

## APLIKASI ALGORITMA FUZZY C-MEANS CLUSTERING UNTUK PENGELOMPOKKAN LULUSAN

Abdul Aziz, S.Si, M.Si.

### Abstrak

Teknik *Fuzzy c-means clustering* termasuk dalam salah satu keluarga clustering. Seperti teknik clustering lainnya, teknik inipun mencoba mengelompokkan sejumlah objek. Pada artikel ini dengan menggunakan teknik *fuzzy c-means clustering* akan mencoba mengelompokkan objek para lulusan yang akan dibagi kedalam 3 cluster, adapun besar pangkat bobot dipilih  $m = 2$ , iterasi maksimum pada pelaksanaan perhitungan adalah 10, criteria penghentian iterasi adalah bila selisih antara 2 solusi yang berurutan telah bernilai kurang dari  $10^{-5}$ , hasil dari perhitungan adalah didapatkannya pusat cluster atau center, derajat keanggotaan atau markis U serta nilai fungsi tujuan atau ObjFcn. Hasil klasifikasi ini diharapkan dapat memberikan gambaran mengenai lulusan suatu program studi atau perguruan tinggi berdasarkan kualitas lulusan. dinyatakan oleh IPK dan lama studi

**Kata kunci:** *Fuzzy c-means clustering*, pelompokan, lulusan

### 1. Pendahuluan

Lulusan sebagai output terakhir dari sebuah perguruan tinggi, biasa diberi predikat kelulusan. Dasar dari pemberian predikat adalah indeks prestasi kumulatif (IPK). Klasifikasi predikat kelulusan adalah sebagai berikut: IPK 2, 25 – 2,75 memuaskan, 2, 76 – 3, 50 sangat memuaskan, 3, 51 – 4, 00 terpuji. Selain IPK untuk mengukur tingkat keberhasilan lulusan dalam menyerap ilmu pengetahuan adalah lama studi. Semakin tinggi IPK dan semakin kecil lama studi lulusan maka akan semakin baik, sehingga jika IPK dan lama studi dijadikan sebagai dasar klasifikasi, maka dapat mengukur tingkat keberhasilan dan kecepatan lulusan menyelesaikan studinya. Hasil klasifikasi ini diharapkan dapat memberikan gambaran mengenai lulusan suatu program studi atau perguruan tinggi berdasarkan kualitas lulusan. dinyatakan oleh IPK dan lama studi.

Bagaimanakah klasifikasi atau pengelompokkan lulusan itu akan dilakukan? Teknik *Fuzzy c-means clustering* termasuk dalam salah satu keluarga clustering. Seperti teknik clustering lainnya, teknik inipun mencoba mengelompokkan sejumlah objek. Dalam teknik clustering, untuk istilah “kelompok” digunakan kata “cluster”. Pengelompokkan yang diharapkan adalah yang memenuhi sifat sedemikian rupa sehingga objek-objek dalam suatu cluster “semirip” mungkin satu sama lain, dan “seberbeda” mungkin dengan objek-objek yang tidak berada pada cluster tersebut.

Pada umumnya teknik clustering, relasi yang ada diantara suatu objek dengan suatu cluster hanya berada pada 2 kemungkinan ekstrim. Kemungkinan ekstrim pertama, suatu objek adalah anggota penuh dari suatu cluster, atau derajat keanggotaan suatu objek terhadap suatu cluster adalah 100%. Kemungkinan ekstrim kedua, suatu objek sama sekali bukan anggota dari suatu cluster, atau derajat keanggotaan suatu objek terhadap suatu cluster adalah 0 %. Tidak demikian halnya pada teknik *fuzzy c-means clustering*, pada teknik ini reaksi yang ada diantara suatu objek dengan suatu cluster dimungkinkan pula untuk berada diantar kedua ekstrim tersebut, artinya relasi suatu objek terhadap suatu cluster dimungkinkan berderajat keanggotaan 0% atau 100% atau bahkan diantaranya.

Pada artikel ini dengan menggunakan teknik *fuzzy c-means clustering* akan mencoba mengelompokkan objek para lulusan. Adapun masalah yang akan dibahas dalam artikel ini adalah:

1. Bagaimana hasil pengelompokan lulusan dengan menggunakan algoritma *Fuzzy C-Means Clustering*?
2. Dimana titik pusat cluster pada *Fuzzy C-Means Clustering* pada pengelompokan lulusan?

Teknik yang digunakan dalam memecahkan masalah ini adalah dengan menggunakan teknik *Fuzzy C-Means Clustering*. Dengan jenis data yang diambil adalah data IPK dan lama studi dari keempat puluh lulusan Universitas Islam Negeri (UIN) Malang semester ganjil tahun akademik 2005/2006.

## 2. Teknik *Fuzzy c-means clustering*

Bentuk umum dari masalah clustering adalah sebagai berikut:

- Terdapat  $n$  buah objek, yaitu  $0_1, 0_2, \dots, 0_k, \dots, 0_n$ ,
- Objek ke- $k$  ( $0_k$ ) dirincikan oleh  $p$  buah atribut, yang dinyatakan dalam bentuk vector  $x_k = (x_{k1} \ x_{k2} \ \dots \ x_{kj} \ \dots \ x_{kp})$  dengan  $x_{kj}$  menyatakan nilai atribut ke- $j$  dari objek ke- $k$  ( $0_k$ ),
- Berdasarkan ke- $p$  buah atribut yang dimiliki setiap objek, ke- $n$  buah objek ini akan dikelompokkan kedalam  $c$  buah cluster, yaitu cluster-1, cluster-2, ..., cluster- $c$ ,
- Yang harus ditentukan adalah besaran:
  - Matriks  $U = (U_{ik})_{c \times n}$ , yaitu matriks dengan sel pada baris ke- $i$  dan kolom ke- $k$  menyatakan besarnya derajat keanggotaan  $0_k$  terhadap cluster- $i$ , sehingga objek-objek se-cluster akan semirip mungkin satu sama lain, dan seberbeda mungkin dengan objek-objek yang tidak secluster.
  - Besaran vector baris yang terdiri atas  $p$  kolom,  $v_i = (v_{i1} \ v_{i2} \ \dots \ v_{ij} \ \dots \ v_{ip})$ , dengan elemen menyatakan nilai-nilai rata-rata dari atribut ke- $j$  yang dimiliki oleh setiap objek yang merupakan anggota cluster ke- $i$

Adapun range dari indeks-indeks diatas adalah:  $i = 1, 2, \dots, c$ ;  $j = 1, 2, \dots, p$  dan  $k = 1, 2, \dots, n$ .

Bila masalah umum diatas dimodelkan kedalam masalah *fuzzy c-means clustering* maka akan diperoleh masalah optimasi terkendala (*constrained optimization*) berikut ini (Bezdek, 1981):

- Diberikan  $n$  buah vector baris  $x_k = (x_{k1} \ x_{k2} \ \dots \ x_{kj} \ \dots \ x_{kp})$
- Harus ditentukan dua besaran berikut:
  - \* Matriks  $U = (U_{ik})_{c \times n}$ , dan
  - \* Vector baris  $v_i = (v_{i1} \ v_{i2} \ \dots \ v_{ij} \ \dots \ v_{ip})$
- Sehingga nilai fungsional berikut menjadi minimum:

$$J(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2 \tag{1}$$

Dalam hal ini  $m \in [0, 1]$  disebut pangkat bobot (dalam penelitian ini nilai yang umum digunakan, yaitu  $m = 2$ )

- Dan kendala-kendala berikut terpenuhi:

$$U_{ik} \in [0, 1] \tag{2}$$

$$\sum_{i=1}^c u_{ik} = 1, \text{ untuk setiap } k = 1, 2, \dots, n \tag{3}$$

$$0 < \sum_{k=1}^n u_{ik} < n, \text{ untuk setiap } k = 1, 2, \dots, c \tag{4}$$

Pada persamaan (1), besaran  $\|x_k - v_i\|$  menyatakan jarak antara atribut dari objek ke-k ( $x_k$ ) terhadap “titik pusat” dari cluster-i ( $v_i$ ). Mengingat elemen-elemen vector  $v$  menyatakan nilai rata-rata atribut dari semua objek menjadi anggota cluster i, maka vector ini merupakan prototype atau representasi dari seluruh objek yang menjadi anggota cluster-i. Semakin kecil nilai  $\|x_k - v_i\|$ , maka semakin mirip atribut dari objek ke-k dengan prototype (representasi) dari cluster-I, sehingga semakin pantas objek ini untuk menjadi anggota cluster-i, yang berarti semakin besar nilai  $u_{ik}$ . Sebaliknya semakin besar nilai  $\|x_k - v_i\|$  maka semakin kecil nilai  $u_{ik}$ .

Pada kendala (2), nilai keanggotaan objek ke-k ( $o_k$ ) berada pada selang 0 sampai 1. pada kendala (3), untuk setiap objek, jumlah dari seluruh nilai derajat keanggotaannya adalah 1. pada kendala (4), setiap objek pasti menjadi anggota salah satu cluster, namun untuk setiap objek tidak mungkin menjadi anggota dari seluruh cluster.

Penerapan teknik *fuzzy c-means clustering* ini dapat diuraikan sebagai berikut. Terdapat 40 lulusan sebagai sample penelitian, dan setiap lulusan ini dirincikan oleh dua jenis data (atribut) yang dimilikinya, yaitu IPK (berskala 0-4) dan lama studi (dalam tahun). Ke-40 lulusan ini akan dikelompokkan kedalam sekian buah cluster. Kedua atribut dari setiap lulusan inilah yang menjadi dasar pengelompokkan. Pada teknik *fuzzy c-means clustering* terlebih dahulu menentukan besaran banyaknya jumlah cluster yang diharapkan. Dari hasil akhir upaya pengelompokkan diharapkan dapat mengetahui derajat keanggotaan seorang lulusan tertentu terhadap suatu cluster tertentu, Lulusan yang mana sajakah yang akan menjadi anggota dari suatu cluster tertentu, Setelah cluster-cluster dari pada lulusan terbentuk bagaimana gambaran dari cluster-cluster tersebut.

### 3. Data Nilai IPK Lulusan

Data yang terambil disajikan pada tabel berikut:

No sampel	IPK ( $X_i$ )	Lama Studi ( $Y_i$ )
1	2.75	5.50
2	2.76	5.50
3	2.60	5.83
4	2.83	5.83
5	2.81	5.83
6	2.49	6.17
7	2.55	6.17
8	2.53	6.17
9	2.60	6.83
10	2.45	7.25
11	3.49	4.85
12	3.42	4.83
13	3.11	5.17
14	2.76	5.17
15	2.63	5.83
16	2.62	6.42
17	2.63	6.42
18	2.62	6.67
19	3.37	4.17
20	2.96	4.50

21	2.79	4.50
22	3.21	4.83
23	3.05	4.83
24	2.77	6.25
25	3.13	5.42
26	3.72	4.25
27	3.10	5.17
28	2.92	5.42
29	3.12	5.42
30	2.94	5.42
31	2.98	5.42
32	2.86	3.67
33	2.93	7.83
34	2.67	6.83
35	2.84	5.83
36	2.77	4.83
37	2.37	8.17
38	2.72	4.17
39	2.84	5.25
40	2.68	5.33

Tabel 1. IPK dan lama studi 40 lulusan (sumber: Ernawati, 2003)

Data IPK dan lama studi dari setiap mahasiswa dapat dinyatakan sebagai vector baris yang terdiri atas 2 kolom. Sebagai contoh untuk mahasiswa ke-1, ke dua data ini dapat dinyatakan sebagai vektor

$$x_1 = (x_{11} \ x_{12}) = (2.75 \ 5.50), x_2 = (2.76 \ 5.50), \dots, x_{40} = (2.68 \ 5.33)$$

Ke-40 lulusan yang akan menjadi objek dari *teknik fuzzy c-means clustering* dapat dikelompokkan menjadi  $c$  cluster, dengan  $c = 1, 2, \dots, 40$ . mengelompokkan 40 lulusan menjadi 1 atau 40 cluster tidak menimbulkan kesulitan untuk mengidentifikasi lulusan yang mana sajakah yang akan menjadi anggota cluster-cluster tersebut. Tidak demikian halnya bila ke-40 lulusan itu hendak kita kelompokkan kedalam 2, 3, ..., 38 cluster. Pada bagian ini ke-40 lulusan ini akan dikelompokkan menjadi 3 cluster.

#### 4. Hasil Pengelompokan

Misalkan ke-40 lulusan akan dikelompokkan kedalam 3 cluster maka nilai  $c$  adalah 3. Misalkan data pada tabel sebelumnya disimpan diberi nama "DATA.dat" dan disimpan pada drive C, untuk mengolah data ini digunakan dengan bantuan software MATLAB 6.5, dengan perintah sebagai berikut:

```
>> X = load('c:\DATA.dat');
>> [center,U,ObjFcn] = fcm(X,3,[2,10,10^-5])
```

Perintah "X = load('c:\DATA.dat')" akan membaca file dengan nama "DATA.dat" dan komputer akan membacanya dalam variabel X. Perintah "[center,U,ObjFcn] = fcm(X,3,[2,10,10^-5])" akan memasukkan nilai variabel X kedalam fungsi *fcm (fuzzy c-means)* dan data pada variabel ini akan dibagi kedalam 3 cluster, adapun besar pangkat bobot dipilih  $m = 2$ , iterasi maksimum pada

pelaksanaan perhitungan adalah 10, criteria penghentian iterasi adalah bila selisih antara 2 solusi yang berurutan telah bernilai kurang dari  $10^{-5}$ , hasil dari perhitungan adalah didapatkannya pusat cluster atau center, derajat keanggotaan atau markis U serta nilai fungsi tujuan atau ObjFcn . Setelah kedua perintah diberikan , didapat 3 bagian hasil berikut ini

**Hasil pertama**, yaitu hasil perhitungan nilai fungsional, sebagai berikut:

Iterasi count = 1,obj.fcn = 16.098602
Iterasi count = 2,obj.fcn = 12.126401
Iterasi count = 3,obj.fcn = 10.444322
Iterasi count = 4,obj.fcn = 9.553445
Iterasi count = 5,obj.fcn = 8.984033
Iterasi count = 6,obj.fcn = 8.188539
Iterasi count = 7,obj.fcn = 7.629264
Iterasi count = 8,obj.fcn = 7.366269
Iterasi count = 9,obj.fcn = 7.249556
Iterasi count = 10,obj.fcn = 7.201554

Interpretasinya, software MATLAB 6.5 memerlukan iterasi 10 kali sebelum memperoleh solusi optimal bagi nilai fungsional  $J(U,v)$  sebesar 7.201554.

**Hasil kedua**, yaitu hasil perhitungan dari nilai-nilai  $v_{ij}$ , sebagai berikut:

Center =		
2.7173	7.0516	
2.7660	5.7233	
3.0681	4.6110	

Nilai-nilai ini merupakan nilai dari “koordinat” titik pusat ketiga cluster dan memberikan garis besar citra tiap cluster:

- ❖ Untuk cluster-1, “koordinat” dari titik pusat cluster ini adalah  $v_1 = (2.7173 \ 7.0516)$ , yang arti fisisnya, cluster 1 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7173 dan rata-rata lama lulus 7.0516 tahun;
- ❖ Untuk cluster-2, “koordinat” dari titik pusat cluster ini adalah  $v_2 = (2.7660 \ 5.7233)$ , yang arti fisisnya, cluster 3 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7660 dan rata-rata lama lulus 5.7233 tahun; sedangkan
- ❖ Untuk cluster-3, “koordinat” dari titik pusat cluster ini adalah  $v_3 = (3.0681 \ 4.6110)$ , yang arti fisisnya, cluster 3 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7660 dan rata-rata lama lulus 5.7233 tahun

Dari hasil perhitungan dari nilai-nilai  $u_{ik}$  didapat pola yang menarik, bahwa semakin tinggi nilai IPK seorang alumni, semakin singkat masa studinya.

**Bagian ketiga**, yaitu hasil perhitungan dari nilai-nilai  $u_{ik}$ , sebagai berikut:

U =					
Columns 1 through 7					
0.0193	0.0192	0.0246	0.0101	0.0087	
0.0513	0.2188				
0.9285	0.9284	0.9536	0.9801	0.9828	

0.9052	0.7159			
0.0522	0.0523	0.0218	0.0098	0.0084
0.0435	0.0653			
Columns 8 through 14				
0.2231	0.2091	0.8979	0.7622	0.0286
0.0092	0.0470			
0.7102	0.7285	0.0817	0.1699	0.1282
0.0510	0.5441			
0.0666	0.0624	0.0204	0.0679	0.8433
0.9398	0.4089			
Columns 15 through 21				
0.0192	0.5198	0.5197	0.8306	0.0288
0.0036	0.0128			
0.9637	0.4190	0.4193	0.1404	0.0905
0.0154	0.0558			
0.0172	0.0611	0.0610	0.0290	0.8806
0.9811	0.9314			
Columns 22 through 28				
0.0122	0.0090	0.2810	0.0558	0.0505
0.0463	0.0353			
0.0633	0.0516	0.6537	0.7041	0.1450
0.4089	0.8253			
0.9246	0.9394	0.0653	0.2401	0.8045
0.5448	0.1409			
Columns 29 through 35				
0.0547	0.0367	0.0401	0.0623	0.8274
0.9508	0.0109			
0.7104	0.8148	0.7946	0.1690	0.1207
0.0396	0.9783			
0.2349	0.1485	0.1653	0.7687	0.0519
0.0096	0.0107			
Columns 36 through 40				
0.0231	0.7533	0.0325	0.0449	0.0421
0.1430	0.1682	0.1118	0.6374	0.7708
0.8339	0.0785	0.8556	0.3177	0.1871

Nilai-nilai ini seperti telah diungkapkan pada tinjauan pustaka, menggambarkan derajat keanggotaan suatu objek (dalam hal ini lulusan) terhadap ketiga cluster. Untuk lulusan ke-1,2,..., 40 didapatkan vector berikut (lihat kolom ke-1,2,..., 40):

$$\begin{array}{lll}
 u_{11} & 0.0193 & u_{12} & 0.0192 & u_{1,40} & 0.0421 \\
 u_{21} & = 0.9285, & u_{22} & = 0.9284 & \dots, & u_{2,40} & = 0.7708 \\
 u_{31} & 0.0522 & u_{32} & 0.0523 & u_{3,40} & 0.1871
 \end{array}$$

Bagaimana interpretasi dari besaran-besaran ini? Tinjau misalnya untuk lulusan ke-40, ia dapat

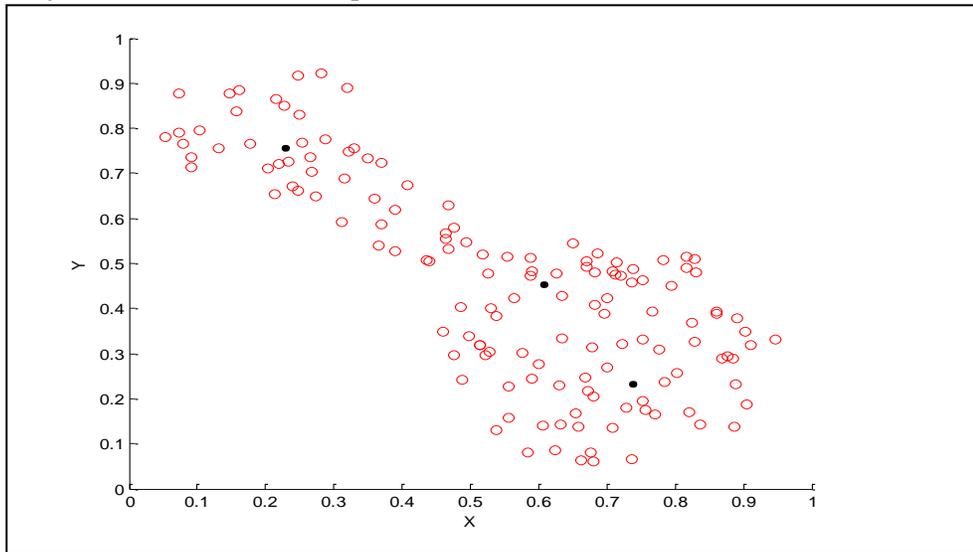
- menjadi anggota cluster-1 dengan derajat keanggotaan 0.0421
- menjadi anggota cluster-2 dengan derajat keanggotaan 0.7708, atau
- menjadi anggota cluster-3 dengan derajat keanggotaan 0.1871

Namun karena derajat keanggotaan terbesarnya adalah terhadap cluster-2, maka lulusan ke-40 ini akan dimasukkan kedalam cluster-2

Hasil selengkapnya pengelompokan ke-40 lulusan kedalam 3 cluster adalah:

- cluster-1 beranggotakan lulusan nomor 10, 11, 16, 17, 18, 33, 34 dan 37,
- cluster-2 beranggotakan lulusan nomor 1,2,3,4,5,6,7,8,9,14, 15, 24, 25,28,29, 30, 31, 35, dan 40,
- cluster-3 beranggotakan lulusan nomor 12, 13, 19, 20,21,22,23, 32, 36 dan 38.

a. Hasil selengkapnya pengelompokan 40 lulusan ke dalam 1-10 cluster. Disajikan dalam bentuk tabel, pada tabel 2



Grafik. Hasil Akhir Clustering untuk 40 lulusan yang dikelompokkan dalam 3 cluster

**Keterangan gambar:**

Dari grafik diatas dapat kita lihat bahwa titik hitam itu merupakan nilai dari “koordinat” titik pusat ketiga cluster. Yang terbagi menjadi tiga kelompok dan masing-masing kelompok mempunyai satu titik pusat (center).

Tabel 2. Hasil Akhir Clustering untuk 40 lulusan yang dikelompokkan kedalam 1,2, ..., 10 Cluster

Jml Clus ter	Anggota cluster ke-									
	1	2	3	4	5	6	7	8	9	10
10	1,2,40	19,26	10,11,18, 34	14,39	33,37	12,13,20, 22,23,36	3,4,5,6 15,35	25,27,28, 29,30,31	21,32, 38	7,8,9 16,17, 24
9	10,11,18, 34	3,6,15	1,2,25,27 , 28,29,30, 31	4,5,35	12,13,20, 21,22,23, 26,36	14,39,40	7,8,9,16, 17,24	33,37	19,32, 38	
8	11,33,37	32	10,16,17, 18,34	20,21,36, 38	12,13,22, 23	1,2,14, 25,27,28, 29,30,31, 39,40	3,4,5,6,7, 8,9,15, 24,35	19,26		
7	10,16,17, 18,34	21,32,38	1,2,14,25 , 27,28,30, 31,39,40	12,13,22, 23,36	3,4,5,6,7, 8,9,15, 24,35	19,26	11,33,37			
6	12,13,20, 21,22,23, 36	3,4,5,6,7, 8,9,15, 24,35	1,2,14, 25,27,28, 29,30,31, 39,40	10,16,17, 18,34	11,33,37	19,26,32, 38				
5	1,2,3,4,5, 6,14,15, 22,25,28, 29,30,31, 39,40	12,13,20, 22,23,27, 36	11,33,37	7,8,9,10, 16,17,18, 24,34	19,21,26, 32,38					
4	12,13,19, 20,21,22, 23,26,32, 36,38	3,6,7,8,9, 10,15,16, 17,18,24, 34	1,2,4,5, 14,25,27, 28,29,30, 31,35,39, 40	1,33,37						
3	10,11,16, 17,18,33, 34,37	1,2,3,4,5, 6,7,8,9,14, 15, 24,25,28, 29,30,31, 35,39,40	12,13,19, 20,21,22, 23,26,27, 32,36,36							
2	1,2,12,13, 14,19,20, 21,22,23, 25,26,27, 28,29,30, 31,32,36, 38,39,40									
1	1-40									

**4. Kesimpulan**

Dari hasil yang telah diperoleh maka dapat diambil kesimpulan bahwa:

1. Hasil perhitungan dari nilai-nilai  $u_{ik}$  Hasil selengkapnya pengelompokan ke-40 lulusan kedalam 3 cluster adalah:
  - ❖ cluster-1 beranggotakan lulusan nomor 10, 11, 16, 17, 18, 33, 34 dan 37,
  - ❖ cluster-2 beranggotakan lulusan nomor 1,2,3,4,5,6,7,8,9,14, 15, 24, 25,28,29, 30, 31, 35, dan 40,
  - ❖ cluster-3 beranggotakan lulusan nomor 12, 13, 19, 20,21,22,23, 32, 36 dan 38.
2. Hasil perhitungan dari nilai-nilai  $v_{ij}$ , sebagai berikut:

Center =	
2.7174	7.0516
2.7661	5.7233
3.0681	4.6110

Nilai-nilai ini merupakan nilai dari “koordinat” titik pusat ketiga cluster dan memberikan garis besar citra tiap cluster:

- ❖ Untuk cluster-1, “koordinat” dari titik pusat cluster ini adalah  $v_1 = (2.7173 \ 7.0516)$ , yang arti fisisnya, cluster 1 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7173 dan rata-rata lama lulus 7.0516 tahun;
- ❖ Untuk cluster-2, “koordinat” dari titik pusat cluster ini adalah  $v_2 = (2.7660 \ 5.7233)$ , yang arti fisisnya, cluster 3 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7660 dan rata-rata lama lulus 5.7233 tahun; sedangkan
- ❖ Untuk cluster-3, “koordinat” dari titik pusat cluster ini adalah  $v_3 = (3.0681 \ 4.6110)$ , yang arti fisisnya, cluster 3 akan beranggotakan lulusan yang dengan rata-rata IPK 2.7660 dan rata-rata lama lulus 5.7233 tahun

**Daftar Pustaka**

Kusumadewi, Sri. 2004. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta. Graha ilmu.

Kusumadewi, Sri. 2002. *Analisis Desain System Fuzzy Menggunakan Tool box Matlab*. Yogyakarta. Graha ilmu