**BAB II**

**TINJAUAN PUSTAKA**

**2.1. Tinjauan Umum**

**2.1.1. Sejarah Universitas Bina Darma**

Universitas Bina Darma berdiri atas penggabungan tiga Sekolah Tinggi berdasarkan Surat Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor : 112/D/0/2002 tanggal 7 Juni 2002 yaitu STMIK Bina Darma (Surat Keputusan Mendikbud RI. Nomor : 027/D/O/1994, tanggal 18 Mei 1994), STIE Bina Darma (SK. Mendikbud RI. Nomor : 046/D/O/1994, tanggal 7 Juli 1994), dan STBA Bina Darma (Surat Keputusan Menteri Pendidikan Nasional Republik Indonesia nomor : 143/D0/2001 tanggal 27 Agustus 2001).

Universitas Bina Darma adalah Perguruan Tinggi Swasta (PTS) yang mengasuh dan mengembangkan ilmu dan keahlian profesional pada tujuh fakultas (Fakultas Ilmu Komputer, Fakultas Bahasa dan Sastra, Fakultas Teknik, Fakultas Ilmu Komunikasi, Fakultas Keguruan dan Ilmu Pendidikan, Fakultas Ekonomi dan Fakultas Psikologi) dengan program studi unggulan tiap fakultas yang berada di Sumatera Selatan. Universitas Bina Darma memiliki komitmen dan tujuan untuk menciptakan lulusan yang siap bekerja dan dapat diterima di kalangan masyarakat. Maka dengan itu Universitas Bina Darma terus mengusahakan untuk mendapatkan Sertifikat dari Internasional *Organization for Standarization* (ISO 9001:2000), dan pada tanggal 7 Juli 2003 telah mendapatkan Sertifikat dengan nomor Registrasi 04100. 30981. Dengan telah ditentukan sistem manajemen mutu ISO 9001 : 2000 di Universitas Bina Darma telah sesuai dengan persyaratan dan peraturan yang telah berlaku, sehingga lulusannya dapat sesuai dengan apa yang telah diharapkan oleh masyarakat.

UBD secara aktif mengembangkan kerja sama di dalam maupun di luar negeri yang saat ini tercatat memiliki perjanjian dengan UBD diantaranya adalah: *Sun Microsystem*,Barring Edu Training Sdn Bhd-Malaysia, *University of Industri* Selangor (UNISEL) Malaysia, NIIT *Antilles* NV – Netherlands, *Pearson VUE Authorised Center-India*, *NPO International Japanese Education Center*, *Cisco Networking Academy, Stiching Hogeschool Zeeland* (HZ)-*Holland*, dan Planet Edupro Indonesia *(University of Cambridge English for Speaker of Other languages (ESOL) Authorised Main Center)*.

**2.1.2. Gambaran Umum Universitas Bina Darma**

 UPT-Sim merupakan unit pelaksana teknis yang memberikan dukungan kepada Universitas Bina Darma (UBD) dalam hal pengembangan dan penyediaan pelayanan ICT. Layanan yang diberikan pada saat ini adalah layanan pengolahan data dan penyediaan akses intra dan internet untuk seluruh civitas akademika UBD. Pada saat ini UPT-SIM mempunyai divisi-divisi antara lain : divisi *maintenance*, divisi pengembangan sistem, divisi pengolahan data, dan divisi *network administrator*. Setiap divisi mempunyai tugasnya masing-masing diantaranya, divisi pengembangan sistem memiliki tugas untuk membangun dan mengembangkan sistem agar dapat mendukung proses akademik dan non akademik di UBD.

 Layanan yang telah diberikan UPT- SIM untuk mendukung proses akademik dan non akademik dilingkungan UBD antara lain : UPT-SIM telah mendukung proses komputerisasi dengan membangun sistem akademik yang meliputi : entri data registrasi untuk mahasiswa baru dan lama, entri data nilai, proses penjadwalan, KHS Intranet dan KRS *Online*. Bukan hanya itu, UPT-SIM telah membangun sistem perhitungan KIDO/KIKAR, HRIS, Laman binadarma.ac.id , *e-Learning* , layanan *email* dan *hosting* untuk beberapa unit kerja seperti perpustakaan (otomasi.binadarma.ac.id dan digilib.binadarma.ac.id), ELC (elc.binadarma.ac.id), laman *web* , BDCTC (bdctc.binadarma.ac.id), dan beberapa laman *web* program studi. Selanjutnya UPT-SIM UBD juga telah memberikan layanan akses data / *internet* bagi staf / dosen dan mahasiswa melalui kabel dan nirkabel (*wireless WIFI*) yang dapat diakses diseluruh lingkungan UBD.

**2.1.3. Visi dan Misi**

**2.1.3.1. Visi**

 Visi Universitas Bina Darma ialah untuk menjadi Universitas Berstandar Internasional Berbasis Teknologi Informasi pada tahun 2020.

**2.1.3.2. Misi**

 Misi utama Universitas Bina Darma ialah menghasilkan lulusan yang cerdas, professional, dan berkarakter yang berdaya saing Internasional. Misi UBD yaitu :

1. Menghasilkan Lulusan yang Professional, Cerdas, dan Berkarakter yang Berdaya Saing Internasional.
2. Menyelenggarakan program pendidikan yang berstandar Internasional.
3. Menyelenggarakan proses pembelajaran yang berstandar Internasional melalui pemanfaatan teknologi informasi.
4. Membangun komunitas intelektual yang berkualitas.
5. Melakukan penelitian yang berstandar Internasional.
6. Melakukan pengabdian guna meningkatkan kemandirian masyarakat.
7. Menyelenggarakan kerja sama dengan pihak yang lain saling menguntungkan.

**2.1.4. Struktur Organisasi Universitas Bina Darma**

****

Sumber: http://binadarma.ac.id/content/120/0/miscuts.html

**Gambar 2.1.** Struktur Organisasi UPT-SIM Universitas Bina Darma

**2.2. Landasan Teori**

**2.2.1. *Data Mining***

*Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam sistem database. *Data Mining* merupakan proses yang menggunakan *machine learning* untuk mengekstraksi , matematika, teknik statistik, kecerdasan buatan, dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini,Luthfi. 2009).

*Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono,2003).

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining*  didorong oleh beberapa faktor, antara lain (Larose, 2005) :

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Adanya peningkatan akses data melalui navigasi *web* dan *intranet*.
3. Penyimpanan data dalam *data* *warehouse*, sehingga semua perusahaan memiliki akses ke dalam *database* yang andal.
4. Perkembangan teknologi perangkat lunak untuk *data mining*.
5. Perkembangan yang hebat dalam pengembangan kapasitas media penyimpanan dan kemampuan komputasi.
6. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar di dalam globalisasi ekonomi.

**2.2.2. Pengelompokkan *Data Mining***

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Kusrini,Luthfi. 2009).

1. Deskripsi, terkadang analis dan peneliti dengan cara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi, estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai variable target sebagai nilai prediksi.
3. Prediksi, prediksi juga sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi, dalam klasifikasi, memiliki target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan menjadi tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. Pengklusteran, pengklusteran merupakan pengamat, pengelompokkan *record*, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang mempunyai kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan *record-record* dalam kluster lain.
6. Asosiasi, asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih sering disebut analisis keranjang belanja.

**2.2.3. Fungsi-fungsi *Data Mining***

Fungsi-fungsi yang umum diterapkan dalam *data mining* :

1. *Association*, adalah proses untuk menemukan aturan assosiatif antara suatu kombinasi item dalam semua waktu.
2. *Secuence*, hampir sama dengan association bedanya secuence diterapkan lebih dari satu periode.
3. *Clustering*, adalah proses pengelompokkan sejumlah data/objek ke dalam kelompok-kelompok data (klaster) sehingga setiap klaster akan berisi data yang saling mirip.
4. *Classification*, merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. *Regretion*, merupakan proses pemetaan data di dalam suatu nilai prediksi.
6. *Forecasting*, adalah proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.
7. *Solution*, adalah proses penemuan akar masalah dan problem solving dari persoalan bisnis yang dihadapi atau paling tidak sebagai informasi pendukung dalam pengambilan keputusan.

**2.2.4. Tugas *Data Mining***

Terdapat empat tugas pokok *data mining*, yaitu:

1. *Predictive Modelling,* digunakan untuk membangun sebuah model untuk target *variable* sebagai fungsi dari *explanatory variable*. *Explanatory variable* dalam hal ini adalah semua atribut yang digunakan untuk melakukan prediksi, sedangkan *variabel* target merupakan atribut yang akan diprediksi nilainya. *Predictive modelling* dibagi menjadi dua tipe yaitu *classification* digunakan untuk memprediksi nilai dari target variable yang *discrete* (diskrit), dan *regression* digunakan untuk memprediksi nilai dari target *variable* yang *continue* (berkelanjutan).
2. *Association Analysis,* berguna untuk menemukan aturan asosiasi yang memperlihatkan kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.
3. *Cluster Analysis,* tidak seperti klasifikasi yang menganalisa kelas data objek yang mengandung label. *Clustering* menganalisa data objek tanpa memeriksa kelas label yang diketahui. Label-label kelas diikutkan dalam *training data*. Karena belum diketahui sebelumnya bahwa *clustering* merupakan proses pengelompokkan sekumpulan objek yang sangat mirip.
4. *Anomaly Detection,* merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar lain objek. *Anomaly* bisa di deteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.

**2.3. Metode *Clustering***

 *Clustering* termasuk metode yang sudah cukup dikenal dan banyak dipakai dalam *data mining*. Sampai saat ini 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 banyak jumlah *cluster* yang optimal dan pengklasteran yang paling baik masih terus dilakukan.

****

**Gambar 2.2.** *Clustering*

 Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data/obyek ke dalam klaster(*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin seperti diilustrasikan pada gambar 2.7.

Metode ini berusaha untuk menempatkan objek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin. Ini berarti bahwa objek dalam satu klastersangat mirip satu sama lain dan berbeda dengan obek dalam klaster-klasteryang lain. Dalam metode ini tidak diketahui sebelumnya berapa jumlah *cluster* dan bagaimana pengelompokannya (Santoso, 2007).

**2.4. Sistem Pendukung Keputusan-SPK (*decision support system-*DSS)**

Merupakan suatu sistem informasi yang interaktif dan fleksibel yang memungkinkan para admin mendapatkan dan memanupulasi informasi pada saat mereka mengambil keputusan. SPK melampaui pemrosesan informasi oleh ahlinya dan memberikan para manajer akses terhadap data yang berguna.

Beberapa karakteristik dari SPK, yaitu :

* 1. *Interaktif* : Para manajer memberikan intruksi yang sederhana dan melihat hasilnya segera. Proses ini berada dibawah pengawasanya langsung.
	2. *Fleksibel*  : suatu SPK dapat menyaring, mengelompokkan kembali, menjumlah, merata-ratakan, dan memanipulasi data dalam berbagai cara. SPK akan memindahkan roda persneling karena berubahnya topik pemakai, kesesuaian informasi terhadap masalah yang sedang dihadapi.
	3. *Orientasi Penemuan* : para manajer dapar mengamati trend, memilah masalah, dan menanyakan pertanyaan :bagaimana jika”.
	4. *Kemampuan Akses* : SPK mudah dipelajari dan digunakan oleh para manajer yang tidak memiliki keahlian komputer.

**2.5. Fuzzy C-Means**

*Cluster* secara umum merupakan wujud himpunan dari bagian suatu himpunan data dan metode *clustering* dapat diklasifikasikan berdasarkan himpunan bagian yang dihasilkan: apakah *fuzzy* atau *crisp (hard)*.

Metode *hard clustering* merupakan model yang berdasar pada teori himpunan klasik, yang mana suatu objek dapat menjadi anggota atau tidak menjadi anggota secara penuh kedalam suatu kelompok. *Hard clustering*  membagi data ke sejumlah himpunan bagian secara eksklusif.

Sebaliknya, metode *fuzzy clustering* mengizinkan objek untuk menjadi bagian dari beberapa kelompok secara bersamaan dengan perbedaan level keanggotaan. (Kusrini,Luthfi. 2009).

 Terdapat banyak algoritma yang digunakan untuk *clustering*, Salah satunya *Fuzzy C-Means* (FCM). *Fuzzy C-Means Clustering* (FCM) merupakan algoritma *clustering* data yang setiap datanya menjadi anggota dari suatu kluster dengan derajat didefinisikan dengan level keanggotaan.

Algoritma FCM berdasarkan minimasi fungsi objektif yang diformulasikan dalam persaman :

$$J\_{m}=\sum\_{i=1 j=1}^{N}\sum\_{j=1}^{c}\left(U\_{ij}\right)^{m} \left‖Z\_{i}-C\_{j}\right‖^{2}$$

Dengan :

*Uij* merupakan level keanggotaan dari Zi di dalam *Cluster* *j* 2

*Zi* merupakan nilai ke-*I* dari d-dimensi data

*Cj* merupakan nilai ke-*j* dari d-dimensi data *cluster* *center*

*M* merupakan sembarang dari bilangan *real* lebih besar dari 1

Selanjutnya algoritma FCM disusun berdasarkan langkah sebagai berikut :

Langkah 1 : Tentukan himpunan data Z. Tentukan jumlah nilai *cluster* yang        diharapkan 1 < c < N, nilai pembobot m > 1, toleransi

 penghentian        €      > 0.

Langkah 2 : Inisialisasi matriks partisi secara acak, U(0)€Mfc. Ulangi untuk                          1=3,4,5

Langkah 3 : Hitung *cluster center* (*means*).

$V\_{i}^{(i)}=\frac{\sum\_{K=1}^{N}\left(U\_{ik}^{(i-1)}\right)^{m}Z\_{k}}{\sum\_{K=1}^{N}\left(U\_{ik}^{(i-1)}\right)^{m}}, 1 \leq i \leq c$

Langkah 4 : Hitung jarak

 $D\_{ikA=\left(Z\_{k}-V\_{i}^{(i)}\right)^{T}A \left(Z\_{k}-V\_{i}^{(i)}\right), 1 \leq i \leq c, 1 \leq k N }^{2}$

Langkah 5 : Perbaharui matriks partisi.

 Untuk 1 ≤ K ≤ N

 Jika $D\_{ikA}>0 untuk i=1,2,…., C$

 $u\_{iK}^{(i)}=\frac{1}{\sum\_{j}^{c}=1 \left(D\_{ikA}/D\_{jKA}\right)^{2/(m-1)}} ,$

 Atau dengan kata lain :

$u\_{iK }^{(i)}=0 jika D\_{ikA }>0 dan u\_{iK}^{(i)} \in \left[0,1\right] dengan \sum\_{i=1}^{c}u\_{iK}^{\left(i\right) }=1$

Ulangi sampai $\left‖U^{\left(i\right)}- U^{(i-1)}\right‖ < \in .$

**2.6. Metode *Knowledge Discovery in Databases* (*KDD)***

Proses KDD merupakan proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database (Azevedo, 2008).

Istilah *data mining* dan *knowledge discovery in databases* (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut mempunyai konsep yang berbeda, tetapi memiliki keterkaitan satu sama lain. Dan salah satu tahapan di dalam proses KDD yaitu *data mining.* Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Kusrini, Luthfi 2009).



**Gambar 2.3.** Tahapan *Knowledge Discovery in Databases*

1. *Data Selection,*pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *KDD* dimulai. Data hasil seleksi yang digunakan pada proses *data mining,* disimpan di dalam suatu berkas, yang terpisah dari basis data operasional.
2. *Pre-processing/Cleaning,*sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses cleaning untuk data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain yakni memeriksa data yang inkonsisten, membuang duplikasi data, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.
3. *Transformation, coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut dapat sesuai untuk proses *data mining.* Proses *coding* dalam KDD adalah proses kreatif dan sangat tergantung pada sejenis 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 di dalam *data mining* sangat bervariasi. Pemilihan metode ataupun 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 adalah bagian dari proses KDD yang disebut *interpretation.* Tahap ini mencakup pemeriksaan apakah dari pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

**2.7. *XLMiner***

 *XLMiner* adalah hanya komprehensive *data mining add-in* pada *excel*. Dengan *regression trees,* *neural nets, and classification and principal components*, *association rules, clustering k-nearest neighbors, discriminant analysis, bayes classifier,* dan lainnya.

 *XLMiner* menyediakan segala yang dibutuhkan untuk sample data dari beberapa sumber *powerpoint*, *spreadsheet* atau *Microsoft/IBM/Oracle database*, mengevaluasi daya prediksi, *fit* *model* *data* *mining*, dan mengevaluasi daya prediksi, mengeksplorasi dan memvisualisasikan data dengan grafik berhubungan dengan *preprocess* dan *cleaning* data.

Gunakan *XLMiner* sesuai dengan kebutuhan karena,

1. *XLMiner Professional* edisi dirancang untuk praktisi dan peneliti yang menangani dataset besar.
2. *XLMiner Education*, dengan batas ukuran khusus pada dataset dan algoritma, hanya tersedia bagi penggunaan intruksional (dosen dan mahasiswa).(Frontline, 2012).

**2.8. Penelitian Sebelumnya**

Penelitian sebelumnya digunakan untuk dapat dijadikan bahan pertimbangan dan diharapkan dapat membantu dalam pembuatan sistem yang baru.

1. *Data Mining* Dengan Teknik *Clustering* Dalam Pengklasifikasian Data Mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara, Lindawati. Banyaknya data yang dipunyai sebuah organisasi bisa menyebabkan kesulitan dalam pengklasifikasian data tersebut untuk kepentingan di organisasi. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia untuk menampung jumlah data yang ingin diklasifikasikan. Selain itu dapat menyebabkan terjadinya kesalahan dalam pengklasifikasian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik *clustering*. Penelitian yang sudah dilakukan untuk mengklasifikasikan data prediksi lama studi mahasiswa di Universitas Bina Nusantara ini ditujukan untuk merancang dan mengimplementasikan Data Mining dengan teknik clustering menggunakan algoritma *self* *organizing maps*. Tujuannya adalah untuk menguji keakuratan *clustering* dapat dengan menggunakan *algoritma self organizing maps*.
2. Penerapan Data Mining Untuk Memprediksi Pemilihan Mata Kuliah Semester Pendek Mahasiswa Universitas Bina Darma, Darmawiansyah 2013. Pemanfaatan teknologi informasi memungkinkan terjadinya data dalam jumlah yang besar. Universitas Bina Darma merupakan organisasi yang memanfaatkan teknologi informasi, khususnya basis data, akan mengalami akumulasi data mahasiswa dalam jumlah besar tiap tahunnya. Basis data mahasiswa binadarma berisi data dalam jumlah besar tetapi belum dimanfaatkan secara optimal. Diperlukan suatu sistem yang bisa dimanfaatkan gunungan data menjadi informasi yang bernilai strategis. Dalam masalah ini dibahas suatu sistem pendukung bagi perguruan tinggi tentang penerapan data mining untuk mempredikis pemilihan mata kuliah semester pendek mahasiswa. Sistem yang dibangun menggunakan teknik *data mining* bertujuan untuk menggali dan menemukan pola-pola yang tersembunyi dalam pengolahan data mata kuliah semester pendek mahasiswa tiap pertahunnya. Untuk memperoleh data yang real dalam penerapan data mining dalam mata perkuliahan mahasiswa pada data base menggunakan *metode clustering*. Hasil dari proses *mining* ini dapat membantu untuk mengetahui statistik mata kulia semester pendek mahasiswa pertahunnya. Informasi yang dihasilkan dapat digunakan sebagai analisis dalam pengambilan keputusan.