

**PEMILIHAN *CLUSTER* OPTIMUM PADA *FUZZY C-MEANS*  
(STUDI KASUS: PENGELOMPOKAN KABUPATEN/KOTA  
DI PROVINSI JAWA TENGAH BERDASARKAN INDIKATOR  
INDEKS PEMBANGUNAN MANUSIA)**

**Sarita Budiyan Purnamasari<sup>1</sup>, Hasbi Yasin<sup>2</sup>, Triastuti Wuryandari<sup>3</sup>**

<sup>1</sup>Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

<sup>2,3</sup>Staf Pengajar Jurusan Statistika FSM Universitas Diponegoro

**ABSTRACT**

Cluster analysis is a process of separating the objects into groups, so that the objects that belong to the same group are similar to each other and different from the other objects in another group. One method of clustering is Fuzzy C-Means (FCM). FCM is used because each data in a cluster determined by a degree of membership that have value between 0 and 1. This research use two kinds of distance, Manhattan and Euclidean. To determine the proper distance in clustering district / city in Central Java based on indicators of Human Development Index (HDI), we have to calculate the ratio of the standard deviation, where the smaller value indicates a better clustering. While the optimum number of groups obtained from the minimum value of Xie Beni. Variables that used in this research are the indicators of HDI in 2012 for district / city in Central Java, consists of: Life Expectancy Value (years), Literacy Rate (percent), Average Length of School (years), and Purchasing Power Parity (thousands rupiah). The results from this research are the distance that gives a better quality is Euclidean and the optimum cluster given when the number of cluster is five with the smallest value of Xie Beni is 0,50778.

**Keywords:** cluster analysis, Fuzzy C-Means (FCM), HDI, optimum cluster.

## 1. PENDAHULUAN

Pembangunan nasional Indonesia menempatkan rakyat sebagai titik sentral pembangunan. Untuk dapat ikut berpartisipasi dalam proses pembangunan, tentunya dibutuhkan masyarakat yang unggul dari segi kuantitas dan kualitas. Sehingga, dirumuskan suatu konsep baru dalam mengukur pembangunan yang berorientasi pada manusia (BPS, 2012).

Indeks Pembangunan Manusia (IPM) merupakan tolok ukur keberhasilan pembangunan manusia yang mencakup tiga dimensi kehidupan manusia, yaitu peluang hidup (diukur dengan indikator Angka Harapan Hidup dalam tahun), pengetahuan (diukur dengan indikator Angka Melek Huruf dalam persen dan Rata-rata Lama Sekolah dalam tahun), dan hidup layak (diukur dengan indikator Paritas Daya Beli per bulan dalam ribuan rupiah). Pada tahun 2012, IPM Jawa Tengah menempati peringkat ke-15 secara nasional dengan IPM 73,36 (BPS, 2013).

Pengelompokan wilayah kabupaten/kota di Jawa Tengah perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran program pemerintah untuk meningkatkan angka pembangunan manusia berdasarkan indikator pembentuk IPM. Algoritma pengelompokan yang dapat digunakan salah satu diantaranya adalah *Fuzzy C-Means*.

*Fuzzy C-means Clustering* (FCM) adalah teknik pengelompokan dimana keberadaan tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan (Kusumadewi, 2002). Dalam penelitian ini akan dilakukan pengelompokan FCM dengan jarak Manhattan dan Euclidean. Kemudian akan dipilih jarak yang menghasilkan kualitas pengelompokan terbaik dengan memperhatikan nilai rasio simpangan baku. Penentuan jumlah *cluster*

optimum pada jarak terpilih dilakukan dengan menggunakan perhitungan nilai validitas Xie Beni.

## 2. TINJAUAN PUSTAKA

### 2.1. Konsep Pembangunan Manusia

Pembangunan manusia adalah pembangunan yang berpusat pada manusia, yang menempatkan manusia sebagai tujuan akhir dan bukan hanya sebagai alat pembangunan. Konsep pembangunan manusia yang direkomendasikan oleh UNDP mencakup empat indikator, yaitu kesetaraan, produktivitas, pemberdayaan, dan berkelanjutan (BPS, 2012).

### 2.2. Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) merupakan tolok ukur keberhasilan pembangunan manusia yang mencakup tiga dimensi kehidupan manusia, yaitu peluang hidup (diukur dengan indikator Angka Harapan Hidup dalam tahun), pengetahuan (diukur dengan indikator Angka Melek Huruf dalam persen dan Rata-rata Lama Sekolah dalam tahun), dan hidup layak (diukur dengan indikator Paritas Daya Beli perbulan dalam ribuan rupiah) (BPS, 2012).

### 2.3. Indikator IPM

Indikator indeks pembangunan manusia ada empat, yaitu (BPS, 2012):

1. Angka Harapan Hidup (AHH) merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidupnya.
2. Angka Melek Huruf (AMH) adalah persentase penduduk usia 15 tahun ke atas yang bisa membaca dan menulis.
3. Rata-rata Lama Sekolah (RLS) adalah lama sekolah (tahun) penduduk usia 15 tahun ke atas.
4. Kemampuan Daya Beli Penduduk atau *Purchasing Power Parity* (PPP) mencerminkan kemampuan masyarakat secara ekonomi dalam memenuhi kebutuhan konsumsinya.

### 2.4. Penyusunan IPM

Formula yang digunakan untuk menghitung indeks indikator IPM yaitu (BPS, 2012):

$$\text{Indeks } X_i = \frac{X_i - X_{i.min}}{X_{i.max} - X_{i.min}}$$

Keterangan:  $X_i$  = Indikator IPM ke-i  
 $X_{i.min}$  = Nilai minimum dari indikator ke-i  
 $X_{i.max}$  = Nilai maksimum dari indikator ke-i

Selanjutnya nilai IPM dapat dihitung dengan formula:

$$IPM = \frac{1}{3} [X_1 + X_2 + X_3]$$

Keterangan:

$X_1$  = Indeks harapan hidup

$X_2$  = Indeks pendidikan =  $\frac{2}{3}$  (indeks melek huruf) +  $\frac{1}{3}$  (indeks rata-rata lama sekolah)

$X_3$  = Indeks hidup layak

## 2.5. Analisis Cluster

Menurut Prasetyo (2012), *clustering* atau pengelompokan adalah proses pemisahan/pemecahan/segmentasi data ke dalam sejumlah kelompok (*cluster*) agar objek-objek yang tergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek dalam kelompok lainnya.

## 2.6. Konsep Jarak

Jarak merupakan konsep penting dalam pengembangan metode pengelompokan. Jarak antara titik A ke titik B didefinisikan sebagai  $d(A,B)$  (Santosa, 2007).

### 1. Jarak Manhattan

Formula jarak Manhattan atau *Cityblock* mencari jarak terpanjang atau jarak yang ditempuh oleh dua titik. Formula jarak Manhattan dinyatakan sebagai berikut:

$$d(A,B) = \sum_{i=1}^n (|x_i - y_i|)$$

### 2. Jarak Euclidean

Formula jarak Euclidean merupakan formula jarak yang paling sering digunakan dalam analisis pengelompokan, karena dalam perhitungan jarak Euclidean adalah mencari jarak terpendek dari dua titik dengan prinsip orthogonal (tegak lurus). Formula jarak Euclidean dinyatakan sebagai berikut:

$$d(A,B) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

## 2.7. Fuzzy C-Means (FCM)

*Fuzzy C-Means* (FCM) adalah suatu teknik kluster data yang keberadaan tiap-tiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaan. Metode FkM merupakan pengembangan dari metode non-hierarki *k-Means cluster*, karena pada awalnya ditentukan dahulu jumlah kelompok atau kluster yang akan dibentuk. Kemudian dilakukan iterasi sampai mendapatkan keanggotaan kelompok (Kusumadewi, 2002).

## 2.8. Algoritma FCM

Langkah-langkah dalam algoritma *Fuzzy C-Means* (Kusumadewi, 2002):

1. Menentukan data yang akan di-*cluster*  $X_{ij}$  yaitu data sampel ke- $i$  ( $i=1,2,\dots,n$ ) dan variabel ke- $j$  ( $j=1,2,\dots,p$ ).
2. Menentukan jumlah kluster ( $c$ ), pangkat pembobot ( $m>1$ ), maksimum iterasi, error terkecil yang diharapkan ( $\epsilon$ ), fungsi objektif awal ( $P^{(0)}=0$ ).
3. Membangkitkan bilangan random  $u_{ik}$ ,  $i=1,2,\dots,n$ ;  $k=1,2,3,\dots,c$  dengan syarat  $\sum_{k=1}^c u_{ik} = 1$
4. Menghitung pusat kluster ke- $k$  ( $v_{kj}$ ), dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,p$

$$v_{kj} = \frac{\sum_{i=1}^n (u_{ik})^m x_{ij}}{\sum_{i=1}^n (u_{ik})^m}$$

5. Menghitung perubahan nilai keanggotaan  $u_{ik}$

$$u_{ik}^{(t)} = \left[ \frac{\left[ \sum_{j=1}^p d_{ik}^2 \right]^{\frac{1}{m-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^p d_{jk}^2 \right]^{\frac{1}{m-1}}} \right]^{-1}$$

6. Menghitung fungsi objektif pada iterasi ke-t

$$P^{(t)} = \sum_{i=1}^n \sum_{k=1}^c (u_{ik})^m d_{ik}^2(x_i, v_k)$$

7. Memeriksa kondisi berhenti (konvergen):

Jika  $|P_t - P_{t-1}| < \varepsilon$  atau  $t >$  maksimum iterasi, maka iterasi berhenti

Jika tidak, maka  $t = t+1$ , mengulangi kembali ke langkah 4

## 2.9. Formula Jarak Terbaik dan Cluster Optimum

Pemilihan jarak yang menghasilkan kualitas pengelompokan terbaik dilakukan dengan memperhatikan nilai rasio rata-rata simpangan baku dalam kelompok dan simpangan baku antar kelompok. Rata-rata simpangan baku di dalam kelompok dinyatakan dengan:

$$S_W = \frac{1}{c} \sum_{k=1}^c S_k$$

Sedangkan simpangan baku antar kelompok dinyatakan sebagai:

$$S_B = \left[ \frac{1}{c-1} \sum_{k=1}^c (\bar{X}_k - \bar{X})^2 \right]^{\frac{1}{2}}$$

dengan  $c$  sebagai jumlah *cluster*,  $S_k$  merupakan simpangan baku di dalam *cluster* ke- $k$ ,  $\bar{X}_k$  sebagai rata-rata *cluster* ke- $k$  dan  $\bar{X}$  adalah rata-rata dari semua *cluster* (Bunkers dan Miller, 1996). Jarak yang dipilih adalah yang memberikan nilai rasio  $S_W/S_B$  terkecil.

Penentuan jumlah *cluster* optimum pada jarak terpilih dilakukan dengan menggunakan perhitungan nilai validitas Xie Beni. Perhitungan Xie Beni dilakukan dengan membandingkan nilai kepadatan dan keterpisahan.

$$\pi (\text{Kepadatan}) = \frac{\sum_{i=1}^n \sum_{k=1}^c (u_{ik})^m d_{ik}^2(x_i, v_k)}{n}$$

$$d_{\min} (\text{Keterpisahan}) = \min_{i \neq k} d_{ik}^2(v_i, v_k)$$

dimana  $d_{ik}^2(x_i, v_k)$  adalah jarak data terhadap pusat *cluster* dan  $d_{ik}^2(v_i, v_k)$  adalah jarak dari pusat *cluster*  $v_k$  ke pusat *cluster*  $v_i$ . Semakin kecil nilai Xie Beni maka semakin bagus hasil *cluster* yang telah dilakukan (Duo, et al, 2007).

## 3. METODOLOGI PENELITIAN

### 3.1. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah. Variabel yang digunakan dalam pengelompokan adalah indikator-indikator Indeks Pembangunan Manusia (IPM)

kabupaten/kota di Provinsi Jawa Tengah tahun 2012 yang meliputi Angka Harapan Hidup-AHH ( $X_1$ ), Angka Melek Huruf-AMH ( $X_2$ ), Rata-rata Lama Sekolah-RLS ( $X_3$ ), dan Pengeluaran per kapita yang disesuaikan-PPP ( $X_4$ ).

### 3.2. Langkah Analisis

Metode analisis yang digunakan untuk mencapai tujuan penelitian diuraikan sebagai berikut:

1. Menyiapkan data yang digunakan dalam peng-*cluster*-an
2. Membangun matriks  $X$  yang merupakan nilai standarisasi data penelitian
3. Menentukan jumlah *cluster*, pangkat pembobot ( $m=2$ ), error terkecil diharapkan ( $\epsilon=10^{-5}$ ) dan fungsi objektif awal ( $P^0=0$ )
4. Membangkitkan derajat keanggotaan awal
5. Menghitung pusat *cluster*, memperbaharui nilai keanggotaan, menghitung fungsi objektif iterasi ke- $t$ , menghitung rasio simpangan baku dan nilai Xie Beni untuk jarak Manhattan
6. Jika  $|P^{(t)} - P^{(t-1)}| < 10^{-5}$ , maka iterasi berhenti. Jika tidak ulangi langkah ke-5
7. Mengulangi langkah 4 sampai dengan langkah 6 menggunakan jarak Euclidean
8. Membandingkan nilai rasio simpangan baku untuk jarak Manhattan dan Euclidean, kemudian dipilih jarak terbaik, yaitu yang memberikan nilai rasio simpangan baku minimum
9. Menentukan jumlah *cluster* optimum pada jarak terpilih dengan memperhatikan nilai Xie Beni yang minimum

## 4. HASIL DAN PEMBAHASAN

### 4.1. Analisis Hasil Pengelompokan dengan FCM

*Clustering* menggunakan metode FCM akan dilakukan dengan menggunakan dua formula jarak, yaitu Manhattan dan Euclidean. Untuk masing-masing jarak kemudian dilakukan *clustering* mulai dari jumlah *cluster* 2 sampai dengan jumlah *cluster* 5. Adapun hasil *clustering* adalah sebagai berikut:

**Tabel 1.** Perbandingan Hasil Fungsi Objektif FCM dengan Jarak Manhattan dan Euclidean

Manhattan	Euclidean
Jumlah cluster 2	
stop iterasi = 41	stop iterasi = 23
$P^{(41)} = 188,72744$	$P^{(23)} = 62,62420$
Jumlah cluster 3	
stop iterasi = 35	stop iterasi = 27
$P^{(35)} = 114,09329$	$P^{(27)} = 37,99416$
Jumlah cluster 4	
stop iterasi = 76	stop iterasi = 35
$P^{(76)} = 76,60879$	$P^{(35)} = 25,66074$
Jumlah cluster 5	
stop iterasi = 42	stop iterasi = 67
$P^{(42)} = 58,15257$	$P^{(67)} = 19,83431$

#### 4.2. Penentuan Formula Jarak Terbaik

Pemilihan jarak yang menghasilkan kualitas pengelompokan terbaik dilakukan dengan memperhatikan nilai rasio rata-rata simpangan baku dalam kelompok dan simpangan baku antar kelompok yang minimum.

**Tabel 2.** Hasil Perhitungan Rasio Simpangan Baku

Jenis Jarak	$S_w/S_B$				
	AHH	AMH	RLS	PPP	Average
<b>jumlah cluster=2</b>					
Manhattan	2,60500	1,08170	0,46550	0,75260	1,22620
Euclidean	1,24699	1,29868	0,70366	0,70444	0,98844
<b>Jumlah Cluster=3</b>					
Manhattan	0,88778	0,61508	0,50315	0,70125	0,67682
Euclidean	1,16883	0,87317	0,45174	0,62714	0,72556
<b>Jumlah Cluster=4</b>					
Manhattan	0,88124	0,51008	0,57814	0,60827	0,64443
Euclidean	0,85513	0,50725	0,55202	0,59652	0,62773
<b>Jumlah Cluster=5</b>					
Manhattan	1,08940	0,52547	0,51923	0,36012	0,62355
Euclidean	0,60396	0,45736	0,34225	0,63949	0,51077

Dari Tabel 2, terlihat bahwa rata-rata formula jarak Euclidean memberikan nilai Rasio Simpangan Baku yang lebih kecil dibandingkan formula jarak Manhattan untuk jumlah *cluster* yang sama. Hal ini menjelaskan bahwa pengelompokan dengan formula jarak Euclidean memberikan hasil pengelompokan yang lebih baik.

#### 4.3. Penentuan Jumlah Cluster Optimum

Jumlah *cluster* optimum diberikan ketika nilai Xie Beni minimum.

**Tabel 3.** Nilai Xie Beni

Jumlah Cluster	Xie Beni
2	0,70242
3	0,59439
4	0,57915
5	0,50778

Berdasarkan Tabel 3 dapat dilihat bahwa jumlah *cluster* optimum diberikan ketika jumlah *cluster* = 5 dengan nilai Xie Beni terkecil dibanding jumlah kelompok yang lain.

#### 4.4. Interpretasi Hasil Pengelompokan

Menggunakan formula jarak Euclidean dengan jumlah *cluster* 5, dilakukan kembali pengelompokan kabupaten/kota di Jawa Tengah berdasarkan indikator Indeks Pembangunan Manusia menggunakan FCM. Setelah *cluster* terbentuk, terhadap seluruh objek penelitian (35 kabupaten/kota) diambil rata-rata dari masing-masing indikator IPM Provinsi Jateng, yaitu AHH, AMH, RLS dan PPP ( $\bar{X}$ ). Selanjutnya untuk masing-masing *cluster* juga diambil rata-rata untuk variabel AHH, AMH, RLS dan PPP ( $\bar{X}_c$ ). Setiap variabel di dalam *cluster*, jika  $\bar{X}_c > \bar{X}$  diberikan tanda (+), sedangkan jika  $\bar{X}_c < \bar{X}$  maka diberikan tanda (-).

**Tabel 4.** Hasil Pengelompokan Kabupaten/Kota di Jawa Tengah menggunakan FCM

C	Nama Kab/Kota	Karakteristik			
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	Kab.Cilacap, Kab.Purbalingga, Kab.Purworejo, Kab.Wonosobo, Kab.Boyolali, Kab.Jepara, Kab.Demak, Kab.Semarang, Kab.Temanggung, Kab.Batang	+	+	-	-
2	Kab.Banyumas, Kab.Magelang, Kab.Sukoharjo, Kab.Rembang, Kab.Kudus, Kota Pekalongan, Kota Tegal	-	+	+	+
3	Kab.Klaten, Kab.Wonogiri, Kab.Karanganyar, Kab.Sragen, Kab.Blora, Kab.Pati	+	-	-	+
4	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang	+	+	+	+
5	Kab.Banjarnegara, Kab.Kebumen, Kab.Grobogan, Kab.Kendal, Kab.Pekalongan, Kab.Pemalang, Kab.Tegal, Kab.Brebes	-	-	-	-

Berdasarkan Tabel 4 terlihat bahwa *cluster* 1 beranggotakan 10 kabupaten. *Cluster* 1 merupakan *cluster* dengan nilai rata-rata *cluster* lebih besar dibanding nilai rata-rata Provinsi Jawa Tengah pada variabel AHH dan AMH. Sedang untuk variabel RLS dan PPP, *cluster* ini memiliki nilai rata-rata *cluster* lebih kecil dibanding nilai rata-rata Provinsi Jawa Tengah.

*Cluster* 2 beranggotakan 5 kabupaten dan 2 kota. *Cluster* 2 merupakan *cluster* dengan nilai rata-rata *cluster* lebih besar dibanding nilai rata-rata Provinsi Jawa Tengah pada variabel AMH, RLS dan PPP. Sedang untuk variabel AHH, *cluster* ini memiliki nilai rata-rata *cluster* lebih kecil dibanding nilai rata-rata Provinsi Jawa Tengah.

*Cluster* 3 beranggotakan 6 kabupaten. *Cluster* 3 merupakan *cluster* dengan nilai rata-rata *cluster* lebih besar dibanding nilai rata-rata Provinsi Jawa Tengah pada variabel AHH dan PPP. Sedang untuk variabel AMH dan RLS, *cluster* ini memiliki nilai rata-rata *cluster* lebih kecil dibanding nilai rata-rata Provinsi Jawa Tengah.

*Cluster* 4 beranggotakan 4 kota. *Cluster* ini merupakan *cluster* yang terbaik karena memiliki nilai rata-rata *cluster* lebih besar dibanding nilai rata-rata Provinsi Jawa Tengah di semua indikator IPM.

*Cluster* 5 beranggotakan 8 kabupaten. *Cluster* 5 merupakan *cluster* dengan rata-rata *cluster* lebih kecil dibanding rata-rata Provinsi Jawa Tengah di semua indikator IPM.

## 5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan pengelompokan kabupaten/kota di Provinsi Jateng berdasarkan indikator IPM menggunakan metode *Fuzzy C-Means* adalah sebagai berikut:

1. Berdasarkan kualitas ketepatan pengelompokan menggunakan rasio simpangan baku dalam *cluster* dan antar *cluster*, pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator IPM menggunakan FCM lebih tepat menggunakan formula jarak Euclidean.
2. Menggunakan nilai Xie dan Beni didapatkan jumlah *cluster* yang optimum untuk FCM dengan jarak terpilih adalah 5 *cluster*.
3. Berdasarkan hasil pengelompokan kabupaten/kota di Provinsi Jawa Tengah menurut indikator IPM diketahui bahwa *cluster* 4 merupakan *cluster* yang terbaik karena memiliki nilai rata-rata *cluster* lebih besar dibanding nilai rata-rata Provinsi Jawa Tengah di semua indikator IPM. Sedangkan *cluster* 5 merupakan *cluster* terburuk karena memiliki nilai rata-rata *cluster* lebih kecil dibanding nilai rata-rata Provinsi Jawa Tengah di semua indikator IPM.

## 6. DAFTAR PUSTAKA

BPS. 2012. *Indeks Pembangunan Manusia 2010-2011*.

BPS. 2013. *Indikator Utama Sosial, Politik, dan Keamanan Jawa Tengah 2012*.

Bunkers, M.J dan Miller, J.R. 1996. "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique", *Journal of Climate*. Vol.9, pp. 130-146.

Duo, C., Xue, L. dan Du-Wu, C. 2007. "An Adaptive Cluster Validity Index for the Fuzzy C-Means", *International Journal of Computer Science and Network Security*, Vol.7 No.2, pp.146-156.

Kusumadewi, S. 2002. *Analisis & Desain Sistem Fuzzy*. Yogyakarta: Graha Ilmu.

Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi Offset.

Santosa, B. 2007. *Data Mining: Teori dan Aplikasi*. Yogyakarta: Graha Ilmu.