

**PENERAPAN *DATA MINING* UNTUK MENGOLAH INFORMASI KONSENTRASI  
KEAHLIAN DENGAN METODE *CLUSTERING* PADA  
UNIVERSITAS BINA DARMA**

**Kikie Riesky Andini<sup>1</sup>, M. Akbar<sup>2</sup>, Helda Yudiastuti<sup>3</sup>**

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

**Jalan Jenderal Ahmad Yani No.12 Palembang**

**kikieandini@gmail.com<sup>1</sup>, akbar@mail.binadarma.ac.id<sup>2</sup>,**

**helda@mail.binadarma.ac.id<sup>3</sup>**

---

**Abstract :** *The development of computer technology is growing rapidly along with the development of software and hardware used. One in the field of education, a lot of computers play an important role in it. Advances in technology and information can be seen with the use of computers as a tool in learning, with the aim to facilitate the teaching and learning activities so that time is used more quickly, easily, efficiently. Practice of assembling a computer activities are now using equipment such as computer hardware components, but there are still obstacles in the practice because of the limitations of the equipment. Therefore, the author seeks to facilitate the user in practice by designing and building tools that directly show on the computer. In this research made a stout device simulation tools for vocational subjects Bina Darma ICT-based multimedia in hopes to assist and facilitate the students to learn.*

**Keywords:** *multimedia, ICT, teaching tools, simulation tools*

**Abstrak :** *Clustering* merupakan pengelompokan data tanpa berdasarkan kelas data tertentu. Klasterisasi dapat dipakai untuk memberikan label pada kelas data yang belum diketahui. Prinsip dari klasterisasi adalah memaksimalkan kesamaan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/klaster. Dalam penerapan *data mining* dengan metode *clustering* ini pada Universitas Bina Darma yang merupakan organisasi yang memanfaatkan teknologi informasi, khususnya basis data, akan mengalami akumulasi data mahasiswa dalam jumlah besar sehingga untuk mengolah informasi konsentrasi keahlian berdasarkan nilai, sistem yang dibangun dengan menggunakan teknik *data mining* bertujuan untuk menggali dan menemukan pola-pola yang tersembunyi dalam pengolahan data konsentrasi keahlian mahasiswa tiap per tahunnya agar memperoleh data yang real. Berdasarkan uraian di atas, maka penulis berkeinginan membangun aplikasi *data mining* untuk mengolah informasi konsentrasi keahlian pada Universitas Bina Darma Program Studi Teknik Informatika, sehingga mahasiswa mendapatkan alternatif informasi untuk menentukan konsentrasi keahlian yang lebih mudah dan cepat.

**Kata Kunci:** *teknologi informasi, data mining, clustering*

---

## **1. PENDAHULUAN**

### **1.1.Latar Belakang**

Dengan kemajuan teknologi informasi dewasa ini, kebutuhan akan informasi yang

akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan waktu mendatang.

Penerapan *Data Mining* Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode *Clustering* Pada

Universitas Bina Darma

(Kikie Riesky Andini) 1

Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering kali informasi tersebut masih harus di gali ulang dari data yang jumlahnya sangat besar. Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstrak pengetahuan dari data. Metode tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar.

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan *data mining*. Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga.

Banyaknya data mahasiswa Universitas Bina Darma di dalam database, sehingga mengalami penumpukan data yang sangat besar, disini terdapat bermacam fakultas yang terdapat di universitas bina darama dan dilihat dari para

mahasiswa banyak mendapatkan nilai di bawah rata-rata maka dari itu para mahasiswa harus mengikuti semester pendek untuk memperbaiki nilai. Di setiap tahunnya semester pendek ini dilakukan maka dari itu perlu adanya suatu solusi untuk melihat statistik dalam pengambilan mata kuliah semester pendek ini dan agar bisa mengelompokkan mana yang paling banyak para mahasiswa mengambil mata kuliah semester pendek. Oleh karena itu, dengan melakukan prediksi pemilihan mata kuliah semester pendek dapat diketahui tingkat per mata kuliah semester pendek melalui teknik data mining.

Program studi teknik informatika mempunyai konsentrasi keahlian seperti *IT infrastructure, software engineering, dan database management system* yang sesuai dengan kurikulum program studi teknik informatika di Universitas Bina Darma. Sistem pemilihan konsentrasi keahlian dapat ditentukan berdasarkan minat mahasiswa yaitu, mahasiswa dapat menentukan mata kuliah konsentrasi berdasarkan konsentrasi keahlian yang dipilih, setelah itu mahasiswa dapat melakukan *entry* mata kuliah konsentrasi keahlian yang terdapat pada program studi teknik informatika dan dapat dilihat Kartu Rencana Studi (KRS) di *website* Universitas Bina Darma.

Hasil dari Kartu Rencana Studi (KRS) mahasiswa yang mengambil konsentrasi keahlian dapat dikelompokkan sesuai

dengan mata kuliah konsentrasi keahlian, sehingga dapat diketahui konsentrasi keahlian yang paling banyak dimintai oleh mahasiswa, dengan itu dapat ditemukan suatu solusi untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengolahan pola.

Berdasarkan uraian di atas, maka penulis melakukan penelitian untuk mengolah informasi konsentrasi keahlian pada Universitas Bina Darma Program Studi Teknik Informatika, sehingga mahasiswa mendapatkan solusi alternatif informasi untuk mengajukan konsentrasi keahlian dengan efektif dan efisien. Maka di dalam penulisan skripsi ini digunakan sebuah judul yaitu **“Penerapan *Data Mining* Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode *Clustering* Pada Universitas Bina Darma”**.

### **1.2. Batasan Masalah**

Agar penelitian ini lebih terarah dan tidak menyimpang, maka permasalahan yang ada pada penelitian dibatasi pada :

1. Data yang digunakan data program studi teknik informatika tahun angkatan 2009 dan 2010.

2. Menampilkan informasi calon mahasiswa yang mengambil konsentrasi keahlian pada Universitas Bina Darma Program Studi Teknik Informatika dengan metode *clustering*.

### **1.3. Tujuan Penelitian**

Adapun tujuan dalam penelitian ini adalah antara lain :

1. Memberikan informasi dari data mahasiswa yang diolah dalam bentuk grafik yang informatif dan mudah dianalisis.
2. Memudahkan mahasiswa dalam menentukan pemilihan konsentrasi keahlian berdasarkan nilai dan minat mahasiswa.

### **1.4. Manfaat Penelitian**

Adapun manfaat adanya penelitian ini adalah sebagai berikut :

1. Bagi Universitas Bina Darma Untuk mengetahui tingkat statistik mahasiswa dalam pemilihan konsentrasi keahlian program studi teknik informatika di Universitas Bina Darma dan dapat menjadi bahan pertimbangan guna perkembangan sistem yang lebih lanjut sehingga dapat membantu kinerja karyawan.
2. Bagi Penulis Diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di

Penerapan *Data Mining* Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode *Clustering* Pada

Universitas Bina Darma

dalam gudang data sehingga menjadi informasi berharga khususnya dalam penerapan *data mining* yang dapat digunakan dalam kegiatan masyarakat nantinya.

### 3. Bagi Pembaca

Diharapkan dengan adanya penulisan ini dapat bermanfaat sebagai sumber pembelajaran untuk penelitian selanjutnya serta untuk menambah wawasan.

## 2. METODOLOGI PENELITIAN

### 2.1. Metode Penelitian

Metode pada penelitian ini menggambarkan metode deskriptif yaitu penelitian yang berusaha menggambarkan dan menginterpretasikan objek sesuai dengan apa adanya berdasarkan fakta-fakta mengenai data Mahasiswa di Universitas Bina Darma Palembang. Penelitian deskriptif bertujuan untuk menggambarkan secara sistematis fakta dan karakteristik objek atau subjek yang diteliti secara tepat.

### 2.2. Metode Pengumpulan Data

Metode pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data beserta sumbernya. Data yang dinilai objektif dan relevan dengan inti permasalahan penelitian merupakan indikator keberhasilan dari penelitian. Metode pengumpulan data dalam penelitian ini dilakukan dengan cara sebagai berikut :

1. Metode *Observasi*, metode observasi ini yang dibahas tentang mengelolah informasi

konsentrasi keahlian di program studi teknik informatika Universitas Bina Darma Palembang, hasil dari obeservasi ini mendapatkan data mahasiswa, data mata kuliah, dan data Kartu Rencana Studi (KRS) yang akan mengambil konsentrasi keahlian.

2. Wawancara, Merupakan teknik pengumpulan data dengan cara mengadakan tanya jawab atau wawancara kepada bagian pengolahan data dan bagian IT Universitas Bina Darma Palembang.
3. Metode Studi Pustaka, metode yang dilakukan dengan mencari dan mengumpulkan materi-materi yang berkaitan dengan penyusunan skripsi, seperti informasi dari buku-buku, dokumentasi, *internet* dan skripsi yang telah selesai dikerjakan. Materi-materi yang tersedia dapat digunakan sebagai bahan pembelajaran dan referensi pembuatan aplikasi.

### 2.3. Metode Analisis Data

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan tahapan *Knowledge Discovery in Databases (KDD)* yang terdiri dari beberapa tahapan, yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*.

## 3. PROSES DATA MINING

### 3.1. Data Selection

Data yang digunakan dalam penelitian berasal dari Universitas Bina Darma Palembang

yaitu data Mahasiswa Teknik Informatika Universitas Bina Darma Palembang tahun 2009 dan 2010, terdiri dari beberapa tabel antara lain tabel Mahasiswa, Tabel Khs, Tabel Matkul. Dari data Mahasiswa mempunyai data 697 record, Tabel Khs mempunyai 20,646 record sedangkan untuk tabel Matkul mempunyai 365 record selama 2 tahun. Dari semua atribut yang ada pada tabel Matkul dan tabel Khs terdapat 4 atribut yang akan digunakan dalam proses *knowledge discovery in databases (kdd)*. Atribut tersebut yaitu :

1. kd\_matkul merupakan atribut yang terdapat pada tabel khs yang berperan sebagai *primary key* dalam menghubungkan tabel khs dan tabel matkul.
2. nilai merupakan atribut yang terdapat pada tabel khs yang berisi mengenai nilai mahasiswa.
3. kd matkul merupakan atribut yang terdapat pada tabel matkul yang akan dihubungkan dengan tabel khs.
4. matkul merupakan atribut yang terdapat pada tabel matkul yang berisikan matakuliah yang diambil oleh para mahasiswa.

### **3.2. Preprocessing**

Pada tahapan *preprocessing* ini akan dilakukan proses integrasi data untuk menghubungkan tabel khs dan tabel matkul, selanjutnya dilakukan proses data *cleaning* untuk menghasilkan *dataset* yang bersih

sehingga dapat digunakan dalam tahap berikutnya yaitu *mining* dengan tujuan untuk melihat prediksi pemilihan konsentrasi jurusan terhadap mahasiswa teknik informatika Bina Darma Palembang. Berikut merupakan penjelasan dari kedua proses di atas, Berikut langkah-langkah yang dimaksud diatas :

#### **3.2.1. Data Cleaning**

Data *Cleaning*, tahap pembersihan data merupakan tahap awal dari proses KDD. Seluruh atribut yang ada pada *dataset* di atas selanjutnya akan diseleksi untuk mendapatkan atribut-atribut yang berisi nilai yang relevan. Tidak *redundant* dan tidak *missing value*, dimana syarat tersebut merupakan syarat awal yang harus dikerjakan dalam *data mining* sehingga akan diperoleh *dataset* yang bersih untuk digunakan pada tahap *mining* data. Data dikatakan *missing value* bila atribut-atribut dalam *dataset* tidak berisi nilai atau kosong, sedangkan data dikatakan *redundant* jika dalam satu *dataset* yang sama terdapat lebih dari satu *record* yang berisi nilai yang sama.

*Dataset* Mahasiswa Universitas Bina Darma Palembang, pada tahap ini akan diseleksi dari data yang mengandung *missing value* dan data dengan nilai yang *redundant*. *Databases* Mahasiswa Universitas Bina Darma Palembang



*mining* dengan menggunakan teknik *Clustering*, data mahasiswa yang telah dilakukan proses *transformasi* di atas siap untuk dilakukan proses selanjutnya.

## 4. HASIL

### 4.1. Hasil Data Mining

*Data mining* merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pemilihan teknik dan algoritma yang tepat sangat bergantung pada proses KDD secara keseluruhan. Pada penelitian ini penerapan data mining menggunakan teknik *Clustering* dan algoritma *K-Means* untuk menemukan informasi mengenai data mahasiswa.

### 4.2. Penerapan Metode *Clustering* dengan Algoritma *K-means*

Setelah melakukan proses transformasi data ke dalam bentuk data yang sesuai untuk penerapan data *mining* dengan teknik *Clustering* maka tahapan ini dapat dilakukan. *Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar dari pada kesamaan tersebut dengan data kelompok lainnya.

Dalam tahapan penemuan aturan *clustering* ini, langkah yang dilakukan adalah mengembangkan aturan yang mungkin akan menjadi informasi data mahasiswa. Dalam

tahapan awal untuk menghasilkan informasi aturan *clustering* dari setiap data mahasiswa dihitung dengan *dataset*. Dalam terminologi pengolahan pola, kolom dalam matriks disebut *patterns* atau *objects*, baris disebut *features* atau *attribute*, dan *Z* disebut *patterns* atau *data matrix*. Arti kolom dan baris dalam *Z* bergantung pada konteks. *Cluster* secara umum merupakan wujud himpunan bagian dari suatu himpunan data dan metode *clustering* dapat diklasifikasikan berdasarkan himpunan bagian yang dihasilkan : apakah *fuzzy* atau *crisp (hard)*. Metode *hard clustering* merupakan model yang berdasarkan pada teori himpunan klasik, yang mana suatu objek menjadi anggota atau tidak menjadi anggota secara penuh ke dalam suatu kelompok. *Hard clustering* membagi data ke dalam sejumlah himpunan bersamaan dengan perbedaan level keanggotaan.

Berikut merupakan proses *mining* untuk menentukan informasi mengenai data mahasiswa berdasarkan data yang sudah ada yang telah dipilih. Untuk proses yang dilakukan *clustering* mempunyai 6 langkah proses dalam menentukan nilai yang telah dipilih. Berikut langkah-langkah dari *clustering* :

Algoritma FCM didasarkan pada minimasi fungsi objektif yang diformulasikan dalam persamaan :

$$J_m = \sum_{i=1}^N \sum_{j=1}^c (U_{ij})^m \|z_i - c_j\|^2$$

Dengan :

Penerapan *Data Mining* Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode *Clustering* Pada

Universitas Bina Darma

$U_{ij}$  merupakan level keanggotaan dari  $Z_i$  dalam Cluster  $j$

$Z_i$  merupakan nilai data ke- $i$  dari  $d$ -dimensi data

$C_j$  merupakan nilai ke- $j$  dari  $d$ -dimensi cluster center

$m$  merupakan sembarang bilangan real lebih besar dari 1

Selanjutnya algoritma FCM disusun dengan langkah sebagai berikut :

Langkah 1 : Tentukan himpunan data  $Z$ . Tentukan jumlah cluster yang diharapkan  $1 < c < N$ , nilai pembobot  $m > 1$ , toleransi penghentian  $\epsilon > 0$ .

Langkah 2 : Inisialisasi matriks partisi secara acak,  $U(0) \in M_{fc}$ .  
Ulangi untuk  $l=3,4,5$

Langkah 3 : Hitung cluster center (means).

$$V_i^{(l)} = \frac{\sum_{k=1}^N (U_{ik}^{(l-1)})^m Z_k}{\sum_{k=1}^N (U_{ik}^{(l-1)})^m}, 1 \leq i \leq c$$

Langkah 4 : Hitung jarak

$$D_{ikA}^2 = (Z_k - V_i^{(l)})^T A (Z_k - V_i^{(l)}), 1 \leq i \leq c, 1 \leq k \leq N$$

Langkah 5 : Perbaharui matriks partisi.

Untuk  $1 \leq k \leq N$

Jika  $D_{ikA} > 0$  untuk  $i = 1, 2, \dots, C$

$$u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA} / D_{jKA})^{2/(m-1)}}$$

Atau dengan kata lain :

$$u_{ik}^{(l)} = 0 \text{ jika } D_{ikA} > 0 \text{ dan } u_{ik}^{(l)} \in [0,1] \text{ dengan } \sum_{i=1}^c u_{ik}^{(l)} = 1$$

Ulangi sampai  $\|U^{(l)} - U^{(l-1)}\| < \epsilon$ .

Untuk proses awal maka algoritma k-means yang ada akan dilakukan sebagai berikut :  
Diberikan data sebagai data awal.

$$Z = \begin{pmatrix} 40 & 50 & 60 & 70 & 80 & 90 & 100 \\ 40 & 50 & 60 & 70 & 80 & 90 & 100 \end{pmatrix}$$

Akan dilaksanakan clustering dengan FCM.

Langkah 1 :

- Jumlah cluster yang di harapkan  $c = 5$
- Nilai pembobot  $m = 3$
- Toleransi penghentian  $\epsilon = 1.10^{-6}$
- Maksimum iterasi = 100

Langkah 2 : inisialisasi matriks partisi awal secara acak.

$$V = \begin{pmatrix} 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \\ 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \\ 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \\ 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \\ 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \end{pmatrix}$$

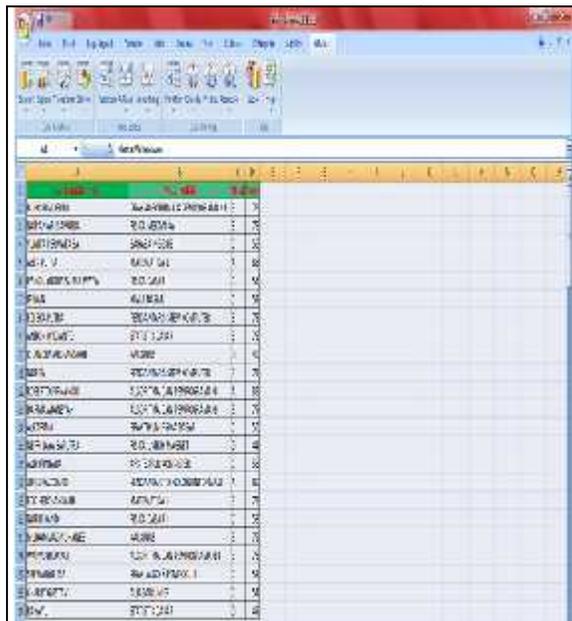
$$\frac{0,5 \times 40 + 0,5 \times 50 + 0,5 \times 60 + 0,5 \times 70 + 0,5 \times 80 + 0,5 \times 90 + 0,5 \times 100}{6} = 66,67$$

$$V_2 = \frac{0,5 \times 40 + 0,5 \times 50 + 0,5 \times 60 + 0,5 \times 70 + 0,5 \times 80 + 0,5 \times 90 + 0,5 \times 100}{6} = 1,5$$

$$V = \begin{pmatrix} 3,33 & 4,16 & 5,00 & 5,83 & 6,66 & 7,5 & 8,33 \\ 3,33 & 4,16 & 5,00 & 5,83 & 6,66 & 7,5 & 8,33 \end{pmatrix}$$

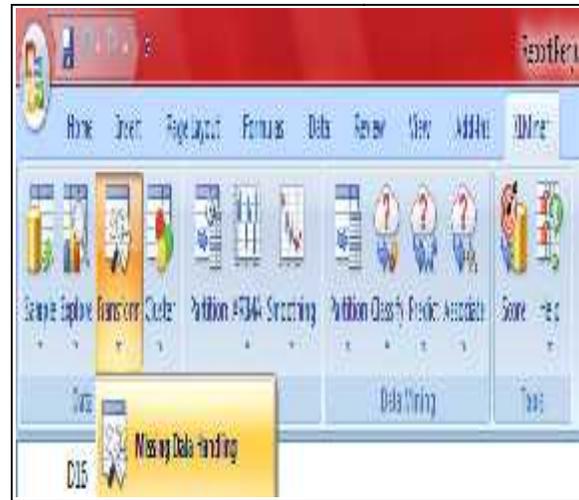
### 4.3. Proses Data Mining Menggunakan XLMiner

Setelah dijelaskan proses penerapan data mining dengan teknik clustering secara teoritis pada penjelasan di atas, maka kali ini akan di jelaskan proses data mining secara aplikatif dimana proses data mining yang akan dilakukan menggunakan software data mining XLMiner V4.0.2. Seperti yang telah dijelaskan pada bab sebelumnya XLMiner merupakan add in yang bekerja pada Microsoft Excel. Berikut merupakan tampilan awal Microsoft Excel 2007 yang telah ditambahkan add in XLMiner.



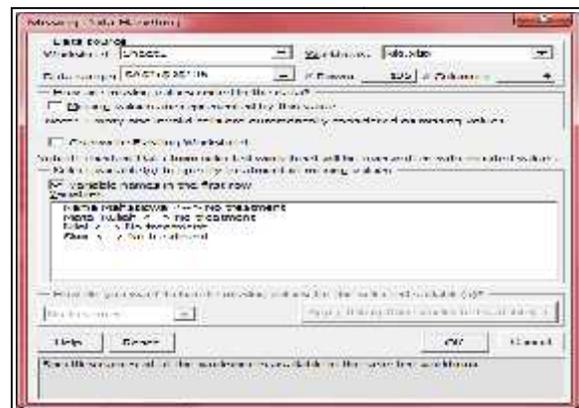
Gambar 4.1. Microsoft Excel 2007 dengan XLMiner

Pada gambar 4.1 terlihat bahwa tampilan Microsoft Excel 2007 yang telah ditambahkan add in XLMiner sama saja dengan tampilan Microsoft Excel 2007 pada umumnya, hanya saja terdapat menu tambahan yaitu menu XLMiner seperti pada gambar 4.1.



Gambar 4.2. Menu XLMiner

Setelah memilih menu clustering seperti pada gambar 4.2, maka langkah yang dilakukan selanjutnya yaitu mengatur proses dan hasil yang diinginkan seperti pada gambar 4.3



Gambar 4.3 Setting Missing Data Handling

Penerapan Data Mining Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode Clustering Pada

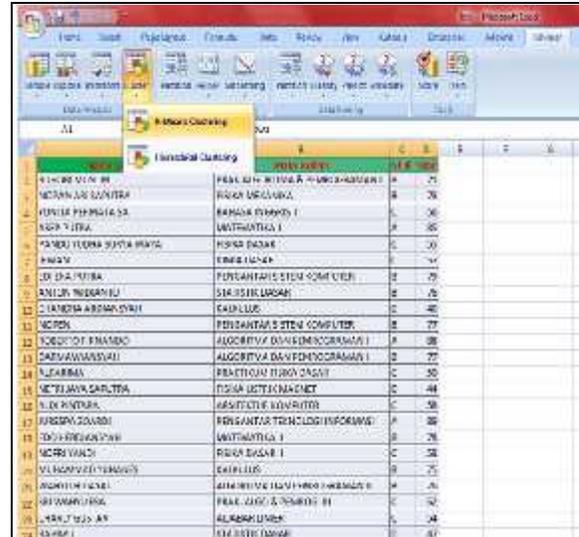
Dari pengaturan yang dilakukan pada gambar 4.3 dimana dari 106 item nama mahasiswa dimana ditentukan *missing data handling* (merupakan batas transformasi pada *XLMiner*), penggunaan *missing data handling* dengan batasan pada *XLMiner* bertujuan untuk mendapatkan data yang telah di transformasikan ke dalam *XIMiner*. Dari pengaturan yang dilakukan pada gambar 4.3 didapatkan hasil seperti pada tabel 4.1.

**Tabel 4.1.** Hasil *Missing Data Handling* Menggunakan *XLMiner*

Nama Mahasiswa	Nilai	Skor
ERKANBAR	C	50
DELLORI MURIM	C	77
ROHAN ARI SAPUTRA	C	70
YUNITA PERMATA SA	C	58
ASEPTI LITRA	A	55
PANDU YULHA SUTERA WAYA	C	55
BOYAN	C	57
EBEKA LITRA	C	70
WILSON WINDHITO	H	70
CHANDRA ARDIANSYAH	C	40
NORTN	C	77
ROBERTO FERNANDO	A	00
GARMAWANUSYAH	C	77
ALHARIMA	C	50
NEPTI LAYLA SAJUTRA	C	44
ALDI HINIKHA	C	58
JUNESTA ADANID	A	80
EDU HENDIANSYAH	B	78
RUFU YAND	C	58
MUHAMMAD SUHARTO	C	77
WALDO IBRAHIM	C	70
SRIWA NULISA	C	58
SHARLY GUSTAV	C	54
RAJAYU	C	47
ROSEN SADIHAB	C	51
IWAN SEBAYAN	C	50
SALIM	C	50
HUJI WULANDI	C	41

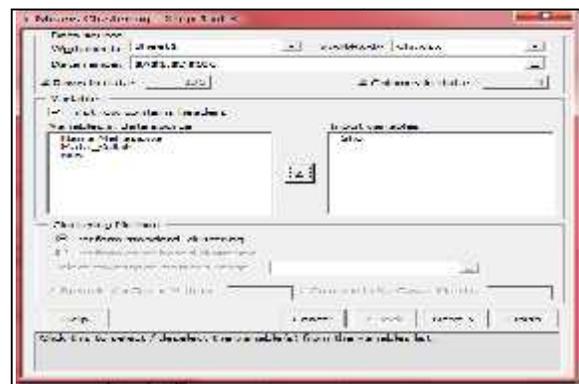
Hasil dari proses *data mining* menggunakan *XLMiner* ditampilkan *worksheet* baru dengan nama *Missing* mahasiswa seperti gambar 4.3. informasi yang dihasilkan di tampilkan dalam 1 tabel dimana tabel tersebut memberikan informasi mengenai data yang digunakan. Untuk menentukan *missing data handling* selanjutnya sama seperti proses yang telah dilakukan seperti langkah-langkah yang di atas karena data yang sudah di *missing data handling* sama hasilnya.

1. Data mahasiswa yang telah di transformasikan dan siap untuk menentukan algoritma *K-Means*, lihat pada gambar 4.4.



**Gambar 4.4.** Proses Untuk Menentukan *K-Means*

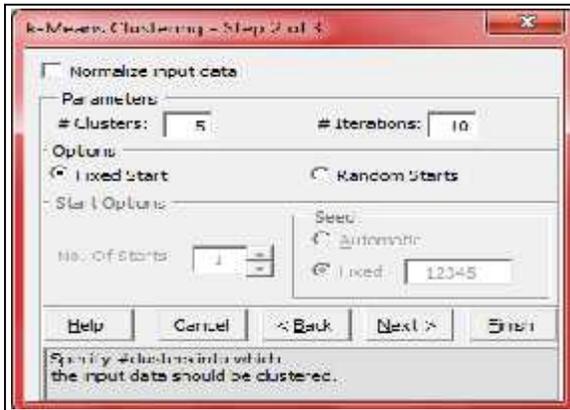
2. Langkah yang selanjutnya penulis menentukan *K-Means Clustering*, input *variables* dengan memilih jumlah dari data mahasiswa, lalu klik *next* seperti gambar 4.5.



**Gambar 4.5.** *K-Means Clustering*

3. Untuk menentukan jumlah *cluster* yang diinginkan, di sini penulis menentukan 5

jumlah *cluster* karena jumlah *cluster* menentukan hasil dari nilai data mahasiswa, *iterations* nya 10, lalu klik *finish* lihat gambar 4.6.



**Gambar 4.6.** Menentukan Jumlah *Cluster*

4. Apabila semua step *clustering* dilakukan maka pilih *finish* untuk menampilkan *K-Means input*. Di *K-Means* output menampilkan beberapa tabel yang berisikan, **tabel *input***, **tabel *cluster centers***, **tabel *data summary***, **tabel *elapsed time***, **tabel *predicted***.

**Tabel 4.2.** *XLMiner : K-Means Clustering*

Output Navigator		
<a href="#">inp.xls</a>	<a href="#">Cluster Centers</a>	<a href="#">Data Summ</a>
<a href="#">Elapsed Time</a>	<a href="#">Predicted Clusters</a>	<a href="#">Random Starts Summary</a>

5. Input yang dihasilkan dari proses *XIMiner : K-Means Clustering*

Penerapan *Data Mining* Untuk Mengolah Informasi Konsentrasi Keahlian Dengan Metode *Clustering* Pada

Universitas Bina Darma

**Tabel 4.3.** Tabel *Input*

Input	
Input file	File: inp (Sheet1) (xls)
Records in the input data	105
Input variables removed	Nil
Variables	
Selected variables	1
Deleted variables	Nil
Parameters/Options	
# Clusters	5
Start Option	Fixed Start
# Iterations	10
Show data summary	Yes
Show distance from cluster center	Yes

6. Dari proses yang telah ditentukan maka didapat jumlah *cluster* seperti tabel 4.4.

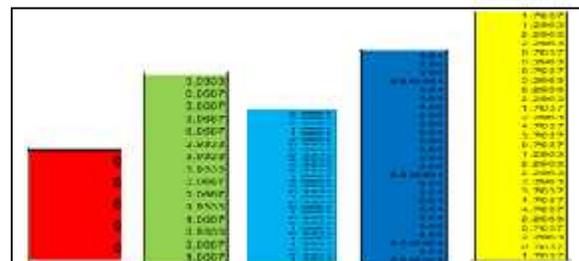
**Tabel 4.4.** Jumlah *Cluster* 1-5 dan *Cluster Centers*

Cluster	Size
Cluster-1	3
Cluster-2	43.9333
Cluster-3	75.3333
Cluster-4	25.90
Cluster-5	66.7637

Distance between cluster centers	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0	61.8511	26.1555	84.648	89.1117
Cluster 2	43.9333	0	27.4	15.4567	41.1634
Cluster 3	75.3333	21.4	0	18.2521	17.5734
Cluster 4	25.90	11.0467	7.6128	0	21.1471
Cluster 5	66.7637	42.7724	1.1704	23.7427	0

Tabel *cluster center* di atas menjelaskan bahwa *cluster* yang diperoleh atau yang telah ditentukan ada **5 cluster** banyak *cluster*, untuk *cluster* 2 yang jumlahnya **43.9333** yang didapat dari banyak di setiap *cluster* yang didapat dari ***Predicted Clusters***.

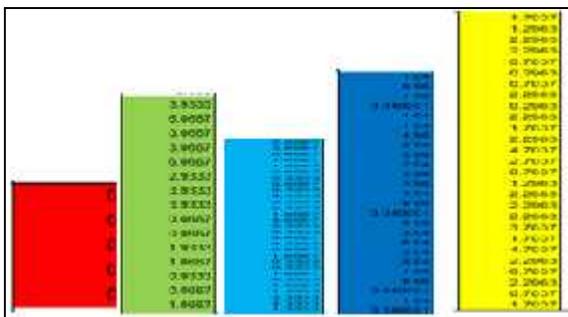


Penjelasan tabel diatas adalah dari *cluster* 1 ini diberikan warna merah pada jumlah yang

untuk mencari jumlah *cluster* yang didapat di tabel **cluster center**.

Dari proses yang dilakukan dalam *XLMiner* maka seluruh dari algoritma *K-Means Clustering*, maka informasi yang didapatkan dari jumlah *Cluster* yang ditentukan dapat dilihat pada tabel 4.6.

**Tabel 4.5.** Tabel *Cluster 2, Cluster 3, Cluster 4, Cluster 5*



Penjelasan dari beberapa tabel yaitu *cluster 2*, *cluster 3*, *cluster 4*, *cluster 5* diberikan warna yang berbeda seperti *cluster 2* di beri warna **hijau**, *cluster 3* warna **biru**, *cluster 4* **biru tua**, *cluster 5* **uning**. Dari tabe-tabel *cluster* ini diberikan warna yang berbeda agar dapat mempermudah pada saat mencari jumlah *cluster center* dan dapat membedakan mana *cluster1* sampai *cluster 5*.

7. Selanjutnya ke tabel data *summary* dari tabel ini terdapat beberapa atribut yaitu atribut *cluster* di mana terdapat *cluster-cluster* yang telah dibuat sebelumnya, atribut *#Obs* atribut ini menjelaskan banyak *cluster* dari setiap *record* dan yang terakhir *atribut Average distance in cluster* maksud dari

atribut ini adalah mencari nilai rata-rata di dalam *cluster*.

**Tabel 4.6.** *Summary*

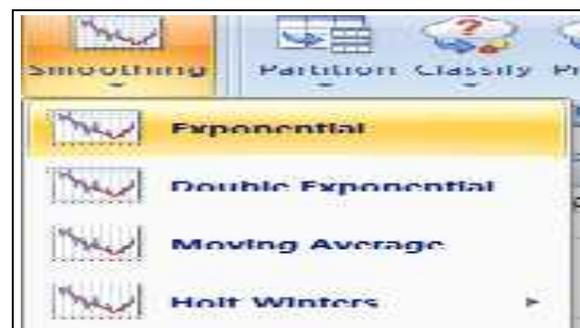
Data summary		
Cluster	#Obs	Average distance in cluster
Cluster-1	5	0
Cluster-2	10	2.047
Cluster-3	35	2.182
Cluster 4	28	3.528
Cluster 5	27	1.937
Overall	105	2.12

8. Waktu dalam proses *K-Means Clustering*

**Tabel 4.7.** *Elapsed Time*



*Exponential* untuk melakukan proses data membentuk grafik dari hasil data mahasiswa, lihat perintah yang telah dilakukan pada gambar 4.7.



**Gambar 4.7.** Proses Untuk Menampilkan Grafik

Dari proses *Exponential Smoothing* maka aplikasi *XLMiner* akan menampilkan perintah

untuk menentukan grafik yang akan dihasilkan dari data mahasiswa.



**Gambar 4.8.** Exponential Smoothing

Setelah menjelaskan tabel *Predicted Clusters* yang terakhir tabel *Fitted Model*.

**Tabel 4.8.** Tabel *Fitted Mode*

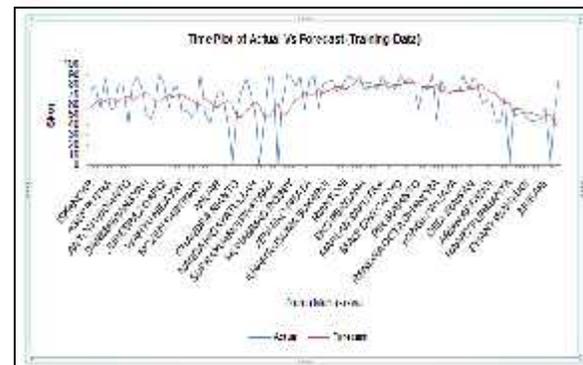
Nama Mahasiswa	Actual	Forecast	Residuals
KIKI SAMP	60	60	0
ELIENI MUDA M	70	61.1	-8.9
SOPAN ARIGARITRA	70	64.72	-5.72
YUNITA TERMATA DA	50	52.278	-2.278
ASE TIU IISA	35	36.3018	-1.3018
KEBI YUTINA KEIYAWAN A	35	36.3018	-1.3018
REZKA	10	10.94918	-0.94918
TRIF CA P. TRA	70	63.33778	-6.66222
ANTON WIDANTO	70	66.40007	-3.40007
CHARLINA A. SARYA PAF	40	38.267178	-1.732822
NOHAN	70	60.80878	-9.19122
KELI REI. NEMANI	100	101.49887	-1.49887
SARIZAWANGYANI	77	76.043991	-0.956009
AL-CABWA	50	71.451158	-21.451158
TEPHI JAYA SARU IISA	44	67.142125	-23.142125
AL E. PRIN. PA	30	30.678204	-0.678204
JURFARA DOARD	69	67.01409	-1.98591
EDO. IERDIANISYALI	70	67.021055	-3.021055
TOFTIYANDI	55	65.273458	-10.273458

Dari penjelasan tabel di atas bahwa tabel ini hasil akhir yang telah diproses dalam *data mining*. Di dalam tabel mempunyai beberapa atribut yaitu atribut nama mahasiswa, atribut *actual*, atribut *forecast*, atribut *residuals*. Untuk atribut nama mahasiswa ini berisi seluruh jumlah data nama-nama mahasiswa seluruh angkatan 2009 dan 2010 jurusan teknik informatika, atribut *actual* suatu nilai yang benar-benar fakta yang telah diproses sebelumnya, atribut *residuals* atribut ini hasil

sisia dari atribut *actual* dikurang *forecast* maka dapatlah hasil nilai *residuals* nya.

#### 4.4. Interpretation /Evaluation

Pola informasi yang dihasilkan dari proses *data mining* yang diterapkan berupa grafik seperti gambar 4.9.



**Gambar 4.9.** Grafik Dari Data Mahasiswa Tahun 2009

Hasil akhir dari tabel *Fitted Mode* yang dihasilkan berupa grafik. Di grafik *Time Plot of Actual Vs Forecast (Training Data)* ini memberikan informasi tingkat banyaknya mahasiswa yang mengambil konsentrasi keahlian di bidangnya masing-masing, di grafik ini mempunyai beberapa proses yaitu *actual* diberikan warna biru di mana proses yang dapat memberikan informasi mahasiswa yang mengambil konsentrasi keahlian. Dari grafik di atas dapat disimpulkan bahwa mahasiswa yang mengambil konsentrasi jurusan dapat berdasarkan nilai dan minat mahasiswa, berdasarkan nilai yaitu nilai diantara 60-70 dapat mengambil konsentrasi jurusan *database*

*management system* , nilai diantara 70-80 dapat mengambil konsentrasi jurusan *software engineering* sedangkan nilai diantara 80-90 dapat mengambil konsentrasi jurusan *IT infrastructure*. Dari grafik ini menunjukkan bahwa informasi prediksi konsentrasi keahlian yang paling banyak diambil mahasiswa adalah konsentrasi jurusan *database management system* sebanyak 80%.

Dari seluruh penjelasan yang telah dipaparkan baik teoritis maupun menggunakan aplikasi *XLMiner* merupakan informasi hasil prediksi, keputusan yang diambil oleh mahasiswa akan menentukan peningkatan mahasiswa yang mengambil konsentrasi keahlian yang telah dirancang agar berjalan dengan baik berdasarkan minat di bidangnya masing-masing sehingga dapat menciptakan sumber daya manusia yang berkualitas nantinya.

## 5. SIMPULAN

Berdasarkan dari penelitian yang telah dilaksanakan dan sudah diuraikan dalam penerapan *data mining* dari data mahasiswa teknik informatika di Universitas Bina Darma Palembang , maka penulis dapat menarik kesimpulan sebagai berikut :

1. Penerapan *data mining* dengan teknik *clustering* dan algoritma *K-Means* yang dilakukan menghasilkan sebuah informasi mengenai data mahasiswa dalam mengambil keputusan untuk menentukan konsentrasi keahlian.
2. Dalam penerapan data mining ini dapat memberikan informasi konsentrasi keahlian mana yang paling banyak yang diambil oleh mahasiswa, dan dapat menjadi suatu keputusan atau pertimbangan agar Universitas Bina Darma lebih baik lagi untuk ke depannya.
3. Semakin kecil batasan nilai dari data mahasiswa yang mengambil keputusan dalam menentukan konsentrasi keahlian yang ditentukan maka semakin sedikit pula pola/aturan yang dapat dihasilkan dan waktu yang diperlukan lebih sedikit.
4. Perhitungan yang dilakukan secara teoritis dan aplikatif menghasilkan nilai dari data mahasiswa yang mengambil keputusan dalam menentukan konsentrasi keahlian yang ditentukan.
5. Waktu yang diperlukan untuk pemrosesan tergantung pada spesifikasi komputer, jumlah mahasiswa yang diolah dan jumlah item yang terlibat sehingga perumusan yang tepat untuk perhitungan waktu belum dapat ditentukan

## DAFTAR RUJUKAN

- Azevedo, A. Santos & Manuel F. (2008), *KDD, SEMMA AND CRIPS-DM: A PARALLEL OVERVIEW*, IADIS. ISBN: 978-972-8924-63-8.
- Frontline Solvers.(2012), *XLMiner Data Mining add-in for Excel*. Diakses 10 Mei 2013, dari [http://www.solver.com/xlminer\\_data\\_mining](http://www.solver.com/xlminer_data_mining).

Kusrini & Luthfi, E. Taufiq.(2009), *Algorima Data Mining*, Penerbit Andi, Yogyakarta.

Larose, Daniel T . (2005), *Discovering Knowledge in Data: An Introduction to Data Mining* , John Willey & Sons, Inc.

Pramudiono, I.(2003), Pengantar *Data Mining*. Diakses 10 Mei 2013, dari <http://ikc.depsos.go.id/umum/iko-datamining.php>.

Santoso, Budi.(2007), *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.