

Perbandingan Metode C-Means dan Fuzzy C-Means Dalam Pengelompokan Wilayah Desa/Kelurahan di Kabupaten Kutai Kartanegara

Comparison C-Means Method and Fuzzy C-Means (FCM) Method in Grouping of Village/Urban Village in Kutai Kartanegara Regency

Nissa Irabawati¹, Sri Wahyuningsih², Rudy Ramadani Syoer²

¹Mahasiswa Program Studi Statistika Fakultas MIPA Universitas Mulawarman

²Dosen Program Studi Statistika Fakultas MIPA Universitas Mulawarman

Email: irabawatinissa@yahoo.com¹

Abstract

Cluster analysis is a multivariate statistical technique that has the main purpose to classify objects based on common characteristics. With this analysis, the object will be grouped such that each object is the closest similarity to other objects are in the same group. In the clustering process by using no hierarchical C-Means formation of partition is done such that each object explicitly declared as a member of one group and not a member of any other group. But sometimes can not put an object just in one partition, because in fact the object is located between two or more other partitions, so it needs to be weighted based on its fuzzy membership level. In this way, it is to define a method in the formation of the group will be more flexible. The concept is called fuzzy clustering, the fuzzy way each object can be members of multiple groups. The difference lies in the assumptions used as a basis for allocation. One technique that is not part of the method of using the hierarchical nature of fuzzy clustering technique is using Fuzzy C-Means (FCM). This study will examines comparative method C-Means and FCM clustering in a case study, namely the grouping of the village/urban village in Kutai Kartanegara regency based on the characteristics of facilities/infrastructure and socio-economic factors of the population. The results showed that in some respects, FCM was superior than the C-Means, especially in generating the minimum of objective function, the computation time and ratio value S_w and S_b . Based on the similarity matrix eigen value and the index value Xie and Beni (XB) concluded that the most optimal number of groups is 5 (five) groups.

Keyword : Cluster analysis, C-Means, FCM, grouping region, Xie and Beni index (XB).

PENDAHULUAN

Analisis kelompok merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya.

Secara umum terdapat dua metode dalam analisis kelompok yaitu: metode hierarki dan metode tak berhierarki. Hasil *clustering* dengan metode berhierarki secara umum membentuk diagram pohon (*tree diagram*) atau *dendrogram* yang menggambarkan pengelompokan objek berdasarkan jarak. Metode tak berhierarki disebut juga metode partisi (*partitional methods*) termasuk di dalamnya adalah *C-Means*.

Pada proses *clustering* dengan menggunakan metode tak berhierarki (*C-Means*) pembentukan partisi dilakukan sedemikian rupa sehingga setiap objek berada tepat pada satu partisi. Akan tetapi pada suatu saat secara alami hal ini sering tidak dapat dilakukan untuk menempatkan suatu objek tepat pada satu partisi, karena sebenarnya objek tersebut terletak diantara dua atau lebih partisi yang lain. Pengelompokan dengan mempertimbangkan tingkat keanggotaan yang mencakup himpunan *fuzzy* sebagai dasar

pembobotan disebut dengan *fuzzy clustering* (Bezdek, 1981). Metode ini merupakan pengembangan dari metode partisi cara tegas (*C-Means*) dengan melakukan pembobotan *fuzzy* yang memungkinkan objek untuk bisa bergabung ke setiap kelompok yang ada. Salah satu teknik yang merupakan bagian dari metode tak berhierarki dengan menggunakan logika *fuzzy* adalah teknik pengelompokan menggunakan metode *Fuzzy C-Means* (FCM). Secara umum, metode FCM berbasis pada fungsi objektif yang diperoleh dari penghitungan jarak terhadap pusat kelompok. Dengan teknik ini, objek akan cenderung menjadi anggota suatu kelompok dimana objek tersebut memiliki derajat keanggotaan yang tertinggi terhadap kelompoknya. Metode FCM sering digunakan dalam pengelompokan, karena metode ini memberikan hasil yang halus dan cukup efektif untuk meningkatkan homogenitas tiap kelompok yang dihasilkan (Shihab, 2000).

Isu kemiskinan merupakan fokus perhatian pembangunan di hampir semua negara. Di Indonesia masalah kemiskinan merupakan salah satu permasalahan utama pemerintah yang diprioritaskan dalam menyusun strategi pembangunan.

Berkaitan dengan upaya pengentasan kemiskinan, Badan Pusat Statistik (BPS) berperan dalam menyediakan data tentang kemiskinan, antara lain dengan melakukan penghitungan desa tertinggal sebagai proksi identifikasi daerah kantong-kantong kemiskinan.

Penentuan desa tertinggal seringkali dilakukan dalam rangka menetapkan penyaluran bantuan pemerintah. Untuk itu penetapan status desa tertinggal diharapkan dapat menjadi manifestasi suatu ukuran yang mengidentifikasi daerah kantong-kantong kemiskinan. Untuk mengatasi hal tersebut, maka dibutuhkan sebuah kegiatan evaluasi dan studi kasus yang dapat mengelompokkan desa tertinggal untuk mengetahui karakteristik dalam bidang kemiskinan. Sehingga dapat menentukan desa mana saja yang diprioritaskan untuk mendapatkan bantuan dari pemerintah.

Pada penelitian ini akan dilakukan pengelompokan wilayah desa/kelurahan di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur dengan menerapkan metode *C-Means* dan FCM berdasarkan data hasil pendataan Potensi Desa (PODES) 2011, kemudian dibandingkan hasil pengelompokan tersebut.

Analisis Kelompok

Analisis kelompok merupakan suatu teknik statistik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang dimilikinya. Dengan analisis ini, objek akan dikelompokkan sedemikian rupa sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam suatu kelompok yang sama. Kelompok-kelompok yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi.

Ukuran Kemiripan

Dalam Agusta (2007) ada beberapa cara yang dapat digunakan untuk mengukur jarak antar data ke pusat kelompok (*centroid*) salah satunya adalah *euclidean*. Ukuran jarak ini merupakan ukuran jarak yang paling terkenal yang didefinisikan sebagai berikut:

$$D_{ik} = D(x_k, v_i) = \sum_{i=1}^c \|x_k - v_i\| = \sqrt{\sum_{i=1}^c (x_k - v_i)^2} \quad (1)$$

Metode C-Means

C-Means merupakan salah satu metode data *clustering* tak berhierarki yang berusaha mempartisi data kedalam bentuk satu atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan kedalam kelompok lain

Pada metode *C-Means*, pengalokasian kembali data ke dalam masing-masing kelompok didasarkan pada perbandingan jarak antara data dengan *centroid* setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai *centroid* terdekat dengan data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut:

$$a_{ik} = \begin{cases} 1, d = \min \{D^2(x_k, v_i)\} \\ 0, d \text{ lainnya} \end{cases} \quad (2)$$

Untuk metode *C-Means*, fungsi objektif yang digunakan adalah sebagai berikut:

$$J = \sum_{k=1}^n \sum_{i=1}^c a_{ik} D^2(x_k, v_i) \quad (3)$$

a_{ik} mempunyai nilai 0 atau 1. Apabila suatu data merupakan anggota suatu kelompok maka nilai $a_{ik} = 1$ dan jika sebaliknya, $a_{ik} = 0$

Pengelompokan data dengan metode *C-Means* secara umum dilakukan dengan algoritma sebagai berikut:

1. Menentukan jumlah kelompok yang ingin dibentuk.
2. Mengalokasikan data kedalam kelompok secara acak.
3. Menghitung nilai *centroid* masing-masing kelompok dengan persamaan berikut:

$$v_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{kj} \quad (4)$$

dimana:

v_{ij} = *Centroid*/rata-rata kelompok ke- i

n_i = Jumlah objek yang menjadi anggota kelompok ke- i

i, k = Indeks dari kelompok

j = Indeks dari variabel

x_{kj} = Data ke- k yang ada di dalam kelompok tersebut untuk variabel ke- j

4. Mengalokasikan masing-masing data ke *centroid* terdekat dan membuat *centroid* baru.
5. Kembali ke langkah 3, apabila masih terdapat perpindahan data dari satu kelompok ke kelompok yang lain, atau apabila perubahan pada nilai fungsi objektif masih di atas nilai *threshold* yang ditentukan. Nilai *threshold* ($<$) adalah suatu nilai yang sangat kecil mendekati 0, (misal 10^{-6}).

(Prasetyo, 2012)

Metode Fuzzy C-Means (FCM)

Metode *Fuzzy C-Means* (FCM) (Klawonn, 2007) merupakan salah satu metode pengelompokan yang dikembangkan dari metode *C-Means* dengan menerapkan sifat *fuzzy*

keanggotaannya sehingga data dapat menjadi anggota dari semua kelas atau kelompok terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga tingkat keberadaan data dalam suatu kelompok ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 (Luthfi, 2007). Metode FCM memperkenalkan suatu variabel w yang merupakan *weighting exponent* dari *membership function*. Variabel ini dapat mengubah besar pengaruh dari *membership function*, dalam proses pengeompokkan menggunakan metode FCM, w mempunyai wilayah nilai lebih besar dari 1 ($w > 1$). *Membership function* untuk suatu data ke suatu kelompok tertentu. *Membership function* mempunyai jangkauan nilai $0 \leq \tilde{\sim}_{ik} \leq 1$. Nilai w yang umumnya digunakan adalah 2 (Bezdek and Dunn, 1975).

Tahapan dari metode FCM untuk pengelompokkan data secara umum dirumuskan sebagai berikut (Kusumadewi dan Purnomo, 2010):

1. Menentukan *initial value* yaitu jumlah kelompok (c), pangkat (w), maksimum iterasi (MaxIter), error terkecil yang diharapkan (ϵ), fungsi obyektif awal ($P_0=0$), iterasi awal ($t=1$)
2. Membangkitkan bilangan *random* $(\tilde{\sim}_{ik})_{c \times n}$, $i=1,2,\dots,c$; $k=1,2,\dots,n$; $c \times n$ menunjukkan ukuran matriks sebagai elemen-elemen matriks partisi awal U . Matriks partisi (U) pada pengelompokan *fuzzy* memenuhi kondisi pada persamaan:

$$\tilde{\sim}_{ik} \in [0,1], 1 \leq i \leq c, 1 \leq k \leq n \quad (5)$$

$\tilde{\sim}_{ik}$ adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu kelompok.

Hitung jumlah setiap kolom:

$$Q_i = \sum_{k=1}^c \tilde{\sim}_{ik} \quad (6)$$

dimana:

$k = 1,2,\dots,n$

Q_i = Hasil penjumlahan derajat keanggotaan seluruh kelompok objek ke- k

$\tilde{\sim}_{ik}$ = Derajat keanggotaan pada kelompok ke- i untuk objek ke- k

Setelah itu hitung elemen matriks yang ternormalisasi dengan rumus:

$$\tilde{\sim}_{ik_{norm}} = \frac{\tilde{\sim}_{ik}}{Q_i} \quad (7)$$

dengan $\tilde{\sim}_{ik_{norm}}$ merupakan derajat keanggotaan pada data kelompok ke- i untuk objek ke- k setelah normalisasi.

Bentuk matriks partisi awal yang telah ternormalisasi sebagai berikut:

$$\begin{bmatrix} \tilde{\sim}_{11_{norm}} & \tilde{\sim}_{12_{norm}} & \cdots & \tilde{\sim}_{1n_{norm}} \\ \tilde{\sim}_{21_{norm}} & \tilde{\sim}_{22_{norm}} & \cdots & \tilde{\sim}_{2n_{norm}} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\sim}_{c1_{norm}} & \tilde{\sim}_{c2_{norm}} & \cdots & \tilde{\sim}_{cn_{norm}} \end{bmatrix} \quad (8)$$

3. Menghitung pusat kelompok v_{ij} dengan $i=1,2,\dots,c$; dan $j=1,2,\dots,m$; menggunakan persamaan berikut:

$$v_{ij} = \frac{\sum_{k=1}^n (\tilde{\sim}_{ik_{norm}})^w x_{kj}}{\sum_{k=1}^n (\tilde{\sim}_{ik_{norm}})^w} \quad (9)$$

Persamaan untuk proses iterasi pusat kelompok:

$$v_{ij} = \frac{\sum_{k=1}^n (\tilde{\sim}_{ik_{iter}}^{t-1})^w x_{kj}}{\sum_{k=1}^n (\tilde{\sim}_{ik_{iter}}^{t-1})^w} \quad (10)$$

Hasil perhitungan v_{ij} ditulis dalam bentuk matriks berukuran $c \times m$ sebagai berikut:

$$v^t = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ v_{c1} & v_{c2} & \cdots & v_{cm} \end{bmatrix} \quad (11)$$

4. Memperbaiki derajat keanggotaan setiap data dalam setiap kelompok dengan menghitung perubahan matriks partisi:

$$\tilde{\sim}_{ik_{iter}}^t = \frac{\left[\sum_{j=1}^m (x_{kj} - v_{ij}^t)^2 \right]^{\frac{-1}{w-1}}}{\sum_{i=1}^c \left[\sum_{j=1}^m (x_{kj} - v_{ij}^t)^2 \right]^{\frac{-1}{w-1}}} \quad (12)$$

dengan $\tilde{\sim}_{ik_{iter}}^t$ merupakan derajat keanggotaan pada data kelompok ke- i untuk objek ke- k dalam proses iterasi ke- t . Derajat keanggotaan yang telah diperbaiki ini merupakan elemen matriks U^t berukuran $c \times n$. Bentuk matriksnya adalah sebagai berikut:

$$U^t = \begin{bmatrix} \tilde{\sim}_{11_{iter}}^t & \tilde{\sim}_{12_{iter}}^t & \cdots & \tilde{\sim}_{1n_{iter}}^t \\ \tilde{\sim}_{21_{iter}}^t & \tilde{\sim}_{22_{iter}}^t & \cdots & \tilde{\sim}_{2n_{iter}}^t \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\sim}_{c1_{iter}}^t & \tilde{\sim}_{c2_{iter}}^t & \cdots & \tilde{\sim}_{cn_{iter}}^t \end{bmatrix} \quad (13)$$

Dalam menghitung $\tilde{\sim}_{ik_{iter}}^t$ pada iterasi ke- t (elemen matriks U^t), maka nilai v_{ij} yang digunakan adalah v_{ij} pada iterasi ke- t pula (elemen matriks v^t).

- Menghitung nilai fungsi objektif yang digunakan sebagai syarat perulangan untuk mendapatkan pusat kelompok yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke kelompok mana pada step akhir. Persamaan untuk nilai awal fungsi objektif:

$$J_t = \sum_{k=1}^n \sum_{i=1}^c \left[\sum_{j=1}^m (x_{kj} - v_{ij}^t)^2 \right] * (\sim_{ik_{norm}})^2 \quad (14)$$

Untuk $t=1$

Persamaan untuk proses iterasi fungsi objektif:

$$J_t = \sum_{k=1}^n \sum_{i=1}^c \left[\sum_{j=1}^m (x_{kj} - v_{ij}^t)^2 \right] * (\sim_{ik_{iter}})^2 \quad (15)$$

Untuk $t = 1, 2, \dots$

Untuk menghitung nilai fungsi objektif pada iterasi ke- t (P_t), pusat kelompok v_{ij} yang digunakan adalah v_{ij} iterasi ke- t (elemen matriks v^t) dan derajat keanggotaan $\sim_{ik_{iter}}$ iterasi ke- $(t-1)$ yaitu elemen matriks U^{t-1} .

- Mengecek kondisi iterasi berhenti:

Jika $(|J_t - J_{t-1}| < \epsilon)$ atau $(t > \text{MaxIter})$ maka iterasi dihentikan, jika tidak maka iterasi dinaikkan $t = t+1$ dengan mengulangi langkah ke-4 untuk memulai iterasi baru.

Evaluasi Hasil Pengelompokkan

Untuk melihat kinerja kedua algoritma metode pengelompokkan yang digunakan, dapat dilihat melalui kriteria nilai simpangan baku, yaitu nilai simpangan baku dalam kelompok (S_w) dan nilai simpangan baku antar kelompok (S_b). Simpangan baku dalam kelompok diharapkan memiliki nilai yang minimum, dan simpangan baku antar kelompok memiliki nilai maksimum (Bunkers and Miller, 1996). Formulasnya sebagai berikut:

$$S_{wj} = \frac{1}{c} \sum_{k=1}^c S_{kj} \quad (16)$$

$$S_{bj} = \sqrt{\frac{1}{c-1} \sum_{k=1}^c |x_{kj} - \bar{x}_j|^2} \quad (17)$$

dimana:

j = banyaknya atribut/variabel

S_{kj} = Simpangan baku di dalam kelompok variabel ke- j

\bar{x}_{kj} = Rata-rata dari kelompok variabel ke- j

Semakin kecil nilai S_w dan semakin besar nilai S_b , maka metode pengelompokkan memiliki kinerja yang semakin baik, sehingga digunakan rasio antara S_w dan S_b . Rasio S_w/S_b yang terkecil menunjukkan ketepatan pengelompokkan yang paling baik.

Penentuan Jumlah Kelompok Optimal

Penentuan jumlah kelompok didasarkan pada dua hal, yang pertama adalah dengan membatasi jumlah kelompok yang terbentuk melalui proporsi nilai eigen matriks *similarity* dari objek yang akan dikelompokkan. Yang kedua adalah melakukan kontrol dengan indeks XB.

Sesuai dengan namanya Indeks XB ditemukan oleh Xie dan Beni yang pertama kali dikemukakan pada tahun 1991. Indeks ini dapat digunakan untuk metode *hard partition* seperti *C-Means* maupun FCM. Kriterianya banyak kelompok optimum diberikan oleh nilai XB yang minimum pada lembah pertama.

$$XB(c) = \frac{\sum_{i=1}^c \sum_{k=1}^n (\sim_{ik})^w D_{ik}^2(x_k, v_i)}{n \min_{i,k} \|\mathbf{v}_k, \mathbf{v}_i\|^2} \quad (18)$$

Dengan c menyatakan banyak kelompok, \sim_{ik} adalah derajat keanggotaan, D_{ik}^2 adalah jarak data dengan pusat kelompok, n merupakan jumlah objek penelitian, $\min_{i,k} \|\mathbf{v}_k, \mathbf{v}_i\|^2$ menyatakan jarak minimum antara pusat kelompok v_k dan v_i .

Analisis Faktor

Analisis faktor merupakan salah satu teknik statistik untuk menyederhanakan deskripsi dari suatu set data (variabel) yang banyak dan saling berkorelasi menjadi set data yang ringkas dan tidak lagi berkorelasi. Analisis ini berguna untuk meneliti keterkaitan peubah-peubah dalam satu set data. Analisis faktor pada dasarnya bertujuan untuk mendapatkan sejumlah kecil faktor yang memiliki sifat berikut (Johnson and Wichern, 2007):

- Mampu menerangkan semaksimal mungkin keragaman data,
- Terdapat kebebasan antar faktor,
- Tiap faktor dapat diinterpretasikan sejelas-jelasnya.

Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder hasil pendataan Potensi Desa (PODES) 2011 Kabupaten Kutai Kartanegara, data tersebut diperoleh dari Badan Pusat Statistik (BPS) Provinsi Kalimantan Timur (Kaltim). Dari data PODES, akan diambil sebanyak 18 variabel yang dapat menggambarkan karakteristik daerah, yaitu:

Faktor Sarana dan Prasarana/Akses

X_1 = Ketersediaan sarana pendidikan (jumlah TK/sederajat).

X_2 = Ketersediaan sarana pendidikan (jumlah SD/sederajat).

X_3 = Ketersediaan tenaga kesehatan (jumlah bidan).

X_4 = Ketersediaan tenaga kesehatan (jumlah mantri).

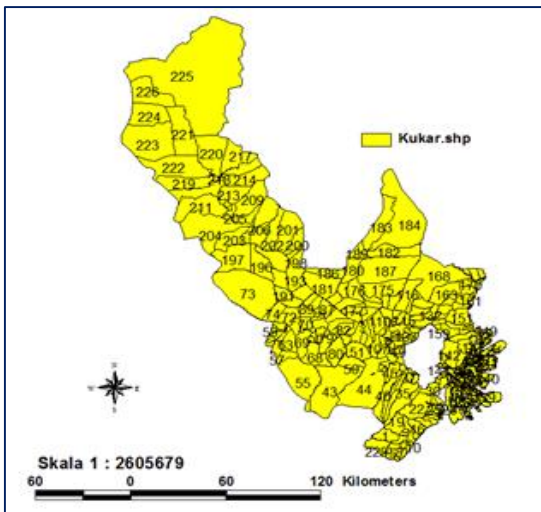
X_5 = Ketersediaan sarana kesehatan (jumlah

- puskesmas).
- X_6 = Jumlah keluarga berlangganan telepon kabel.
- X_7 = Jumlah minimarket.
- X_8 = Jumlah toko/warung kelontong.
- X_9 = Jumlah warung/kedai makanan minuman.
- X_{10} = Jumlah restoran/rumah makan.
- X_{11} = Jumlah penginapan/hostel/motel/losmen/wisma.
- X_{12} = Jumlah tempat ibadah (masjid).

Faktor Sosial Ekonomi Penduduk

- X_{13} = Jumlah keluarga pertanian.
- X_{14} = Jumlah keluarga pengguna listrik PLN.
- X_{15} = Jumlah keluarga tinggal di bantaran tepi sungai.
- X_{16} = Jumlah keluarga tinggal di pemukiman kumuh.
- X_{17} = Jumlah penderita gizi buruk selama 3 tahun terakhir.
- X_{18} = Jumlah penerima kartu JAMKESMAS/JAMKESDA.

Berikut adalah peta Kabupaten Kutai Kartanegara



Gambar 1. Peta desa/kelurahan Kabupaten Kutai Kartanegara

Adapun tahapan analisis data dalam penelitian ini adalah:

1. Langkah awal dengan melakukan seleksi variabel, mendeskripsikan data, kemudian melakukan analisis faktor, untuk mereduksi variabel yang ada sehingga dihasilkan nilai skor-skor faktor.
2. Kemudian menyajikan data dalam matriks berukuran $n \times m$ dengan n menyatakan jumlah objek penelitian yang diteliti, yaitu 227 desa/kelurahan dan m menyatakan jumlah variabel penelitian, yaitu jumlah skor faktor terbentuk.
3. Meng-input data.
4. Menormalisasi data berdasarkan range-nya dengan bantuan Software Matlab 7.8.

5. Menggunakan metode pengelompokkan tak berhierarki dengan menerapkan metode *C-Means*.
6. Menggunakan metode pengelompokkan tak berhierarki dengan menerapkan metode FCM.
7. Mengevaluasi hasil pengelompokkan melalui nilai S_w sesuai persamaan (16) dan nilai S_b sesuai persamaan (17), untuk mempermudah dalam membandingkan nilai tersebut maka digunakan rasio antara S_w/S_b , nilai terkecil menunjukkan ketepatan pengelompokkan yang paling baik.

Berdasarkan langkah 5, 6, dan 7 akan didapatkan ukuran hasil pengelompokkan seperti nilai fungsi objektif, waktu komputasi, dan rasio S_w/S_b . Kemudian berdasarkan beberapa kriteria ukuran hasil pengelompokkan tersebut, maka dapat diputuskan metode terbaik mana yang akan direkomendasikan untuk analisis selanjutnya.

8. Tahap selanjutnya adalah menentukan atau merekomendasikan jumlah kelompok optimal, caranya adalah dengan melihat nilai eigen matriks *similarity* (angka nilai eigen dibawah 1 tidak digunakan dalam menghitung jumlah kelompok yang terbentuk) kemudian dikontrol dengan menghitung nilai indeks XB sesuai persamaan (18). Jumlah kelompok optimal biasanya diberikan pada saat indeks validitas kelompok mencapai kondisi nilai minimum pertama di lembah pertama yang didapatkan (Syoyer, 2011).

Hasil dan Pembahasan
Statistika Deskriptif

Tabel 1. Statistika Deskriptif

Variabel	N	Rata-Rata	Standar Deviasi	Missing
X_1	227	1,5991	1,5118	0
X_2	227	2,1718	1,4938	0
X_3	227	1,2643	1,3924	0
X_4	227	1,7137	2,9286	0
X_5	227	1,8678	0,3394	0
X_6	227	20,8106	85,2463	0
X_7	227	0,1498	0,6942	0
X_8	227	29,8678	45,7970	0
X_9	227	7,1982	12,6435	0
X_{10}	227	0,1630	0,6684	0
X_{11}	227	0,3568	1,5141	0
X_{12}	227	2,5154	2,2740	0
X_{13}	227	329,0000	286,1391	0
X_{14}	227	466,4449	621,7253	0
X_{15}	227	130,5991	223,4931	0
X_{16}	227	4,6960	47,9432	0
X_{17}	227	0,3700	1,7586	0
X_{18}	227	295,7445	727,8187	0

Berdasarkan nilai pada Tabel 1, jika dilihat kolom N terlihat angka untuk semua variabel adalah sama yaitu berjumlah 227 yang berarti

bahwa pada variabel X_1 hingga variabel X_{18} datanya lengkap terisi. Hal ini bisa juga dilihat pada kolom *missing*, pada bagian untuk jumlah nyata dimana variabel X_1 hingga X_{18} bernilai 0, dan begitu juga pada bagian persen untuk menghitung persentase pada semua variabel bernilai 0%. Hal ini menunjukkan bahwa tidak terdapat satupun variabel yang mengalami *missing value*.

Kolom rata-rata dan standar deviasi menunjukkan nilai statistik dasar, yakni rata-rata dan standar deviasi untuk setiap variabel, yang dihitung dari jumlah data yang valid (tidak *missing*). Untuk variabel X_1 , nilai rata-ratanya adalah 1,5991 dengan standar deviasi 1,5118. Demikian seterusnya untuk data yang lain.

Reduksi Variabel dengan Analisis Faktor

Berdasarkan matriks *Anti-Image (Anti Image Matrices)* yang menunjukkan nilai, terdapat 3 variabel yaitu variabel X_{11} dengan nilai $MSA=0,494$, X_{16} dengan nilai $MSA=0,434$, X_{17} dengan nilai $MSA =0,366$, yang memiliki nilai MSA dibawah 0,5 yang berarti variabel tersebut tidak dapat diprediksi dan dianalisis lebih lanjut atau dikeluarkan dari analisis, karena lebih dari satu variabel yang mempunyai nilai $MSA < 0,5$ maka yang dikeluarkan adalah variabel dengan MSA terkecil yaitu variabel X_{17} . Kemudian proses pengujian tetap diulangi lagi sampai tidak ada variabel yang memiliki nilai $MSA < 0,5$. Setelah pengulangan, sudah tidak ada lagi variabel yang nilai $MSA < 0,5$ dan menghasilkan nilai KMO sebesar 0,8075 dengan signifikansi *Barlett's test* sebesar 0,0000 (berarti variabel sudah memadai untuk dianalisis lebih lanjut. Berdasarkan nilai KMO dan *Barlett's test* tersebut, dapat disimpulkan bahwa semua variabel layak untuk dianalisis dengan variabel yang dapat diolah lebih lanjut sebanyak 16 buah variabel.

Tabel 2. Nilai KMO dan *Bartlett's Test*

<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</i>			0,7558
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi-Square</i>		1577,8157
	<i>Df</i>		153
	<i>P-value</i>		0,0000

Berdasarkan Tabel 3, hanya terdapat 5 faktor yang memiliki nilai eigen lebih dari 1, ini berarti bahwa kelima faktor tersebut memberikan kontribusi keragaman yang besar, dan faktor yang memiliki nilai eigen kurang dari 1 dianggap memiliki kontribusi keragaman yang kurang. Hal ini juga menunjukkan bahwa 5 faktor adalah paling bagus untuk meringkas ke-16 variabel tersebut.

Tabel 3. Nilai Eigen dan Persentase Variansi

Faktor	Nilai Eigen	Persentase Variansi (%)	Kumulatif (%)
1	5,2332	32,7077	32,7077
2	1,5451	9,6571	42,3648
3	1,3352	8,3447	50,7095
4	1,2713	7,9455	58,6551
5	1,0571	6,6071	65,2622
6	0,8945	5,5907	70,8529
7	0,8032	5,0198	75,8727
8	0,7133	4,4579	80,3306
9	0,6375	3,9842	84,3149
10	0,5144	3,2149	87,5298
11	0,4483	2,8018	90,3316
12	0,4166	2,6039	92,9355
13	0,3723	2,3268	95,2623
14	0,3064	1,9148	97,1771
15	0,2704	1,6899	98,8670
16	0,1813	1,1330	100,0000

Merekomendasikan Metode Pengelompokan Terbaik

Tabel 4. Pengelompokan dengan Metode *C-Means*

Kelompok	Fungsi Objektif Akhir	Waktu Komputasi (Detik)	Rasio Sw/Sb
2	12,8326	6,8569	0,4196
3	9,1117	8,1267	0,4980
4	7,2415	8,0158	0,4464
5	6,6964	7,2568	0,4496
6	6,1069	8,4896	0,4755
7	5,6526	7,3077	0,4803
8	4,8726	7,5401	0,4963
9	4,3820	8,9192	0,4915
10	4,2909	8,0067	0,3112

Tabel 5. Pengelompokan dengan Metode *FCM*

Kelompok	Fungsi Objektif Akhir	Waktu Komputasi (Detik)	Rasio Sw/Sb
2	7,3706	6,2666	0,3957
3	4,8496	7,8613	0,4648
4	3,4725	7,4524	0,4371
5	2,7091	6,6658	0,4112
6	2,2377	6,4237	0,3883
7	1,9085	6,6814	0,3856
8	1,6670	6,0576	0,3915
9	1,4604	6,8606	0,3310
10	1,2992	6,6067	0,1181

Tabel 4 dan Tabel 5 menyajikan hasil dari ukuran kriteria pengelompokan. Pada metode *C-Means* dan *FCM*, nilai fungsi objektifnya menunjukkan bahwa semakin turun apabila jumlah kelompok bertambah. Metode *C-Means* memberikan nilai fungsi objektif minimum saat jumlah kelompok 10 dengan nilai sebesar 4,2909, begitu juga dengan metode *FCM* nilai fungsi objektif minimum saat jumlah kelompok 10 dengan nilai sebesar 1,2992. Untuk jumlah

kelompok dari $c=2$ sampai dengan $c=10$, FCM memberikan nilai fungsi objektif yang lebih kecil dari pada *C-Means*. Fungsi objektif merupakan fungsi tujuan yang akan bernilai minimum pada saat model sudah konvergen, artinya model sudah dalam bentuk terbaiknya.

Berdasarkan waktu komputasi, *C-Means* memerlukan waktu yang lebih lama dibandingkan FCM. Sebagai contohnya untuk jumlah kelompok $c=2$, waktu komputasi yang dibutuhkan *C-Means* adalah 6,8569 detik sedangkan FCM 6,2666 detik, demikian untuk jumlah kelompok $c=3$ sampai dengan $c=10$. *C-Means* memerlukan waktu yang lebih lama dibanding dengan FCM.

Tahapan terakhir dalam analisis pengelompokan adalah melakukan mengukur kinerja hasil pengelompokan. Kelompok yang baik adalah kelompok yang terdiri atas objek-objek yang homogen, dan heterogen terhadap anggota kelompok lain. Tabel 4 dan Tabel 5 menjelaskan perbandingan rasio dari hasil pengelompokan metode *C-Means* dan FCM. Semakin kecil rasio antara S_w dan S_b , maka metode yang digunakan memiliki kinerja yang baik, yang dapat membentuk kelompok dengan homogenitas yang maksimum dalam kelompok, dan homogenitas yang minimum antar kelompok. Berdasarkan kriteria ini, FCM memberikan nilai rasio S_w/S_b yang lebih kecil dibanding *C-Means* pada jumlah kelompok $c=2$ sampai dengan $c=10$.

Kesimpulan yang dapat diambil berdasarkan beberapa kriteria hasil pengelompokan yang telah didapat, bahwa berdasarkan perbandingan tersebut hasil pengelompokan dengan menggunakan metode *C-Means* dan FCM dengan kriteria fungsi objektif minimum, waktu komputasi yang minimum, dan rasio S_w/S_b yang minimum, maka dalam penelitian ini, penulis merekomendasikan penggunaan metode FCM untuk analisis kelompok.

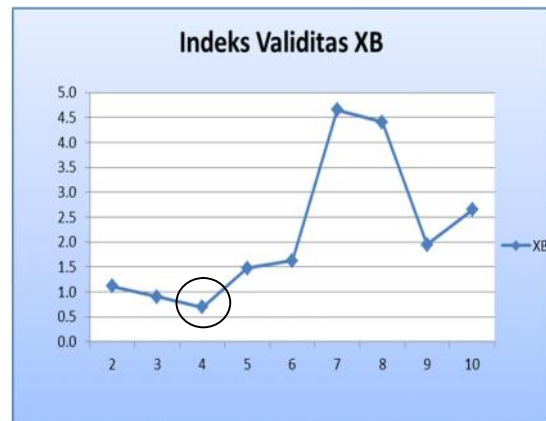
Merekomendasi Jumlah Kelompok Optimal

Setelah direkomendasikan metode FCM sebagai metode pengelompokan terbaik, langkah selanjutnya adalah menentukan jumlah kelompok optimal/terbaik. Dari nilai pada Tabel 3, dapat dilihat bahwa proporsi 65,2622% diberikan oleh lima nilai eigen, maka kesimpulan sementara yang dapat diambil adalah pengelompokan optimum yang terbentuk adalah sebanyak lima kelompok. Angka nilai eigen dibawah 1 tidak digunakan dalam menghitung jumlah kelompok yang terbentuk

Selanjutnya pembatasan jumlah pengelompokan ini kemudian dikontrol dengan indeks XB. Perhitungan akan dicobakan untuk jumlah kelompok 2 sampai dengan 10. Untuk menghitung indeks XB digunakan persamaan (19), yaitu fungsi objektif dibagi dengan banyak sampel kali jarak minimum antar pusat kelompok.

Tabel 6. Indeks XB 2 s/d 10 Kelompok

Jumlah Kelompok	Fungsi Objektif	Jarak	
		Minimum Pusat Kelompok	Indeks XB
2	7,3706	0,0290	1,1196
3	4,8496	0,0236	0,9068
4	3,4725	0,0220	0,6964
5	2,7091	0,0081	1,4792
6	2,2377	0,0061	1,6255
7	1,9085	0,0018	4,6518
8	1,6669	0,0017	4,4107
9	1,4604	0,0033	1,9500
10	1,2992	0,0022	2,6450



Gambar 2. Grafik indeks validitas kelompok

Jumlah kelompok optimal umumnya ditunjukkan pada saat indeks validitas kelompok mencapai kondisi nilai minimum pertama di lembah pertama yang didapatkan. Berdasarkan Tabel 6, terlihat bahwa nilai XB minimal dilembah pertama yaitu sebesar 0,6964, yaitu pada saat jumlah kelompok sebanyak 4 (Gambar 2).

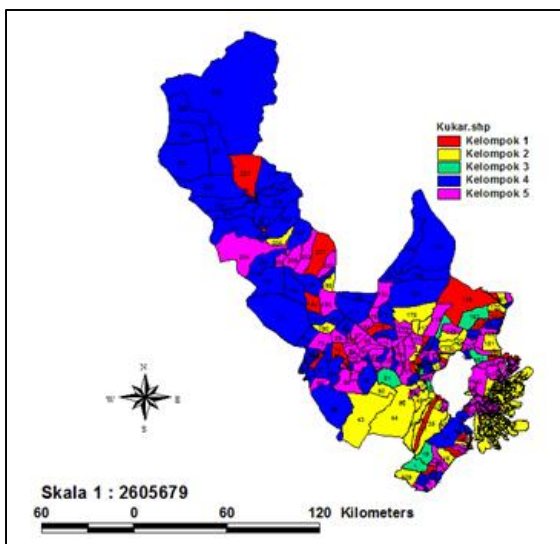
Pengelompokan dengan 4 kelompok menghasilkan *range* yang cukup lebar sehingga informasi pengelompokan terlalu sedikit dan kurang detail. Abonyi dan Feil (2007) dalam Syoer (2011) menyebutkan: pertama, tidak ada indeks yang terbaik, untuk bisa dijadikan “acuan” untuk penentuan jumlah kelompok. Kedua, sesuai keperluan untuk penelitian, atau sesuai dengan kasus yang akan digunakan untuk tujuan peneliti. Sehingga berdasarkan dua hal tersebut dan hasil pada Tabel 6, untuk mendapatkan hasil pengelompokan yang detail dan informasi lebih banyak, maka dalam penelitian ini peneliti memilih 5 kelompok sesuai rekomendasi dari nilai eigen. Jumlah kelompok tersebut dirasa sudah cukup untuk menjelaskan keragaman dan karakteristik kelompok data.

Interpretasi Hasil Pengelompokan

Tabel 7. Hasil Pengelompokan dengan Metode FCM

Kelompok	Kode Desa/Kelurahan
1	12,17,25,29,31,38,41,62,70,76,85,99, 104,114,128,149,153,162,164, 168,178,194,201,220,227. 16,20,24,28,35,37,39,40,42,43,44, 45,47,50,100,106,112,113,118,127,
2	129,130,132,133,139,140,141,144, 146,150,151,155,167,170,175,190, 198,205.
3	1,19,48,51,96,97,98,101,136,152, 163. 4,5,6,8,9,10,11,22,26,27,30,33,46, 52,55,57,58,59,63,64,65,66,67,69, 71,73,74,78,80,81,90,94,95,102, 103,115,117,119,122,123,124,125,
4	131,135,137,159,160,161,165,176, 177,181,182,183,184,185,186,187, 189,191,192,195,196,197,199,203, 207,209,210,211,212,213,214,215, 216,217,218,219,221,222,223,224, 225,226. 2,3,7,13,14,15,18,21,23,32,34,36, 49,53,54,56,60,61,68,72,75,77,79, 82,83,84,86,87,88,89,91,92,93,105,
5	107,108,109,110,111,116,120,121, 126,134,138,142,143,145,147,148, 154,156,157,158,166,169,171,172, 173,174,179,180,188,193,200,202, 204,206,208.

Kelompok 1 terdiri dari 25 desa/kelurahan, kelompok 2 terdiri dari 38 desa/kelurahan, kelompok 3 terdiri dari 11 desa/kelurahan, kelompok 4 terdiri dari 84 desa/kelurahan, dan kelompok 5 terdiri dari 69 desa/kelurahan.



Gambar 3. Peta desa/kelurahan Kabupaten Kutai Kartanegara pengelompokan dengan metode FCM dengan jumlah kelompok sebanyak 5

Gambar 3. menunjukkan peta desa/kelurahan di Kabupaten Kutai Kartanegara hasil

pengelompokan dengan metode FCM dengan jumlah kelompok sebanyak 5 kelompok. Berdasarkan peta wilayah tersebut terlihat bahwa kelompok 1 ditandai dengan warna merah, kelompok 2 ditandai dengan warna kuning, kelompok 3 ditandai dengan warna hijau, kelompok 4 ditandai dengan biru, dan kelompok 5 ditandai dengan warna ungu. Adapun karakteristik masing-masing kelompok dapat digambarkan dari pusat kelompoknya (*centroid*) seperti terlihat pada Tabel 8.

Tabel 8. Karakteristik Kelompok Berdasarkan Centroid-nya Metode FCM

Variabel	Centroid				
	Kel 1	Kel 2	Kel 3	Kel 4	Kel 5
SF 1	0,4061	0,1665	0,2690	0,1692	0,1685
SF 2	0,2518	0,3273	0,2175	0,1114	0,1995
SF 3	0,1191	0,1397	0,1292	0,1111	0,1289
SF 4	0,1544	0,2068	0,7918	0,2069	0,2076
SF 5	0,4602	0,5415	0,4901	0,5239	0,5242
Rata-rata	0,2783	0,2764	0,3795	0,2245	0,2458

Karakteristik kelompok dinilai berdasarkan *ranking* kelompoknya, yaitu dari *ranking* kelompok tertinggi ke *ranking* terendah. Berdasarkan nilai rata-rata *centroid*, kelompok 4 merupakan kelompok yang paling rendah *ranking*-nya, sedangkan kelompok 3 merupakan kelompok yang paling tinggi. Susunan *ranking* berdasarkan rata-rata *centroid*-nya adalah sebagai berikut:

$$v = \begin{pmatrix} 0,4061 & 0,2518 & 0,1191 & 0,1544 & 0,4602 \\ 0,1665 & 0,3273 & 0,1397 & 0,2068 & 0,5415 \\ 0,2690 & 0,2175 & 0,1292 & 0,7918 & 0,4901 \\ 0,1692 & 0,1114 & 0,1111 & 0,2069 & 0,5239 \\ 0,1685 & 0,1995 & 0,1289 & 0,2076 & 0,5242 \end{pmatrix} \begin{matrix} \rightarrow \text{Kelompok 1} \rightarrow \text{Rangking 2} \\ \rightarrow \text{Kelompok 2} \rightarrow \text{Rangking 3} \\ \rightarrow \text{Kelompok 3} \rightarrow \text{Rangking 1} \\ \rightarrow \text{Kelompok 4} \rightarrow \text{Rangking 5} \\ \rightarrow \text{Kelompok 5} \rightarrow \text{Rangking 4} \end{matrix}$$

Berdasarkan *ranking* kelompok tersebut, maka dapat diinterpretasikan karakteristik dari tiap-tiap kelompok yang ada, berikut adalah interpretasi tiap-tiap kelompok yang diurut berdasarkan *ranking*-nya:

Kelompok 3:

Kelompok ini beranggotakan 11 desa/kelurahan. Merupakan desa yang maju dilihat dari faktor sarana/prasarana desa dan faktor sosial ekonomi penduduk desa. Kelompok ini memiliki nilai rata-rata pusat kelompok paling tinggi. Artinya, kelompok ini terdiri dari desa-desa yang paling maju dibanding kelima desa lainnya.

Kelompok 1:

Kelompok ini beranggotakan 25 desa/kelurahan. Merupakan desa yang sedang maju dilihat dari faktor sarana/prasarana desa dan faktor sosial ekonomi penduduk desa. Bila dilihat dari rata-rata pusat kelompok, desa-desa pada kelompok ini merupakan desa-desa yang tingkat kemajuannya dibawah dari kelompok 3.

Kelompok 2:

Kelompok ini beranggotakan 38 desa/kelurahan. Merupakan desa yang cukup maju, secara keseluruhan jika dilihat dari rata-rata pusat kelompok, desa-desa pada kelompok ini tingkat kemajuannya masih dibawah dari kelompok 3 dan kelompok 1.

Kelompok 5:

Kelompok ini beranggotakan 69 desa/kelurahan. Merupakan desa yang kurang maju, jika dilihat dari pusat kelompok, desa-desa pada kelompok ini memiliki tingkat kemajuan dibawah kelompok 1, kelompok 2 dan kelompok 3 namun diatas kelompok 4.

Kelompok 4:

Kelompok ini beranggotakan 84 desa/kelurahan. Merupakan desa yang tertinggal jika dibanding desa-desa pada kelompok 1, kelompok 2, kelompok 3 dan kelompok 5 jika dilihat dari faktor sarana/prasarana dan faktor sosial ekonomi penduduknya.

Kesimpulan

Berdasarkan hasil analisis pada penelitian ini, maka dapat diperoleh kesimpulan sebagai berikut:

1. Berdasarkan beberapa kriteria hasil pengelompokan dari dua metode yaitu *C-Means* dan *FCM clustering*, maka direkomendasikan metode *FCM* lebih baik dalam pengelompokan wilayah desa/kelurahan di Kabupaten Kutai Kartanegara. *FCM* memiliki nilai yang lebih baik dibanding *C-Means* berdasarkan kriteria fungsi objektif, waktu komputasi dan rasio nilai S_w/S_b . Nilai fungsi objektif pada metode *FCM* dari $c=2$ sampai dengan $c=10$ lebih kecil dibanding *C-Means*, begitu pula waktu komputasi *FCM* secara umum memerlukan waktu yang lebih singkat dibanding dibanding *C-Means*. Adapun berdasarkan rasio nilai S_w/S_b , *FCM* memiliki nilai yang lebih kecil untuk semua jumlah kelompok dibandingkan dengan *C-Means clustering*.
2. Hasil pengelompokan 227 desa/kelurahan di Kabupaten Kutai Kartanegara dengan menggunakan metode *FCM* sebagai rekomendasi metode terbaik dalam pengelompokan dengan jumlah kelompok yang terbentuk sebanyak 5 (lima) kelompok. Adapun karakteristik masing-masing kelompok dinilai berdasarkan *ranking* kelompoknya, yaitu dari *ranking* tertinggi ke *ranking* terendah. Berdasarkan nilai rata-rata *centroid*-nya, maka kelompoknya adalah sebagai berikut: kelompok 3 beranggotakan 11 desa/kelurahan merupakan desa maju, kelompok 1 beranggotakan 25 desa/kelurahan merupakan desa yang

sedang maju, kelompok 2 beranggotakan 38 desa/kelurahan merupakan desa cukup maju, kelompok 5 beranggotakan 69 desa/kelurahan yang merupakan desa kurang maju, kelompok 4 beranggotakan 84 desa/kelurahan yang merupakan desa tertinggal. Kemajuan dan ketertinggalan desa pada penelitian ini dibentuk berdasarkan faktor sarana/prasarana dan faktor sosial ekonomi penduduk desa.

Daftar Pustaka

- Agusta, Y. 2007. *K-Means Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3, hal 47-60.
- Bezdek, J.C dan Dunn, J.C. (1975). *Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions*. *IEEE Transactions on Computers*, hal 835-838.
- Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York : Plenum Press.
- Bunkers, M.J dan Miller, J.R. 1996. *Definition of Climate Regions in Northern Plains Using an Objective Cluster Modification Technique*. *Journal of Climate*, No.9, hal 130-146.
- Johnson, R.A dan Wichern, D.W. 2002. *Applied Multivariate Statistical Analysis - Fifth Edition*. New Jersey : Prentice-Hall, Inc.
- _____. 2007. *Applied Multivariate Statistical Analysis - Sixth Edition*. New Jersey : Prentice-Hall, Inc.
- Kusumadewi, Sri dan Purnomo, Hadi. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta : Graha Ilmu.
- Luthfi, Taufik, Emha. 2007. *Fuzzy C-Means Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)*. SNT 2007.
- Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta : Penerbit ANDI.
- Shihab, A.I. 2000. *Fuzzy Clustering Algorithm and Their Application to Medical Image Analysis*. [Dissertation]. London : University of London.
- Syoer, R.R. 2011. *Analisis Kelompok dengan Algoritma Fuzzy C-Means dan Gath-Geva Clustering (Studi Kasus: Pengelompokan Desa/Kelurahan di Kabupaten Kutai Kartanegara)*. [Tesis]. Surabaya : ITS.
- Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta : Penerbit ANDI.

