Analisis Kinerja Algoritma *Fuzzy C-Means* dan K-*Means* pada Data Kemiskinan

Aniq Noviciatie Ulfah*[1], Shofwatul 'Uyun[2]
[1,2]Program Studi Teknik Informatika Universitas Islam Negeri Sunan Kalijaga
Jl. Marsda Adisucipto No.1 Yogyakarta 55281
Telp (0274) 519739, fax (0274) 540971
e-mail: *[1]noviciatie.binti.akhmad@gmail.com, [2]shofwatul.uyun@uin-suka.ac.id

Abstrak

Salah satu upaya untuk mewujudkan program pemerintah Kabupaten Gunungkidul dalam rangka untuk pengentasan kemiskinan adalah dengan melakukan pendataan data kemiskinan warganya. Pemerintah telah merumuskan pendataan dengan melakukan pembobotan terhadap 15 indikator kedalam 3 kelompok. Banyaknya data dan indikator yang harus digunakan tentunya akan menimbulkan kesulitan dalam pelaksanaannya, tidak efektif dan kurang obyektif. Oleh karena itu diperlukan otomatisasi dalam proses clustering data kemiskinan. Penelitian ini bertujuan untuk menganalisis kinerja antara algoritma FCM dan K-means yang diimplementasikan pada data kemiskinan di Desa Girijati Purwosari menjadi 3 cluster. Beberapa tahapan yang harus dilakukan sebelum dilakukan clustering, terebih dahulu dilakukan prapengolahan yaitu data cleaning dan data transformation untuk selanjutnya dilakukan clustering menggunakan kedua algoritma tersebut. Hasil perhitungan digunakan untuk membandingkan antara algoritma FCM dengan K-Means. Kesesuaian data antara algoritma FCM dengan perhitungan indikator kemiskinan di Desa Girijati sebesar 50% dan untuk algoritma K-Means sebesar 83,33%. Algoritma K-Means lebih tepat digunakan pada pengelompokan data kemiskinan berdasarkan ketiga kriteria kemiskinan dibandingkan algoritma FCM.

Kata kunci*—Clustering, Data Kemiskinan, FuzzyC-Means, K-Means.*

Abstract

One of the local government of Gunungkidul efforts to realize program in order to alleviate poverty is to perform the data collection poverty of its citizens. The local government of Gunungkidul has formulated a collection by weighting against 15 indicators into three groups. The amount of data and indicators to be used will certainly lead to difficulties in implementation, ineffective and less objective. Therefore we need automation in the process of clustering data on poverty. This study aims to analyze the performance of the FCM algorithm and K-means are implemented in the data on poverty in Girijati Purwosari village into 3 clusters. Some of the steps that must be done prior to clustering, first performed pretreatment includes data cleaning and data transformation for clustering is then performed using the second algorithm. The suitability of data between FCM algorithm and the calculation of poverty indicators in the Girijati village is 50 % and for the K - Means algorithm is 83.33 % . Therefore, K- Means algorithm is more appropriately used in data classification of poverty based on the three criteria of poverty, beside FCM algorithm.

Keyword*—Clustering, The poverty, FuzzyC-Means, K-Means.*

1. PENDAHULUAN

Kemiskinan di Yogyakarta menurut Badan Pusat Statistik (BPS) pada tahun 2012 tertinggi se-Jawa melebihi DKI Jakarta, Banten dan Jawa Tengah yang mencapai 15,88 persen sedangkan tingkat kemiskinan masyarakat Jawa Tengah hanya mencapai 14,98 persen, Jawa Timur 13,08 persen, Jawa Barat 9,89 persen, Banten 5,71 persen dan DKI jakarta hanya 3,7 persen [1]. Sebagai upaya pengentasan kemiskinan pemerintahan Kabupaten Gunungkidul membuat suatu penanggulangan kemiskinan dengan membuat rumusan yaitu membuat indikator-indikator kemiskinan serta mengklasifikasikannya kedalam tiga kategori yaitu keluarga tidak miskin, keluarga miskin dan keluarga sangat miskin berdasarkan 15 indikator kemiskinan [2]. Pendataan kemiskinan dilakukan secara berjenjang mulai dari tinggat RT, Pedukuhan, Desa, Kecamatan, sampai pada tingkat Kabupaten. Desa Girijati yang teletak di Kabupaten Gunungkidul merupakan bagian dari pendataan kemiskinan yang dilakukan oleh BAPPEDA Kabupaten Gunungkidul. Untuk mempermudah pengklusteran diperlukan suatu sistem yang dapat mengkluster data kemiskinan tersebut. Beberapa penelitian telah melakukan kajian terhadap beberapa algoritma clustering, antara lain K-Means oleh [3,4,5] yang menyatakan bahwa algoritma dari K-Means memiliki waktu proses yang lebih cepat dari pada *hierarchical clustering* (jika k kecil) dengan jumlah variabel yang besar dan menghasilkan *cluster* yang lebih rapat. Fuzzy C-means juga telah dianalisis kinerjanya oleh [6,7,8] bahwa algoritma FCM memiliki waktu proses lebih cepat dibandingkan *Agglomerative Hierarchical Clustering* dan hasil *clustering* lebih mudah untuk diinterpretasikan.

Oleh karena itu pada penelitin ini menggunakan algoritma *Fuzzy C-Means* dan *K-Means*untuk mengkluster data kemiskinan menjadi tiga kategori dengan 15 indikator yang diperoleh dari 10 aspek kemiskinan menurut BAPPEDA yaitu aspek penghasilan, aspek pangan, aspek sandang, aspek papan, aspek air bersih, aspek kesehatan, aspek pendidikan, aspek kekayaan, aspek penerangan dan aspek jumlah jiwa. Indikator-indikator tersebut digunakan sebagai variabel dari *clustering* data kemiskinan tersebut. Kedua algoritma tersebut digunakan untuk membagi data kedalam beberapa kelompok (grup atau kluster atau segmen) yang tiap kluster dapat ditempati beberapa anggota bersama-sama. Setiap objek dilewatkan pada grup yang paling mirip dengannya. Kedua algoritma tersebut merupakan algoritma *unsupervised learning* dimana data yang diolah tidak memerlukan pembelajaran terlebih dahulu.

2. DATA DAN METODE PENELITIAN

Data yang digunakan pada penelitian ini diambil dari data keluarga sejahtera Desa Girijati, Purwosari, Gunungkidul dengan 15 indikator kemiskinan yang terbagi menjadi 10 aspek [9], antara lain : penghasilan, pangan, sandang, papan, air bersih, kesehatan, pendidikan, kekayaan, penerangan dan jumlah jiwa. Untuk memperkuat analisis data juga dilakukan wawancara dengan Bapak Joko Hardiyanto selaku KASUBID Statistik, Bapak Sungkem selaku KASI KESOS Kecamatan Purwosari dan Bapak Budi Suryono selaku kepala Desa Girijati terkait sasaran pendataan kemiskinan, kriteria, indikator kemiskinan dan perhitungan data-data kemiskinan. Penelitian ini bertujuan untuk melakukan clustering data kemiskinan menggunakan algoritma FCM dan k-means clustering serta menganalisis kinerja untuk kedua algoritma tersebut. Untuk mengukur kinerja dari kedua algoritma tersebut dilakukan dengan membandingkannya dengan hasil klasifikasi yang dilakukan secara manual berdasarkan 15 indikator kemiskinan menurut BAPPEDA tahun 2008. Penelitian ini terdiri dari beberapa tahapan, antara lain : prapengolahan, proses *clustering* dan perbandingan hasil clustering.

2.1 Prapengolahan

Data yang diperoleh masih dalam bentuk *Microsoftaccess* serta masih terdapat data-data yang tidak terisi dengan lengkap, oleh karena itu perlu dilakukan prapengolahan terlebih dahulu

sebelum data dikluster. Data diubah menjadi data yang ber-*extensi* .sql yang berisi beberapa tabel yang berhubungan dengan program Pendataan Ketenagakerjaan Gunungkidul. Selanjutnya dibuat satu tabel lagi yang berisi data-data yang berhubungan dengan proses perhitungan *clustering* data kemiskinan yaitu id, Nomerator, nama KK dan 15 indikator kemiskinan. Untuk kepentingan clustering data harus dalam bentuk numerik atau angka, sehingga semua data harus dikonversi terlebih dahulu, selain itu kelimabelas indikator yang digunakan pada penelitian ini perlu dilakukan *data cleaning* dan *data transformation.* Salah satu contoh data sebelum dilakukan perubahan kedalam bentuk numeris ditunjukkan pada Tabel 1.

Tabel 1. Contoh salah satu data indikator kemiskinan yang dilakukan perubahan bentuk dari frase ke numeris

No	Aspek	Kriteria	Numerisasi
1	Pangan	Berapa kali makan dalam sehari seluruh keluarga a) Kurang dari 2 kali b) Dua kali atau lebih	 1 2
2	Pangan	Konsumsi daging/susu/telor/ikan/ayam dalam seminggu seluruh anggota keluarga a) Tidak pernah b) Satu kali c) Dua kali atau lebih	 1 2 3

Setelah dilakukan prapengolahan, format data kemiskinan yang sudah siap untuk diolah ditunjukkan pada Tabel 2.

Tabel 2. Contoh data hasil proses prapengolahan dan siap untuk dilakukan clustering

No	Nomerator	Indikator														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	764042	900000	2	2	3	1	3	2	1	2	3	4	3	2	3	3
2	764043	450000	2	2	3	1	3	2	1	2	3	4	3	2	3	2

2.2 Proses Clustering

Penelitian ini menggunakan dua algoritma clustering dengan dua pendekatan yang berbeda, yaitu pendekatan *hard* dan *fuzzy* clustering yang akan dijelaskan lebih detail pada subbab 2.2.1. dan 2.2.2.

2.2.1 Fuzzy C-Means (FCM)Clustering

Algoritma FCM telah dikenalkan oleh [10] dan merupakan metode clustering dengan pendekatan fuzzy, artinya setiap data yang dicluster memungkinkan menjadi anggota lebih dari satu cluster. Konsep dasar FCM adalah menentukan pusat *cluster*, pada kondisi awal pusat *cluster* ini masih belum akurat dan setiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster.* Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap data secara berulang untuk mendapatkan posisi pusat cluster yang tepat berdasarkan pada minimisasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster*.

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} \times x_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \quad (1)$$

dengan k=1,2,….,c; dan j=1,2,…,m

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}([\sum_{j=1}^{m}(X_{ij} - V_{kj})^2](\mu_{ik})^{w}) \quad (2)$$

$$\mu_{ik} = \frac{[\sum_{i=1}^{n}(X_{ij}-V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}[\sum_{j=0}^{m}(X_{ij}-V_{ij})^2]^{\frac{-1}{w-1}}} \quad (3)$$

dengan i = 1,2,3,…,n; dan k = 1,2,3,…,c.

Algoritma FCM :

[1] Input data yang akan dicluster X, berupa matriks berukuran n x m (n = jumlah sampel data, m = atribut setiap data). Xij = data sampel ke-i (i = 1,2,3…,n), atribut ke-j (j = 1,2,3,…,m).
[2] Tentukan : jumlah cluster (c),pangkat (w), maksimum iterasi (MaxIter), error terkecil yang diharapkan (ξ), fungsi objektif awal (P0 = 0) dan iterasi awal (t=1).
[3] Bangkitkan bilangan random μik , dimana i=1,2,3,…,n; k=1,2,….,c; sebagai elemen-elemen matriks partisi awal U.
[4] Hitung pusat cluster ke-k dengan persamaan (1)
[5] Hitung fungsi obyektif pada iterasi ke–t, Pt atau dengan cara menghitung hasil maksimal dari selisih matriks partisi akhir-dikurangi matriks partisi awal menggunakan persamaan (2)
[6] Hitung perubahan matriks partisi dengan persamaan (3)
[7] Cek kondisi berhenti :
    a. Jika (|Pt-Pt-1| <ξ) atau (t > MaxIter) maka berhenti.
    b. Jika tidak : t=t+1, ulangi langkah ke -4.

Contoh hasil clustering menggunakan algoritma FCM dengan 5 kali iterasi dan 3 *cluster* ditunjukkan pada Tabel 3 Perubahan kluster pada FCM didasarkan pada derajat keanggotaan yang terbesar. Jika derajat keanggotaan terbesar terletak pada kolom pertama, hal itu menunjukkan bahwa data tersebut termasuk ke dalam *cluster* pertama, begitu seterusnya. Sebagai contoh pada data pertama di Tabel 3 yang memiliki derajat keanggotaan untuk kluster 1, 2 dan 3 adalah 0.26661900, 0.19149878 dan 0.07512022, sehingga data tersebut masuk dalam kluster yang pertama. Dalam FCM satu data dapat masuk dalam lebih dari satu *cluster*, sehingga kemungkinan satu KK bisa di kategorikan dalam dua atau tiga kriteria kemiskinan dengan nilai keanggotaan yang dapat digunakan sebagai pembeda antara anggota cluster yang satu dengan yang lain.

Tabel 3. Contoh hasil perhitungan menggunakan algoritma FCM

Derajat keanggotaan			Max	Cluster		
C1	C2	C3		1	2	3
0.26661900	0.19149878	0.07512022	0.26661900	*		
0.86809009	0.07904618	0.78904391	0.86809009	*		
0.18817855	0.21732199	0.02914245	0.21732199		*	
0.55897680	0.05708580	0.61606260	0.61606260			*
0.25794736	0.24453871	0.50248607	0.50248607			*
0.75466121	0.10686331	0.86152452	0.86152452			*

2.2.2 K-Means Clustering

K-means clustering termasuk dalam *partitioning clustering* yang disebut juga *exclusive clustering* yang memisahkan data ke k daerah bagian yang terpisah dan setiap data harus termasuk ke dalam cluster tertentu dan memungkinkan setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-means merupakan algoritma yang sangat terkenal karena kemudahan dan kemampuannya untuk melakukan pengelompokan data besar dengan sangat cepat [11].

Algoritma k-means clustering :

[1] Inisialisasi cluster centroid $\mu_1, \mu_2, \dots, \mu_k \in R^n$ secara random

[2] Repeat until convergence :

{

fori , $c^{(i)} := \arg\min_j \left|\left|x^{(i)} - \mu_j\right|\right|^2$

for j , $\mu_j := \frac{\sum_{i=1}^{m} 1\{c^{(i)}=j\}x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)}=j\}}$

}

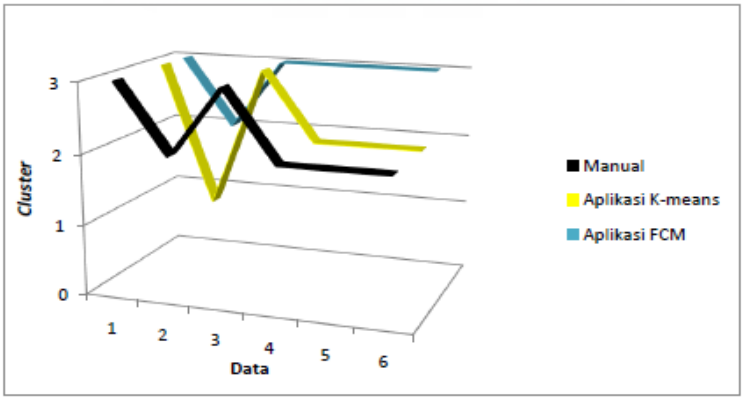
Contoh hasil perhitungan menggunakan algoritma K-Means ditunjukkan pada Tabel 4. Contoh hasil akhir dengan 5 kali iterasi dan 3 *cluster* ditunjukkan pada Tabel 4.Perubahan kluster pada *clustering* menggunakan *K-Means* didasarkan pada derajat keanggotaan yang terkecil. Jika derajat keanggotaan terkecil terletak pada kolom pertama, hal ini menunjukkan bahwa data tersebut masuk ke dalam *cluster* pertama, begitu seterusnya. Sebagai contoh data pertama pada Tabel 4menunjukkan bahwa derajat keanggotaan untuk kluster 1, 2 dan 3 adalah 666666.67, 450000 dan 50000 maka data satu tersebut masuk dalam cluster yang ke-3.

2.3 Perbandingan Hasil Clustering

Hasil clustering menggunakan agoritma FCM dan K-means selanjutnya dibandingkan dengan hasil clustering yang dilakukan secara manual berdasarkan beberapa indikator yang telah digunakan oleh pemerintah sesuai BAPPAEDA tahun 2008 yang ditunjukkan pada Gambar 1. Gambar 1 menunjukkan hasil perbandingan perhitungan menggunakan algoritma FCM ditunjukkan dengan warna biru ,*K-Means* dengan warna kuning dan perhitungan Manual dengan warna hitam. Perhitungan manual dengan perhitungan menggunakan algoritma FCM memiliki kesesuaian sebesar 50% dan hasil perhitungan manual dengan perhitungan algoritma *K-Means* sebesar 83,33%.

Tabel 4.Contoh hasil perhitungan menggunakan algoritma K-Means

Derajat keanggotaan			Min	Cluster		
C1	C2	C3		1	2	3
666666.67	450000	50000	50000			*
216666.67	0	500000	0		*	
766666.67	550000	50000	50000			*
16666.667	200000	700000	16666.667	*		
83333.333	300000	800000	83333.333	*		
66666.667	150000	650000	66666.667	*		



Gambar 1. Perbandingan Hasil Perhitungan Manual dan Sistem untuk Algoritma FCM dan *K-Means*

3. HASIL DAN PEMBAHASAN

Penelitian ini mengunakan 9 skenario untuk pengujian algoritma FCM dan 7 skenario untuk pengujian algoritma *K-Means*. Pengujian kedua algoritma tersebut menggunakan data kemiskinan desa Girijati, Purwosari, Gunungkidul pada tahun 2009, 2010, 2012 dan 2013. Jumlah data dari masing-masing tahun adalah 347, 13,16 dan 23.

3.1. Fuzzy C-Means

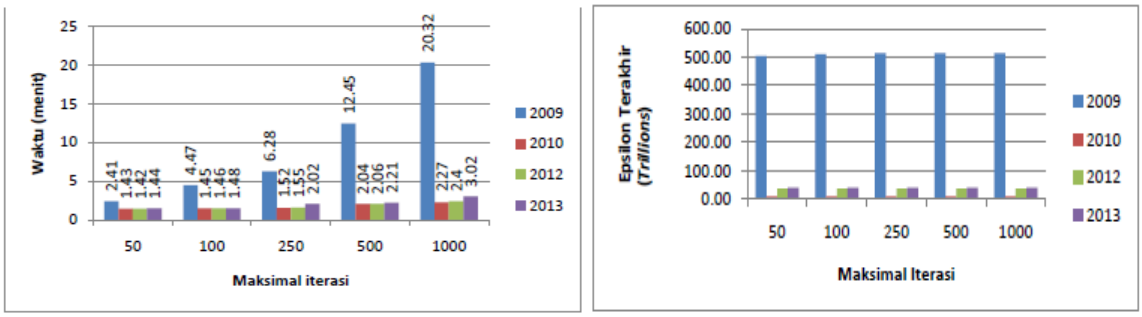
Pengujian menggunakan algoritma FCM menggunakan 3 parameter yaitu maksimal iterasi, pangkat dan *epsilon*. Secara umum skenario-skenario pengujian ditunjukan pada Tabel 5.

Tabel 5.Ringkasan Skenario untuk Algoritma FCM

Skenario	Maksimal Iterasi	Pangkat	*Epsilon*
1	50	2	10^{-5}
2	100	2	10^{-5}
3	250	2	10^{-5}
4	500	2	10^{-5}
5	1000	2	10^{-5}
6	50	5	10^{-5}
7	100	5	10^{-5}
8	50	100	10^{-5}
9	100	100	10^{-5}

3.1.1. Skenario 1 – 5

Percobaan dengan sKenario 1 sampai dengan 5 dilakukan untuk mengetahui pengaruh banyaknya iterasi dan data dengan waktu yang dibutuhkan sistem untuk mengolah data dan nilai *epsilon*.Begitu juga semakin banyak data dan maksimal iterasi maka semakin tinggi nilai *epsilon*-nya. Hasil percobaan ditunjukkan pada Gambar 2 (a) dan (b), dapat dilihat bahwa semakin banyak data dan jumlah iterasi maka akan mempengaruhi waktu yang diperlukan untuk proses perhitungan algoritma FCM. Dapat dilihat juga bahwa untuk kesemua pengujian maksimal iterasinya sama dengan inisialisasi iterasi awal, hal ini menunjukan bahwa proses berhenti dikarenakan sudah mencapai batas maksimal iterasi yang ditentukan diawal bukan karena nilai *epsilon*nya. Percobaan hanya dilakukan mencapai maksimal iterasi 1000 saja, ini dikarenakan hasil yang diperoleh semakin jauh dari yang diharapkan. Hal ini dapat dilihat dari data yang diperoleh bahwa pusat kluster yang nantinya akan dijadikan acuan sebagai kriteria tidak sesuai dari masing - masing indikator dan efisiensi waktu. Hasil untuk pusat kluster masih jauh dari yang diharapkan, Hal ini terjadi karena data yang ada pada kriteria 2 – 14 kurang bervariasi.

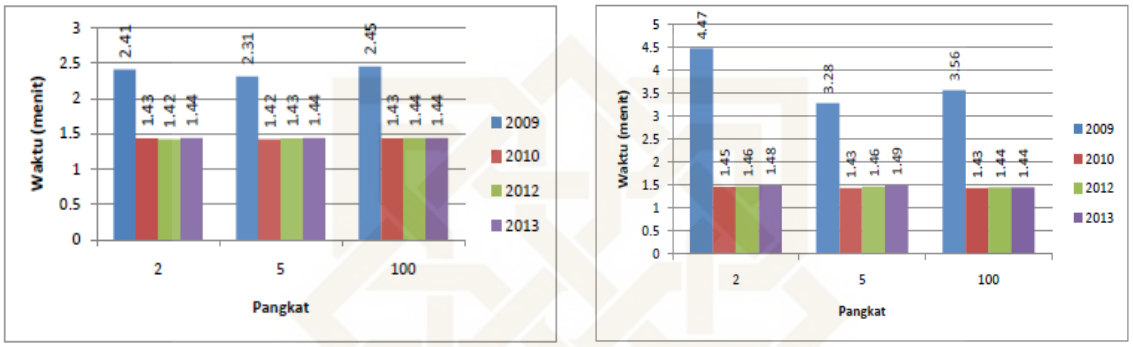


(a) (b)

Gambar 2 (a) Waktu yang diperlukan sistem untuk parameter maksimal iterasi, (b) Hasil *epsilon* terakhir untuk parameter maksimal iterasi

3.1.2. Skenario 1, 6 dan 8

Percobaan dengan sKenario 1, 6 dan 8 dilakukan untuk mengetahui pengaruh parameter pangkat terhadap waktu yang dibutuhkan sistem untuk mengolah data dan nilai *epsilon*. Hasil clustering yang ditunjukkan pada Gambar 3 (a) dan (b)menunjukkan bahwa banyaknya pangkat kurang mempengaruhi waktu yang diperlukan untuk proses perhitungan yang dilakukan. Hal ini dilihat pada nilai pangkat 5 mengalami penurunan waktu yang diperlukan, akan tetapi pada pangkat 100 waktu yang diperlukan bertambah kembali. Hal ini kemungkinan terjadi karena pengaruh nilai *random* untuk partisi awal. Nilai *random* partisi awal juga sangat berpanguruh terhadap banyaknya iterasi, hal itu disebabkan nilai dari fungsi objektif dapat berubah sesuai nilai partisi awal.



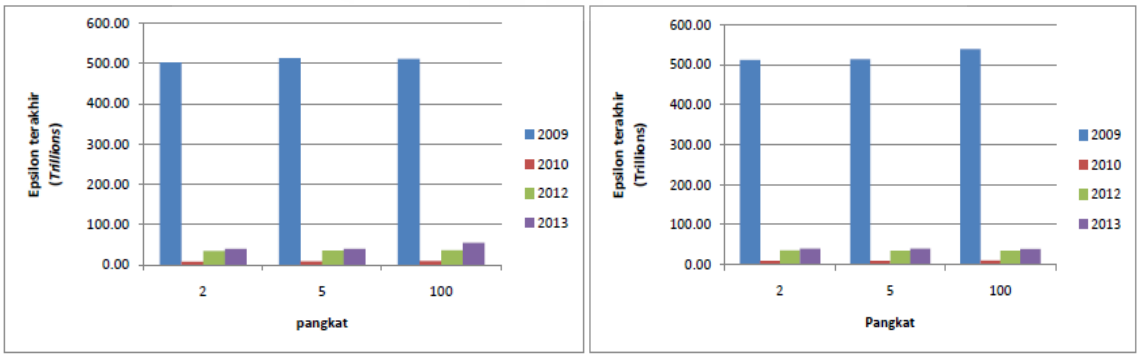
(a) (b)

Gambar 3. Hasil *clustering* terkait waktu yang dibutuhkan oleh system dengan parameter pangkat dengan maksimal iterasi 50 (a) dan 100 (b)

Berdasarkan dari Gambar 3 (a) dan (b)menunjukkan bahwa nilai pangkat kurang mempengaruhi waktu yang diperlukan untuk menyelesaikan proses perhitungan pada algoritma FCM.

3.1.3. Skenario 1, 6 dan 8

Percobaan dengan skenario 1, 2, 6, 7, 8 dan 9 dilakukan untuk mengetahui pengaruh banyak data, iterasi dan pangkat dengan waktu yang dibutuhkan sistem untuk mengolah data dan nilai *epsilon*. Dari Gambar 4 (a) dapat dilihat bahwa parameter pangkat mempengaruhi nilai *epsilon* akhir. Semakin tinggi pangkat maka nilai *epsilon* semakin besar dan semakin jauh dari nilai *epsilon* awal yang ditetapkan. Begitu juga untuk hasil clustering dengan maksimal iterasi 100 menunjukkan bahwa parameter pangkat berpengaruh terhadap nilai akhir *epsilon* yang ditunjukkan pada Gambar 4 (b).



(a) (b)

Gambar 4. Hasil *clustering* terkait nilai *Epsilon* Terakhir untuk Parameter Pangkat dengan maksimal iterasi 50 (a) dan 100 (b)

Semakin banyak data dan jumlah iterasi maka akan mempengaruhi waktu yang diperlukan untuk proses perhitungan algoritma FCM. Semakin banyak data dan maksimal iterasi maka semakin tinggi nilai *epsilon*-nya. Dari sekian percobaan yang dilakukan, data - data pada proses pengujian belum mencapai pada hasil yang diinginkan. Hal ini dikarenakan beberapa faktor yang mempengaruhinya antara lain proses perhitungan dipaksa berhenti berdasarkan maksimal iterasi yang telah ditentukan, bukan karena nilai *epsilon* yang ditentukan sehingga kemungkinan data dapat berpindah kluster masih bisa terjadi. Dari penjelasan – penjelasa di atas maka dapat disimpulkan bahwa algortima *Fuzzy C-Means* kurang sesuai untuk *clustering* data kemiskinan di Desa Girijati, mengingat hasil yang diperoleh masih jauh dari harapan. Algortima *Fuzzy C-Means* tidak sesuai untuk data Kemiskinan di Kabupaten Gunungkidul disebabkan data yang kurang bervariasi. Pengaruh inisialisasi pangkat dan maksimal iterasi berpengaruh pada waktu yang diperlukan untuk proses perhitungan menggunakan algoritma *Fuzzy C-Means*.

3.2. K-Means

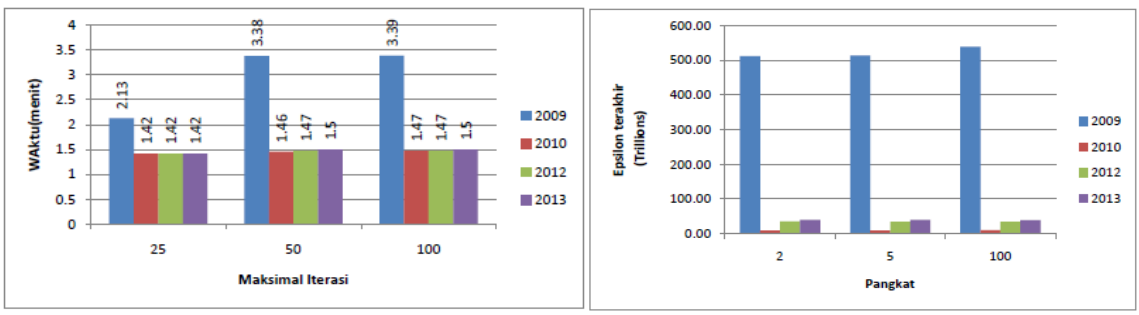
Data yang digunakan untuk pengujian algoritma *K-Means* adalah data kemiskinan di Kabupaten Gunungkidul pada tahun 2009, 2010, 2012 dan 2013 dengan banyak data adalah 347, 13, 16 dan 23. Parameter untuk skenario – skenario ditunjukan pada Tabel 6.

Tabel 6. Ringkasan Skenario untuk Algoritma *K-Means*

Skenario	Maksimal Iterasi	Threshold			
		2009	2010	2011	2012
1	25	10^{-5}	10^{-5}	10^{-5}	10^{-5}
2	50	10^{-5}	10^{-5}	10^{-5}	10^{-5}
3	100	10^{-5}	10^{-5}	10^{-5}	10^{-5}
4	25	10^{-5}	10^{-5}	10^{-5}	10^{-5}
5	100	10^{-5}	10^{-5}	10^{-5}	10^{-5}
6	25	$1{,}37 \times 10^{8}$	$1{,}53 \times 10^{6}$	$7{,}15 \times 10^{6}$	$6{,}51 \times 10^{6}$
7	100	$1{,}37 \times 10^{8}$	$1{,}53 \times 10^{6}$	$7{,}15 \times 10^{6}$	$6{,}51 \times 10^{6}$

3.2.1. Skenario 1-3

Percobaan dengan skenario 1 – 3 dilakukan untuk mengetahui pengaruh banyak iterasi dan data dengan waktu yang dibutuhkan sistem untuk mengolsssah data. Gambar 5 (a) menunjukkan banyaknya iterasi dan data yang mempengaruhi waktu untuk menyelesaikan proses perhitungan algoritma *K-Means*. Semakin banyak iterasi maka semakin lama juga waktu yang diperlukan. Dari Gambar 5 (b) dapat disimpulkan bahwa banyaknya iterasi tidak berpengaruh pada nilai *epsilon* akhir. Hal ini terjadi karena nilai *epsilon* dari percobaan yang dilakukan sama untuk semua percobaan yang dilakukan. Kemungkinan ini terjadi karena pada iterasi 25 kebawah perhitungannya sudah berhenti. Karena nilai *epsilon* awal terlalu kecil proses tetap berjalan sampai maksimal iterasi terpenuhi.



(a) (b)

Gambar 5 (a) Hasil Waktu yang Diperlukan Sistem untuk Parameter Maksimal Iterasi, (b) Hasil *Epsilon* Terakhir untuk Parameter Maksimal Iterasi

3.1.1. Skenario 6 Dan 7

Dari permasalahan di atas maka penulis meneliti sampai iterasi keberapakah proses perhitungan berhenti. Maka percobaan skenario 6 dan 7 dilakukan untuk mengetahui iterasi berapa proses perhitungan berhenti, sehingga penelitian ini mencoba untuk melakukan perhitungan dengan menggunakan maksimal iterasi adalah 50 dan threshold adalah 1,37 x 108 untuk tahun 2009, 1,55 x 106 untuk tahun 2010, 7,15 x 106 untuk tahun 2012, dan 6,51 x 106 untuk tahun 2013.

Tabel 7. Hasil Iterasi Terakhir pada Algoritma *K-Means* dengan Parameter Maksimal Iterasi 50

Tahun	Iterasi akhir
2009	12
2010	4
2011	3
2012	9

Tabel 7 menunjukkan bahwa proses berhenti pada iterasi ke-12 untuk tahun 2009, pada tahun 2010 berhenti pada iterasi ke-4, pada tahun 2012 berhenti pada iterasi ke-312, pada tahun 2013 berhenti pada iterasi ke-9. Percobaan dilakukan juga untuk mengetahui pengaruh nilai *random* objek *cluster* awal terhadap lama waktu dan hasil yang diperoleh. Hasil percobaan pada skenario 4, 5 dan 7 menunjukkan bahwa nilai random sedikit berpengaruh terhadap waktu dan maksimal iterasi yang diperlukan, akan tetapi tidak berpengaruh terhadap hasil yang diperoleh.

Berdasarkan hasil pengujian yang sudah dilakukan dengan skenario-skenario yang sudah ditentukan didapatkan hasil bahwa beberapa parameter saling mempengaruhi hasil yang diperoleh. Seperti halnya percobaan pada algortima *Fuzzy C-Means* bahwa banyak data dan inisialisasi maksimal iterasi awal dapat mempengaruhi lama waktu proses perhitungan, semakin besar data dan maksimal iterasi awal maka akan semakin lama waktu yang diperlukan.

4. KESIMPULAN

Berdasarkan hasil analisa dan pembahasan, diperoleh kesimpulan sebagai berikut :

1. Data kemiskinan di Desa Girijati, Purwosari, Gunungkidul, Yogyakarta memiiki data yang kurang bervariasi untuk kelima belas indikatornya, hal itu menyebabkan hasil clustering menjadi sensitif terhadap perubahan nilai parameter. Dari percobaan yang dilakukan, waktu clustering yang diperlukan algoritma FCM relatif lebih lama serta membutuhkan iterasi lebih banyak dibandingkan dengan algoritma *K-Means*. Algoritma FCM lebih cocok diterapkan pada data yang lebih variatif.
2. Pengujian dilakukan dengan membandingkan hasil clustering kedua algoritma tersebut dengan pengelompokan berdasarkan aturan BAPPEDA 2008. Hasil clustering menggunakan algoritma FCM memiliki tingkat akurasi hanya 50%, sedangkan untuk algoritma K-means memiliki tingkat akurasi lebih baik yaitu 83.33%.

DAFTAR PUSTAKA

[1] Republika, 2013, Republika Online. http://www.republika.co.id/berita/nasional/jawa-tengah-diy-nasional/13/01/02/mfzoyv-tingkat-kemiskinan-di-diy-tertinggi-sejawa [diakses tanggal 26 November 2013].

[2] BAPPEDA, 2008, Petunjuk Teknis Pendataan Ketenegakerjaan dan Keluarga Sejahtera. Kabupaten Gunungkidul, Pemerintahan Kabupaten Gunungkidul BAPPEDA.

[3] Kardi dalam Susanto, A.R., 2013, *Sistem Pendukung Keputusan Pengadaan Buku Perpustakaan STIKOM Surabaya Menggunakan Metode K-Means Clustering*. Makalah TA. Surabaya, STIKOM Surabaya.

[4] Helmiah, 2013, *Sisitem Pendukung Keputusan untuk Pengkatogorian IPK dan Llama Studi Alumni Menggunakan Metode K-Means*. Skripsi, Yogyakarta: UII Teknik Informatika.

[5] Pamungkas, A., 2010, *Perbandingan Distance Space Manhattan(Cityblock) dengan Ecludiean pada Algoritma K-Means Clustering Studi Kasus : Data Balita di Wilayah Kecamatan Melati, Sleman*. Skripsi, Yogyakarta, AKAKOM Sekolah Tinggi Manajemen Informatika dan Komputer

[6] Susilowati, R., 2012, *Clustering Data Pasien Menggunakan Fuzzy C-Means dan Aglomerative Hierarchical Clustering*. Skripsi, Yogyakarta, UIN Sunan Kalijaga Teknik Informatika

[7] Pebrianto, R., 2011, *Aplikasi Clustering Dengan Menggunakan Metode Fuzzy C-Means.* Skripsi, Yogyakarta, UII Teknik Informatika.

[8] Pahri, A.N.I., 2012, *Pengelompokan Uji Laboratorium sebagai Penunjang Diagnosa Demam Berdarah Menggunakan Fuzzy C-Means*. Skripsi, Yogyakarta, UII Teknik Informatika.

[9] BPS, 2011, Kemiskinan dan Ketimbangan Pendapatan : Pengukuran, Relevasi dan Pemanfaatan. In Nasional, D.S.K., ed. *Workshop Evaluasi Data Podes*. Bandung, BPS.

[10] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2), 191-203.

[11] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, 100-108.