81 + 77}{10}, \right. \\ \left. \frac{76 + 75 + 75 + 79 + 75 + 75 + 76 + 77 + 79 + 77}{10}, \frac{80 + 81 + 75 + 75 + 75 + 80 + 77 + 79 + 84 + 80}{10} \right)$$

$$C1 = (78.5; 76.2; 75.3; 79.9; 76.9; 78.4; 76.4; 78.6)$$

Karena C2 hanya memiliki 2 anggota *cluster* baru menjadi

$$\left(\frac{77 + 82}{2}, \frac{84 + 82}{2}, \frac{78 + 81}{2}, \frac{85 + 91}{2}, \frac{92 + 90}{2}, \frac{89 + 82}{2}, \frac{77 + 79}{2}, \frac{82 + 91}{2} \right)$$

$$C2 = (79.5; 83; 79.5; 88; 91; 85.5; 78; 86.5)$$

Iterasi 2 (Contoh Kasus)

Ulangi langkah ke 2 (kedua) hingga posisi data tidak mengalami perubahan.
Pengulangan langkah ke 2 hingga posisi data tidak mengalami perubahan.

a. Perhitungan Jarak dari data ke 1 terhadap pusat *cluster*

$$C1 = (78.5; 76.2; 75.3; 79.9; 76.9; 78.4; 76.4; 78.6)$$

$$C1 = \sqrt{(78.5 - 79)^2 + (76.2 - 75)^2 + (75.3 - 75)^2 + (79.9 - 85)^2 + (76.5 - 76)^2 + (78.4 - 78)^2 + (76.4 - 76)^2 + (78.6 - 80)^2}$$

$$= 5.5$$

$$C2 = \sqrt{(79.5 - 79)^2 + (83 - 75)^2 + (79.5 - 75)^2 + (88 - 85)^2 + (91 - 76)^2 + (85.5 - 78)^2 + (78 - 76)^2 + (86.5 - 80)^2}$$

$$= 20.5$$

b. Perhitungan Jarak dari data ke 2 terhadap pusat *cluster*

$$C2 = (79.5; 83; 79.5; 88; 91; 85.5; 78; 86.5)$$

$$C1 = \sqrt{(78.5 - 84)^2 + (76.2 - 76)^2 + (75.3 - 79)^2 + (79.9 - 87)^2 + (76.5 - 76)^2 + (78.4 - 77)^2 + (76.4 - 75)^2 + (78.6 - 81)^2}$$

$$= 7.9$$

$$C2 = \sqrt{(79.5 - 84)^2 + (83 - 76)^2 + (79.5 - 79)^2 + (88 - 77)^2 + (91 - 76)^2 + (85.5 - 77)^2 + (78 - 75)^2 + (86.5 - 81)^2}$$

$$= 22.9$$

c. Perhitungan seterusnya sampai Jarak dari data ke 12 terhadap pusat *cluster*. Sehingga hasil perhitungan jarak selengkapnya adalah :

$$D^1 =$$

Tabel 2.5 Jarak terhadap *Cluster*

1	2	3	4	5	6	7	8	9	10	11	12	Pusat <i>Cluster</i>
5.5	7.9	21.1	14.08	23.3	9	10.6	6.7	5.1	5.6	10	3.4	C1
20.5	22.9	7.2	23.13	7.2	26.6	19.5	23.3	19.8	24.9	23.8	21	C2

d. Lakukan pengelompokan data kembali sehingga dihasilkan matrik yang dimisalkan dengan G2.

$$G^1 =$$

Tabel 2.6 *Group* Jarak terhadap *Cluster*

1	2	3	4	5	6	7	8	9	10	11	12	Pusat <i>Cluster</i>
1	1	0	1	0	1	1	1	1	1	1	1	C1
0	0	1	0	1	0	0	0	0	0	0	0	C2

e. Karena G1 = G2 dimana anggota yang sama, maka tidak perlu dilakukan iterasi/perulangan lagi. Dan sampai disini hasil *Clustering* sudah mencapai stabil dan Konvergen.

Kesimpulan :

Hasil *Clustering* adalah sebagai berikut:

Cluster 1 : Siswa 1, 2, 4, 6, 7, 8, 9, 10, 11, 12

Cluster 2 : Siswa 3 dan 5

2.5 Bahasa Pemrograman PHP

PHP adalah singkatan dari "PHP: *Hypertext Preprocessor*", yang merupakan sebuah bahasa *scripting* yang terpasang pada HTML. Sebagian besar sintaks mirip dengan bahasa C, Java dan Perl, ditambah beberapa fungsi PHP yang spesifik. Tujuan utama penggunaan bahasa ini adalah untuk memungkinkan perancang web menulis halaman web dinamik dengan cepat. Program php harus diterjemahkan oleh *web-server* sehingga menghasilkan kode html yang dikirim ke *browser* agar dapat ditampilkan. Program ini dapat berdiri sendiri ataupun disisipkan diantara kode-kode html sehingga dapat langsung ditampilkan bersama dengan kode-kode html tersebut. File html yang telah dibubuhi program php harus diganti ekstensinya menjadi .php3 atau .php.

PHP pertama kali dibuat oleh Rasmus Lerdroft, seorang *programmer C*. Versi terbaru, yaitu PHP 4.0 keluar pada tanggal 22 Mei 2000 merupakan versi yang lebih lengkap lagi dibandingkan dengan versi sebelumnya. Tujuan dari bahasa *scripting* ini adalah untuk membuat aplikasi-aplikasi yang dijalankan diatas teknologi *web*. Dalam hal ini, aplikasi pada umumnya akan memberikan hasil pada *web browser*, tetapi prosesnya secara keseluruhan dijalankan *web server*. Kekuatan yang paling utama PHP adalah pada konektivitasnya dengan *system* database didalam *web*. Sistem *database* yang dapat didukung oleh PHP adalah: Oracle, MySQL, Sybase, Postgre SQL, dan lainnya. PHP dapat berjalan diberbagai *system* operasi seperti windows 98/NT, UNIX/LINUX, solaris maupun macintosh. Keunggulan lainnya dari PHP adalah bahwa PHP juga mendukung komunikasi dengan layanan seperti protocol IMAP, SNMP, NNTP, POP3 dan bahkan HTTP (Mulyanto, 2008).

2.6 Penelitian Terkait

Tabel 2.7 Penelitian Terkait

No	Judul	Nama	Tahun	Kesimpulan
1.	Penerapan Metode Klustering Dengan Menggunakan Algoritma <i>K-MEANS</i> Untuk Prediksi Kelulusan Mhasiswa Pada Program Studi Teknik Informatika Strata Satu	Gita Premashanti Trayasiwi	2011	Algoritma <i>K-MEANS</i> sangat membantu dalam menentukan kelulusan mahasiswa yang dapat mengelompokkan kedalam beberapa kluster berdasarkan jarak terdekat dengan beberapa parameter yaitu SKS dan IPK.
2.	<i>Clustering</i> Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode <i>K-MEANS</i>	Silvi Agustina	2010	Pada penelitian ini didapatkan tingkat akurasi 92,82% yang menunjukkan bahwa program ini dapat dijadikan acuan dalam klasterisasi kualitas beras.
3.	Perancangan Aplikasi <i>K-MEANS</i> Untuk Pengelompokan Mahasiswa STMIK ELRAHMA Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan dan IPK	Andri Syafrianto	2011	Berdasarkan penelitian yang dilakukan, bahwa algoritma <i>K-Means</i> dapat digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan frekuensi berkunjung ke perpustakaan dalam kurun waktu yang ditentukan.