

# PENGELOMPOKAN MINAT BELAJAR MAHASISWA MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE CLUSTERING

Marlindawati <sup>1</sup>, Andri <sup>2</sup>

Dosen Universitas Bina Darma <sup>1,2</sup>

Jalan Jenderal Ahmad Yani No.12 Palembang

Pos-el : [marlinda@gmail.com](mailto:marlinda@gmail.com) <sup>1</sup>, [andri@mail.binadarma.ac.id](mailto:andri@mail.binadarma.ac.id) <sup>2</sup>

---

*Abstrak* : Mengikuti perkembangan teknologi informasi yang berhubungan dengan media penyimpanan saat ini, data dapat disimpan dalam bentuk digital. Dalam sebuah perguruan tinggi jumlah mahasiswa akan terus bertambah seiring berjalannya waktu. Hal ini akan menyebabkan data yang tersimpan terutama data yang berhubungan dengan akademik mahasiswa akan terus bertambah. Jika tidak dimanfaatkan dengan maksimal maka akan terjadi kemubaziran terhadap data tersebut. Dengan memanfaatkan teknik data mining maka didapat informasi yang berguna dari tumpukan data yang ada. Data mining merupakan kegiatan yang bertujuan untuk menggali informasi dari tumpukan data yang besar. Data Mining merupakan proses analisis data menggunakan perangkat lunak untuk menemukan pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat menemukan pengetahuan guna mendukung keputusan. Dalam penelitian ini akan dilakukan mining data yang berhubungan dengan akademik mahasiswa untuk dapat mengetahui minat belajar dari mahasiswa yang ada dalam perguruan tinggi. Penelitian ini akan menghasilkan sebuah model yang dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajarnya. Teknik data mining yang digunakan dalam penelitian ini adalah teknik clustering.

Kata Kunci: teknologi informasi, data mining, clustering

---

## I. PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi terutama yang berhubungan dengan media penyimpanan saat ini berkembang begitu pesat. Media penyimpanan data dapat digunakan oleh sebuah perguruan tinggi untuk menyimpan data-data yang berhubungan dengan akademik mahasiswa. Data bagi suatu perguruan tinggi sangat bermanfaat jika diolah lebih lanjut untuk menghasilkan sebuah informasi yang dapat digunakan dalam pengambilan keputusan bagi pucuk pimpinan perguruan tinggi tersebut.

Data merupakan sumber utama yang dapat dimanfaatkan untuk menghasilkan informasi. Informasi begitu penting bagi sebuah perguruan tinggi untuk dapat tetap bertahan dalam menghasilkan lulusan yang terbaik.

Pengolahan data yang maksimal maka akan menghasilkan informasi yang berkualitas.

Seiring dengan berjalannya waktu, data dalam sebuah perguruan tinggi akan terus bertambah sehingga akan menjadi tumpukan data yang menggunung. Jika tidak dikelola dengan baik maka data tersebut hanya berupa data yang tidak ada manfaatnya. Dengan memanfaatkan teknik data mining maka akan didapatkan informasi yang berguna dari tumpukan data yang ada.

Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Menurut Pramudiono (2007) data mining dapat juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang

selama ini tidak diketahui secara manual dari kumpulan data.

Data mining dapat dimanfaatkan untuk menggali informasi dari tumpukan data akademik yang berhubungan dengan mahasiswa untuk menemukan informasi atau pola dari minat belajar mahasiswa. Informasi yang didapat dari teknik mining data ini adalah sebuah keputusan yang dapat dimanfaatkan pihak yang berwenang dalam perguruan tinggi untuk mengarahkan konsentrasi mahasiswa tersebut.

Berdasarkan fenomena inilah maka akan dilakukan sebuah penelitian yang bertujuan untuk membuat sebuah model data mining yang dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajarnya.

## **1.2 Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah bagaimana membuat model data mining yang digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajar matakuliah pilihan pada program studi sistem informasi dengan menggunakan metode clustering?

## **1.3 Batasan Masalah**

Agar penelitian lebih terarah dan memudahkan dalam pembahasan, maka perlu adanya pembatasan masalah, yaitu:

1. Data yang digunakan dalam penelitian ini adalah data mahasiswa program studi sistem informasi fakultas Ilmu Komputer Universitas Binadarma
2. Data penelitian berasal dari data mahasiswa program studi Sistem Informasi tahun angkatan 2011-2013.

3. Teknik data mining yang digunakan yaitu *Clustering*.

## **1.4 Tujuan Penelitian dan Manfaat Penelitian**

### **1.4.1 Tujuan Penelitian**

Tujuan dari penelitian ini adalah membuat model data mining untuk melakukan pengelompokan mahasiswa program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Binadarma Palembang berdasarkan minat belajarnya.

### **1.4.2 Manfaat Penelitian**

Penelitian diharapkan dapat menghasilkan sebuah model data mining yang dapat dimanfaatkan untuk menentukan konsentrasi minat belajar mahasiswa yang ada pada program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Binadarma Palembang.

## **II. METODOLOGI PENELITIAN**

### **2.1 Metode Pengumpulan Data**

Metode yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Pengamatan (*Observasi*)
2. Wawancara (*Interview*)
3. Studi Pustaka

### **2.2 Data Warehouse**

Menurut W.H. Inmon dan Richard D.H (1994), *data warehouse* adalah koleksi data yang mempunyai sifat berorientasi subjek, terintegrasi, time-variant, dan bersifat tetap dari kumpulan data dalam mendukung proses pengambilan keputusan manajemen.

Menurut Vidette Poe (1997), *data warehouse* merupakan *database* yang bersifat analisis dan *read only* yang digunakan sebagai fondasi dari sistem penunjang keputusan.

Menurut Paul Lane (2002), *data warehouse* merupakan *database* relasional yang didesain lebih kepada *query* dan analisis daripada proses transaksi, biasanya mengandung *history* data dari proses transaksi dan bisa juga data dari sumber lainnya.

Menurut Poniah (2001) *data warehouse* adalah suatu paradigma baru dilingkungan pengambilan keputusan strategik. *Data warehouse* bukan suatu produk tetapi suatu lingkungan dimana *user* dapat menemukan informasi strategik. *Data warehouse* adalah kumpulan data-data logik yang terpisah dengan *database* operasional dan merupakan suatu ringkasan. Karakteristik *data warehouse* terdiri dari:

#### 1. Berorientasi subyek

*Data warehouse* adalah tempat penyimpanan berdasarkan subyek bukan berdasarkan aplikasi. Subyek merupakan bagian dari suatu perusahaan. Contoh subyek pada perusahaan manufaktur adalah penjualan, konsumen, inventori, dan lain sebagainya. Gambar dibawah ini merupakan perbedaan mengenai *data warehouse* dan *database* operasional.

#### 2. Data yang terintegrasi

Sumber data yang ada dalam *data warehouse* tidak hanya berasal dari *database* operasional (*internal source*) tetapi juga berasal dari data diluar sistem (*external source*). Data pada sumber berbeda dapat di-*encode* dengan cara yang berbeda.

#### 3. *Nonvolatile*

Data dalam *database* operasional akan secara berkala atau periodik dipindahkan kedalam *data warehouse* sesuai dengan jadwal yang sudah ditentukan. Misal perhari, perminggu, perbulan, dan lain sebagainya. Sekali masuk ke dalam *data warehouse*, data adalah *read-only*.

#### 4. *Time-Variant*

Sistem operasional mengandung data yang bernilai sekarang sedangkan data dalam *data warehouse* mengandung data tidak hanya data terkini tetapi juga data *history* yang akan digunakan dalam analisis dan pengambilan keputusan. Waktu adalah dimensi penting yang harus didukung oleh semua *data warehouse*. Data untuk analisis dari berbagai sumber berisi berbagai nilai waktu, misalkan harian, mingguan, dan bulanan.

#### 5. Ringkas

Jika diperlukan, data operasional dikumpulkan ke dalam ringkasan-ringkasan.

#### 6. *Granularity*

Pada sistem operasional data dibuat secara *real-time* sehingga untuk mendapatkan informasi langsung dilakukan proses *query*. Pada *data warehouse* pada menganalisis harus memperhatikan *level-of-detail* misalkan perhari, ringkasan perbulan, ringkasan per-tiga-bulan.

#### 7. Tidak ternormalisasi

Data di dalam sebuah *data warehouse* biasanya tidak ternormalisasi dan sangat redundan.

*Data warehouse* mengandung beberapa elemen penting antara lain (Mallach, 2000):

1. Sumber data yang digunakan oleh *data warehouse*, database transaksional dan sumber data eksternal.
2. Proses ETL (*Extraction, Transformation, Loading*) dari sumber data ke *database data warehouse*.
3. Membuat suatu ringkasan atau *summary* terhadap *data warehouse* misalkan dengan menggunakan fungsi agregat.
4. Metadata, metadata mengacu data tentang data. Metadata menguraikan struktur dan beberapa arti tentang data, dengan demikian mendukung penggunaan efektif atau tidak efektif dari data.
5. *Database data warehouse*. *Database* ini berisi data yang detail dan ringkasan data dari data yang ada di dalam *data warehouse*. Karena *data warehouse* tidak digunakan dalam proses transaksi individu, maka *databasenya* tidak perlu diorganisasikan untuk akses transaksi dan untuk pengambilan data, melainkan dioptimisasikan untuk pola akses yang berbeda di dalam analisis.
6. *Query Tools* yaitu dengan OLAP (*Online Analytical Processing*) dan *data mining*. *Tool* untuk *query* ini meliputi antarmuka pengguna akhir dalam mengajukan pertanyaan kepada *database*, dimana proses ini disebut sebagai *On-line Analytical Processing* (OLAP). *Tool* ini juga terdiri dari *tool* otomatis yang menemukan pola-pola di dalam data, yang sering disebut sebagai *data mining*. *Data warehouse* harus memiliki salah satu dari kedua tipe ini atau malah kedua-duanya.

7. *User*, merupakan pengguna yang memanfaatkan *data warehouse* tersebut.

Proses ETL (*Extraction, Transformation, dan Loading*) merupakan proses yang harus dilalui dalam pembentukan *data warehouse*, yaitu:

1. *Extraction*

Ekstraksi dilakukan dari sumber data yang digunakan melalui proses pemilihan data yang kemudian disimpan pada basis data sementara. Dalam penelitian ini, basis data sementara ditempatkan pada penyimpanan basis data, mesin dan *platform* yang sama dengan yang digunakan untuk *data warehouse* nantinya.

2. *Cleaning*

Proses *cleaning* adalah proses untuk membersihkan data yang rangkap, tidak konsisten, rusak, tidak benar (bernilai salah), data yang hilang (kosong), atau data dengan spasi yang berlebihan.

3. Transformasi

Proses transformasi yang dilakukan dibagi berdasarkan dua *level* yaitu *record level* dan *field level*. Pada proses ini dilakukan proses pemilihan, penggabungan dan agregasi untuk mendapatkan data ringkasan sesuai dengan dimensi yang akan dibuat. Proses transformasi ke *field* yang baru dapat dilakukan dengan (menggunakan *script*) tertentu untuk melakukannya.

4. *Loading*

Proses terakhir yang perlu dilakukan adalah proses pemuatan data (*loading*).

Data yang digunakan pada tahap ini merupakan data dari proses-proses yang dilakukan sebelumnya yaitu ekstraksi, pembersihan (*cleaning*) dan transformasi untuk kemudian dimasukkan ke dalam *data warehouse*. Cara pemuatan data ke dalam *data warehouse* adalah dengan menggunakan *script* yang dijalankan secara periodik.

### 2.3 Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Data mining dapat juga disebut proses menemukan suatu pola dalam data. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan dan data yang dibutuhkan dalam jumlah besar (Witten, 2005).

Menurut Davies (2004), data mining memiliki beberapa karakteristik, yaitu:

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. Data mining biasanya menggunakan data yang sangat besar yang digunakan untuk membuat hasil yang lebih dipercaya.
- c. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi

Data mining juga disebut dengan *Knowledge Discovery in Database (KDD)*, yaitu kegiatan yang meliputi pengumpulan, pemakaian data untuk menemukan keteraturan, pola dan hubungan dalam set data yang berukuran besar.

Keluaran data mining dapat dipakai untuk pengambilan keputusan.

*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006)

*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)

Secara garis besar *data mining* dapat dikelompokkan menjadi 2 kategori utama, yaitu (Tan *et al*, 2005) :

1. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

### 2.4 Tahapan dalam Data Mining

Menurut Han (2006) terdapat enam tahap dalam data mining:

- a. Pembersihan data (*cleaning*)  
Data *cleaning* merupakan proses untuk menghilangkan noise serta data yang tidak konsisten atau tidak relevan.
- b. Integrasi data (*integration*)

Integrasi data bertujuan untuk menggabungkan data dari berbagai sumber basisdata dalam satu basisdata yang baru.

- c. Seleksi data (selection)  
Seleksi data bertujuan untuk memilih data-data apa saja yang akan diproses selanjutnya.
- d. Transformasi data (transformation)  
Transformasi data bertujuan untuk mengubah data kedalam format tertentu yang sesuai untuk proses selanjutnya.
- e. Proses mining (data mining)  
Proses mining adalah proses untuk menerapkan metode data mining yang bertujuan untuk menemukan pengetahuan yang baru.
- f. Evaluasi pola (pattern evaluation)  
Evaluasi bertujuan untuk mengidentifikasi pola-pola yang menarik kedalam knowledge base yang ditemukan.
- g. Presentasi pengetahuan (knowledge presentation)  
Presentasi pengetahuan merupakan penyajian pengetahuan untuk memperoleh pengetahuan yang diperoleh pengguna.

## 2.6 *Business Intelligence Development Studio (BIDS)*

Business Intelligence Development Studio (BIDS) adalah Tool yang digunakan untuk mendesain, membuat, dan mengetes package. BIDS menyediakan sebuah lingkungan pengembangan yang berorientasi grafis yang membantu proses pengkopian, pengaturan dan pembuatan package menggunakan menu dan

kotak Tool dengan metode drag-and-drop. beberapa tools yang ada dalam BIDS yang digunakan dalam penelitian meliputi:

- a. SQL Server Integration Service (SSIS)  
SQL Server Integration Service merupakan sebuah tools yang digunakan untuk melakukan proses ETL (Extract, Transform, and Load).
- b. SQL Server Analysis Service (SSAS)  
SSAS merupakan tools yang digunakan untuk melakukan proses OLAP dan Data Mining.

## 2.7 *Clustering*

*Clustering* atau analisis kelompok merupakan pekerjaan mengelompokkan data (objek) berdasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Tan, 2006).

Tujuan dari pengelompokan ini adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok yang lain.

Tujuan pekerjaan pengelompokan (clustering) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan (Prasetyo, 2012). Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata, standar deviasi),

pelabelan kelas pada setiap kelompok yang untuk kemudian digunakan sebagai data latu klasifikasi, dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokkan biasanya adalah mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

## 2.8 K-Means

Dalam statistic dan mesin pembelajaran, pengelompokan K-Means merupakan metode analisis kelompok yang merngarah pada pemartisian N objek pengamatan ke dalam K kelompok (cluster) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma Expectation-Maximization untuk Gaussian Mixture dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan ang dilakukan oleh kedua algoritma. Pengelompokan data dengan metode K-Means ini secara umum dilakukan dengan langkah berikut (Prasetyo, 2012):

1. Tentukan jumlah kelompok
2. Alokasikan data ke dalam kelompok secara acak
3. Hitung pusat kelompok (sentroid/rata-rata) dari data yang ada di masing-masing kelompok
4. Alokasikan masing-masing data ke sentroid/rata-rata tersebut
5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok, atau

apabila ada perubahan nilai sentroid diatas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

## 2.9 Rancangan Database

- a. tb\_mhs

**Tabel 1.** Tabel Mahasiswa

Nama	Type	Width
nim*	varchar	9
nama	varchar	40
jenis_kelamin	varchar	2
temp_lahir	varchar	50
tgl_lahir	date	10
kd_progdi	varchar	3
alamat	varchar	250
asal_sek	varchar	20
kota	varchar	25
pek_ortu	varchar	20

- b. tb\_khs

**Tabel 2.** Tabel Kartu Hasil Studi (KHS)

Nama	Type	Width
nim*	varchar	9
kd_mk	varchar	10
sms	varchar	1
sms_pendek	varchar	1
tahun_akademik	varchar	9
kelas	varchar	5
kd_dosen	varchar	5
kd_progdi	varchar	3
tugas	int	6
kuis	int	6
mid	int	6
semester	int	6

nilai_angka	int	6
nilai_huruf	varchar	1

c. tb\_mk

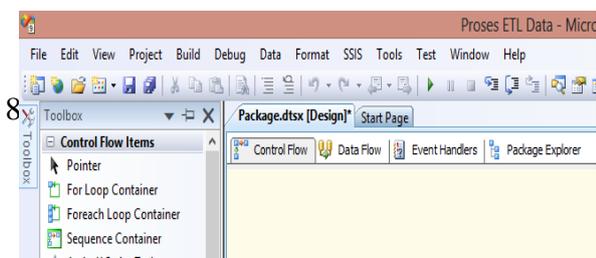
**Tabel 3.** Tabel Matakuliah

Nama	Type	Width
kd_mk*	varchar	10
nm_mk	varchar	30
sms	varchar	1
sks	int	1
kd_progdi	varchar	3

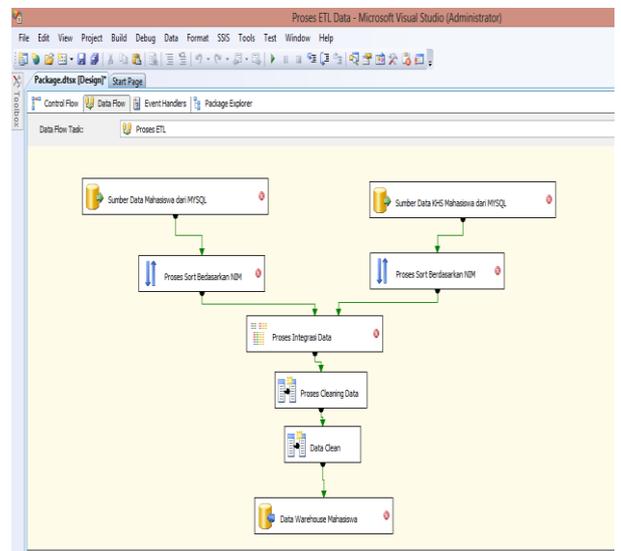
## 2.12 ETL (Extract, Transform and Load)

ETL merupakan proses yang bertujuan untuk melakukan pembentukan data warehouse yang bersumber dari data mentah dalam format database MySQL. Data yang di ekstrak akan dilakukan proses transform dan load data kedalam format baru yaitu format mdb. Proses ETL ini dilakukan dengan menggunakan tools SQL Server Integration Service (SSIS) yang ada dalam Business Intelligence Development Studio (BIDS). ETL yang dilakukan dalam penelitian dapat dilihat pada gambar berikut ini:

### 1. Pembuatan Control Objek



## 2. Proses ETL



### 2.13 Seleksi Data (Data Selection)

Sumber data yang digunakan dalam penelitian berasal dari data mahasiswa tahun angkatan 2011 sampai dengan 2013 pada jurusan sistem informasi. Beberapa jenis data diperoleh dari sistem yang berjalan namun hanya data mahasiswa dan data IPK saja yang digunakan untuk penelitian, dikarenakan informasi yang terkandung di dalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penentu dalam pengelompokan data keluaran yang diinginkan.

Data yang diperoleh adalah sebanyak 1.000 record data yang berasal dari tb\_mhs, 4.1367 record data yang berasal dari tb\_khs dan 417 record berasal dari tb\_mk. Dataset mahasiswa terdiri dari 23 atribut yang menjelaskan identitas diri mahasiswa dan informasi tentang keadaan mahasiswa yang bersangkutan. Atribut data mahasiswa diambil dari penggabungan beberapa tabel yang terdiri dari tb\_mhs, tb\_khs, dan tb\_mk. Atribut tersebut diantaranya adalah nim, nama, jenis\_kelamin, temp\_lahir, tgl\_lahir, kd\_progdi, alamat, asal\_sek, kota, pek\_ortu, kd\_mk, sms, sms\_pendek, tahun\_akademik, kelas, kd\_dosen, kd\_progdi, tugas, kuis, mid, semester, nilai\_angka, nilai\_huruf, dan sks. Sedangkan dataset IPK hanya terdiri dari 5 atribut, dimana memberikan informasi mengenai prestasi akademik dan beban studi yang diambil mahasiswa yang bersangkutan. Atribut tersebut diantaranya adalah nim, nama, sum(sks), total\_nilai, dan IPK. Berikut ini adalah tahap-tahap yang akan digunakan dalam penelitian ini.

a. tb\_mhs

b. tb\_khs

Gambar 2. Record Tabel Kartu Hasil Studi

c. tb\_mk

2.14 Praproses Data

1 Pembersihan Data (Data Cleaning)

Tahap kedua pada proses data mining adalah cleaning data yaitu melakukan pembersihan data terhadap noise yang ditemukan berupa missing value, inkonsisten data, dan redundant data.



Berdasarkan hasil penelitian dan pembahasan yang telah dijelaskan dapat disimpulkan:

Penelitian ini menghasilkan Perancangan dan Implementasi Aplikasi Data Mining untuk mengelompokan minat belajar mahasiswa dalam menentukan konsentrasi matakuliah pilihan program studi sistem inforasi Fakultas Ilmu Komputer Universitas Bina Darma Palembang.

## DAFTAR PUSTAKA

- Eko Prasetyo, 2012, Data Mining Konsep dan Aplikasi menggunakan MatLab, Andi, Yogyakarta.
- Fatansyah, 2004. *Basis Data*, Bandung: Informatika Bandung.
- Lane, Paul. 2002. *Oracle9i Data Warehousing Guide*. Oracle Corporation.
- Mallach, Efrem G. 2000. *Decision Support and Data Warehouse Systems*. Singapore: Irwin McGraw Hill.
- Ponniah, Paulraj. 2001. *Data Warehouse Fundamentals: a Comprehensive Guide for IT Professional*. New York : John Wiley & Sons.
- Vedette Poe, 1997. *Building A Data Warehouse for Decision Support (2nd Edition)*, Prentice Hall
- William H. Inmon, Richard D. Hackathorn. 1994: *Using the Data Warehouse*, John Wiley & Son's