

**PERANCANGAN APLIKASI K-MEANS SEBAGAI PENENTU KONSENTRASI  
BAGI MAHASISWA INFORMATIKA UMS**

Makalah

Program Studi Informatika  
Fakultas Komunikasi dan Informatika



**Diajukan oleh :**

*DIAN SETIAWAN*

*YUSUF SULISTYO NUGROHO, S.T, M.ENG*

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**JUNI 2015**

**HALAMAN PERSETUJUAN**

Publikasi Ilmiah dengan judul :

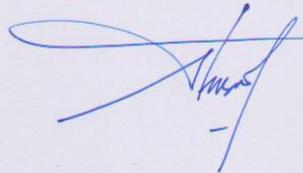
**PERANCANGAN APLIKASI *K-MEANS* SEBAGAI PENENTU KONSENTRASI  
BAGI MAHASISWA INFORMATIKA UMS**

Ini telah diperiksa dan disetujui pada :

Hari : Jum'at

Tanggal : 12 Juni 2015

Pembimbing

A handwritten signature in blue ink, appearing to be 'Yusuf Sulisty Nugroho', written over a horizontal line.

**Yusuf Sulisty Nugroho, S.T., M. Eng**  
NIK : 119

**HALAMAN PENGESAHAN**

Publikasi ilmiah dengan judul :

**PERANCANGAN APLIKASI K-MEANS SEBAGAI PENENTU KONSENTRASI  
BAGI MAHASISWA INFORMATIKA UMS**

Yang dipersiapkan dan disusun oleh :

Dian Setiawan

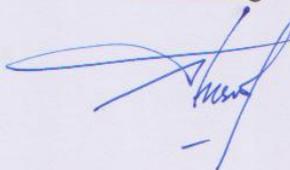
L200110020

Telah disetujui pada :

Hari : Jum'at

Tanggal : 12 Juni 2015

Pembimbing



**Yusuf Sulistyono, S.T, M.Eng**

**NIK : 1197**

Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar sarjana

Tanggal.....

Mengetahui,

Ketua Program Studi

Informatika



**Dr. Heru Supriyono, M.Sc.**

**NIK : 970**

**PERANCANGAN APLIKASI *K-MEANS* SEBAGAI PENENTU KONSENTRASI  
BAGI MAHASISWA INFORMATIKA UMS**

Dian Setiawan, Yusuf Sulistyono Nugroho

Program Studi Informatika, Fakultas Komunikasi dan Informatika,  
Universitas Muhammadiyah Surakarta  
Email: [dianst92@gmail.com](mailto:dianst92@gmail.com)

**ABSTRACT**

Department of Informatics in UMS has three concentrations. Due to these three concentrations, the students are required to choose one of these three concentrations at the end of the fourth semester. Currently, students determine the concentration based on the students' own wishes without a system which gives the students consideration in selecting their concentration. By designing *clustering k-means* application, the researcher expects it can provide the students source of information to determine their class concentration. The variables used are the value of the dominant subjects of each concentration ranging from semester 1 to 4, the variables of information system and enterprise (Basic Web Programming, Algorithms and Programming, Introduction to Information System, Data Base System) concentrations. Computer network and multimedia (Data Communication, Computer Network, Practicum Computer Network, Operation System). Software engineering and animation (linear algebra and matrix, Algorithms and Data Structures, Discrete Structure 2, Digital system). The result is that the students of 2011 academic year data are distributed into 3 cluster, cluster 1 with centroid point (3,333 ; 3,548 ; 3,098) and cluster 2 with centroid point (0,915 ; 1,110 ; 0,773), also cluster 3 with centroid point (2,682 ; 3,221 ; 1,880). The members of cluster 1 are recommended to take the Computer Network and Multimedia concentration, the members of cluster 2 are recommended to take the Software engineering and animation concentration then the members of cluster 3 are recommended to take the Information System and Enterprise concentration.

**Keyword : Class Concentration, Clustering Method, Data Mining, K-Means Algorithms, Majoring.**

## ABSTRAKSI

Program Studi Informatika di UMS memiliki tiga konsentrasi jurusan. Dengan adanya tiga konsentrasi penjurusan maka mahasiswa diwajibkan untuk memilih salah satu dari tiga konsentrasi yang ada, pada akhir semester 4. Saat ini dalam menentukan konsentrasi hanya berdasarkan dari keinginan mahasiswa itu sendiri tanpa adanya suatu sistem yang memberikan pertimbangan bagi mahasiswa dalam memilih konsentrasi jurusan. Dengan melakukan perancangan aplikasi *clustering k-means*, diharapkan dapat memberikan sumber informasi penentu kelas konsentrasi bagi mahasiswa. Variabel yang digunakan nilai mata kuliah yang dominan pada masing-masing konsentrasi dari semester 1 sampai 4, variabel konsentrasi Sistem Informasi dan Enterprise (Pemrograman Web Dasar, Algoritma dan Pemrograman, Pengantar Sistem Informasi, Sistem Basis Data). Jaringan Komputer dan Multimedia (Komunikasi Data, Jaringan Komputer, Prak. Jaringan Komputer, Sistem Operasi). Rekayasa Perangkat Lunak dan Animasi (Aljabar Linier dan Matriks, Algoritma dan Struktur Data, Struktur Diskret 2, Sistem Digital). Hasil *clustering* data mahasiswa 2011 dengan algoritma *k-means* diperoleh data *cluster* 1 dengan titik *centroid* (3,333 ; 3,548 ; 3,098) dan *cluster* 2 dengan titik *centroid* (0,915 ; 1,110 ; 0,773) serta *cluster* 3 dengan titik *centroid* (2,682 ; 3,221 ; 1,880). Pada anggota *cluster* 1 maka akan direkomendasikan mengambil konsentrasi Jaringan Komputer dan Multimedia, anggota *cluster* 2 direkomendasikan untuk konsentrasi Rekayasa Perangkat Lunak dan Animasi sedangkan anggota *cluster* 3 direkomendasikan mengambil konsentrasi Sistem Informasi dan Enterprise.

**Kata Kunci** : *Algoritma K-Means, Data Mining, Jurusan, Kelas Konsentrasi, Metode Clustering.*

## PENDAHULUAN

Universitas Muhammadiyah Surakarta merupakan satu dari beberapa instansi perguruan tinggi yang bergerak di bidang pendidikan yang di dalamnya terdapat 12 Fakultas, 64 Program Studi dan 4 Program Profesi. Informatika adalah satu dari sekian program studi yang berdiri sejak tahun 2007 serta memiliki tiga konsentrasi pilihan yaitu Sistem Informasi dan Enterprise, Jaringan Komputer dan Multimedia, serta Rekayasa Perangkat Lunak dan Animasi. Dengan adanya konsentrasi penjurusan yang ada mahasiswa diwajibkan untuk memilih salah satu dari konsentrasi jurusan yang ada. Penentuan konsentrasi dilaksanakan pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5.

Pada saat ini dalam menentukan konsentrasi jurusan hanya berdasarkan dari keinginan mahasiswa tanpa adanya suatu sistem yang memberikan pertimbangan atau aturan yang bisa dijadikan sebagai bahan acuan bagi mahasiswa dalam memilih konsentrasi jurusan sesuai dengan kemampuan akademik yang dimilikinya, pemilihan konsentrasi jurusan tanpa adanya pertimbangan yang matang sesuai

dengan kemampuan akademik yang dimiliki akan berdampak pada mahasiswa itu sendiri yang merasa salah mengambil konsentrasi dan menyebabkan kesulitan dalam penyerapan materi-materi perkuliahan sesuai dengan konsentrasi yang dipilihnya.

Berdasarkan dari permasalahan tersebut, maka dalam penelitian ini menggunakan metode *Clustering K-Means* diharapkan bisa dijadikan sebagai sumber informasi penentu kelas konsentrasi bagi mahasiswa.

## TINJAUAN PUSTAKA

### 1. Telaah Penelitian

Menurut Nugroho dan Setyawan (2014), data yang berlimpah bisa dimanfaatkan untuk *data mining* dalam rangka pengelolaan yang lebih baik dan pelaksanaan pembelajaran yang efektif. Salah satunya adalah Fakultas Komunikasi dan Informatika (FKI) UMS yang memiliki sebanyak 2358 mahasiswa termasuk yang sudah lulus sebanyak 600-700 mahasiswa. Klasifikasi masa studi terhadap data lulusan mahasiswa FKI UMS menggunakan metode *Decision Tree* dengan algoritma C4.5. Dari total 2358 data diambil sebanyak 341 data

mahasiswa yang sudah lulus. Atribut yang digunakan jurusan sekolah, jenis kelamin, asal sekolah, rerata jumlah SKS per semester, dan peran menjadi asisten. Hasil penelitian mengindikasikan variabel yang digunakan sebagai pertimbangan bagi fakultas untuk memperoleh tingkat masa studi efektif yaitu rerata SKS yang diambil mahasiswa.

## 2. Landasan Teori

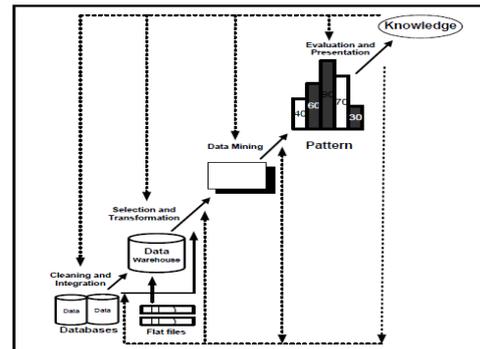
### a. Kelas Konsentrasi

Menurut Kamus Besar Bahasa Indonesia (2014), konsentrasi adalah pemusatan perhatian atau pikiran pada suatu hal, sedangkan Kelas Konsentrasi adalah suatu penjurusan yang menitik beratkan pada satu pembelajaran di bidang tertentu (KBBI, 2014).

### b. Data Mining

Nugroho (2014) mendefinisikan *data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata. Informasi yang dihasilkan diperoleh dengan cara mengekstrasi dan mengenali pola yang menarik dari suatu data.

Dalam *data mining* pengelompokan bisa dilakukan, tujuannya untuk mengetahui pola universal data-data yang ada (Prasetyo, 2013). langkah-langkah melakukan *data mining*.



**Gambar 1** Langkah-langkah *data mining* (Fadli, 2011).

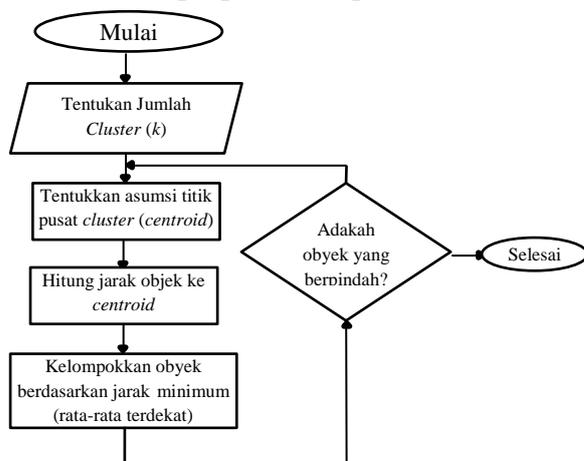
### c. Clustering

*Clustering* pengelompokan sejumlah data atau obyek ke dalam *cluster* (*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. Dalam metode *clustering* ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster*. Dalam metode ini tidak diketahui sebelumnya berapa jumlah *cluster* dan bagaimana pengelompokannya (Santoso, 2007).

### d. Algoritma *K-Means*

Algoritma *K-Means* pertama kali diperkenalkan oleh J. MacQueen pada tahun 1967, salah

satu algoritma *clustering* sangat umum yang mengelompokkan data sesuai dengan karakteristik atau ciri-ciri bersama yang serupa. Grup data ini dinamakan sebagai *cluster*. Data di dalam suatu *cluster* mempunyai ciri-ciri (karakteristik, atribut, properti) serupa.



**Gambar 2** Flowchart K-Means

Keterangan :

1. Tentukan jumlah *cluster* (k).
2. Alokasikan data ke dalam *cluster*.
3. Hitung *centroid*/rata-rata dari data yang ada di masing-masing *cluster*.
4. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat.
5. Kembali ke step 3, apabila masih ada perubahan *cluster* atau nilai *centroid* (Nugroho, 2014).

#### e. NetBeans

Syarief (2012) mendefinisikan NetBeans adalah Platform aplikasi kerangka kerja Desktop yang sangat lengkap untuk membuat aplikasi

desktop serta untuk mengembangkan berbagai aplikasi GUI dan Console berbasis Java. NetBeans adalah salah satu aplikasi IDE yang digunakan untuk menulis, mengkompile, mencari kesalahan, dan menyebarkan program. NetBeans mencakup *compiler* atau *builder* dan *debugger* internal sehingga mempermudah perancangan program (Hartati, 2007).

## METODE PENELITIAN

### 1. Identifikasi Masalah

Pada saat ini untuk menentukan konsentrasi jurusan hanya berdasarkan pada keinginan mahasiswa saja tanpa adanya suatu sistem yang bisa memberikan bahan pertimbangan dalam menentukan konsentrasi jurusan. Akibatnya banyak mahasiswa yang mengalami kesulitan saat mengambil konsentrasi yang mereka pilih.

Dengan adanya salah satu teknik *data mining clustering k-means* yang akan memberikan kelompok-kelompok data yang bisa dijadikan bahan pertimbangan mahasiswa untuk menentukan konsentrasi berdasarkan nilai

akademik yang berhubungan dengan konsentrasi yang ada.

## 2. Penentuan Kebutuhan dan Data

Setelah menganalisa identifikasi masalah tahap dalam menentukan data yang dibutuhkan untuk proses *clustering* yaitu data-data nilai yang akan digunakan dalam proses penelitian ini diperoleh dari IT UMS yaitu data nilai mahasiswa dari semester 1 sampai 4.

**Tabel 1** Daftar Atribut

Mata Kuliah	Semester
Pemrograman Web Dasar	I
Algoritma dan Pemrograman	I
Pengantar Sistem Informasi	III
Sistem Basis Data	IV
Komunikasi Data	II
Jaringan Komputer	III
Prak. Jaringan Komputer	III
Sistem Operasi	III
Aljabar Linier dan Matriks	I
Algoritma dan Struktur Data	II
Struktur Diskret 2	III
Sistem Digital	IV

Penjelasan dari atribut yang dipilih berdasarkan (RMP, 2014).

## 3. Kelengkapan data dan kebutuhan

Beberapa tahapan proses *data mining* antara lain yaitu:

### a. Data Cleaning

Pada proses data *cleaning* perlu dilakukan karena terdapat beberapa

data yang mempunyai isian tidak sempurna seperti ada sel yang tidak punya nilai, data mengandung *error*, ada *outlier* serta ada inkonsistensi pada isian sel. Sel yang tidak mempunyai nilai bisa diabaikan, diisi dengan rata-rata data, diisi dengan rata-rata kelompok, dan estimasi (perkiraan).

### b. Penentuan Sampel

Dalam penelitian penentuan sampel merupakan kebutuhan utama untuk memperoleh data yang menggambarkan realitas yang akan dijadikan fokus dalam penelitian dengan menggunakan rumus *slovin* sebagai berikut:

$$n = \frac{N}{1+Ne^2} \dots\dots\dots (1)$$

Keterangan :

$n$  = Jumlah *sample*

$N$  = Jumlah populasi

$e^2$  = Toleransi ketidaktelitian

## 4. Perancangan Aplikasi *Data Mining*

Dalam proses penelitian perlu adanya perancangan aplikasi *data mining* dengan metode *clustering k-means* agar memudahkan penulis. Adapun beberapa tahapan-tahapan yang diperlukan dalam prosesnya, yaitu :

**a. Perubahan Parameter Nilai dalam format Numerik (angka)**

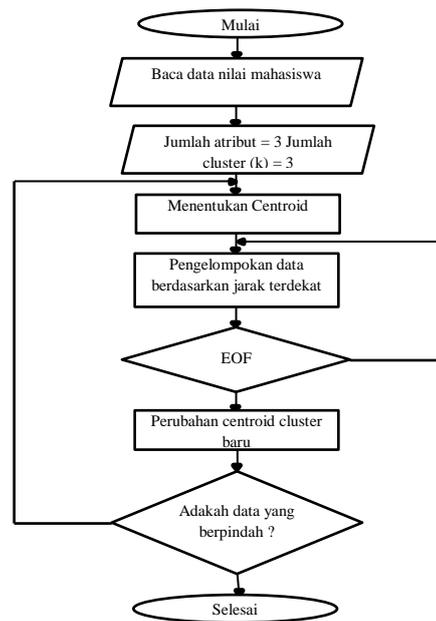
Data nilai mata kuliah yang awalnya berupa huruf akan diolah untuk *clustering* dengan algoritma *k-means* harus diubah ke dalam bentuk angka agar dapat diproses. Tabel 2 adalah daftar perubahan parameter nilai.

**Tabel 2** Perubahan Parameter Nilai

NILAI	ANGKA
A	4
AB	3.5
B	3
BC	2.5
C	2
D	1
E/-	0

**b. Menentukan Metode**

Dalam menentukan metode untuk penelitian penulis menggunakan metode *clustering k-means*. Beberapa kelebihan dari *k-means* yaitu dalam prosesnya cepat dan dinilai cukup efisien untuk proses pengelompokan dalam jumlah data yang banyak. Selain itu, dalam iterasinya, algoritma ini akan berhenti dalam kondisi *optimum local* (Andayani, 2007). Secara umum urutan proses *clustering* dengan algoritma *k-means* ditunjukkan pada gambar 3.



**Gambar 3** Flowchart Proses *K-Means*

Beberapa metode yang digunakan untuk menghitung selisih jarak adalah *Euclidean distance*. Adapun persamaanya adalah sebagai berikut:

$$d(x, y) = ||x - y||^2$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (2)$$

Keterangan :

- $d(x,y)$  = Jarak objek antara  $X_i$  dan  $Y_i$
- $n$  = Dimensi data
- $X_i$  = Koordinat dari objek  $X_i$  pada dimensi  $i$
- $Y_i$  = Koordinat dari objek  $Y_i$  pada dimensi  $i$

## HASIL DAN PEMBAHASAN

### 1. Pembahasan *Clustering K-Means*

#### a. Menentukan Variabel

Tabel variabel mata kuliah yang digunakan:

**Tabel 3** Mata Kuliah

Vr	Mata Kuliah	Konsentrasi
X1	Pemrograman Web Dasar	Sistem Informasi dan Enterprise
X2	Algoritma dan Pemrograman	
X3	Pengantar Sistem Informasi	
X4	Sistem Basis Data	
X5	Komunikasi Data	Jaringan Komputer dan Multimedia
X6	Jaringan Komputer	
X7	Prak. Jaringan Komputer	
X8	Sistem Operasi	
X9	Aljabar Linier dan Matriks	Rekayasa Perangkat Lunak dan Animasi
X10	Algoritma dan Struktur Data	
X11	Struktur Diskret 2	
X12	Sistem Digital	

#### b. Menentukan Sampel

Berdasarkan jumlah mahasiswa reguler jurusan Informatika UMS dari tahun 2007 sampai dengan tahun 2013 yaitu sebanyak 1173 mahasiswa, yang dinotasikan dalam jumlah populasi ( $N$ ). Dengan toleransi ketidakteelitian sebesar 5% dinotasikan dalam ( $e$ ).

$$n = \frac{N}{1 + Ne^2}$$

$$n = 1173 / 1 + 1173 \times (0,05)^2$$

$$n = 1173 / 1 + 1173 \times 0,0025$$

$$n = 1173 / 1 + 2,9325$$

$$n = 1173 / 3,9325$$

$$n = 298,2835 \text{ sampel.}$$

Dari hasil perhitungan untuk menentukan sampel maka dibulatkan untuk data yang digunakan yaitu data mahasiswa angkatan 2011 dan data mahasiswa angkatan 2012 dengan jumlah 312 mahasiswa.

#### c. Inisialisasi Data Nilai

Perubahan format nilai dikarenakan metode *clustering k-means* bekerja pada tipe data berupa angka. Data yang sudah diubah dikelompokkan berdasarkan mata kuliah yang dominan ke masing-masing konsentrasi meliputi 12 mata kuliah yang akan dikelompokkan dan dicari rata-rata nilainya (Rata-rata SI, Rata-rata Jarkom, dan Rata-rata RPL). Kemudian di-*cluster* menjadi 3 *group*, berikut adalah cara mencari nilai rata-rata.

$$X_S = \frac{X1+X2+X3+X4}{4}$$

= Nilai Rata-Rata SI

$$X_J = \frac{X5+X6+X7+X8}{4}$$

= Nilai Rata-Rata JARKOM

$$X_R = \frac{X9+X10+X11+X12}{4}$$

= Nilai Rata-Rata RPL

Tabel 4 menunjukkan beberapa data nilai rata-rata konsentrasi.

**Tabel 4** Nilai Rata-rata Konsentrasi

Nim	Dataset	X <sub>S</sub>	X <sub>J</sub>	X <sub>R</sub>
L200110001	1	3,13	3,50	2,63
L200110002	2	2,88	3,25	3,38
L200110003	3	2,50	2,63	1,88
L200110004	4	2,38	3,38	2,25
L200110005	5	3,25	3,63	3,25
L200110006	6	3,25	3,63	3,63
L200110007	7	2,25	3,50	1,50
L200110008	8	1,63	1,00	0,75
L200110009	9	2,75	3,38	1,88
L200110010	10	2,63	3,63	2,38
.....	.....	.....	.....	.....

**2. Proses Manual Clustering K-Means**

Tahapan proses perhitungan secara manual yaitu:

a. Menentukan jumlah *cluster*

Dalam menentukan jumlah *cluster* dari data-data yang ada akan dibuat menjadi 3 *cluster* misalkan A, B, C bila dibuat indeks yaitu 1, 2, 3 agar lebih mudah dalam proses perhitungan manualnya.

b. Menghitung nilai *centroid* awal

Dalam menghitung nilai *centroid* awal menggunakan nilai *mean*. Nilai masing-masing kolom rata-rata konsentrasi diambil sampel 10 *dataset* saja adalah:

**Tabel 5** Nilai rata-rata

Variabel	X <sub>S</sub>	X <sub>J</sub>	X <sub>R</sub>
Rata-rata	2,665	3,153	2,353

Sehingga nilai *centroid* awal A, B, C ditunjukkan pada tabel 6.

**Tabel 6** Nilai *Centroid* awal

	Centroid		
A	2,665	3,153	2,353
B	2,665	3,153	2,353
C	2,665	3,153	2,353

c. Menghitung *dataset* ke 1

K	Centroid			Dataset ke 1			
A	2,665	3,153	2,353	1	2	3	Cluster
B	2,665	3,153	2,353				
C	2,665	3,153	2,353				
				Euclidean Distance			
ke	X <sub>S</sub>	X <sub>J</sub>	X <sub>R</sub>	A	B	C	
1	3,13	3,5	2,63	0,643	0,643	0,643	1
K	Centroid						
A	3,13	3,5	2,63				
B	2,665	3,153	2,353				
C	2,665	3,153	2,353				

**Gambar 4** Hasil *Centroid* dan *Cluster Dataset* ke 1

Cara menentukan jarak dengan teknik *euclidean distance* yaitu :

$$A = \sqrt{(3,13 - 2,665)^2 + (3,5 - 3,153)^2 + (2,63 - 2,353)^2} = 0,643$$

$$B = \sqrt{(3,13 - 2,665)^2 + (3,5 - 3,153)^2 + (2,63 - 2,353)^2} = 0,643$$

$$C = \sqrt{(3,13 - 2,665)^2 + (3,5 - 3,153)^2 + (2,63 - 2,353)^2} = 0,643$$

Maka hasil jarak terkecil pertama antara nilai *centroid* dengan objek yaitu A sehingga dataset ke-1 termasuk *cluster* 1. Hasil *centroid* yaitu:

$$B = \frac{3,13}{1}, \frac{3,5}{1}, \frac{2,63}{1} = 3,13 . 3,5 . 2,63$$

B dan C belum ada sehingga nilai *centroid*-nya tetap.

d. Menghitung *dataset* ke 2

K	Centroid		
A	3,13	3,5	2,63
B	2,665	3,153	2,353
C	2,665	3,153	2,353

Dataset ke 2							
Dataset				1	2	3	Cluster
ke	X <sub>S</sub>	X <sub>J</sub>	X <sub>R</sub>	A	B	C	
1	3,13	3,5	2,63				1
2	2,88	3,25	3,38	0,829	1,054	1,054	1
				Euclidean Distance			

K	Centroid		
A	3,005	3,375	3,005
B	2,665	3,153	2,353
C	2,665	3,153	2,353

**Gambar 5** Hasil *Centroid* dan *Cluster Dataset* ke 2

Perhitungan jarak dengan teknik *euclidean*.

$$A = \sqrt{(2,88 - 3,13)^2 + (3,25 - 3,5)^2 + (3,38 - 2,63)^2} = 0,829$$

$$B = \sqrt{(2,88 - 2,665)^2 + (3,25 - 3,153)^2 + (3,38 - 2,353)^2} = 1,054$$

$$C = \sqrt{(2,88 - 2,665)^2 + (3,25 - 3,153)^2 + (3,38 - 2,353)^2} = 1,054$$

Maka hasil terkecil pertama yaitu A sehingga dataset ke-2 termasuk *cluster* 1 dan hasil *centroid* yaitu :

$$A = \frac{3,13+2,88}{2}, \frac{3,5+3,25}{2}, \frac{2,63+3,38}{2} = 3,005 . 3,375 . 3,005$$

B serta C belum ada sehingga nilai *centroid*-nya tetap.

e. Menghitung *dataset* ke 3

K	Centroid		
A	3,005	3,375	3,005
B	2,665	3,153	2,353
C	2,665	3,153	2,353

Dataset ke 3							
Dataset				1	2	3	Cluster
ke	X <sub>S</sub>	X <sub>J</sub>	X <sub>R</sub>	A	B	C	
1	3,13	3,5	2,63				1
2	2,88	3,25	3,38				1
3	2,5	2,63	1,88	1,441	0,724	0,724	2
				Euclidean Distance			

K	Centroid		
A	3,005	3,375	3,005
B	2,5	2,63	1,88
C	2,665	3,153	2,353

**Gambar 6** Hasil *Centroid* dan *Cluster Dataset* ke 3

Dalam menentukan *cluster* perhitungannya sama yaitu dimulai dari perhitungan jarak dengan titik pusat atau *centroid* terhadap objek data, berikut adalah cara menghitung jarak dengan teknik *euclidean*:

$$A = \sqrt{(2,5 - 3,005)^2 + (2,63 - 3,375)^2 + (1,88 - 3,005)^2} = 1,441$$

$$B = \sqrt{(2,5 - 2,665)^2 + (2,63 - 3,153)^2 + (1,88 - 2,353)^2} = 0,724$$

$$C = \sqrt{(2,5 - 2,665)^2 + (2,63 - 3,153)^2 + (1,88 - 2,353)^2} = 0,724$$

Maka hasil jarak terkecil (pertama) yaitu B sehingga *dataset* ke 3 masuk ke dalam kelas atau *cluster* 2 dan hasil perhitungan *centroid*-nya yaitu :

$$A = \frac{3,13+2,88}{2}, \frac{3,5+3,25}{2}, \frac{2,63+3,38}{2} = 3,005 . 3,375 . 3,005$$

$$B = \frac{2,5}{1}, \frac{2,63}{1}, \frac{1,88}{1} = 2,5 . 2,63 . 1,88$$

C belum ada sehingga nilai *centroid*-nya tetap.

f. Hasil *centroid* akhir

Dalam perhitungan *centroid* dengan nilai rata-rata inti dari perhitungannya adalah adanya *update centroid* tiap data ke - n yang berdasarkan jumlah dari kelas atau *cluster* yang ada.

### 3. Implementasi *Clustering K-Means* dengan Netbeans 7.4

a. Menu Utama

Menu utama dalam aplikasi ini terdiri menu *clustering* dan menu *about* serta ada tombol keluar. *Background* yang digunakan adalah map cover dari informatika agar lebih menarik dalam tampilannya.



**Gambar 7** Tampilan Utama

**b. Menu About**

Menu *about* sebagai menu tambahan saja yaitu berisi tentang informasi dari mahasiswa, dosen pembimbing, dewan penguji I dan II.



**Gambar 8** Menu About

**c. Menu Clustering**

Pada menu hasil aplikasi *clustering k-means* berupa hasil akhir nilai *centroid* atau nilai titik pusat yang akan menghasilkan nilai *cluster* berdasarkan *Euclidean distance* (jarak terkecil) dari objek data ke titik pusat (*centroid*) yang akan menghasilkan kelas

konsentrasi. Hasil *clustering* ditunjukkan pada gambar 9.



**Gambar 9** Hasil Clustering

Berdasarkan hasil *clustering* maka diperoleh *centroid* akhir yaitu: **Tabel 7** Hasil *Centroid* akhir data 2011

Cluster	Centroid		
1	3,362	3,560	3,123
2	0,494	0,640	0,544
3	2,629	3,148	1,881

Dari nilai *centroid* untuk *cluster* 1, 2 dan 3 diasumsikan masing-masing konsentrasinya yaitu *cluster* 1 (Jaringan Komputer dan Multimedia), *cluster* 2 (Rekayasa Perangkat Lunak dan Animasi) dan *cluster* 3 (Sistem Informasi dan Enterprice). Asumsi ini diperoleh dari rata-rata nilai variable masing konsentrasi dan hasil final *centroid*.

**Tabel 8** Menentukan Konsentrasi dan *Cluster* data 2011

Konsentrasi	Rata-rata	Cluster
Jaringan Komputer dan Multimedia	3,081	1
Rekayasa Perangkat Lunak dan Animasi	2,360	2
Sistem Informasi dan Enterprice	2,764	3

## KESIMPULAN

1. Dengan melakukan perancangan aplikasi *Clustering K-Means* dapat menggali informasi-informasi serta memberikan kelompok-kelompok data yang bisa dijadikan sebagai sumber informasi dalam proses menentukan konsentrasi jurusan berdasarkan *cluster* yang didapat dari jarak terkecil objek data dengan titik pusat (*centroid*).
2. Berdasarkan sampel data sebanyak 312 data dibagi menjadi 2 angkatan yaitu angkatan informatika 2011 dan 2012. Data mahasiswa informatika angkatan 2011 sebanyak 154 data diperoleh data *cluster* sebanyak 78 data masuk ke *cluster* 1 dengan titik pusat atau *centroid* (3,333 ; 3,548 ; 3,098) dan 17 data masuk ke *cluster* 2 dengan titik *centroid* (0,914 ; 1,109 ; 0,773) serta 60 data masuk ke *cluster* 3 dengan titik *centroid* (2,68 ; 3,220 ; 1,879). Sedangkan untuk data informatika angkatan 2012 diperoleh data *cluster* sebanyak 69 data masuk ke *cluster* 1 dengan titik *centroid* (3,359 ; 3,576 ; 3,188), dan 18 data masuk ke *cluster* 2 dengan titik *centroid* (0,944 ; 1,382 ; 1,261) serta 64 data masuk ke *cluster* 3 dengan titik *centroid*

(2,485 ; 3,367 ; 2,250). Pada anggota data yang masuk *cluster* 1 maka akan direkomendasikan untuk masuk ke kelas konsentrasi Jaringan Komputer dan Multimedia, untuk anggota *cluster* 2 maka akan direkomendasikan untuk masuk ke kelas konsentrasi Rekayasa Perangkat Lunak dan Animasi sedangkan untuk anggota *cluster* 3 akan merekomendasikan untuk masuk ke kelas konsentrasi Sistem Informasi dan Enterprise.

## DAFTAR PUSTAKA

- Andayani, Sri. (2007). *Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means*. Jurnal, semnasIF 2007. Yogyakarta: Fakultas Matematika dan Pendidikan Matematika Universitas Negeri Yogyakarta.
- Fadli, Ari. (2011). *Konsep Data Mining*. Diakses dari: Komunitas Elearning Ilmu Komputer <<http://ilmukomputer.org/category/datamining/>> [diakses tanggal 16 September 2014].
- Hartati, Sri Wijono. (2007). *Pemrograman GUI Swing Java Dengan Netbeans 5*. Yogyakarta: Penerbit Andi.
- Kamus Besar Bahasa Indonesia. (2014). *Pengertian Kelas Konsentrasi*. Diakses dari: <<http://kbbi.web.id/>> [diakses tanggal 14 September 2014].
- Nugroho, Yusuf Sulistyio dan Setyawan. (2014). *Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika*. Jurnal Komuniti, Vol. 6, No. 1 Maret 2014: 84-91. Surakarta: Fakultas Komunikasi dan informatika Universitas Muhammadiyah Surakarta.
- Nugroho, Yusuf Sulistyio. (2014). *Modul Praktikum Data Mining Berdasarkan Kurikulum 2013*. Surakarta: Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta.
- Prasetyo, Eko. (2013). *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Rencana Mutu Pembelajaran Universitas Muhammadiyah Surakarta. (2014). *Penjelasan masing-masing atribut*. Diakses dari: <<http://rmp.ums.ac.id/>> [diakses tanggal 14 September 2014].
- Santoso. (2007). *Pengertian K-Means Clustering*. Diakses dari: <[eprints.undip.ac.id/23168/1/TA\\_NUXON\\_J2F005280.pdf](http://eprints.undip.ac.id/23168/1/TA_NUXON_J2F005280.pdf)> [diakses tanggal 13 September 2014].
- Syarief, Mulkam. (2012). *Berbagai Project Java dengan IDE NetBeans*. Yogyakarta: Penerbit Andi.

## **BIODATA PENULIS**

Nama : Dian Setiawan

Tempat, Tanggal Lahir : Sragen, 16 Januari 1992

Jenis Kelamin : Laki-Laki

Agama : Islam

Jurusan : Informatika

Peguruan Tinggi : Universitas Muhammadiyah Surakarta

Alamat : Jl. A. Yani Tromol Pos 1 Pabelan, Kartasura

Telp./Fax : (0271)717417, 719483 / (0271)714448

Alamat Rumah : Gendol RT03, Mojokerto, Kedawung, Sragen

No. HP : 085642455877

Alamat e-mail : [dianst92@gmail.com](mailto:dianst92@gmail.com)